Distortion Exponent in Fading MIMO Channels with Time-Varying Side Information

Iñaki Estella Aguerri and Deniz Gündüz

Imperial College London, UK

Email: {i.estella-aguerri12, d.gunduz }@imperial.ac.uk

Abstract

Transmission of a Gaussian source over a time-varying multiple-input multiple-output (MIMO) channel is studied under strict delay constraints. Availability of a correlated side information at the receiver is assumed, whose quality, i.e., correlation with the source signal, also varies over time. A block-fading model is considered for both the states of the time-varying channel and the time-varying side information; and perfect channel and side information state information at the receiver is assumed, while the transmitter only has a statistical knowledge. The high SNR performance, characterized by the distortion exponent, is studied for this joint source-channel coding problem. An upper bound on the achievable distortion exponent is derived by providing the side information state to the transmitter, while the channel state remains unknown. For achievability, transmission schemes based on separate source and channel coding, uncoded transmission, joint decoding, as well as hybrid digital-analog transmission are considered. Multi-layer schemes, which transmit successive refinement layers of the source, are also proposed, based on progressive as well as superposed transmission with joint decoding. The optimal distortion exponent is characterized for the single-input multiple-output (SIMO) and multiple-input single-output (MISO) scenarios by showing that the distortion exponent achieved by multi-layer superpositon encoding with joint decoding meets the proposed upper bound. In the MIMO scenario, the optimal distortion exponent is characterized in the low bandwidth ratio regime, and it is shown that the multi-layer superposition encoding performs very close to the upper bound in the high bandwidth expansion regime as well.

Index Terms

Distortion exponent, time-varying channel and side information, multiple-input multiple-output (MIMO), joint source-channel coding, joint decoding, broadcast codes, successive refinement.

I. Introduction

Many applications in wireless networks require the transmission of a source signal over a fading channel, i.e., multimedia signals over cellular networks or the accumulation of local measurements at a fusion center in sensor networks, to be reconstructed with the minimum distortion possible at the destination. In many practical scenarios, the destination receives additional correlated side information about the underlaying source signal, either form other transmitters in the network or through its own sensing devices. For example, measurements from other sensors at a fusion center, signals from repeaters in digital TV broadcasting or relay signals in mobile networks.

The theoretical benefits of having correlated side information at the receiver for source encoding are well known [1]. However, similar to estimating the channel state information at the transmitter, it is costly to provide an estimate of the available side information to the transmitter, or may even be impossible in uncoordinated scenarios. Without the knowledge of the channel and the side information states, a transmitter needs to transmit in a manner that can adapt dynamically to the time-varying channel and side information qualities without knowing their realizations.

Here, we consider the joint source-channel coding problem of transmitting a Gaussian source over a multiple-input multiple-output (MIMO) block-fading channel when the receiver has access to time-varying correlated side information. Both the channel and the side-information are assumed to follow block-fading models, whose states are unknown at the transmitter. Moreover, strict delay constraints apply requiring the transmission of a block of source samples, for which the side-information quality state is constant, over a block of the channel, during which the channel state is constant. The two blocks do not necessarily have the same length, and their ratio is defined as the *bandwidth ratio* between the channel and the source bandwidths.

When the knowledge of the channel and side information states is available at both the transmitter and the receiver (CSI-TR), Shannon's separation theorem applies [2], assuming that the channel and source blocks are sufficiently long. In this case, the optimal performance is achieved by first compressing the source with an optimal source code and transmitting the compressed bits with a capacity achieving channel code. However, the optimality of source-channel separation does not extend to non-ergodic scenarios such as the model studied in this paper.

This problem has been studied extensively in the literature in the absence of correlated side information at the receiver [3]–[5]. Despite the ongoing efforts, the minimum achievable average distortion remains an open problem; however, more conclusive results on the performance can be obtained by studying the *distortion exponent*, which characterizes the exponential decay of the expected distortion in the high SNR regime [6]. The distortion exponent has been studied for parallel fading channels in [7], for the relay channel in [8], for point-to-point MIMO channels in [9], for channels with feedback in [10], for the two-way relay channel in [11], for the interference channel in [12], and in the presence of side information that might be absent in [13]. In the absence of side information at the receiver, the optimal distortion exponent in MIMO channels is known in some regimes of operation, such as the large bandwidth regime [9] and the low bandwidth regime [14]. However, the general problem remains open. In [9] digital multi-layer superposition transmission schemes are shown to achieve the optimal distortion exponent for high bandwidth ratios in MIMO systems. The optimal distortion exponent in the low bandwidth regime is achieved through hybrid digital-analog transmission [9], [14]. In [15], superposition multi-layer schemes are shown to achieve the optimal distortion exponent some other bandwidth ratios as well. Overall, multi-layer transmission has been shown to achieve the largest distortion exponents among the existing schemes in the literature.

The source coding version of our problem, in which the encoder and decoder are connected by an error-free finite-capacity channel, is studied in [16]. The single-input single-output (SISO) model in the presence of a time-varying channel and side information is considered for matched bandwidth ratios in [17], where uncoded transmission is shown to achieve the minimum expected distortion for certain side information fading gain distributions, while separate source and channel coding is shown to be suboptimal in general. A scheme based on joint decoding at the receiver, called JDS, is also proposed in [17], and it is shown to outperform separate source and channel coding by exploiting the joint quality of the channel and side information states.

Our goal in this work is to find tight bounds on the distortion exponent when transmitting a Gaussian source over a time-varying MIMO channel in the presence of time-varying correlated side information at the receiver[1]. We first derive an upper bound on the distortion exponent by providing the channel state information to the encoder. Then, we consider single-layer encoding schemes based on separate source and channel coding (SSCC), joint decoding (JDS), uncoded transmission and hybrid digital-analog transmission. Motivated by the improvements provided by multi-layer transmission in [9], we then consider two different multi-layer joint decoding schemes based on successive refinement of the source followed either by progressive transmission over the channel (LS-JDS), or by superposing JDS codes in a broadcast fashion (BS-JDS), and show that these schemes achieve the best distortion exponents.

The main results of this work can be summarized as follows:

- We first derive an upper bound on the distortion exponent by providing both the channel and the side information states to the encoder. Then, a tighter upper bound is obtained by providing only the channel state to the encoder.
- We characterize the distortion exponent achieved by JDS. While this scheme achieves a lower expected distortion than SSCC, we show that it does not improve the distortion exponent.
- We then consider a hybrid digital-analog scheme (HDA-WZ) that combines JDS with an analog layer. We show that HDA-WZ outperforms JDS not only in terms of the average distortion, but also the distortion exponent.
- We extend JDS by considering multi-layer transmission, where each layer carries successive refinement information for the source sequence. We consider both the progressive (LS-JDS) and superposition (BS-JDS) transmission of these layers, and derive the respective achievable distortion exponent expressions.
- We show that BS-JDS achieves the optimal distortion exponent for SISO/SIMO/MISO systems,

[1]Preliminary results have been published in the conference version of this work in [18] for SISO channels and in [19] and [20] for MIMO channels.

thus characterizing the optimal distortion exponent in these scenarios. We also show that HDA-WZ achieves the optimal distortion exponent in SISO channels as well.

- In the general MIMO setup, we characterize the optimal distortion exponent in the low bandwidth ratio regime, and show that it is achievable by both HDA-WZ and BS-JDS. In addition, we show that in certain regimes of operation, LS-JDS outperforms all the other proposed schemes.

We will use the following notation in the rest of the paper. We denote random variables with upper-case letters, e.g., X, their realizations with lower-case letters, e.g., x, and the sets with calligraphic letters, e.g. $\mathcal{A}$. We denote $\mathrm{E}_X[\cdot]$ as the expectation with respect to X, and $\mathrm{E}_{\mathcal{A}}[\cdot]$ as the expectation over the set $\mathcal{A}$. We denote random vectors as $\mathbf{X}$ with realizations $\mathbf{x}$ and random matrix by H and realizations $\mathbf{H}$. We denote by $\mathbb{R}_n^+$ the set of positive real numbers, and by $\mathbb{R}_n^{++}$ the set of strictly positive real numbers in $\mathbb{R}^n$, respectively. We define $(x)^+ = \max\{0, x\}$. Given two functions $f(x)$ and $g(x)$, we use $f(x) \doteq g(x)$ to denote the exponential equality $\lim_{x\to\infty} \frac{\log f(x)}{\log g(x)} = 1$, while $\dot{\geq}$ and $\dot{\leq}$ are defined similarly.

The rest of the paper is organized as follows. The problem statement is given in Section II. Then, known results on the diversity multiplexing tradeoff are provided in Section III, which will be used later in the paper. Two upper bounds on the distortion exponent are derived in Section IV. Various single-layer achievable schemes are studied in Section V, while multi-layer schemes are considered in Section VI. The characterization of the optimal distortion exponent for certain regimes is relegated to Section VII. Finally, the conclusions are presented in Section VIII.

II. Problem Statement

We wish to transmit a zero mean, unit variance real Gaussian source sequence $X^m \in \mathbb{R}^m$ of independent and identically distributed (i.i.d.) random variables, i.e., $X_i \sim \mathcal{N}(0, 1)$, over a complex MIMO block Rayleigh-fading channel with M_t transmit and M_r receiver antennas, as shown in Figure 1. In addition to the channel output, time-varying correlated source side information is also available at the decoder. Time-variations in the source side information are assumed to follow a block fading model as well. The channel and the side information states are assumed to be constant for the duration of one block, and independent of each other, and among different blocks. We assume that each source block is composed of m source samples, which, due to the delay limitations of the underlying application, are

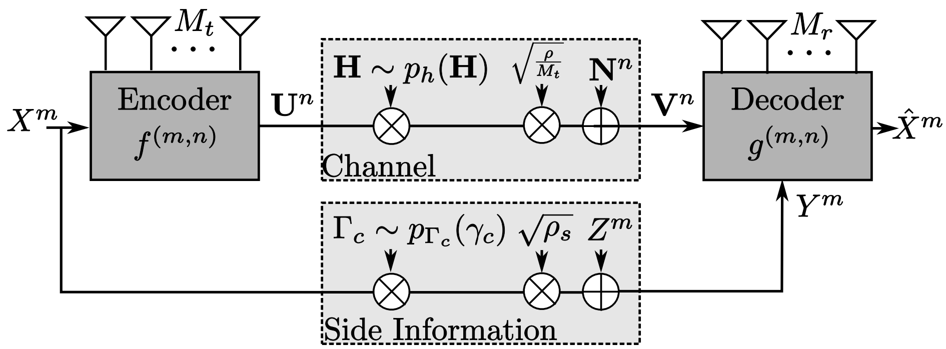


Fig. 1. Block diagram of the joint source-channel coding problem with fading channel and side information qualities.

supposed to be transmitted over one block of the channel, which consists of n channel uses. We define the *bandwidth ratio* of the system as[2]

$$b \triangleq \frac{2n}{m} \quad \textit{complex channel dimension per real source sample}.$$

The encoder maps each source sequence X^m to a channel input sequence $\mathbf{U}^n \in \mathbb{C}^{M_t \times n}$ using an encoding function $f^{(m,n)} : \mathbb{R}^m \to \mathbb{C}^{M_t \times n}$ such that the average power constraint is satisfied: $\sum_{i=1}^{n} \mathrm{Tr}\{\mathrm{E}[\mathbf{U}_i^H \mathbf{U}_i]\} \leq M_t n$. The memoryless slow fading channel is modeled as

$$\mathbf{V}_i = \sqrt{\frac{\rho}{M_t}} \mathsf{H} \mathbf{U}_i + \mathbf{N}_i, \qquad i = 1, ..., n,$$

where $\mathsf{H} \in \mathbb{C}^{M_r \times M_t}$ is the channel matrix with i.i.d. zero mean complex Gaussian entries, i.e., $h_{ij} \sim \mathcal{CN}(0,1)$, whose realizations are denoted by $\mathbf{H}$, $\rho \in \mathbb{R}^+$ is the average signal to noise ratio (SNR) in the channel, and $\mathbf{N}_i$ models the additive noise with $\mathbf{N}_i \sim \mathcal{CN}(0, \mathbf{I})$. We define $M^* = \max\{M_t, M_r\}$ and $M_* = \min\{M_t, M_r\}$, and consider $\lambda_{M_*} \geq \cdots \geq \lambda_1 > 0$ to be the eigenvalues of $\mathbf{H}\mathbf{H}^H$.

In addition to the channel output $\mathbf{V}^n = [\mathbf{V}_1, ..., \mathbf{V}_n] \in \mathbb{C}^{M_r \times n}$, the decoder observes $Y^m \in \mathbb{R}^m$, a randomly degraded version of the source sequence:

$$Y^m = \sqrt{\rho_s} \Gamma_c X^m + Z^m,$$

[2]This scaled definition is done for consistency of results with previous works in the distortion exponent literature, which use real/real or complex/complex sources and channels [9].

where Γ_c models Rayleigh fading[3] in the quality of the side information satisfying $\mathrm{E}[\Gamma_c^2] = 1$, $\rho_s \in \mathbb{R}^+$ models the average quality of the side information, and $Z_j \sim \mathcal{N}(0,1)$, $j = 1, ..., m$, models the noise. We define the *side information gain* as $\Gamma \triangleq \Gamma_c^2$, and its realization as γ. Then, Γ follows an exponential distribution with probability density function (pdf):

$$p_\Gamma(\gamma) = e^{-\gamma}, \qquad \gamma \geq 0.$$

In this work, we assume that the receiver knows the side information and the channel realizations, γ and $\mathbf{H}$, while the encoder is only aware of their distributions. The decoder reconstructs the source sequence $\hat{X}^m = g^{(m,n)}(\mathbf{V}^n, Y^m, \mathbf{H}, \gamma)$ with a mapping $g^{(m,n)} : \mathbb{C}^{n\times M_r} \times \mathbb{R}^m \times \mathbb{C}^{M_t \times M_r} \times \mathbb{R} \to \mathbb{R}^m$. The distortion between the source sequence and the reconstruction is measured by the quadratic average distortion $D \triangleq \frac{1}{m}\sum_{i=1}^m (X_i - \hat{X}_i)^2$.

We are interested in characterizing the minimum *expected distortion*, $\mathrm{E}[D]$, where the expectation is taken with respect to the source, the side information and channels state realizations, as well as the noise terms, and expressed as

$$ED^*(\rho, \rho_s, b) \triangleq \lim_{\substack{n,m\to\infty,\\ 2n\leq mb.}} \min_{f^{(m,n)}, g^{(m,n)}} \mathrm{E}[D].$$

In particular, we are interested in characterizing the optimal performance in the high SNR regime, i.e., when $\rho, \rho_s \to \infty$. We define x as a measure of the average *side information quality* in the high SNR regime, as follows:

$$x \triangleq \lim_{\rho\to\infty} \frac{\log \rho_s}{\log \rho}.$$

The performance measure we consider is the *distortion exponent*, defined as

$$\Delta(b, x) \triangleq -\lim_{\rho,\rho_s\to\infty} \frac{\log \mathrm{E}[D]}{\log \rho},$$

[3]The assumption of a real source sequence X^m and a real fading coefficient Γ_c is made in order to allow a degradation model possible. That is, the side-information qualities can be ordered among different channel states. Complex source and fading side information sequences would not allow an ordering in the quality of the side information sequences.

III. Diversity-Multiplexing Tradeoff

Here we digress shortly from the distortion exponent problem introduces above, and briefly talk about another, more commonly used, performance measure in the high SNR regime, that will be instrumented in our analysis. The diversity-multiplexing tradeoff (DMT) measures the tradeoff between the rate and reliability in the transmission of a message over a MIMO fading channel in the asymptotic high SNR regime. Hence, the DMT is a performance measure for the channel coding problem over block-fading channels. In this section we briefly review some known results on the DMT, which will be useful in the distortion exponent analysis. We refer the reader to [21] for a more detailed exposition of the DMT.

For a family of channel codes with rate $R = r \log \rho$, where r is the *multiplexing gain*, the *diversity gain* is defined as

$$d(r) = -\lim_{\rho \to \infty} \frac{\log P_e(\rho)}{\log \rho},$$

where $P_e(\rho)$ is the probability of decoding error of the channel code. For each r, the supremum of the diversity gain $d(r)$ over all coding schemes is given by $d^*(r)$. The DMT for a MIMO channel is given as the solution to the following problem [21],

$$\begin{aligned} d^*(r) = \inf_{\boldsymbol{\alpha}^+} & \sum_{i=1}^{M_*} (2i - 1 + M^* - M_*)\alpha_i \\ \text{s.t. } r \geq & \sum_{i=1}^{M_*} (1 - \alpha_i), \end{aligned} \tag{1}$$

where $\boldsymbol{\alpha}^+ \triangleq \{(\alpha_1, ..., \alpha_{M_*}) \in \mathbb{R}^{M_*} : 1 \geq \alpha_1 \geq ... \geq \alpha_{M_*} \geq 0\}$. The DMT obtained from (1) is a piecewise-linear function connecting the points $(k, d^*(k))$, $k = 0, ..., M_*$, where $d^*(k) = (M^* - k)(M_* - k)$. More specifically, for $r \geq M_*$, we have $d^*(r) = 0$, and for $0 \leq r \leq M_*$ satisfying $k \leq r \leq k+1$ for some $k = 0, 1, ..., M_* - 1$, the DMT curve is characterized by

$$d^*(r) \triangleq \Phi_k - \Upsilon_k (r - k), \tag{2}$$

where we have defined

$$\Phi_k \triangleq (M^* - k)(M_* - k) \quad \text{and} \quad \Upsilon_k \triangleq (M^* + M_* - 2k - 1). \tag{3}$$

IV. Distortion Exponent Upper Bound

In this section we derive two upper bounds on the distortion exponent by extending the two bounds on the expected distortion ED^* obtained in [17] to the MIMO setup with bandwidth mismatch, and analyzing their high SNR behavior.

A. Fully informed encoder upper bound

The first upper bound, which we denote as the *fully informed encoder upper bound*, is obtained by providing the transmitter with both the channel state $\mathbf{H}$ and the side information state γ. At each realization, the problem reduces to the static setup studied in [2], and source-channel separation theorem applies; that is, the concatenation of a Wyner-Ziv source code with a capacity achieving channel code is optimal at each realization. Averaging the achieved distortion over the realizations of the channel and side information states, the expected distortion is found as

$$ED_{\text{inf}}(\rho, \rho_s, b) = \mathrm{E}_{\mathbf{H},\Gamma}\left[\frac{1}{1+\rho_s\gamma}2^{-b\mathcal{C}(\mathbf{H})}\right],$$

where $\mathcal{C}(\mathbf{H})$ is the capacity of the MIMO channel in bits/channel use.

Following similar derivations in [9] and [17], we find an upper bound on the distortion exponent, stated in the following lemma.

Lemma 1. *The distortion exponent is upper bounded by the fully informed encoder upper bound, given by*

$$\Delta_{\text{inf}}(b, x) = x + \Delta_{\text{MIMO}}(b), \tag{4}$$

where

$$\Delta_{\text{MIMO}}(b) \triangleq \sum_{i=1}^{M_*} \min\{b, 2i - 1 + M^* - M_*\}. \tag{5}$$

B. Partially informed encoder upper bound

A tighter upper bound can be constructed by providing the transmitter with only the channel state realization $\mathbf{H}$ while the side information state γ remains unknown. We call this the *partially informed encoder upper bound*. The optimality of separate source and channel coding is shown in [17] when the side information fading gain distribution is discrete, or continuous and quasiconcave for $b = 1$. The proof easily extends to the non-matched bandwidth ratio setup and, since in our model $p_\Gamma(\gamma)$ is exponential, and hence, is continuous and quasiconcave, separation is optimal at each channel block.

As shown in [16], [17], if $p_\Gamma(\gamma)$ is monotonically decreasing, the optimal source encoder ignores the side information completely, and the side-information is used only at the decoder for source reconstruction[4]. Concatenating this side-information-ignorant source code with a channel code at the instantaneous capacity, the minimum expected distortion at each channel state $\mathbf{H}$ is given by

$$D_{\text{op}}(\rho, \rho_s, b, \mathbf{H}) = \frac{1}{\rho_s} e^{\frac{2^{b\mathcal{C}(\mathbf{H})}}{\rho_s}} E_1\left(\frac{2^{b\mathcal{C}(\mathbf{H})}}{\rho_s}\right),$$

where $E_1(x)$ is the exponential integral given by $E_1(x) = \int_x^\infty t^{-1} e^t dt$. Averaging over the channel state realizations, the expected distortion is lower bounded as

$$ED^*_{\text{pi}}(\rho, \rho_s, b) = \mathrm{E}_{\mathbf{H}}[D_{\text{op}}(\rho, \rho_s, b, \mathbf{H})]. \tag{6}$$

An upper bound on the distortion exponent is found by analyzing the high SNR behavior of (6) as given in the next theorem.

Theorem 1. *Let $l = 1$ if $x/M_* < M^* - M_* + 1$, and let $l \in \{2, ..., M_*\}$ be the integer satisfying $2l - 3 + M^* - M_* \leq x/M_* < 2l - 1 + M^* - M_*$ if $M^* - M_* + 1 \leq x/M_* < M^* + M_* - 1$. The*

[4] We note that when the distribution of the side information is not Rayleigh, the optimal encoder follows a different strategy. For example, for quasiconcave continuous distributions the optimal source code compresses the source aiming at a single target side information state. See [17] for details.

distortion exponent is upper bounded by

$$\Delta_{up}(b,x)=\begin{cases} x & \textit{if } 0\leq b<\frac{x}{M_*},\\ bM_* & \textit{if } \frac{x}{M_*}\leq b<M^*-M_*+1,\\ x+d^*\left(\frac{x}{b}\right) & \textit{if } 1+M^*-M_*\leq b<2l-1+M^*-M_*,\\ x+d^*\left(\frac{x}{b}\right) & \textit{if } 2l-1+M^*-M_*\leq b<\frac{x}{M_*-k},\\ \Delta_{\text{MIMO}}(b) & \textit{if } \frac{x}{M_*-k}\leq b<M^*+M_*-1,\\ x+d^*\left(\frac{x}{b}\right) & \textit{if } M^*+M_*-1\leq b, \end{cases}$$

where $k\in\{l,...,M_*-1\}$ is the integer satisfying $2k-1+M^*-M_*\leq b<2k+1+M^*-M_*$.

If $x/M_*\geq M^*+M_*-1$, then

$$\Delta_{up}(b,x)=x+d^*\left(\frac{x}{b}\right),$$

where $d^*(r)$ is the DMT characterized in (2)-(3).

Proof: The proof is given in Appendix I. ■

Comparing the two upper bounds in Lemma 1 and Theorem 1, we can see that the latter is always tighter. When $x>0$, the two bounds meet only at the two extremes, when either $b=0$ or $b\to\infty$. Note that these bounds provide the achievable distortion exponents when either both states (Lemma 1) or only the channel state (Theorem 1) is available at the transmitter, also characterizing the potential gains from channel state feedback in fading joint source-channel coding problems.

V. Single Layer Transmission

In this section, we propose transmission schemes consisting of a single-layer code, and analyze their achievable distortion exponent performance.

A. Separate source and channel coding scheme (SSCC)

SSCC is optimal in the presence of CSI-TR, as described in Section IV-A. When CSI-TR is not available, the binning and the channel coding rates have to be designed solely based on the statistics of the channel and the side information. Thus, the transmission suffers from two separate outage events: outage in channel decoding and outage in source decoding [17]. It is shown in [17, Corollary 1] that, for monotonically decreasing pdfs, such as $p_\Gamma(\gamma)$ considered here, the expected distortion is minimized by avoiding outage in source decoding, that is, by not using binning. Therefore, the optimal SSCC scheme compresses the source sequence at rate R_s ignoring the side information, and transmits the compressed bits over the channel with a channel code with rate R_c such that $\frac{b}{2}R_c = R_s$.

At the encoder, the quantization codebook consists of 2^{mR_s} length-m codewords, $W^m(i)$, $i = 1, ..., 2^{mR_s}$, generated through a 'test channel' given by $W = X + Q$, where $Q \sim \mathcal{N}(0, \sigma_Q^2)$, and is independent of X. The quantization noise variance is such that $R_s = I(X;W) + \epsilon$, for an arbitrarily small $\epsilon > 0$, i.e., $\sigma_Q^2 = (2^{2(R_s-\epsilon)} - 1)^{-1}$. For the channel code, a Gaussian channel codebook with 2^{nRc} length-n codewords $\mathbf{U}^n(s)$ is generated independently with $\mathbf{U} \sim \mathcal{CN}(0, \mathbf{I})$, and each codeword $\mathbf{U}^n(s)$, $s \in [1, ..., 2^{nR_c}]$, is assigned to a quantization codeword $W^m(i)$. Given a source sequence X^m, the encoder searches for a quantization codeword $W^m(i)$ jointly typical with X^m, and transmits the corresponding channel codeword $\mathbf{U}(i)$.

The decoder recovers the digital codeword with high probability if $R_c < I(\mathbf{U}, \mathbf{V})$. An outage is declared whenever due to the channel randomness, the channel rate R_c is above the capacity and the codeword cannot be recovered. Then, the outage event is given by

$$\mathcal{O}_s = \{\mathbf{H} : R_c \geq I(\mathbf{U}; \mathbf{V})\}, \tag{7}$$

where $I(\mathbf{U};\mathbf{V}) = \log\det(\mathbf{I} + \frac{\rho}{M_*}\mathbf{H}\mathbf{H}^H)$.

If W^m is successfully decoded, the source sequence is estimated with an MMSE estimator using the quantization codeword and the side information sequence, i.e., $\hat{X}_i = \mathrm{E}[X_i|W_i, Y_i]$, and reconstructed with a distortion $D_d(bR_c/2, \gamma)$, where

$$D_d(R, \gamma) \triangleq (\rho_s\gamma + 2^{2R})^{-1}. \tag{8}$$

If there is an outage over the channel, only the side information is used in the source reconstruction and the corresponding distortion is given by $D_d(0, \gamma)$. The probability of outage depends only on the channel state $\mathbf{H}$. The expected distortion for SSCC can be written as

$$ED_s(bR_c) = \mathrm{E}_{\mathcal{O}_s^c}[D_d(bR_c/2,\Gamma)] + \mathrm{E}_{\mathcal{O}_s}[D_d(0,\Gamma)] \tag{9}$$

$$= (1-P_o(\mathbf{H}))\mathrm{E}_\Gamma[D_d(bR_c/2,\Gamma)] + P_o(\mathbf{H})\mathrm{E}_\Gamma[D_d(0,\Gamma)],$$

where $P_o(\mathbf{H}) \triangleq \Pr\{R_c \geq \log\det(\mathbf{I} + \frac{\rho}{M_*}\mathbf{H}\mathbf{H}^H)\}$ is the probability of channel outage.

In the next theorem, the distortion exponent achievable by SSCC is provided.

Theorem 2. *The achievable distortion exponent for SSCC, $\Delta_s(b,x)$, is given by*

$$\Delta_s(b,x) = \max\left\{x, b\frac{\Phi_k + k\Upsilon_k + x}{\Upsilon_k + b}\right\}, \qquad \textit{for } b \in \left[\frac{\Phi_{k+1}+x}{k+1}, \frac{\Phi_k + x}{k}\right), k = 0,1,...,M_* - 1,$$

where Φ_k and Υ_k are as defined in (3).

Proof: See Appendix II. ∎

The illustration of the achievable distortion exponent by the SSCC scheme and its comparison with other transmission techniques and the proposed upper bound is deferred to Section V-E.

B. Joint decoding scheme (JDS)

In this section, we consider a joint source-channel coding scheme, which, by joint decoding of the channel and the source codewords, reduces the outage probability. It uses no explicit binning at the encoding, and the success of decoding depends on the joint quality of the channel and the side information states. This scheme is considered in [17] for a SISO system, and is shown to outperform SSCC at any SNR and to achieve the optimal distortion exponent in certain regimes.

At the encoder, we generate a codebook of 2^{mR_j} length-m quantization codewords $W^m(i)$ and an independent Gaussian codebook of size $2^{n\frac{b}{2}R_j}$ with length-n codewords $\mathbf{U}(i) \in \mathbb{C}^{M_t \times n}$ with $\mathbf{U} \sim \mathcal{CN}(0,\mathbf{I})$, such that $\frac{b}{2}R_j = I(X;W)+\epsilon$, for an arbitrarily small $\epsilon > 0$. Given a source outcome X^m, the transmitter finds the quantization codeword $W^m(i)$ jointly typical with the source outcome and transmits the corresponding channel codeword $\mathbf{U}(i)$. Joint typicality decoding is performed such that the decoder looks for an index i for which both $(\mathbf{U}^n(i), V^n)$ and $(Y^m, W^m(i))$ are jointly typical. Then the outage event is

$$\mathcal{O}_j = \left\{ (\mathbf{H}, \gamma) : I(X; W|Y) \geq \frac{b}{2} I(\mathbf{U}; \mathbf{V}) \right\}, \tag{10}$$

where $I(\mathbf{U}; \mathbf{V}) = \log\det(\mathbf{I} + \frac{\rho}{M_*}\mathbf{H}\mathbf{H}^H)$ and $I(X; W|Y) = \frac{1}{2}\log(1 + \frac{2^{R_j - \epsilon} - 1}{\gamma\rho_s + 1})$.

Similarly to SSCC, if there is no outage the source is reconstructed using both the quantization codeword and the side information sequence with an MMSE estimator, while only the side information is used in case of an outage.

The joint decoding produces a binning-like decoding: only some $\mathbf{V}^n$ are jointly typical with $\mathbf{U}(s)$, generating a virtual bin of W^m codewords from which only one is jointly typical with Y^m with high probability. The size of those bins depends on the particular realizations of **H** and Γ unlike in a Wyner-Ziv scheme, in which the bin sizes are chosen in advance. Since the outage event depends jointly on the channel and the side information states $(\mathbf{H}, \gamma)$, the expectation over the states is not separable as in (9). Then, the expected distortion for JDS is expressed as

$$ED_j(R_j) = \mathrm{E}_{\mathcal{O}_j^c}\left[D_d\left(\frac{b}{2} R_j, \Gamma \right) \right] + \mathrm{E}_{\mathcal{O}_j}[D_d(0, \Gamma)].$$

JDS reduces the probability of outage, and hence, the expected distortion compared to SSCC. However, the next theorem reveals that both schemes achieve the same distortion exponent.

Theorem 3. *The distortion exponent of the JDS scheme, $\Delta_j(b, x)$ is the same as that of SSCC characterized in Theorem 2, i.e., $\Delta_j(b, x) = \Delta_s(b, x)$.*

Proof: See Appendix III. ■

Although JDS and SSCC achieve the same distortion exponent in the current setting, JDS is shown to achieve larger distortion exponents than SSCC in general [17]. A comparison between the two schemes is deferred to Section V-E.

C. *Uncoded transmission*

Uncoded transmission is a robust joint source-channel coding scheme that is known to have a gradual degradation with worsening channel quality. Uncoded transmission is suboptimal even in a point-to-point

SISO channel in the presence of side information; however the benefits of uncoded transmission have been illustrated in the presence of time-varying side information in [17]. It is shown in [17] that in a fading SISO channel with $b = 1$, uncoded transmission is exactly optimal in terms of the expected distortion when the side information gain follows a monotonically decreasing pdf, such as $p_\Gamma(\gamma)$ in our model. However, for general MIMO channels and bandwidth ratios, it falls short of the optimal performance, since it cannot fully exploit the additional degrees-of-freedom in the system.

In uncoded transmission, the source samples are used directly as the channel inputs. Since the channel is complex, we reorder the source sequence as $\mathbf{X}_c^{\frac{m}{2}} \in \mathbb{C}^{\frac{m}{2}}$ given by

$$\mathbf{X}_c^{\frac{m}{2}} = \frac{1}{\sqrt{2}} \Big([X_1, ..., X_{\frac{m}{2}}] + j[X_{\frac{m}{2}+1}, ..., X_m] \Big)^T, \tag{11}$$

where $j = \sqrt{-1}$. In the transmission we consider M_* of the M_t transmit antennas since only M_* samples are effectively transmitted at each channel use, because rank$\{\mathbf{H}\} \leq M_*$.

For $bM_* \leq 1$, the channel input $\mathbf{U}^n$ is generated scaling the first nM_* source samples of $X_c^{\frac{m}{2}}$ and mapping them to the channel input as $\mathbf{U}^n = [\mathbf{X}_{c,1}^{M_*}, \mathbf{X}_{c,M_*+1}^{2M_*}, .., \mathbf{X}_{c,(n-1)M_*+1}^{nM_*}]^T$. At reception, the transmitted nM_* source samples are reconstructed with an MMSE estimator using $\mathbf{V}^n$ and Y^{nM_*}, while the remaining $\frac{m}{2} - nM_*$ source samples that have not been transmitted, are estimated using only $Y_{nM_*+1}^m$. For $bM_* \geq 1$, the whole source sequence is transmitted in the first $\frac{m}{2M_*}$ channel uses scaling the power by bM_*, and reconstructed at the decoder using an MMSE estimator. The minimum average distortion achieved by uncoded transmission with average power P at state $(\mathbf{H}, \gamma)$ is given by

$$D_u(P, \gamma, \mathbf{H}) \triangleq \sum_{i=i}^{M_*} \frac{1}{1 + P\mu_i\rho + \gamma\rho_s}, \tag{12}$$

where $\mu_1 \geq \cdots \geq \mu_{M_*} \geq 0$ are the ordered eigenvalues of the matrix $\mathbf{H}_{M_*}\mathbf{H}_{M_*}^H$, where $\mathbf{H}_{M_*}$ is the submatrix of $\mathbf{H}$ formed by the M_* columns corresponding to the antennas effectively used for transmission. Then, the expected distortion is found as

$$ED_u = \begin{cases} bM^*\mathrm{E}[D_u(1, \mathsf{H}, \Gamma)] + (1 - bM_*)\mathrm{E}[D_u(0, \mathsf{H}, \Gamma)] & \text{if } bM_* < 1, \\ \mathrm{E}[D_u(bM^*, \mathsf{H}, \Gamma)] & \text{if } bM_* \geq 1. \end{cases}$$

The distortion exponent for uncoded transmission is obtained similarly to $\Delta_d(b,x)$ and is given in the next theorem without proof.

Theorem 4. *The distortion exponent for uncoded transmission, $\Delta_u(b,x)$ is given by*

$$\Delta_u(b,x) = \begin{cases} x & \text{if } bM_* < 1, \\ \max\{1,x\} & \text{if } bM_* \geq 1. \end{cases}$$

It is evident from Theorem 3 that uncoded transmission cannot exploit the available degrees-of-freedom in the system (multiple antennas, channel bandwidth), and its distortion exponent performance on its own is very poor compared to digital and hybrid schemes. In Section V-E, the performance of uncoded transmission will be compared to the proposed achievable schemes and upper bounds.

D. HDA Wyner-Ziv coding (HDA-WZ)

In this section we consider a hybrid digital-analog (HDA) scheme that quantizes the source sequence, uses a scaled version of the quantization error as the channel input, and exploits joint decoding at the decoder. This scheme is introduced in [22] and named as HDA Wyner-Ziv Coding (HDA-WZ), and shown to be optimal in static SISO channels in the presence of side information for $b = 1$. HDA-WZ is considered in [17] in the SISO fading setup with $b = 1$, and it is shown to achieve the optimal distortion exponent for a wide family of side information distributions. In this paper, we propose a generalization of HDA-WZ in [17] to the MIMO channel and to bandwidth ratios satisfying $b \geq 1/M^*$.

For $b \leq 1/M_*$, we ignore the available side information and use the hybrid digital-analog scheme proposed in [14]. In this scheme, which we denote by superposed HDA (HDA-S), the source sequence is divided and transmitted using two layers. The first layer transmits a part of the source sequence in an uncoded fashion, while the second layer digitally transmits the rest of the source samples. The two layers are superposed and the available power is allocated among them to maximize the achievable distortion exponent. At the destination, the digital layer is decoded treating the uncoded layer as noise. Then, the source sequence is reconstructed using both layers. The distortion exponent achievable by HDA-S is given by $\Delta_h(b,x) = bM_*$ for $b \leq 1/M_*$ [14].

HDA-S can be modified to include joint decoding and to use the available side information at the reconstruction to reduce the expected distortion. However, as we will show in Section VII-A, if $0 \leq b \leq x/M_*$, simple MMSE estimation of the source sequence is sufficient to achieve the optimal distortion exponent, given by $\Delta^*(b,x) = x$, and if $x/M_* \leq b \leq 1/M_*$, HDA-S achieves the optimal distortion exponent. Therefore, considering HDA-S with joint decoding will not improve the distortion exponent in this regime.

Lemma 2. *The distortion exponent achievable by HDA-S is given by $\Delta_h(b,x) = bM_*$ if $b \leq 1/M_*$.*

Next, we consider the HDA-WZ scheme for $bM_* > 1$. At the encoder, consider a quantization codebook of 2^{mR_h} length-m codewords $W^m(s)$, $s = 1, ..., 2^{mR_h}$, with a test channel $W = X + Q$, where $Q \sim \mathcal{N}(0, \sigma_Q^2)$ is independent of X, and quantization noise variance is chosen such that $\frac{R_h}{2} = I(W;X) + \epsilon$, for an arbitrarily small $\epsilon > 0$, i.e., $\sigma_Q^2 \triangleq (2^{R_h - \epsilon} - 1)^{-1}$. Then, each W^m is reordered into length-$\frac{m}{2M_*}$ complex codewords $\mathbf{W}(s) = [\mathbf{W}_1(s), ..., \mathbf{W}_{\frac{m}{2M_*}}(s)] \in \mathbb{C}^{\frac{m}{2M_*} \times M_*}$, where $\mathbf{W}_i(s)$, $i = 1, ..., \frac{m}{2M_*}$, is given by

$$\mathbf{W}_i(s) = \frac{1}{\sqrt{2}} \left([W_{iM_*+1}(s); ...; W_{(i+1)M_*}(s)] + j[W_{(i+1)M_*+1}(s); ...; W_{2iM_*}(s)] \right)^T,$$

Similarly, we can reorder X^m and Q^m, and define $\mathbf{X}_i$ and $\mathbf{Q}_i$ for $i = 1, ..., \frac{m}{2M_*}$.

We then generate 2^{mR_h} independent auxiliary random vectors $\mathbf{T} \in \mathbb{C}^{\left(n - \frac{m}{2M_*}\right) \times M_*}$ distributed as $\mathbf{T}_i \sim \mathcal{CN}(0, \mathbf{I})$, for $i = 1, ..., n - \frac{m}{2M_*}$ and assign one to each $\mathbf{W}(s)$ to construct the codebook of size 2^{mR_h} consisting of the pairs of codewords $(\mathbf{W}(s), \mathbf{T}(s))$, $s = 1, ..., 2^{mR_h}$. For a given source sequence X^m, the encoder looks for the s^*-th codeword $\mathbf{W}(s^*)$ such that $(\mathbf{W}(s^*), X^m)$ are jointly typical. A unique s^* is found if $M_* R_h > I(\mathbf{W}; \mathbf{X})$. Then, the pair $(\mathbf{W}(s^*), \mathbf{T}(s^*))$ is used to generate the channel input, which is scaled to satisfy the power constraint:

$$\mathbf{U}_i = \begin{cases} \sqrt{\frac{1}{\sigma_Q^2}} [\mathbf{X}_i - \mathbf{W}_i(s^*)], & \text{for } i = 1, ..., \frac{m}{2M_*}, \\ \mathbf{T}_{i - \frac{m}{2M_*}}(s^*), & \text{for } i = \frac{m}{2M^*} + 1, ..., n. \end{cases}$$

Basically, in the first block of $\frac{m}{2M_*}$ channel accesses we transmit a scaled version of the error of the quantization codeword $\mathbf{Q}_i$ in an uncoded fashion, while in the second block of $n - \frac{m}{2M_*}$ accesses we transmit a digital codeword.

The decoder looks for an index s such that $\mathbf{W}(s)$, Y^m and the channel output corresponding to the uncoded input, $\mathbf{V}_W^{\frac{m}{2M_*}} \triangleq [\mathbf{V}_1, \ldots, \mathbf{V}_{m/2M_*}]$, are jointly typical, while simultaneously $\mathbf{T}(s)$ is jointly typical with the channel output that corresponds to the coded input block, $\mathbf{V}_T^{n-\frac{m}{2M_*}} \triangleq [\mathbf{V}_{m/2M_*+1}, \ldots, \mathbf{V}_n]$. Let $\mathbf{Y}_i = [Y_{(i-1)M_*+1}, ..., Y_{iM_*}]^H$, for $i = 1, ..., \frac{m}{M_*}$, be blocks of Y^m. At the receiver, decoding is successful with high probability if

$$I(\mathbf{W}; \mathbf{X}) < M_* R_h < I(\mathbf{WT}; \mathbf{VY}) \tag{13}$$

The outage event is obtained in Appendix IV-A as

$$\mathcal{O}_h = \left\{ (\mathbf{H}, \gamma) : I(\mathbf{W}, \mathbf{X}) \geq I(\mathbf{W}; \mathbf{V}_W \mathbf{Y}) + (bM_* - 1) I(\mathbf{T}; \mathbf{V}_T) \right\}, \tag{14}$$

where $I(\mathbf{T}; \mathbf{V}_T) = \log\det(\mathbf{I} + \frac{\rho}{M_*}\mathbf{HH}^H)$ and,

$$I(\mathbf{W}; \mathbf{V}_W \mathbf{Y}) = \log \left(\frac{(\xi(1+\sigma_Q^2))^{M_*} \det(\mathbf{I} + \frac{\rho}{M_t}\mathbf{HH}^H))}{\det(\mathbf{I} + \sigma_Q^2(\frac{\rho}{M_t}\mathbf{HH}^H + \xi\mathbf{I}))} \right), \tag{15}$$

where $\xi \triangleq 1 + \rho_s \gamma$.

If $\mathbf{W}^{\frac{m}{2M_*}}$ is successfully decoded, each X^n is reconstructed with an MMSE estimator using $\mathbf{V}$ and Y^m with a distortion

$$D_h(\sigma_Q^2, \mathbf{H}, \gamma) = \frac{1}{M_*} \sum_{i=1}^{M_*} \left(1 + \rho_s \gamma + \frac{1}{\sigma_Q^2} \left(1 + \frac{\rho}{M_*} \lambda_i \right) \right)^{-1}. \tag{16}$$

The derivation of (16) is found in Appendix IV-B.

If an outage occurs and $\mathbf{W}$ is not decoded, only Y^m is used in the reconstruction, since $\mathbf{U}^n$ is uncorrelated with the source sequence by construction, and so is $\mathbf{V}^n$. Using an MMSE estimator, the achievable distortion is given by $D_d(0, \gamma)$. Then, the expected distortion for HDA-WZ is found as

$$ED_h(R_h) = \mathrm{E}_{\mathcal{O}_h^c}[D_h(\sigma_Q^2, \mathbf{H}, \Gamma)] + \mathrm{E}_{\mathcal{O}_h}[D_d(0, \Gamma)].$$

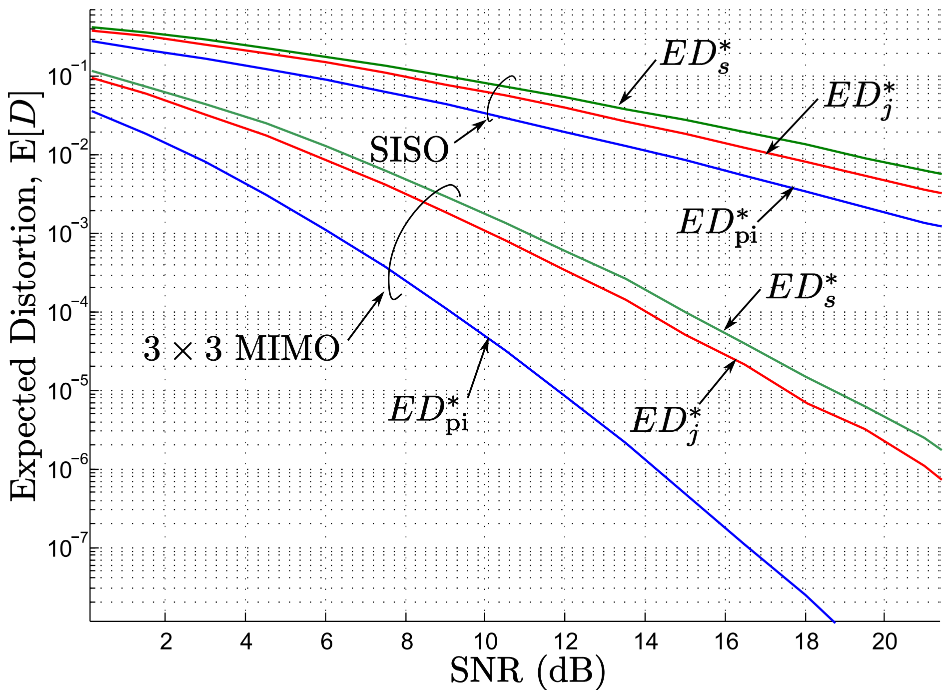


Fig. 2. Minimum expected distortion achievable by SSCC and JDS for a SISO and a 3×3 MIMO channel for $b = 2$ and $x = 1$. The partially informed encoder bound is also included.

The distortion exponent of HDA-WZ, $\Delta_h(b, x)$, is characterized in the next theorem.

Theorem 5. *Let $bM_* > 1$. The distortion exponent achieved by HDA-WZ, $\Delta_h(b, x)$, is given by*

$$\Delta_h(b,x) = \begin{cases} x & \text{if } \frac{1}{M_*} \leq b < \frac{x}{M_*}, \\ 1 + \frac{(bM_*-1)(\Phi_k + k\Upsilon_k - 1 + x)}{bM_* - 1 + M_*\Upsilon_k}, & \text{if } b \in \left[\frac{\Phi_{k+1}-1+x}{k+1} + \frac{1}{M_*}, \frac{\Phi_k - 1 + x}{k} + \frac{1}{M_*}\right), \\ & \text{for } k = 0, \ldots, M_* - 1. \end{cases}$$

Proof: See Appendix IV-C. ■

E. Comparison of single-layer transmission schemes

Here, we compare the performance of the single-layer schemes presented in this section. Figure 2 shows the expected distortion achievable by SSCC and JDS schemes and the partially informed encoder

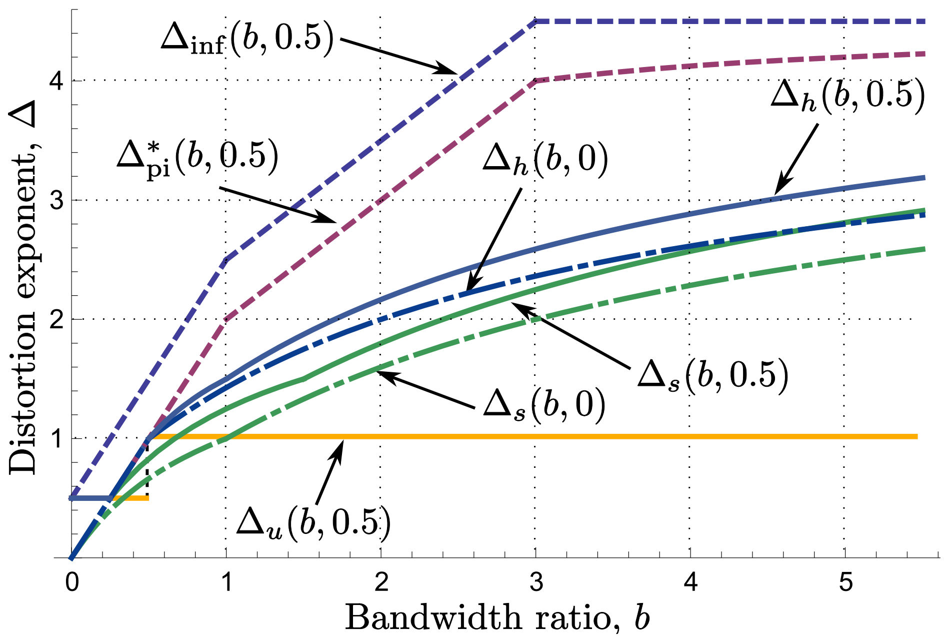


Fig. 3. Distortion exponents upper bounds and lower bounds for single-layer schemes in function of b for $x = 0.5$ and 2×2 MIMO. The performance of these schemes is also shown in the absence of side information, i.e., $x = 0$.

lower bound on the expected distortion in a SISO and a 3×3 MIMO setup for $b = 2$. It is observed that JDS outperforms SSCC in both SISO and MIMO scenarios. We also observe that both SSCC and JDS fall short of the expected distortion lower bound, ED^*_{pi}. Moreover the gap increases with the number of degrees-of-freedom in the system. We note that not only the gap between the achievable distortion values increase, but also the gap between the slopes of the curves, which means that the proposed transmission schemes perform especially poorly in the high SNR regime.

To illustrate this, we compare the distortion exponent achieved by SSCC, JDS, uncoded transmission, HDA-S and HDA-WZ in Figure 3 in a 2×2 MIMO channel, as a function of the bandwidth ratio. In the figure, we consider both a side information quality of $x = 0.5$, and the case without side information, i.e., $x = 0$. First, we note that, as discussed in Section VII-A, for $b \leq 0.25$, the upper bound is achieved by simply estimating the source sequence based on the side information, while for $0.25 \leq b \leq 0.5$, the upper bound is achieved by the S-HDA scheme, ignoring the side information. For larger bandwidth ratios, HDA-WZ improves upon SSCC and JDS, while uncoded transmission achieves the optimal distortion exponent for $b = 0.5$ and then saturates, becoming highly suboptimal for large b values. Note that uncoded transmission outperforms SSCC and JDS for the range $0.5 \leq b \lesssim 0.7$ if $x = 0.5$, and for $0.5 < b < 1$ if $x = 0$. In general we observe that single-layer schemes are not capable of fully exploiting the available degrees-of-freedom in the system, and their distortion exponent performance falls short of the proposed upper bounds, especially in the large bandwidth regime. This motivates us to consider other achievability techniques in order to achieve higher distortion exponent values.

VI. Multi-layer Transmission

In the previous section, we have observed that the distortion exponent achievable with single-layer schemes is far from the upper bound, especially in the high bandwidth regime. Here, we consider multi-layer schemes to improve the achievable distortion exponent in this regime. Multi-layer transmission is proposed in [9] to combat channel fading by transmitting multiple layers that carry successive refinements of the source [23]. At the receiver, as many layers as possible are decoded depending on the channel state. The better the channel state, the more layers can be decoded and the smaller is the distortion at the receiver. We propose to use successive refinement codewords that exploits the side information at the destination [24], and the extension of the JDS schemes to *progressive multi-layer JDS transmission* and *broadcast strategy with JDS*, and derive the corresponding distortion exponents.

A. Progressive multi-layer JDS transmission (LS-JDS)

In this section we consider the progressive transmission of multiple layers over the channel. The refinement codewords are transmitted one after the other over the channel using the JDS scheme introduced in Section V-B. Similarly to [9], we assume that each layer is allocated the same time resources (or number of channel accesses). In the limit of infinite layers, this assumption does not incur a loss in performance.

At the encoder, we generate L Gaussian quantization codebooks, each with 2^{mR_l} codewords W_l^m and $bR_l/2L = I(X; W_l|W_1^{l-1}) + \epsilon/2$, for $l = 1, ..., L$, with an arbitrarily small $\epsilon > 0$, such that each Gaussian codebook is a refinement for the previous layers [24]. The quantization codewords W_l^n are generated with a test channel given by $W_l = X + \sum_{i=l}^{L} Q_i$, for $l = 1, ..., L$, where $Q_l \sim \mathcal{N}(0, \sigma_l^2)$ are independent of X and of each other. Note that $Y - X - W_L - W_{L-1} - \cdots - W_1$ form a Markov chain. As shown in Appendix V-A, for a given rate tuple $\mathbf{R} \triangleq [R_1, ..., R_L]$, with $R_1 \geq \cdots R_L \geq 0$, the quantization noise variances satisfy

$$\sum_{i=l}^{L} \sigma_i^2 = (2^{\sum_{i=1}^{l}(\frac{b}{L}R_i - \epsilon)} - 1)^{-1}, \qquad l = 1, ..., L. \tag{17}$$

We generate L independent channel codebooks, each with $2^{n\frac{bR_l}{2L}}$ length-$\frac{n}{L}$ codewords $\mathbf{U}_l^{n/L} \in \mathbb{C}^{M_t \times n/L}$ with $\mathbf{U}_{l,i} \sim \mathcal{CN}(0, \mathbf{I})$. Each successive refinement codeword is transmitted using JDS as in Section V-B. At the destination, the decoder successively decodes each refinement codeword using joint decoding from the first layer up to the L-th layer. Then, l layers will be successfully decoded if

$$I(X; W_l | Y, W_1^{l-1}) < \frac{b}{2L} I(\mathbf{U}; \mathbf{V}) \leq I(X; W_{l+1} | Y, W_1^l),$$

that is, l layers are successfully decoded while there is an outage in decoding the $(l+1)$-th layer. Let us define the outage event, for $l = 1, ..., L$, as follows

$$\mathcal{O}_l^{ls} \triangleq \left\{ (\mathbf{H}, \gamma) : I(X, W_l | Y, W_1^{l-1}) \geq \frac{b}{2L} I(\mathbf{U}; \mathbf{V}) \right\}, \tag{18}$$

where $I(\mathbf{U}, \mathbf{V}) = \log\det\left(\mathbf{I} + \frac{\rho}{M_*}\mathbf{H}\mathbf{H}^H\right)$, and, with $R_0 \triangleq 0$,

$$I(X; W_l | W_1^{l-1}, Y) = \frac{1}{2} \log \left(\frac{2^{\sum_{i=1}^{l} \frac{b}{L} R_i} + \gamma\rho_s}{2^{\sum_{i=1}^{l-1} \frac{b}{L} R_i} + \gamma\rho_s} \right).$$

The details of the derivation are given in Appendix V-B. Due to the successive refinability of the Gaussian source, provided l layers have been successfully decoded, even in the presence of side information [24], the receiver reconstructs the source with a MMSE estimator using the side information and the decoded layers with a distortion given by $D_d(\sum_{i=1}^{l} bR_l/2L, \gamma)$. The expected distortion can be expressed as follows.

$$ED_{ls}(\mathbf{R}) = \sum_{l=0}^{L} \mathrm{E}_{(\mathcal{O}_l^{ls})^c \bigcap \mathcal{O}_{l+1}^{ls}} \left[D_d \left(\sum_{i=1}^{l} \frac{bR_l}{2L}, \gamma \right) \right]. \tag{19}$$

The distortion exponent achieved by LS-JDS is given next.

Theorem 6. *Let us define*

$$\phi_k \triangleq M^* - M_* + 2k - 1, \quad M_k \triangleq M_* - k + 1, \tag{20}$$

and the sequence $\{c_i\}$ *as*

$$c_0 = 0, \quad c_i = c_{i-1} + \phi_i \ln\left(\frac{M_i}{M_i - 1}\right),$$

for $i = 1, ..., M_* - 1$, *and* $c_{M_*} = \infty$.

The distortion exponent achieved by LS-JDS with infinite number of layers is given by $\Delta^*_{ls}(b, x) = x$ *if* $b \leq x/M_*$, *and if*

$$c_{k-1} + \frac{x}{M_k} < b \leq c_k + \frac{x}{M_k - 1},$$

for some $k \in \{1, ..., M_*\}$, *the achievable distortion exponent is given by*

$$\Delta^*_{ls}(b, x) = x + \sum_{i=1}^{k-1} \phi_i + M_k \phi_k \times \left(1 - e^{-\frac{b(1-\kappa^*) - c_{k-1}}{\phi_k}}\right),$$

where

$$\kappa^* = \frac{\phi_k}{b} \mathcal{W}\left(\frac{e^{\frac{b - c_{k-1}}{\phi_k}} x}{M_k \phi_k}\right),$$

and $\mathcal{W}(z)$ *is the Lambert W function, which gives the principal solution for* w *in* $z = we^w$.

Proof: See Appendix V-B. ■

The proof of Theorem 6 indicates that the distortion exponent for LS-JDS is achieved by allocating an equal rate among the first $\kappa^* L$ layers to guarantee that the distortion exponent is at least x. Then, the rest of the refinement layers, are used to further increase the distortion exponent with the corresponding rate allocation. Note that for $x = 0$, we have $\kappa^* = 0$, and Theorem 6 boils down to Theorem 4.2 in [9].

B. Broadcast strategy with JDS (BS-JDS)

In this section, we consider using the broadcast strategy in which the successive refinement layers are transmitted by superposition, and are decoded one by one with joint decoding. The receiver decodes as many layers as possible using successive joint decoding, and reconstructs the source sequence using the successfully decoded layers and the side information sequence.

At the encoder, we generate L Gaussian quantization codebooks, at rates $\frac{b}{2}R_l = I(X; W_l|W_1^{l-1}) + \epsilon/2$, $l = 1, ..., L$, $\epsilon > 0$, as in Section VI-A, and L channel codebooks $\mathbf{U}_l^n$, $l = 1, \ldots, L$, i.i.d. with $\mathbf{U}_{l,i} \sim \mathcal{CN}(0, \mathbf{I})$. Let $\boldsymbol{\rho} = [\rho_1, ..., \rho_L, \rho_{L+1}]^T$ be the power allocation among channel codebooks such that $\rho = \sum_{i=1}^{L+1} \rho_i$. We consider a power allocation strategy, such that $\rho_l = \rho^{\xi_{l-1}} - \rho^{\xi_l}$ with $1 = \xi_0 \geq \xi_1 \geq \ldots \geq \xi_L \geq 0$, and define $\boldsymbol{\xi} \triangleq [\xi_1, ..., \xi_L]$. In the last layer, the layer $L+1$, Gaussian i.i.d. noise sequence with distribution $\tilde{\mathbf{N}}_i \sim \mathcal{CN}(0, \mathbf{I})$ is transmitted using the remaining power $\rho_{L+1} \triangleq \rho^{\xi_L}$ for mathematical convenience. Then, the channel input $\mathbf{U}^n$ is generated as the superposition of the L codewords, $\mathbf{U}_l^n$ with the corresponding power allocation $\sqrt{\rho_l}$ as

$$\mathbf{U}^n = \frac{1}{\sqrt{\rho}} \sum_{j=1}^{L} \sqrt{\rho_j} \mathbf{U}_j^n + \sqrt{\rho^{\xi_L}} \tilde{\mathbf{N}}^n.$$

At the receiver, successive joint decoding is used from layer 1 up to layer L, considering the posterior layers as noise. Layer $L+1$, containing the noise, is ignored. The outage event at layer l, provided $l-1$ layers have been decoded successfully, is given by

$$\mathcal{O}_l^{bs} = \left\{ (\mathbf{H}, \gamma) : \frac{b}{2} I(\mathbf{U}_l; \mathbf{V}|\mathbf{U}_1^{l-1}) \leq I(X; W_l|Y, W_1^{l-1}) \right\}. \tag{21}$$

If l layers are decoded, the source is reconstructed at a distortion $D_d(\sum_{i=1}^{l} bR_i, \gamma)$ with an MMSE estimator, and the expected distortion is found as

$$ED_{bs}(\mathbf{R}, \boldsymbol{\xi}) = \sum_{l=1}^{L} \mathrm{E}_{\mathcal{O}_{k+1}^{bs}} \left[D_d \left(\sum_{i=0}^{l} \frac{b}{2} R_i, \Gamma \right) \right],$$

where $\mathbf{R} \triangleq [R_1, ..., R_L]$ and $\mathcal{O}_{L+1}^{bs}$ is the set of states in which all the L layers are successfully decoded.

The problem of optimizing the distortion exponent for BS-JDS for L layers, which we denote by

$\Delta_{bs}^{L}(b,x)$, can be formulated as a linear optimization program over the multiplexing gains $\mathbf{r} \triangleq [r_1, ..., r_l]$, where $R_l = r_l \log \rho$ for $l = 1, ..., L$, and the power allocation $\boldsymbol{\xi}$, as shown in (70) in Appendix VI-A, and can be efficiently solved numerically. In general, the performance of BS-JDS is improved by increasing the number of layers L, and an upper bound on the performance, denoted by $\Delta_{bs}^{*}(b,x)$, is given in the limit of infinite layers, i.e., $L \to \infty$, which can be approximated by numerically solving $\Delta_{bs}^{L}(b,x)$ with a large number of layers. However, obtaining a complete analytical characterization of $\Delta_{bs}^{L}(x,b)$ and $\Delta_{bs}^{*}(b,x)$ is in general complicated. In the following, we fix the multiplexing gains, and optimize the distortion exponent over the power allocation. While fixing the multiplexing gains is potentially suboptimal, we obtain a closed form expression for an achievable distortion exponent, and analytically evaluate its limiting behavior. We shall see that as the number of layers increases, this analytical solution matches the numerically optimized distortion exponent.

First, we fix the multiplexing gains as $\hat{\mathbf{r}} = [\hat{r}_1, ..., \hat{r}_L]$ where $\hat{r}_l = [(k+1)(\xi_{l-1} - \xi_l) - \epsilon_1]$ for $l = 1, ..., L$, for some $\epsilon_1 \to 0$, and optimize the distortion exponent over $\boldsymbol{\xi}$. The achievable distortion exponent is given in the next theorem.

Theorem 7. *Let us define*

$$\eta_k \triangleq \frac{b(k+1) - \Phi_{k+1}}{\Upsilon_k} \quad \text{and} \quad \Gamma_k \triangleq \frac{1 - \eta_k^{L-1}}{1 - \eta_k}. \tag{22}$$

The distortion exponent $\hat{\Delta}_{bs}^{L}(b,x)$ is achievable by BS-JDS with L layers and multiplexing gain $\hat{\mathbf{r}}$, and is given by $\hat{\Delta}_{bs}^{L}(b,x) = x$ for $bM_* \leq x$ and by

$$\hat{\Delta}_{bs}^{L}(b,x) = x + \Phi_k - \frac{\Upsilon_k(\Upsilon_k(x + \Phi_k) + xb(k+1)\Gamma_k)}{(\Upsilon_k + b(1+k))(\Upsilon_k + b(1+k)\Gamma_k) - b(k+1)\Phi_k\Gamma_k}, \tag{23}$$

for

$$b \in \left[\frac{\Phi_{k+1} + x}{k+1}, \frac{\Phi_k + x}{k}\right), \quad k = 0, ..., M_* - 1.$$

Proof: See Appendix VI. ■

An upper bound on the performance of BS-JDS with multiplexing gains $\hat{\mathbf{r}}_l$ is obtained for a continuum of infinite layers, i.e., $L \to \infty$.

Corollary 1. *The distortion exponent of BS-JDS with multiplexing gain* $\hat{\mathbf{r}}$ *in the limit of infinite layers,* $\hat{\Delta}_{bs}^{\infty}(b,x)$*, is found, for* $k=0,...,M_*-1$*, by*

$$\hat{\Delta}_{bs}^{\infty}(b,x)=\max\{x,b(k+1)\} \qquad \text{for } b\in\left[\frac{\Phi_{k+1}+x}{k+1},\frac{\Phi_k}{k+1}\right),$$

and

$$\hat{\Delta}_{bs}^{\infty}(b,x)=\Phi_k+x\left(\frac{b(1+k)-\Phi_k}{b(1+k)-\Phi_{k+1}}\right) \qquad \text{for } b\in\left[\frac{\Phi_k}{k+1},\frac{\Phi_k+x}{k}\right).$$

Proof: See Appendix VI-A. ■

The solution in Theorem 7 is obtained by fixing the multiplexing gains to $\hat{\mathbf{r}}$. This is potentially suboptimal since it excludes, for example, the performance of single-layer JDS from the set of feasible solutions. By fixing $\mathbf{r}$ such that $r_2=\cdots=r_L=0$, BS-JDS reduces to single layer JDS and achieves a distortion exponent given in Theorem 3, i.e., $\Delta_j(b,x)$. Interestingly, for b satisfying

$$b\in\left[\frac{\Phi_k}{k},\frac{\Phi_k+x}{k}\right), \qquad k=1,...,M_*-1,$$

single-layer JDS achieves a larger distortion exponent than $\hat{\Delta}_{bs}^{\infty}(b,x)$ in Corollary 1, as shown in Figure 4. Note that this region is empty for $x=0$, and thus, this phenomena does not appear in the absence of side information. The achievable distortion exponent for BS-JDS can be stated as follows.

Lemma 3. *BS-JDS achieves the distortion exponent*

$$\bar{\Delta}_{bs}(b,x)=\max\{\hat{\Delta}_{bs}^{\infty}(b,x),\Delta_j(b,x)\}.$$

Next, we consider the numerical optimization of the distortion exponent $\Delta_{bs}^{L}(b,x)$ and compare it with the distortion exponent achieved by fixing the multiplexing gain. In Figure 4 we show one instance of the numerical optimization of $\Delta_{bs}^{L}(b,s)$ for 3×2 MIMO and $x=0.5$, for $L=2$ and $L=500$ layers. We also include the distortion exponent achievable by single-layer JDS, i.e., when $L=1$, and the exponent achievable by BS-JDS with multiplexing gains $\hat{\mathbf{r}}$, with $L=2$ layers and in the limit of infinite layers, denoted by $\hat{\Delta}_{bs}^{2}(b,x)$ and $\hat{\Delta}_{bs}^{\infty}(b,x)$, respectively. We observe that the numerically optimized distortion exponent improves as the number of layers increases. There is a significant improvement in the distortion exponent just by using two layers in the high bandwidth regime, while this improvement is not so significant for intermediate b values. We also note that there is a tight match between the distortion exponent achievable by Lemma 3 and the one optimized numerically for $L = 500$ layers. For $L = 2$, we observe a tight match between $\hat{\Delta}^2_{bs}(b, x)$ and $\Delta^2_{bs}(b, x)$ in the high bandwidth regime. However, for intermediate bandwidth ratio values, $\hat{\Delta}^2_{bs}(b, x)$ is significantly worse than $\Delta^2_{bs}(b, x)$ and, in general, worse than $\Delta_j(b, x)$. Note that, as expected, if the power allocation and the multiplexing gains are jointly optimized, using two layers provides an improvement on the distortion exponent, i.e., $\Delta^2_{bs}(b, x)$ outperforms $\Delta_j(b, x)$. We also observe that $\hat{\Delta}^2_{bs}(b, x)$ and $\hat{\Delta}_{bs}(b, x)$ are discontinuous at $b = 2.5$, while this discontinuity is not present in the numerically optimized distortion exponents.

Our extensive numerical simulations suggest that, for b values satisfying (24), the performance of $\Delta^L_{bs}(b, x)$ reduces to the distortion exponent achievable by a single layer. We also observe that as the number of layers increases, the difference between $\Delta^L_{bs}(b, x)$ and $\hat{\Delta}^L_{bs}(b, x)$ is reduced, and that the distortion exponent achievable by BS-JDS as stated in Lemma 3, i.e., $\bar{\Delta}_{bs}(b, x)$, is indeed very close to the optimal performance that can be achieved by jointly optimizing the multiplexing gain and the power allocation. In the next section, we will see that in certain cases fixing the diversity multiplexing gain to $\hat{\boldsymbol{r}}$ suffices for BS-JDS to meet the partially informed upper bound in the MISO/SIMO setup, and therefore $\Delta^*_{bs}(b, x) = \hat{\Delta}^{\infty}_{bs}(b, x)$.

VII. Comparison of the Proposed Schemes and Discussion

In this section, we compare the performances of the proposed schemes with each other and with the derived upper bounds. First, we use the upper bound derived in Section IV to characterize the optimal distortion exponent for bandwidth ratio values that satisfy $0 \leq b \leq \max\{M^* - M_* + 1, x\}/M_*$. We show that, when the bandwidth ratio satisfies $0 \leq b \leq x/M_*$, then, the optimal distortion exponent is achieved by ignoring the channel and reconstructing the source sequence using only the side information. If $x/M_* \leq b \leq (M^* - M_* + 1)/M_*$, then the optimal distortion exponent is achieved by ignoring the side information, and employing the optimal transmission scheme in the absence of side information.

Then, we characterize the optimal distortion exponent for MISO/SIMO/SISO scenarios. In MISO/SIMO,

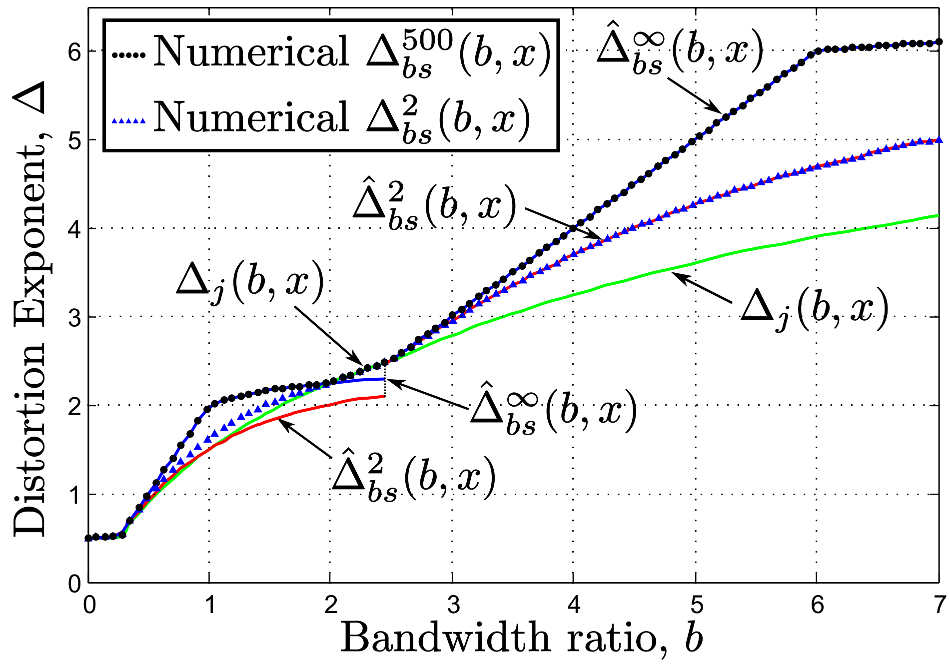


Fig. 4. Distortion exponent achieved by BS-JDS with $L = 1, 2$ and in the limit of infinite layers with respect to the bandwidth ratio b for a 3×2 MIMO system and a side information quality given by $x = 0.5$. Numerical results on the achievable distortion exponent for $L = 2$ and $L = 500$ are also included.

i.e., $M_* = 1$, we show that BS-JDS achieves the partially informed encoder upper bound, thus characterizing the optimal distortion exponent. This extends the result of [9] to the case with time-varying side information. For SISO, i.e., $M^* = M_* = 1$, HDA-WZ also achieves the optimal distortion exponent. For the general MIMO setup, the proposed schemes do not meet the upper bound for $b > 1/M_*$. Nevertheless, multi-layer transmission schemes perform close to the upper bound, especially in the high bandwidth ratio regime.

A. Optimal distortion exponent for low bandwidth ratios

First, we consider the MMSE reconstruction of X^m only from the side information sequence Y^m available at the receiver, i.e., $\hat{X}_i = \mathrm{E}[X_i|Y_i]$. The source sequence is reconstructed with distortion $D_{no}(\gamma) \triangleq (1 + \rho_s \gamma)^{-1}$, and averaging over the side information realizations, the expected distortion is given by $ED_{no} = \mathrm{E}[D_{no}(\Gamma)]$. The achievable distortion exponent is found as $\Delta_{no}(x, b) = x$, which meets the upper bound $\Delta_{up}(x, b)$ for $0 \leq b \leq x/M_*$, characterizing the optimal distortion exponent.

Lemma 4. *For $0 \leq b \leq x/M_*$, the optimal distortion exponent $\Delta^*(b,x) = x$ is achieved by simple MMSE reconstruction of X^m from the side information sequence Y^m.*

Additionally, Theorem 1 reveals that in certain regimes, the distortion exponent is upper bounded by $\Delta_{\text{MIMO}}(b)$, the distortion exponent upper bound in the absence of side information at the destination [9, Theorem 3.1]. In fact, for $x/M_* \leq b \leq M^* - M_* + 1$, we have $\Delta_{up}(x,b) = bM_*$. This distortion exponent is achievable for b satisfying $x/M_* \leq b \leq (M^* - M_* + 1)/M_*$ by ignoring the side information and employing the optimal scheme in the absence of side information, which is given by the multi-layer broadcast transmission scheme considered in [25]. The same distortion exponent is achievable by considering BS-JDS ignoring the side information, i.e., $\Delta^*(b,x) = \hat{\Delta}^L_{bs}(b,0)$. If $x/M_* \leq b \leq 1/M_*$, the optimal distortion exponent is also achievable by HDA-S and $\Delta^*(b,x) = \Delta_h(b,x)$.

Lemma 5. *For $x/M_* \leq b \leq (M^* - M_* + 1)/M_*$, the optimal distortion exponent is given by $\Delta^*(b,x) = bM_*$, and is achievable by ignoring the side information sequence Y^m and using BS-JDS. If $x/M_* \leq b \leq 1/M_*$ the distortion exponent is also achievable by HDA-S.*

In large bandwidth ratio regimes, i.e., for $b > (M^* - M_* + 1)/M_*$, transmission schemes exploiting both the channel output and the side information sequence are required.

B. Optimal distortion exponent for MISO/SIMO/SISO

We first particularize the upper bounds on the distortion exponent to $M_* = 1$. The fully informed encoder upper bound is found as

$$\Delta_{\text{inf}}(b,x) = x + \min\{b, M^*\},$$

and the partially informed encoder upper bound is given by

$$\Delta^*_{up}(b,x) = \begin{cases} \max\{x,b\} & \text{for } b \leq \max\{M^*, x\}, \\ M^* + x\left(1 - \frac{M^*}{b}\right) & \text{for } b > \max\{M^*, x\}. \end{cases}$$

Notice that as the bandwidth ratio increases, the partially informed encoder upper bound $\Delta^*_{up}(b,x)$

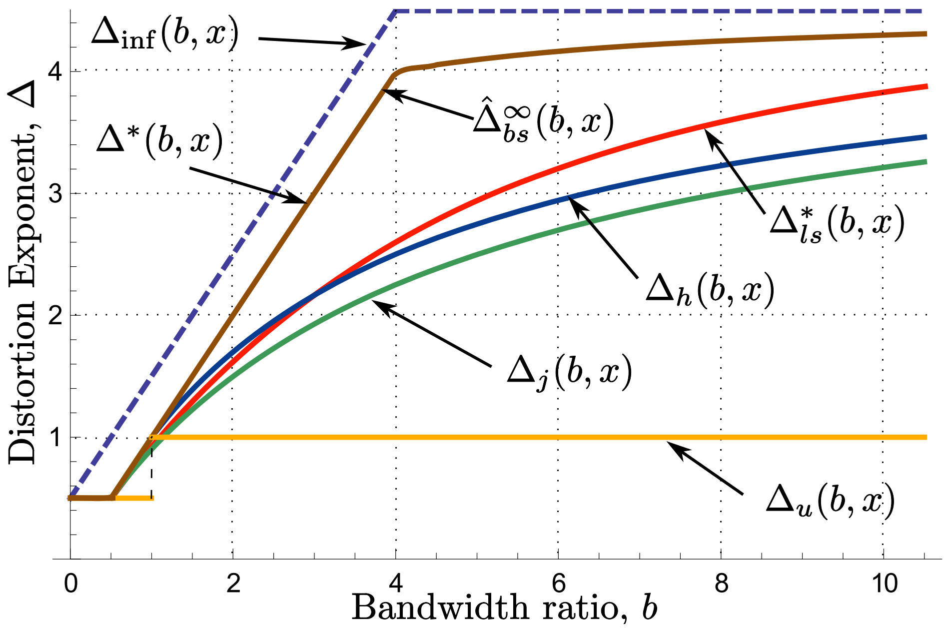


Fig. 5. Distortion exponent $\Delta(b,x)$ with respect to the bandwidth ratio b for a 4×1 MISO system and a side information quality given by $x=0.5$.

converges to the fully informed encoder upper bound $\Delta_{\text{inf}}(x,b)$, i.e., we have $\Delta_{\text{inf}}(\infty,x)=\Delta_{up}^*(\infty,x)=x+M_*$.

Now we particularize the proposed lower bounds to $M_*=1$. The distortion exponent of SSCC and JDS is given by

$$\Delta_j(b,x)=\max\left\{x, b\frac{x+M^*}{b+M^*}\right\},$$

while for uncoded transmission we have

$$\Delta_u(b,x)=\begin{cases} x & \text{if } b<1,\\ \max\{1,x\} & \text{if } b\geq 1.\end{cases}$$

Note that for $b=1$, uncoded transmission meets $\Delta_{up}^*(b,x)=\max\{1,x\}$, while SSCC and JDS are both suboptimal. This observation is valid for general MIMO channels, as well.

The following distortion exponent is achievable by HDA-S for $x\leq b\leq 1$, and by HDA-WZ for $b>1$,

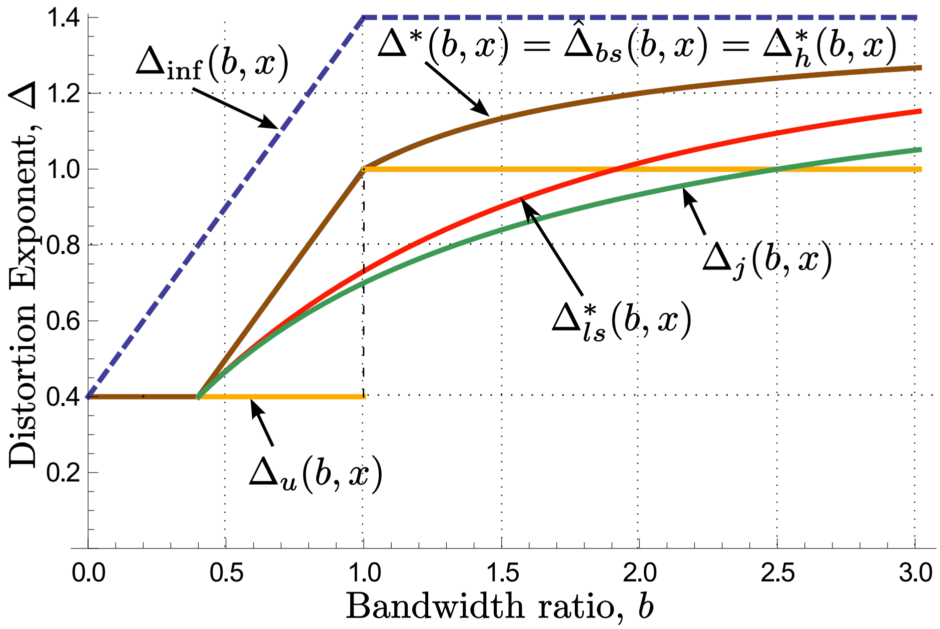


Fig. 6. Distortion exponent in SISO channels as a function of b for $x = 0.4$.

in the MISO/SIMO setup.

$$\Delta_h(b,x) = \begin{cases} \max\{x, b\} & \text{for } b \leq 1, \\ \max\{x, \frac{M^*+(b-1)(M^*+x)}{M^*+b-1}\} & \text{for } b > 1. \end{cases}$$

As seen in Section VII-A, HDA-S meets the partially informed upper bound for $b \leq 1$. HDA-WZ is in general suboptimal.

For the multi-layer transmission schemes, the distortion exponent acheivable by LS-JDS is given by

$$\Delta_{ls}^*(b,x) = x + M^*\left(1 - e^{-\frac{b(1-\kappa^*)}{M^*}}\right), \qquad \kappa^* = \frac{M^*}{b}\mathcal{W}\left(\frac{e^{\frac{b}{M^*}x}}{M^*}\right).$$

As for BS-JDS, considering the achievable rate in Corollary 1, this scheme meets the partially informed encoder lower bound in the limit of infinite layers, i.e., $\hat{\Delta}_{bs}^{\infty}(b,x) = \Delta_{up}^*(b,x)$. This fully characterizes the optimal distortion exponent in the MISO/SIMO setup, as stated in the next theorem.

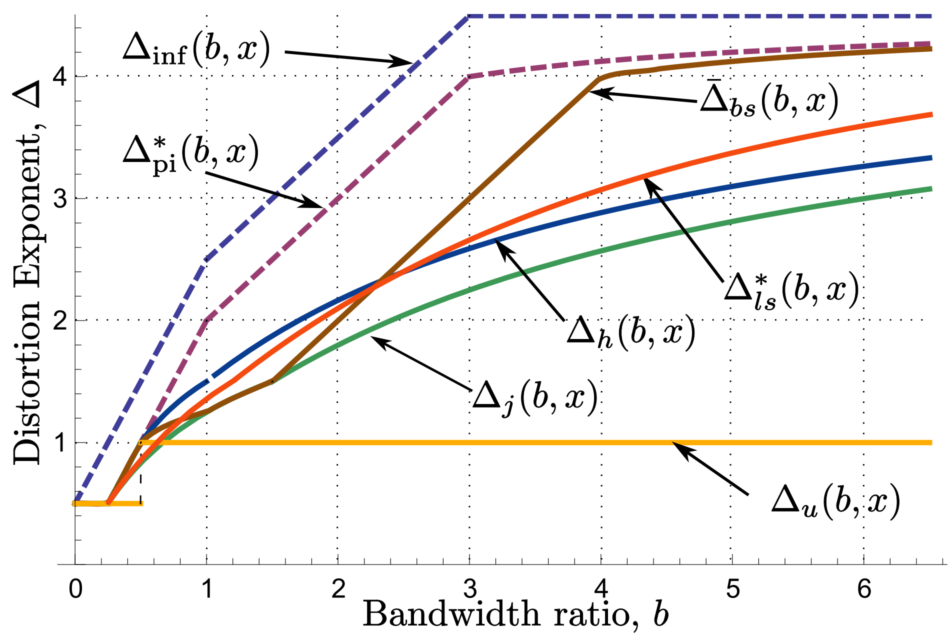


Fig. 7. Distortion exponent $\Delta(b,x)$ with respect to the bandwidth ratio b for a 2×2 MIMO system and a side information quality given by $x = 0.5$.

Theorem 8. *The optimal distortion exponent $\Delta^*(b,x)$ for MISO/SIMO systems is given by*

$$\Delta^*(b,x) = \begin{cases} \max\{x,b\} & \textit{for } b \leq \max\{M^*,x\}, \\ M^* + x\left(1 - \frac{M^*}{b}\right) & \textit{for } b > \max\{M^*,x\}, \end{cases}$$

and is achieved by BS-JDS in the limit of infinite layers.

In Figure 5 we plot the distortion exponent for a MISO/SIMO channel with $M^* = 4$ and $x = 0.5$, with respect to the bandwidth ratio b. We observe that, as given in Theorem 8, BS-JDS achieves the optimal distortion exponent. As discussed in Section V-E, single-layer schemes perform poorly as the bandwidth ratio increases. We observe that HDA-WZ outperforms JDS in all regimes, and, although it outperforms the multi-layer LS-JDS for low b values, LS-JDS achieves larger distortion exponents than HDA-WZ for $b \geq 3$.

In Figure 6, we plot the proposed upper and lower bounds on the distortion exponent for the SISO case and $x = 0.4$. We observe that the performance of the schemes is similar to the MISO/SIMO case. However, the distortion exponent values achievable by LS-JDS are lower than those achievable by

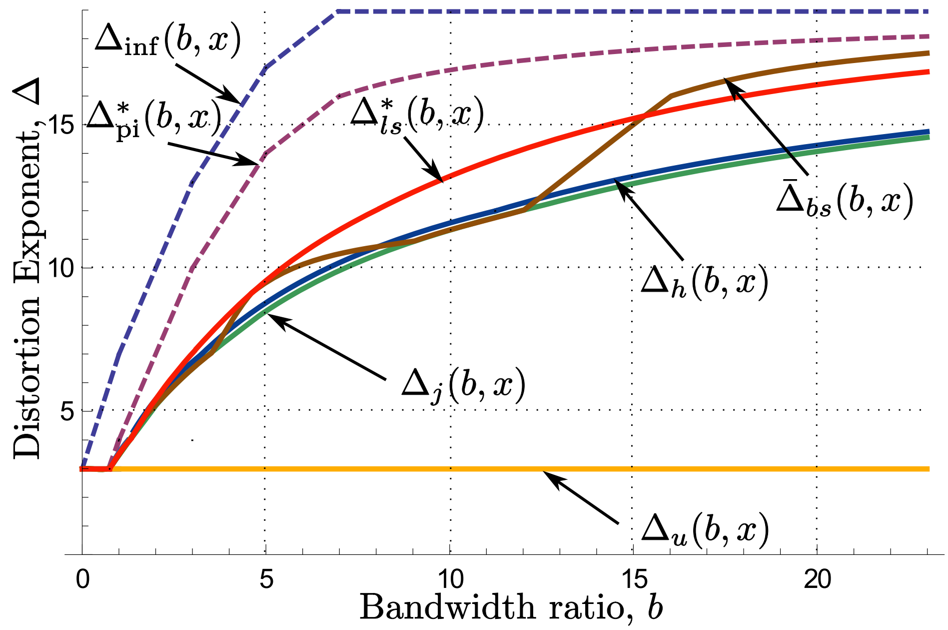


Fig. 8. Distortion exponent $\Delta(b,x)$ with respect to the bandwidth ratio b for a 4×4 MIMO system and a side information quality given by $x=3$.

HDA-WZ for all $b \geq 0.4$. HDA-WZ achieves the optimal distortion exponent for $b \geq 1$.

Lemma 6. *The optimal distortion exponent for SISO channels is achieved by BS-JDS, HDA-WZ and HDA-S.*

C. General MIMO

Here, we consider the general MIMO channel with $M_* > 1$. Figure 7 shows the upper and lower bounds on the distortion exponent derived in the previous sections for a 2×2 MIMO channel with $x = 0.5$. First, it is observed that the optimal distortion exponent is achieved by HDA-S and BS-JDS with infinite layers for $b \leq 0.5$, as expected from Section VII-A, while the other schemes are suboptimal in general. Uncoded transmission also achieves the optimal distortion exponent at $b = 0.5$. This holds for any MIMO system as stated in the following lemma.

Lemma 7. *Uncoded transmission achieves the optimal distortion exponent for* $b = 1/M_*$.

For $0.5 < b \lesssim 2.4$, HDA-WZ is the scheme achieving the largest distortion exponent, and outperforms

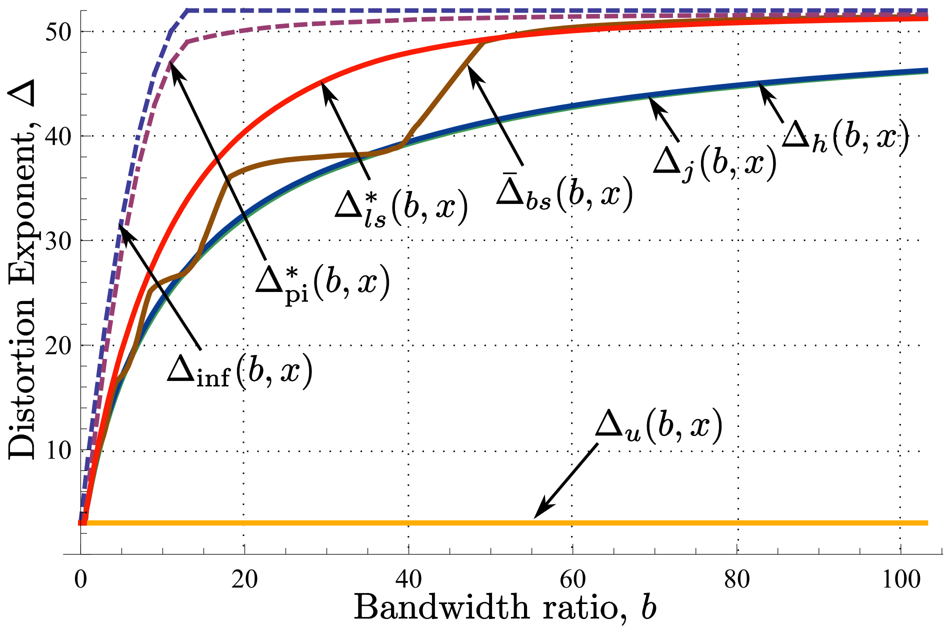


Fig. 9. Distortion exponent $\Delta(b, x)$ with respect to the bandwidth ratio b for a 4×4 MIMO system and a side information quality given by $x = 3$.

BS-JDS, and in particular, when the performance of BS-JDS reduces to the performance of JDS, since HDA-WZ outperforms JDS in general. For larger b values, the largest distortion exponent is achieved by BS-JDS. Note that for $b \geq 4$, $\Delta^*_{bs}(b, 0.5)$ is very close to the partially informed encoder lower bound. We also observe that for $b \gtrsim 2.4$ LS-JDS outperforms HDA-WZ, but it is worse than BS-JDS. This is not always the case, as will be seen next.

In Figure 8, we plot the upper and lower bounds proposed for a 4×4 MIMO channel with $x = 3$. We note that, for $b \leq \max\{1, x\}/M_*$, $\Delta^*(b, 3) = 3$, which is achievable by using only the side information sequence at the decoder. For this setup, LS-JDS achieves the best distortion exponent for intermediate b values, outperforming both HDA-WZ and BS-JDS. Again, in the large bandwidth ratio regime, BS-JDS achieves the best distortion exponent values, and performs close to the upper bound. We note that for high side information quality, the difference in performance between JDS and HDA-WZ decreases. This is in line with the observation that, for $x \geq 1/M_*$, uncoded transmission does not provide any distortion exponent improvement with respect to simple MMSE estimation form the side information sequence. Comparing Figure 7 and Figure 8, we observe that, when the side information quality is high, digital schemes better exploit the degrees-of-freedom of the system than analog schemes.

In Figure 9 we plot the upper and lower bounds for a 7×7 MIMO channel with $x = 3$. In comparison with Figure 8, as the number of antennas increases the difference in performance between JDS and HDA-WZ decreases. This seems to be the case also between BS-JDS and LS-JDS in the high bandwidth regime. However, LS-JDS significantly outperforms LS-JDS for intermediate b values. We also observe that the two proposed upper bounds get closer to each other as the minimum number of antennas M_* increases.

VIII. Conclusions

We have studied the high SNR distortion exponent when transmitting a Gaussian source over a time-varying fading MIMO channel in the presence of time-varying correlated side information at the receiver. We have assumed a block-fading model for both the channel and the side information states, and perfect state information about the time-varying channel and the side information states at the receiver, while the transmitter has only a statistical knowledge. We have derived two upper bounds on the distortion exponent, as well as lower bounds based on separate source and channel coding, joint decoding, uncoded transmission and hybrid digital-analog transmission. We have proposed multi-layer transmission schemes based on progressive transmission with joint decoding as well as superposition with joint decoding. We have considered the effects of the bandwidth ratio and the side information quality on the distortion exponent, and shown that the multi-layer transmission scheme with superposition transmission meets the upper bound in MISO/SIMO/SISO channels, solving the joint source channel coding problem in the high SNR regime. For general MIMO channels, we have characterized the optimal distortion exponent in the low bandwidth regime and shown that the multi-layer scheme based on superposition performs very close to the upper bound in the large bandwidth ratio regime.

Appendix I

Proof of Theorem 1

The exponential integral can be bounded as follows [26, p.229, 5.1.20]:

$$\frac{1}{2}\ln\left(1+\frac{2}{t}\right) < e^{t}E_1(t) < \ln\left(1+\frac{1}{t}\right), \quad t>0. \tag{24}$$

Next, using the lower bound $\ln(1+t) \geq \frac{t}{1+t}$, for $t > -1$, we have

$$\frac{1}{2}\ln\left(1+\frac{2}{t}\right) > \frac{1}{2}\frac{2/t}{1+2/t} = \frac{1}{t+2}. \tag{25}$$

Then, ED^*_{pi} in (6) is lower bounded by

$$ED^*_{\text{pi}}(\rho,\rho_s,b) \geq \int_{\mathbf{H}} \frac{1}{2^{\mathcal{C}(\mathbf{H})}+2\rho_s} p_h(\mathbf{H})d\mathbf{H}. \tag{26}$$

Following [21], the capacity of the MIMO channel is upper bounded as

$$\begin{aligned}\mathcal{C}(\mathbf{H}) &= \sup_{\mathbf{C}_u:\text{Tr}\{\mathbf{C}_u\}\leq M_t} \log\det\left(\mathbf{I}+\frac{\rho}{M_t}\mathbf{H}\mathbf{C}_u\mathbf{H}^H\right)\\ &\leq \log\det\left(\mathbf{I}+\rho\mathbf{H}\mathbf{H}^H\right),\end{aligned} \tag{27}$$

where the inequality follows from the fact that $M_t\mathbf{I}-\mathbf{C}_u \succeq 0$ subject to the power constraint $\text{Tr}\{\mathbf{C}_u\} \leq M_t$, and the function $\log\det(\cdot)$ is nondecreasing on the cone of positive semidefinite Hermitian matrices.

Let $\lambda_{M_*} \geq \cdots \geq \lambda_1 > 0$ be the eigenvalues of matrix $\mathbf{H}\mathbf{H}^H$, and consider the change of variables $\lambda_i = \rho^{-\alpha_i}$, with $\alpha_1 \geq ... \geq \alpha_{M_*} \geq 0$. The joint probability density function (pdf) of $\boldsymbol{\alpha} \triangleq [\alpha_1,...,\alpha_{M_*}]$ is given by [21]:

$$p_A(\boldsymbol{\alpha}) = K^{-1}_{M_t,M_r}(\log\rho)^{M_*}\prod_{i=1}^{M_*}\rho^{-(M^*-M_*+1)\alpha_i}\cdot\left[\prod_{i<j}(\rho^{\alpha_i}-\rho^{\alpha_j})^2\right]\exp\left(-\sum_{i=1}^{M_*}\rho^{\alpha_i}\right), \tag{28}$$

where $K^{-1}_{M_t,M_r}$ is a normalizing constant.

We define the high SNR exponent of $p_A(\boldsymbol{\alpha})$ as $S_A(\boldsymbol{\alpha})$, that is, we have $p_A(\boldsymbol{\alpha}) \doteq \rho^{-S_A(\boldsymbol{\alpha})}$, where

$$S_A(\boldsymbol{\alpha}) \triangleq \begin{cases}\sum_{i=1}^{M_*}(2i-1+M^*-M_*)\alpha_i & \text{if } \alpha_{M_*}\geq 0,\\ \infty & \text{otherwise.}\end{cases} \tag{29}$$

Then, from (26) and (27) we have

$$
\begin{aligned}
ED^*_{\text{pi}}(\rho,\rho_s,b) &\geq \int_{\mathbf{H}} \frac{1}{\prod_{i=1}^{M_*}\left(1+\rho\lambda_i\right)^b+2\rho_s} p_h(\mathbf{H})d\mathbf{H} \\
&= \int_{\boldsymbol{\alpha}} \frac{1}{\prod_{i=1}^{M_*}\left(1+\rho^{1-\alpha_i}\right)^b+2\rho_s} p_A(\boldsymbol{\alpha})d\boldsymbol{\alpha} \\
&\geq \int_{\boldsymbol{\alpha}^+} G_\rho(\boldsymbol{\alpha})p_A(\boldsymbol{\alpha})d\boldsymbol{\alpha}, \qquad (30)
\end{aligned}
$$

where we define

$$
G_\rho(\boldsymbol{\alpha}) \triangleq \left(\prod_{i=1}^{M_*}\left(1+\rho^{1-\alpha_i}\right)^b + 2\rho_s\right)^{-1},
$$

and the set $\boldsymbol{\alpha}^+ \triangleq \{\boldsymbol{\alpha} \in \mathbb{R}^{M_*} : 1 \geq \alpha_1 \geq ... \geq \alpha_{M_*} \geq 0\}$ in (30).

Then, in the high SNR regime we have,

$$
\begin{aligned}
G(\boldsymbol{\alpha}) &\triangleq \lim_{\rho\to\infty} \frac{\log G_\rho(\boldsymbol{\alpha})}{\log\rho} = \lim_{\rho\to\infty} \frac{\log(\rho^{b\sum_{i=1}^{M_*}(1-\alpha_i)^+}+2\rho^x)^{-1}}{\log\rho} \\
&= \begin{cases} -x & \text{if } x > b\sum_{i=1}^{M_*}(1-\alpha_i)^+, \\ -b\sum_{i=1}^{M_*}(1-\alpha_i)^+ & \text{if } x \leq b\sum_{i=1}^{M_*}(1-\alpha_i)^+, \end{cases}
\end{aligned}
$$

where we have used the exponential equalities $1+\rho^{1-\alpha_i} \doteq \rho^{(1-\alpha_i)^+}$, and $\rho_s \doteq \rho^x$.

Therefore, for sufficiently large ρ, we have

$$
\begin{aligned}
ED^*_{\text{pi}}(\rho,\rho_s,b) &\geq \int_{\boldsymbol{\alpha}^+} \exp\left(\frac{\log G_\rho(\boldsymbol{\alpha})}{\log\rho}\log\rho\right) p_A(\boldsymbol{\alpha})d\boldsymbol{\alpha} \\
&\doteq \int_{\boldsymbol{\alpha}^+} \exp\left(G(\boldsymbol{\alpha})\log\rho\right) p_A(\boldsymbol{\alpha})d\boldsymbol{\alpha}.
\end{aligned}
$$

Defining $\Delta^*_{pi}(b,x) = -\lim_{\rho\to\infty}\frac{\log ED^*_{\text{pi}}}{\log\rho}$, the distortion exponent of the partially informed encoder is upper bounded by

$$
\Delta^*_{pi}(b,x) \leq \lim_{\rho\to\infty}\frac{1}{\log\rho}\log\int_{\boldsymbol{\alpha}^+}\exp\left(G(\boldsymbol{\alpha})\log\rho\right)p_A(\boldsymbol{\alpha})d\boldsymbol{\alpha}.
$$

From Varadhan's lemma [27], it follows that the distortion exponent of ED^*_{pi} is upper bounded by the solution to the following optimization problem,

$$\Delta_{up}(b,x) \triangleq \inf_{\boldsymbol{\alpha}^+}[-G(\boldsymbol{\alpha})+S_A(\boldsymbol{\alpha})]. \tag{31}$$

In order to solve (31) we divide the optimization into two subproblems: the case when $x < b\sum_{i=1}^{M_*}(1-\alpha_i)$, and the case when $x \geq b\sum_{i=1}^{M_*}(1-\alpha_i)$. The solution is then given by the minimum of the solutions of these subproblems.

If $x \geq b\sum_{i=1}^{M_*}(1-\alpha_i)$, the problem in (31) reduces to

$$\begin{aligned}\Delta_{up}^1(b,x) = x + &\inf_{\boldsymbol{\alpha}^+} \sum_{i=1}^{M_*}(2i-1+M^*-M_*)\alpha_i \\ &\text{s.t.} \sum_{i=1}^{M_*}(1-\alpha_i) \leq \frac{x}{b}.\end{aligned} \tag{32}$$

The optimization in (32) can be identified with the DMT problem in (1) for a multiplexing gain of $r = \frac{x}{b}$. Next, we give an explicit solution for completeness.

First, if $bM_* \leq x$, the infimum is given by $\Delta_{up}^1(b,x) = x$ for $\boldsymbol{\alpha}^* = 0$. Then, for $k \leq \frac{x}{b} \leq k+1$, for $k = 0, ..., M_*-1$, i.e., $\frac{x}{k+1} \leq b \leq \frac{x}{k}$, the infimum is achieved by

$$\alpha_i^* = \begin{cases} 1 & \text{for } i = 1, ..., M_*-k-1, \\ k+1-\frac{x}{b} & i = M_*-k, \\ 0 & \text{for } i = M_*-k+1, ..., M_*. \end{cases}$$

Substituting, we have, for $k = 0, ..., M_*-1$,

$$\Delta_{up}^1(b,x) = x + \Phi_k - \Upsilon_k\left(\frac{x}{b}-k\right) = x + d^*\left(\frac{x}{b}\right),$$

where Φ_k and Υ_k are defined as in (3).

Now we solve the second subproblem with $x < b\sum_{i=1}^{M_*}(1-\alpha_i)$. Since $1 \geq \alpha_1 \geq ... \geq \alpha_{M_*} \geq 0$ we can rewrite (31) as

$$\Delta_{up}^2(b,x) = \inf_{\boldsymbol{\alpha}^+} bM_* - \sum_{i=1}^{M_*} \alpha_i \phi(i)$$
$$\text{s.t. } \sum_{i=1}^{M_*} \alpha_i < M_* - \frac{x}{b}, \tag{33}$$

where we have defined $\phi(i) \triangleq [b - (2i - 1 + M^* - M_*)]$. Note that $\phi(1) > \cdots > \phi(M_*)$.

First, we note that for $bM_* < x$ there is no feasible solution due to the constraint in (33).

Now, we consider the case $x \leq M_*(1 + M^* - M_*)$. If $\frac{x}{M_*} \leq b < 1 + M^* - M_*$, all the terms $\phi(i)$ multiplying α_i's are negative, and, thus, the infimum is achieved by $\boldsymbol{\alpha}^* = 0$, and is given by $\Delta_{up}^2(b,x) = bM_*$. If $1 + M^* - M_* \leq b < 3 + M^* - M_*$, then $\phi(1)$ multiplying α_1 is positive, while the other $\phi(i)$ terms are negative. Then $\alpha_i^* = 0$ for $i = 2, ..., M_*$. From (33) we have $\alpha_1 \leq M_* - \frac{x}{b}$. If $b \geq \frac{x}{M_*-1}$, the right hand side (r.h.s.) of (33) is greater than one, and smaller otherwise. Then, we have

$$\alpha_1^* = \begin{cases} 1 & \text{if } b \geq \frac{x}{M_*-1}, \\ M_* - \frac{x}{b} & \text{if } b < \frac{x}{M_*-1}. \end{cases}$$

Note that $\alpha_1^* \geq 0$ since $b > \frac{x}{M_*}$.

When $2k - 1 + M^* - M_* \leq b < 2k + 1 + M^* - M_*$ for $k = 2, ..., M_* - 1$, the coefficients $\phi(i)$, $i = 1, ..., k$, associated with the first k α_i terms are positive, while the others remain negative. Then,

$$\alpha_i^* = 0, \quad \text{for } i = k+1, ..., M_*. \tag{34}$$

Since $\phi(i)$, $i = 1, ..., k$, are positive and $\phi(1) > \cdots > \phi(k)$, we have $\alpha_i^* = 1$ for $i = 1, ..., k-1$, and the constraint becomes $\alpha_k < M_* - (k-1) - \frac{x}{b}$. If $b \geq \frac{x}{M_*-k}$, then the r.h.s. is greater than one, and smaller otherwise. In order for the solution to be feasible, we need $\alpha_k \geq 0$, that is, $M_* - (k-1) - \frac{x}{b} \geq 0$. Then we have

$$\alpha_k^* = \begin{cases} 1 & \text{if } b \geq \frac{x}{M_*-k}, \\ M_* - (k-1) - \frac{x}{b} & \text{if } \frac{x}{M_*-(k-1)} \leq b < \frac{x}{M_*-k}. \end{cases} \tag{35}$$

If $b < \frac{x}{M_* - (k-1)}$, the solution in (35) is not feasible. Instead, we have $\alpha_k^* = 0$, since $\phi(k) < \phi(k-1)$, $\alpha_i^* = 0$ for $i = k+1, ..., M_*$, and $\alpha_i^* = 1$, for $i = 1, ..., k-2$. Then, the constraint in (33) is given by $\alpha_{k-1} \leq M_* - (k-2) - \frac{x}{b}$. Since $b < \frac{x}{M_* - (k-1)}$, the r.h.s. is always smaller than one. For the existence of a feasible solution, the r.h.s. is required to be greater than zero. Therefore, we have

$$\alpha_{k-1}^* = M_* - (k-2) - \frac{x}{b}, \qquad \text{if } \frac{x}{M_* - (k-2)} \leq b < \frac{x}{M_* - (k-1)}.$$

In general, iterating this procedure, for

$$\frac{x}{M_* - (j-1)} \leq b < \frac{x}{M_* - j}, \quad j = 1, ..., k,$$

we have

$$\alpha_i^* = \begin{cases} 1 & \text{for } i = 1, ..., j-1, \\ M_* - (j-1) - \frac{x}{b} & \text{for } i = j, \\ 0 & \text{for } i = j+1, ..., M_*. \end{cases} \tag{36}$$

Note that for the case $j = 1$, we have $\alpha_1 = M_* - \frac{x}{b}$, which is always feasible.

We now evaluate (33) with the optimal $\boldsymbol{\alpha}^*$ if $2k - 1 + M^* - M_* \leq b < 2k + 1 + M^* - M_*$ for some $k \in \{2, ..., M_* - 1\}$. For $b \geq \frac{x}{M_* - k}$, we have $\alpha_1 = \cdots = \alpha_k = 1$ and $\alpha_{k+1} = \cdots = \alpha_{M_*} = 0$, and then

$$\Delta_{up}^2(b, x) = \sum_{i=1}^{M_*} \min\{b, 2i - 1 + M^* - M_*\} = \Delta_{\text{MIMO}}(b).$$

For $\frac{x}{M_*} \leq b \leq \frac{x}{M_* - k}$, substituting (36) into (33) we have

$$\begin{aligned} \Delta_{up}^2(b, x) = {} & x + (M^* - M_* - 1 + j)(j-1) \\ & + \left(M_* - (j-1) - \frac{x}{b}\right)(2j - 1 + M^* - M_*), \end{aligned}$$

where

$$\frac{x}{M_* - (j-1)} \leq b \leq \frac{x}{M_* - j}, \quad \text{for some } j \in \{1, ..., k\}.$$

Note that with the change of index $j = M_* - j'$, we have, after some manipulation,

$$\Delta_{up}^2(b, x) = x + (M^* - j')(M_* - j') - \left(\frac{x}{b} - j'\right)(M^* + M_* - 2j' - 1),$$

in the regime

$$\frac{x}{j'+1} \leq b < \frac{x}{j'}, \quad j' = M_* - k, ..., M_* - 1.$$

This is equivalent to the value of the DMT curve in (2) at multiplexing gain $r = \frac{x}{b}$. Then, for $\frac{x}{M_*} \leq b < \frac{x}{M_* - k}$ we have

$$\Delta_{up}^2(b, x) = x + d^*\left(\frac{x}{b}\right).$$

If $b \geq M^* + M_* - 1$, the infimum is achieved by $\alpha_i^* = 1$, for $i = 1, ..., M_* - 1$, and $\alpha_{M_*}^* = 1 - \frac{x}{b}$ if $b \geq x$. If $b < x$, this solution is not feasible, and the solution is given by (36). Therefore, in this regime we also have

$$\Delta_{up}^2(b, x) = x + d^*\left(\frac{x}{b}\right).$$

Putting all these results together, for $x \leq M_*(M^* - M_* + 1)$ we have

$$\Delta_{up}^2(x, b) = \begin{cases} bM_* & \text{for } \frac{x}{M_*} \leq b < M^* - M_* + 1, \\ x + d^*\left(\frac{x}{b}\right) & \text{for } M^* - M_* + 1 \leq b < \frac{x}{M_* - k}, \\ \Delta_{\text{MIMO}}(b) & \text{for } \frac{x}{M_* - k} \leq b < M^* + M_* - 1, \\ x + d^*\left(\frac{x}{b}\right) & \text{for } b \geq M^* + M_* - 1, \end{cases}$$

where $k \in \{1, ..., M_* - 1\}$ is the integer satisfying $2k - 1 + M^* - M_* \leq b < 2k + 1 + M^* - M_*$.

Now, we solve (33) for $M_*(M^* - M_* + 1) \leq x < M_*(M^* + M_* - 1)$. Let $l \in \{2, ..., M_*\}$ be the integer satisfying $M_*(2(l-1) - 1 + M^* - M_*) \leq x < M_*(2l - 1 + M^* - M_*)$. The first interval of b in which a feasible solution exists is given by $\frac{x}{M_*} \leq b < 2l - 1 + M^* - M_*$. From the sign of the coefficients $\phi(i)$ in this interval we have $\alpha_i^* = 0$ for $i = (l+1), ..., M_*$, and $\alpha_i^* = 1$ for $i = 1, ..., l - 1$.

Substituting, the constraint becomes $\alpha_l < M_* - (l-1) - \frac{x}{b}$. If $b > \frac{x}{M_*-l}$ the r.h.s. is larger than one, and $\alpha_l^* = 1$. On the contrary, if $b \leq \frac{x}{M_*-l}$, it is given by $\alpha_l^* = M_* - (l-1) - \frac{x}{b}$ if $b > \frac{x}{M_*-(l-1)}$, so that the r.h.s. of the constraint is larger than zero. Iterating this procedure, the solution for all b values is found following the techniques that lead to (36). In general, for $2k-1+M^*-M_* \leq b < 2k+1+M^*-M_*$, $k = l, ..., M_*-1$ and

$$\frac{x}{M_*-(j-1)} \leq b < \frac{x}{M_*-j}, \quad j = 1, ..., k,$$

we have

$$\alpha_i^* = \begin{cases} 1 & \text{for } i = 1, ..., j-1, \\ M_* - (j-1) - \frac{x}{b} & \text{for } i = j, \\ 0 & \text{for } i = j+1, ..., M_*. \end{cases} \tag{37}$$

The distortion exponent is now obtained similarly to the case $x \leq M_*(M^*-M_*+1)$ in each interval $2k-1+M^*-M_* \leq b < 2k+1+M^*-M_*$ with $k = l, ..., M_*-1$ instead of $k = 1, ..., M_*-1$, and thus, we omit the details. Putting all together, if x satisfies $M_*(2(l-1)-1+M^*-M_*) \leq x < M_*(2l-1+M^*-M_*)$, for some $l \in \{2, ..., M_*\}$, we have

$$\Delta_{up}^2(x,b) = \begin{cases} x + d^*\left(\frac{x}{b}\right) & \text{for } \frac{x}{M_*} \leq b < 2l-1+M^*-M_*, \\ x + d^*\left(\frac{x}{b}\right) & \text{for } 2l-1+M^*-M_* \leq b < \frac{x}{M_*-k}, \\ \Delta_{\text{MIMO}}(b) & \text{for } \frac{x}{M_*-k} \leq b < M^*+M_*-1, \\ x + d^*\left(\frac{x}{b}\right) & \text{for } b \geq M^*+M_*-1. \end{cases}$$

Note that in the case $l = M_*$, we have $\Delta_{up}^2(x,b) = x + d^*\left(\frac{x}{b}\right)$ for any b value.

Finally, the case $x \geq M_*(M^*+M_*-1)$ can be solved simil. Notice that if $\alpha_i^* = 1$, $i = 1, ..., M_*-1$ we have the constraint $\alpha_{M_*} \leq 1 - \frac{x}{b}$, that is, we never have the case $\alpha_{M_*}^* = 1$. Then, the optimal α_i^* are given as in (36), and we have

$$\Delta_{up}^2(x,b) = x + d^*\left(\frac{x}{b}\right) \quad \text{for } \frac{x}{M_*} \leq b.$$

Now, $\Delta_{up}(b,x)$ is given by the minimum of $\Delta^1_{up}(b,x)$ and $\Delta^2_{up}(b,x)$. First, we note that $\Delta_{2up}(b,x)$ has no feasible solution for $bM_* \leq x$, and we have $\Delta_{up}(b,x) = \Delta^1_{up}(b,x) = x$ in this region. For $bM_* > x$, both solutions $\Delta^1_{up}(b,x)$ and $\Delta^2_{up}(b,x)$ coincide except in the range $\frac{x}{M_*-k} \leq b \leq M^* + M_* - 1$. We note that $\Delta^1_{up}(b,x)$ in (32) is linear and increasing in $\boldsymbol{\alpha}$, and hence, the solution is such that the constraint is satisfied with equality, i.e., $x = \sum_{i=1}^{M_*} b(1-\alpha_i)$. That is, $\Delta^2_{up}(b,x) \leq \Delta^1_{up}(b,x)$ whenever both solutions exist in the same $\boldsymbol{\alpha}$ region. Then, the minimizing $\boldsymbol{\alpha}$ will be one such that either $\Delta^1_{up}(b,x) < \Delta^2_{up}(b,x)$, or the one arbitrarily close to the boundary $x = b\sum_{i=1}^{M_*}(1-\alpha_i)^+$, where $\Delta^1_{up}(b,x) = \Delta^2_{up}(b,x)$. Consequently, $\min\{\Delta^1_{up}(b,x), \Delta^2_{up}(b,x)\} = \Delta^1_{up}(b,x)$, whenever they are defined in the same region. Putting all the results together we complete the proof.

Appendix II

Proof of Theorem 2

To derive the distortion exponent of SSCC we first study the exponential behavior of $E_\Gamma[D_d(R,\Gamma)]$ in (9). We consider the change of variables $\gamma = \rho^{-\beta}$, with pdf $p_B(\beta)$ given as in (28) and $S_B(\beta) = \beta$, for $\beta \geq 0$, and $R = r\log\rho$. Then,

$$\begin{aligned} E_\Gamma[D_d(R,\Gamma)] &= \int \frac{1}{\gamma\rho_s + 2^{2R}} p_\Gamma(\gamma) d\gamma \\ &= \int \exp(\log(\rho^{x-\beta} + \rho^{2r})^{-1}) p_B(\beta) d\beta. \end{aligned}$$

In the high SNR regime, we have

$$E_\Gamma[D_d(r\log\rho,\Gamma)] \doteq \int_{\mathbb{R}} \rho^{-\max\{(x-\beta)^+, 2r\}} p_B(\beta) d\beta,$$

where we have used $(1 + \rho^{x-\beta} + \rho^{2r})^{-1} \doteq \rho^{-\max\{(x-\beta)^+, 2r\}}$. Applying Varadhan's lemma we have

$$E_\Gamma[D_d(R,\Gamma)] \doteq \inf_{\beta\in\mathbb{R}^+} \max\{(x-\beta)^+, 2r\} + \beta = \max\{x, 2r\}.$$

Then, for a family of codes with rate $\frac{b}{2}R_c = \frac{b}{2}r_c \log \rho$, (9) is exponentially equivalent to

$$\begin{aligned}ED_s(br_c \log \rho) &= (1 - P_o(\mathbf{H}))\mathrm{E}_\Gamma[D_d(br_c/2 \log \rho, \Gamma)] + P_o(\mathbf{H})\mathrm{E}_\Gamma[D_d(0, \Gamma)] \\ &\doteq (1 - \rho^{-d^*(r_c)})\rho^{-\max\{x, br_c\}} + \rho^{-d^*(r_c)}\rho^{-x} \\ &\doteq \rho^{-\max\{x, br_c\}} + \rho^{-(d^*(r_c)+x)} \\ &\doteq \rho^{-\min\{\max\{x, br_c\}, d^*(r_c)+x\}},\end{aligned}$$

where we have used that the outage probability is exponentially equivalent to the probability of error [21], i.e., $P_o(\mathbf{H}) \doteq \rho^{-d^*(r_c)}$, and $d^*(r_c)$ is the DMT curve characterized in (2).

The best distortion exponent achievable by SSCC, $\Delta_s(b, x)$, is found by maximizing over r_c as follows

$$\Delta_s(b, x) \triangleq \max_{r_c \geq 0}\{\min\{\max\{x, br_c\}, x + d^*(r_c)\}\}. \tag{38}$$

The maximum achieved when the two terms inside $\min\{\cdot\}$ are equal, i.e., $\max\{br_c, x\} = x + d^*(r_c)$. We chose a rate r_c such that $br_c > x$ and $r_c < M_*$, as otherwise, the solution is readily given by $\Delta_s(b, x) = x$. Note that for $bM_* \leq x$ this is never feasible, and thus, $\Delta_s(b, x) = x$, and if $x \geq b \cdot d^*(M_*)$, the intersection is always at $br_c = x$. Assuming $k \leq r_c \leq k+1$, $k = 0, ..., M_*-1$, the optimal r_c satisfies at $br_c = d^*(r_c) + x$, or, equivalently, $br_c = x + \Phi_k - (r_c - k)\Upsilon_k$, and we have

$$r_c^* = \frac{\Phi_k + k\Upsilon_k + x}{\Upsilon_k + b}, \quad \Delta_s(b, x) = br_c^* = b\frac{\Phi_k + k\Upsilon_k + x}{\Upsilon_k + b}.$$

Since solution r_c^* is feasible whenever $k < r_c^* \leq k + 1$, this solution is defined in

$$b \in \left[\frac{\Phi_{k+1} + x}{k+1}, \frac{\Phi_k + x}{k}\right), \quad \text{for } k = 0, ..., M_* - 1, \tag{39}$$

where we have used $\Phi_{k+1} = \Phi_k - \Upsilon_k$. Notice that, whenever $\Delta_s(b, x) \leq x$ in (39), we have $br_c^* \leq x$, which is not feasible, and therefore $\Delta_s(b, x) = x$. Remember that for $bM_* \leq x$ we also have $\Delta_s(b, x) = x$. Putting all these cases together completes the proof of Theorem 2.

APPENDIX III

PROOF OF THEOREM 3

Applying the change of variables $\lambda_i = \rho^{-\alpha_i}$ and $\gamma = \rho^{-\beta}$, and considering a rate $R_j = r_j \log \rho$, $r_j > 0$, the outage event in (10) can be written as

$$\begin{aligned}\mathcal{O}_j &= \left\{ (\mathbf{H}, \gamma) : 1 + \frac{2^{-\epsilon}\rho^{br_j} - 1}{\gamma\rho^x + 1} \geq \prod_{i=1}^{M_*} (1 + \rho\lambda_i)^b \right\} \\ &= \left\{ (\boldsymbol{\alpha}, \beta) : 1 + \frac{2^{-\epsilon}\rho^{br_j} - 1}{\rho^{(x-\beta)} + 1} \geq \prod_{i=1}^{M_*} (1 + \rho^{1-\alpha_i})^b \right\}.\end{aligned}$$

For large ρ, we have

$$\frac{1 + \frac{2^{-\epsilon}\rho^{br_j} - 1}{\rho^{(x-\beta)}+1}}{\prod_{i=1}^{M_*}(1+\rho^{1-\alpha_i})^b} \doteq \frac{1 + \rho^{br_j}\rho^{-(x-\beta)^+}}{\rho^{b\sum_{i=1}^{M_*}(1-\alpha_i)^+}} \doteq \rho^{(br_j - (x-\beta)^+)^+ - b\sum_{i=1}^{M_*}(1-\alpha_i)^+}.$$

Therefore, at high SNR, the achievable expected end-to-end distortion for JDS is found as,

$$\begin{aligned} ED_j(br_j \log\rho) &= \int_{\mathcal{O}_j^c} D_d(br_j/2\log\rho, \rho^{-\beta}) p_A(\boldsymbol{\alpha}) p_B(\beta) d\boldsymbol{\alpha} d\beta \\ &\quad + \int_{\mathcal{O}_j} D_d(0, \rho^{-\beta}) p_A(\boldsymbol{\alpha}) p_B(\beta) d\boldsymbol{\alpha} d\beta \\ &\doteq \int_{\mathcal{A}_j^c} \rho^{-\max\{(x-\beta)^+, br_j\}} \rho^{-(S(\boldsymbol{\alpha})+\beta)} d\boldsymbol{\alpha} d\beta \\ &\quad + \int_{\mathcal{A}_j} \rho^{-(x-\beta)^+} \rho^{-(S(\boldsymbol{\alpha})+\beta)} d\boldsymbol{\alpha} d\beta. \\ &\doteq \rho^{-\Delta_j^1(r_j)} + \rho^{-\Delta_j^2(r_j)} \\ &\doteq \rho^{-\min\{\Delta_j^1(r_j), \Delta_j^2(r_j)\}} \\ &\doteq \rho^{-\Delta_j(r_j)}, \end{aligned} \tag{40}$$

where $D_d(R, \gamma)$ is as defined in (8), and we have used $D_d(r\log\rho, \beta) \doteq \rho^{-\max\{(x-\beta)^+, 2r\}}$. We have also defined the high SNR equivalent of the outage event as

$$\mathcal{A}_j \triangleq \left\{ (\boldsymbol{\alpha}, \beta) : (br_j - (x-\beta)^+)^+ \geq b\sum_{i=1}^{M_*}(1-\alpha_i)^+ \right\}.$$

We have applied Varadhan's lemma to each integral to obtain

$$\Delta_j^1(r_j) \triangleq \inf_{\mathcal{A}_j^c} \max\{(x-\beta)^+, br_j\} + \beta + S_A(\boldsymbol{\alpha}), \tag{41}$$

and

$$\Delta_j^2(r_j) \triangleq \inf_{\mathcal{A}_j} (x-\beta)^+ + \beta + S_A(\boldsymbol{\alpha}). \tag{42}$$

Then, the distortion exponent of JDS is found as

$$\Delta_j(r_j) = \min\{\Delta_j^1(r_j), \Delta_j^2(r_j)\}. \tag{43}$$

We first solve (41). We can constrain the optimization to $\boldsymbol{\alpha} \geq 0$ and $\beta \geq 0$ without loss of optimality, since for $\boldsymbol{\alpha}, \beta < 0$ we have $S_A(\boldsymbol{\alpha}) = S_B(\beta) = +\infty$. Then, $\Delta_j^1(r_j)$ is minimized by $\boldsymbol{\alpha}^* = 0$ since this minimizes $S_A(\boldsymbol{\alpha})$ and enlarges $\mathcal{A}_j^c$. We can rewrite (41) as

$$\begin{aligned} \Delta_j^1(r_j) = \inf_{\beta \geq 0} & \max\{(x-\beta)^+, br_j\} + \beta \\ & \text{s.t. } (br_j - (x-\beta)^+)^+ < bM_*. \end{aligned}$$

If $br_j < (x-\beta)^+$, the minimum is achieved by any $0 \leq \beta < x - r_j b$, and thus $\Delta_j^1(r_j) = x$ for $x > br_j$. If $br_j \geq (x-\beta)^+$, then

$$\begin{aligned} \Delta_j^1(r_j) = \inf_{\beta \geq 0} & \, br_j + \beta \\ & \text{s.t. } br_j - bM_* < (x-\beta)^+ \leq br_j. \end{aligned}$$

If $\beta > x$, the problem is minimized by $\beta^* = x + \epsilon$, $\epsilon > 0$, and $\Delta_j(r_j) = br_j + x + \epsilon$, for $r_j \leq M_*$. For $0 \leq \beta \leq x$, we have $\beta^* = (x - r_j b)^+$, and $\Delta_j^1(r_j) = \max\{br_j, x\}$ if $br_j \leq bM_* + x$. Putting all these together, we obtain

$$\Delta_j^1(r_j) = \max\{br_j, x\} \quad \text{if } br_j \leq x + bM_*. \tag{44}$$

If $br_j > x + bM_*$, $\mathcal{A}_j^c$ is empty, and there is always outage.

Next we solve the second optimization problem in (42). With $\beta = x$, $\Delta_j^2(r_j)$ is minimized and the range of $\boldsymbol{\alpha}$ is enlarged. Then, the problem to solve reduces to

$$\begin{aligned} \Delta_j^2(r_j) &= \inf x + S(\boldsymbol{\alpha}) \\ \text{s.t. } r_j &\geq \sum_{i=1}^{M_*} (1-\alpha_i)^+, \end{aligned}$$

which is the DMT problem in (32). Hence, $\Delta_{2j}(r_j, b) = x + d^*(r_j)$. Bringing all together,

$$\Delta_j(b,x) = \max_{r_j \geq 0} \{\min\{\max\{x, br_j\}, x + d^*(r_j)\}\}. \tag{45}$$

Since $d^*(r_j) = 0$ for $r_j > M_*$, the constraint in (45) can be reduced to $0 \leq r_j \leq M_*$ without loss of optimality since $\Delta_j(b,x) = x$ for any $r_j \geq M_*$. Then, (45) coincides with (38), and thus, SSCC and JDS achieve the same distortion exponent.

APPENDIX IV
PROOF THEOREM 5

In this Appendix we derive the outage region $\mathcal{O}_h$ in (14), and the average distortion expression in (16). Then, using these we obtain the distortion exponent achieved by HDA-WZ.

A. *Outage region for HDA-WZ*

From joint typicality arguments similarly to [22], the decoding of $\mathbf{W}^{\frac{m}{2M_*}}$ is successful with high probability if

$$I(\mathbf{W}^{\frac{m}{2M_*}}; X^m) < \frac{m}{2} R_h < I(\mathbf{W}^{\frac{m}{2M_*}} \mathbf{T}^{n-\frac{m}{2M_*}}; \mathbf{V}^n Y^m). \tag{46}$$

For the left hand side (l.h.s.) of (46) we have

$$I(\mathbf{W}^{\frac{m}{2M_*}}; X^m) = \sum_{i=1}^{\frac{m}{2M_*}} I(\mathbf{W}_i; \mathbf{X}_i) = \frac{m}{2M_*} I(\mathbf{W}; \mathbf{X}) = mI(W;X) < \frac{m}{2} R_h, \tag{47}$$

due to the i.i.d. distribution of the source, $\mathbf{Q}_i$ and $\mathbf{X}_i$. Note that the l.h.s. of (46) always holds since R_h is chosen such that $\frac{R_h}{2} = I(W;X) + \epsilon$.

The r.h.s. of the decoding condition (46) is given by

$$I(\mathbf{W}^{\frac{m}{2M_*}}\mathbf{T}^{n-\frac{m}{2M_*}};\mathbf{V}^n Y^m) \overset{(a)}{=} \sum_{i=1}^{\frac{m}{2M_*}} I(\mathbf{W}_i;\mathbf{V}_{W,i}\mathbf{Y}_i) + \sum_{i=\frac{m}{2M_*}+1}^{n} I(\mathbf{T}_i;\mathbf{V}_{T,i})$$

$$= \frac{m}{2M_*} I(\mathbf{W};\mathbf{V}_W\mathbf{Y}) + \left(n - \frac{m}{2M_*}\right) I(\mathbf{T};\mathbf{V}_T), \tag{48}$$

where (a) follows from the i.i.d. distribution of the implied variables.

Substituting (47) and (48) into (46) and dividing both sides by $m/2M_*$, we obtain the outage condition in (14). Next, we evaluate (14) for Gaussian codewords. We readily obtain $I(\mathbf{V}_T;\mathbf{T}) = \log\left(\det\left(\mathbf{I} + \frac{\rho}{M_*}\mathbf{H}\mathbf{H}^\dagger\right)\right)$. Then, we have $I(\mathbf{W};\mathbf{V}_W\mathbf{Y}) = H(\mathbf{V}_W\mathbf{Y}) - H(\mathbf{V}_W\mathbf{W}\mathbf{Y}) + H(\mathbf{W})$. Let $\mathbf{G} \triangleq [\mathbf{W},\mathbf{V}_W,\mathbf{Y}]^H$. Since $\mathbf{G}$ is a complex multivariate Gaussian random vector, its differential entropy is given by $H(\mathbf{V}_W\mathbf{W}\mathbf{Y}) = \log((2\pi e)^{3M_*}\det(\mathbf{C}_\mathbf{G}))$, where $\mathbf{C}_\mathbf{G} = \mathrm{E}[\mathbf{G}\mathbf{G}^H]$ is given by

$$\mathbf{C}_\mathbf{G} = \begin{bmatrix} \mathbf{I}+\sigma_Q^2\mathbf{I} & \sqrt{\alpha}\sigma_Q^2\mathbf{H}^H & \gamma\sqrt{\rho_s}\mathbf{I} \\ \sqrt{\alpha}\sigma_Q^2\mathbf{H} & \mathbf{I}+\alpha\sigma_Q^2\mathbf{H}\mathbf{H}^H & \mathbf{0} \\ \gamma\sqrt{\rho_s}\mathbf{I} & \mathbf{0} & \xi\mathbf{I} \end{bmatrix},$$

with $\alpha \triangleq \rho/M_t$ and $\xi \triangleq 1+\rho_s\gamma$. By using properties of the determinant of a block matrix and some algebra, we have

$$\det(\mathbf{C}_\mathbf{G}) = \det\left(\mathbf{I} + \alpha\sigma_Q^2\mathbf{H}\mathbf{H}^H + \xi\sigma_Q^2\mathbf{I}\right) = \prod_{i=1}^{M_*}\left(1 + \xi\sigma_Q^2 + \frac{\rho}{M_*}\lambda_i\right).$$

Similarly, we have

$$H(\mathbf{V}_W\mathbf{Y}) = \log\left((2\pi e)^{2M_*}\xi^{M_*}\det\left(\mathbf{I} + \frac{\rho}{M_*}\mathbf{H}\mathbf{H}^H\right)\right),$$
$$H(\mathbf{W}) = \log((2\pi e)^{M_*}(1+\sigma_Q^2)^{M_*}),$$

This completes the characterization of (15).

B. Expected distortion achieved by HDA-WZ

We use an MMSE estimator to reconstruct each source block $\mathbf{X}_i$, $i = 1, ..., \frac{m}{2M_*}$, with the available information, which can be modeled by the linear model as follows:

$$\begin{bmatrix} \mathbf{W}_i \\ \mathbf{V}_i \\ \mathbf{Y}_i \end{bmatrix} = \begin{bmatrix} \mathbf{I} \\ \mathbf{0} \\ \gamma\mathbf{I} \end{bmatrix} \mathbf{X}_i + \begin{bmatrix} \mathbf{Q}_i \\ \sqrt{\alpha}\mathbf{H}\mathbf{Q}_i + \mathbf{N}_i \\ \mathbf{Z}_i \end{bmatrix}.$$

Let $\mathbf{B} \triangleq \begin{bmatrix} \mathbf{I} & \mathbf{0} & \gamma\mathbf{I} \end{bmatrix}^H$ and $\mathbf{S}_i \triangleq \begin{bmatrix} \mathbf{Q}_i & \alpha\mathbf{H}\mathbf{Q}_i + \mathbf{N}_i & \mathbf{Z}_i \end{bmatrix}^H$. Then, the distortion for each source block is found to be given by $\mathrm{Tr}\{\mathbf{D}\} = \frac{1}{M_*}\sum_{i=1}^{M_*} \mathrm{Tr}[\mathbf{I} + \mathbf{B}\mathbf{C}_S\mathbf{B}^H]^{-1}$, where $\mathbf{D}$ is the distortion matrix in the reconstruction of each block, and

$$\mathbf{C}_\mathbf{S} \triangleq \mathrm{E}[\mathbf{S}_i\mathbf{S}_i^H] = \begin{bmatrix} \mathbf{I} & \sqrt{\alpha}\mathbf{H}^H & \mathbf{0} \\ \sqrt{\alpha}\mathbf{H} & \alpha\sigma_Q^2\mathbf{H}\mathbf{H}^H + \mathbf{I} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{I} \end{bmatrix}.$$

Using the block inverse properties, and the singular value decomposition of $\mathbf{H}$ we obtain the expected distortion expression in (16).

C. Distortion exponent achieved by HDA-WZ

The outage region in (14) is given by

$$\mathcal{O}_h = \Bigg\{ (\mathbf{H}, \gamma) : \left(1 + \frac{1}{\sigma_Q^2}\right)^{M_*} \geq \left(((1+\rho_s\gamma)(1+\sigma_Q^2))^{M_*} \cdot \frac{\prod_{i=1}^{M_*}(1 + \frac{\rho}{M_*}\lambda_i)^{bM_*}}{\prod_{i=1}^{M_*}(1 + \frac{\rho}{M_*}\lambda_i + (1+\rho_s\gamma)\sigma_Q^2)} \right) \Bigg\}. \tag{49}$$

Similarly to the analysis of the previous schemes, we consider the change of variables $\lambda_i = \rho^{-\alpha_i}$, and $\gamma = \rho^{-\beta}$, and a rate $R_h = r_h \log \rho$, for $r_h \geq 0$. Then, we start by finding the equivalent outage set in the high SNR regime. We have,

$$\prod_{i=1}^{M_*}\left(1 + \frac{\rho}{M_*}\lambda_i\right)^{bM_*} \doteq \rho^{bM_* \sum_{i=1}^{M_*}(1-\alpha_i)^+},$$

and

$$\prod_{i=1}^{M_*}\left(1+\frac{\rho}{M_*}\lambda_i+(1+\rho_s\gamma)\sigma_Q^2\right) \doteq \prod_{i=1}^{M_*}\left(1+\rho^{1-\alpha_i}+(1+\rho^{x-\beta})\rho^{-r_h}\right)$$
$$\doteq \rho^{\sum_{i=1}^{M_*}\max\{(1-\alpha_i)^+,(x-\beta)^+-r_h\}},$$

where we use $\sigma_Q^2=(2^{R_h-\epsilon}-1)^{-1}=(2^{-\epsilon}\rho^{r_h}-1)^{-1}\doteq\rho^{-r_h}$. For the outage condition in (49), we have

$$\frac{\left(1+\frac{1}{\sigma_Q^2}\right)^{M_*}\prod_{i=1}^{M_*}(1+\frac{\rho}{M_*}\lambda_i+(1+\rho_s\gamma)\sigma_Q^2)}{((1+\rho_s\gamma)(1+\sigma_Q^2))^{M_*}\prod_{i=1}^{M_*}(1+\frac{\rho}{M_*}\lambda_i)^{bM_*}}$$
$$\doteq \frac{\rho^{M_*r_h}\rho^{\sum_{i=1}^{M_*}\max\{(1-\alpha)^+,(x-\beta)^+-r_h\}}}{\rho^{M_*(x-\beta)^+}\rho^{bM_*\sum_{i=1}^{M_*}(1-\alpha)^+}}$$
$$\doteq \rho^{\sum_{i=1}^{M_*}(r_h-(x-\beta)^++(1-\alpha_i))^+-bM_*\sum_1^{M_*}(1-\alpha_i)^+}.$$

Therefore, in the high SNR regime, the set $\mathcal{O}_h$ is equivalent to the set given by

$$\mathcal{A}_h \triangleq \left\{(\boldsymbol{\alpha},\beta)^+ : \sum_{i=1}^{M_*}(r_h-(x-\beta)^++(1-\alpha_i))^+ > bM_*\sum_{i=1}^{M_*}(1-\alpha_i)\right\}.$$

On the other hand, in the high SNR regime, the distortion achieved by HDA-WZ is equivalent to

$$D_h(\sigma_Q^2,\mathbf{H},\gamma)=\frac{1}{M_*}\sum_{i=1}^{M_*}\left(1+\rho_s\gamma+\frac{1}{\sigma_Q^2}\left(1+\frac{\rho}{M_*}\lambda_i\right)\right)^{-1}$$
$$\doteq \sum_{i=1}^{M_*}\left(1+\rho^{x-\beta}+\rho^{r_h+(1-\alpha_i)}\right)^{-1}$$
$$\doteq \rho^{-\min_{i=1,\ldots,M_*}\{\max\{(x-\beta)^+,r_h+1-\alpha_i\}\}}$$
$$\doteq \rho^{-\max\{(x-\beta)^+,r_h+1-\alpha_1\}},$$

where the last equality follows since $\alpha_1\geq\ldots\geq\alpha_{M_*}\geq 0$. Then, in the high SNR regime, the expected distortion for HDA-WZ is given as

$$\begin{aligned}ED_h(r_h\log\rho) &= \int_{\mathcal{O}_h^c} D_h(\sigma_Q^2,\mathbf{H},\gamma)p_h(\mathbf{H})p_\Gamma(\gamma)d\mathbf{H}d\gamma \\ &\quad + \int_{\mathcal{O}_h} D_d(0,\gamma)p_h(\mathbf{H})p_\Gamma(\gamma)d\mathbf{H}d\gamma \\ &\doteq \int_{\mathcal{A}_j^c} \rho^{-\max\{(x-\beta)^+,r_h+(1-\alpha_1)\}}p_A(\boldsymbol{\alpha})p_B(\beta)d\boldsymbol{\alpha}d\beta \\ &\quad + \int_{\mathcal{A}_j} \rho^{-(x-\beta)^+}p_A(\boldsymbol{\alpha})p_B(\beta)d\boldsymbol{\alpha}d\beta.\end{aligned}$$

Similarly to the proof of Theorem 3, applying Varadhan's lemma, the exponent of each integral is found as

$$\Delta_h^1(r_h) = \inf_{\mathcal{A}_h^c} \max\{(x-\beta)^+, r_h+1-\alpha_1\} + S_A(\boldsymbol{\alpha}) + \beta,$$

and

$$\Delta_h^2(r_h) = \inf_{\mathcal{A}_h}(x-\beta)^+ + S_A(\boldsymbol{\alpha}) + \beta, \tag{50}$$

First we solve $\Delta_h^1(r_h)$. The infimum for this problem is achieved by $\boldsymbol{\alpha}^*=0$ and $\beta^*=0$, and is given by

$$\Delta_h^1(r_h) = \max\{x, r_h+1\}, \qquad \text{for } r_h \le M_* b - 1 + x.$$

Now we solve $\Delta_h^2(r_h)$ in (50). By letting $\beta^*=x$, the range of $\boldsymbol{\alpha}$ is enlarged while the objective function is minimized. Thus, the problem reduces to

$$\begin{aligned}\Delta_h^2(r_h) &= \inf x + S(\boldsymbol{\alpha}) \\ &\quad \text{s.t. } r_h > \frac{bM_*-1}{M^*}\sum_{i=1}^{M_*}(1-\alpha_i)^+.\end{aligned}$$

Again, this problem is a scaled version of the DMT curve in (32). Therefore, we have

$$\Delta_h^2(r_h) = x + d^*\left(\left(\frac{bM_*-1}{M_*}\right)^{-1} r_h\right).$$

The distortion exponent is given by optimizing over r_h as

$$\Delta_h(b,x) = \max_{r_h} \min\{\Delta_h^1(r_h), \Delta_h^2(r_h)\}.$$

The maximum distortion exponent is obtained by letting $\Delta_h^1(r_h) = \Delta_h^2(r_h)$. We assume $r_h + 1 > x$ since otherwise $\Delta_h(b,x) = x$, and then, we have $r_h + 1 = x + d^*\left((b - \frac{1}{M_*})^{-1} r_h\right)$. Let $r'_h = r_h(b - \frac{1}{M_*})^{-1}$. Using (2), for $k < r'_h \leq k+1$, $k = 0, ..., M_* - 1$, the problem is equivalent to $r'_h\left(b - \frac{1}{M_*}\right) + 1 = x + \Phi_k - (r'_h - k)\Upsilon_k$, where Φ_k and Υ_k are given as in (3). The r'_h satisfying the equality is given by

$$r'^*_h = \frac{\Phi_k + k\Phi_k - 1 + x}{b - \frac{1}{M_*} + \Phi_k},$$

and the corresponding distortion exponent is found as

$$\Delta_h(b,x) = 1 + \frac{(bM_* - 1)(\Phi_k + k\Upsilon_k - 1 + x)}{bM_* - 1 + M_*\Upsilon_k},$$

for

$$b \in \left[\frac{\Phi_{k+1} - 1 + x}{k+1} + \frac{1}{M_*}, \frac{\Phi_k - 1 + x}{k} + \frac{1}{M_*}\right), \qquad \text{for } k = 0, ..., M_* - 1.$$

Note that we have $r_h^* + 1 > x$ whenever $\Delta_h(b,x) > x$. Otherwise, r_h^* is not feasible and $\Delta_h(b,x) = x$. Note also that if $x \geq bM_*$, the distortion exponent is given by $\Delta_h(b,x) = x$.

APPENDIX V

PROOF OF THEOREM 6

A. Successively refinable codebooks

Consider a successively refinable codebook [23] at rate $\frac{b}{2L}R_l = I(X; W_l | W_1^{l-1}) + \epsilon/2$ for each layer. We have

$$\begin{aligned} I(X; W_l | W_1^{l-1}) &\overset{(a)}{=} I(X; W_1^l) - I(X; W_1^{l-1}) \\ &\overset{(b)}{=} I(X; W_l) - I(X; W_{l-1}), \end{aligned} \tag{51}$$

where (a) is due to the chain rule, and (b) holds form the Markov chain $X - W_l - W_{l-1} - \cdots - W_1$. We have

$$\begin{aligned}\sum_{i=1}^{l}\left(\frac{b}{2L}R_i - \frac{\epsilon}{2}\right) &= \sum_{i=1}^{l} I(X; W_i | W_1^{i-1}) \\ &\overset{(a)}{=} \sum_{i=1}^{l} I(X; W_i) - I(X; W_{i-1}) \\ &= I(X; W_l) \\ &= \frac{1}{2}\log\left(1 + \frac{1}{\sum_{i=l}^{L}\sigma_i^2}\right),\ l = 1, ..., L, \end{aligned} \tag{52}$$

where (a) follows from (51) and $W_0 = \emptyset$ for the case $l = 1$.

B. Distortion exponent achievable by LS-JDS

In this section we obtain the distortion exponent for LS-JDS. Let us define $\bar{R}_1^l \triangleq \sum_{i=1}^{l} R_i$. First, we consider the outage event. For the successive refinement codebook the l.h.s. of (18) is given by

$$\begin{aligned} I(X; W_l | W_1^{l-1}, Y) &\overset{(a)}{=} I(X; W_l | Y) - I(X; W_{l-1} | Y) \\ &\overset{(b)}{=} H(W_l|Y) - H(\overline{Q}_l) - H(W_{l-1}|Y) + H(\overline{Q}_{l-1}) \\ &\overset{(c)}{=} \frac{1}{2}\log\left(\frac{\sum_{i=l-1}^{L}\sigma_i^2}{\sum_{i=l}^{L}\sigma_i^2}\frac{1 + (1+\gamma\rho_s)\sum_{j=l}^{L}\sigma_j^2}{1 + (1+\gamma\rho_s)\sum_{j=l-1}^{L}\sigma_j^2}\right), \end{aligned}$$

where $\overline{Q}_l \triangleq \sum_{i=l}^{L} Q_l$, and (a) is due to the Markov chain $Y - X - W_L - ... - W_1$, and (b) is due to the independence of $\bar{Q}_i$ from X and Y, and finally (c) follows since $H(W_l|Y) = \frac{1}{2}\log\left(\sum_{i=l}^{L}\sigma_i^2 + \frac{1}{1+\gamma\rho_s}\right)$ for $l = 1, ..., L$. We also have

$$I(X; W_1 | Y) = \frac{1}{2}\log\left(1 + \frac{1}{(1+\gamma\rho_s)\sum_{i=1}^{L}\sigma_i^2}\right).$$

Substituting (17) into (52), we have

$$I(X; W_l | W_1^{l-1}, Y) = \frac{1}{2}\log\left(\frac{2^{\sum_{i=1}^{l}\frac{b}{L}R_i - \epsilon} + \gamma\rho_s}{2^{\sum_{i=1}^{l-1}\frac{b}{L}R_i - \epsilon} + \gamma\rho_s}\right).$$

Then, the outage condition in (18) is given by

$$\log\left(\frac{2^{\sum_{i=1}^{l}\frac{b}{L}R_i-\epsilon}+\gamma\rho_s}{2^{\sum_{i=1}^{l-1}\frac{b}{L}R_i-\epsilon}+\gamma\rho_s}\right)\geq\frac{b}{L}\log\prod_{i=1}^{M_*}\left(1+\frac{\rho}{M_*}\lambda_i\right). \tag{53}$$

Therefore, in the high SNR regime, we have, for $l=1,...,L$

$$\begin{aligned}\frac{2^{\sum_{i=1}^{l}(\frac{b}{L}R_i-\epsilon)}+\gamma\rho_s}{2^{\sum_{i=1}^{l-1}(\frac{b}{L}R_i-\epsilon)}+\gamma\rho_s}&\doteq\frac{\rho^{\sum_{i=1}^{l}\frac{b}{L}r_i}+\rho^{x-\beta}}{\rho^{\sum_{i=1}^{l-1}\frac{b}{L}r_i}+\rho^{x-\beta}}\\&\doteq\frac{\rho^{\sum_{i=1}^{l}\frac{b}{L}r_i-(x-\beta)}+1}{\rho^{\sum_{i=1}^{l-1}\frac{b}{L}r_i-(x-\beta)}+1}\\&\doteq\frac{\rho^{(\sum_{i=1}^{l}\frac{b}{L}r_i-(x-\beta))^+}}{\rho^{(\sum_{i=1}^{l-1}\frac{b}{L}r_i-(x-\beta))^+}},\end{aligned} \tag{54}$$

and

$$\frac{b}{L}\log\prod_{i=1}^{M_*}\left(1+\frac{\rho}{M_*}\lambda_i\right)\doteq\rho^{\frac{b}{L}\sum_{i=1}^{M_*}(1-\alpha_i)^+}.$$

The outage set (18) in the high SNR regime is equivalent to

$$\mathcal{A}_l^{ls}\triangleq\left\{(\boldsymbol{\alpha},\beta):\frac{b}{L}\sum_{i=1}^{M_*}[(1-\alpha_i)^+<\left(\sum_{i=1}^{l}\frac{b}{L}r_i-(x-\beta)\right)^+-\left(\sum_{i=1}^{l-1}\frac{b}{L}r_i-(x-\beta)\right)^+\right\}. \tag{55}$$

Now, we study the high SNR behavior of the expected distortion. It is not hard to see that (19) is given by

$$ED_{ls}(\mathbf{R})=\sum_{l=0}^{L}\mathrm{E}_{\mathcal{O}_{l+1}^{ls}}\left[D_d\left(\frac{b}{2L}\bar{R}_1^l,\gamma\right)\right]-\mathrm{E}_{\mathcal{O}_{l}^{ls}}\left[D_d\left(\frac{b}{2L}\bar{R}_1^l,\gamma\right)\right], \tag{56}$$

where $\mathcal{O}_0^{ls}\triangleq\emptyset$ and $\mathcal{O}_{L+1}^{ls}\triangleq\mathbb{R}^{M_*+1}$. For each term in (56), we have

$$\mathrm{E}_{\mathcal{O}_{l+1}^{ls}}\left[D_d\left(\frac{b}{2L}\bar{R}_1^l,\gamma\right)\right]\doteq\int_{\mathcal{A}_{l+1}^{ls}}\rho^{-\max\{\frac{b}{L}\sum_{i=1}^{l}r_l,(x-\beta)^+\}}\rho^{-S_A(\boldsymbol{\alpha})}\rho^{-\beta}d\boldsymbol{\alpha}d\beta, \tag{57}$$

$$\mathrm{E}_{\mathcal{O}_{l}^{ls}}\left[D_d\left(\frac{b}{2L}\bar{R}_1^l,\gamma\right)\right]\doteq\int_{\mathcal{A}_{l}^{ls}}\rho^{-\max\{\frac{b}{L}\sum_{i=1}^{l}r_l,(x-\beta)^+\}}\rho^{-S_A(\boldsymbol{\alpha})}\rho^{-\beta}d\boldsymbol{\alpha}d\beta, \tag{58}$$

where the outage set in the high SNR regime is given by (55).

Applying Varadhan's lemma to (57), the exponential behavior of (57) for $l=0,...,L-1$, is found as the solution to

$$\tilde{\Delta}_l^+ \triangleq \inf_{\mathcal{A}_{l+1}^{ls}} \max\{b/L\bar{r}_1^l, (x-\beta)^+\} + S_A(\boldsymbol{\alpha}) + \beta,$$

where we define $\bar{r}_1^l \triangleq \sum_{i=1}^l r_i$. Similarly, applying Varadhan's lemma to (58), the exponential behavior of (58) for $l = 0, ..., L-1$ is given by

$$\tilde{\Delta}_l \triangleq \inf_{\mathcal{A}_l^{ls}} \max\{b/L\bar{r}_1^l, (x-\beta)^+\} + S_A(\boldsymbol{\alpha}) + \beta.$$

Since $r_1 \leq r_2 \leq \cdots \leq r_L$ we have $\mathcal{A}_l^{ls} \subseteq \mathcal{A}_{l+1}^{ls}$, and therefore $\tilde{\Delta}_l \geq \tilde{\Delta}_l^+$. Then, from (56) we have

$$ED_{ls}(\mathbf{R}) \doteq \sum_{l=0}^{L} \rho^{-\tilde{\Delta}_l^+} - \rho^{-\tilde{\Delta}_l} \doteq \sum_{l=0}^{L} \rho^{-\Delta_l^+}.$$

We define $\Delta_l^{ls}(\mathbf{r}) \triangleq \tilde{\Delta}_l^+$, where $\mathbf{r} \triangleq [r_1, ..., r_L]$. Then, the distortion exponent of LS-JDS is given as follows:

$$\Delta_{ls}^*(b, x) = \max_{\mathbf{r}} \min \Delta_l^{ls}(\mathbf{r}).$$

For $l\!=\!0$, i.e., no codeword is successfully decoded, we have

$$\begin{aligned}\Delta_0^{ls}(\mathbf{r}) &= \inf (x-\beta)^+ + \beta + S_A(\boldsymbol{\alpha}) \\ &\text{s.t. } \frac{b}{L}\sum_{i=1}^{M_*}(1-\alpha_i)^+ < \left(\frac{b}{L}r_1 - (x-\beta)\right)^+.\end{aligned}$$

The infimum is achieved by $\beta = x$ and using the DMT in (1), we have

$$\Delta_0^{ls}(\mathbf{r}) = x + d^*(r_1).$$

The distortion exponent when l layers are successfully decoded is found as

$$\begin{aligned}\Delta_l^{ls}(\mathbf{r}) \;=\; &\inf \max\left\{\frac{b}{L}\bar{r}_1^l, (x-\beta)^+\right\} + \beta + S_A(\boldsymbol{\alpha}) \\ &\text{s.t. } \frac{b}{L}\sum_{i=1}^{M_*}(1-\alpha_i)^+ < \left(\frac{b}{L}\bar{r}_1^{l+1} - (x-\beta)\right)^+ - \left(\frac{b}{L}\bar{r}_1^l - (x-\beta)\right)^+.\end{aligned} \tag{59}$$

If $\frac{b}{L}\bar{r}_1^l \geq x$, the infimum of (59) is obtained for $\beta^* = 0$ and

$$\begin{aligned}\Delta_l^{ls}(\mathbf{r}) = \inf \frac{b}{L}\bar{r}_1^l + S_A(\boldsymbol{\alpha}) \\ \text{s.t.} \sum_{i^1}^{M_*} (\xi_l - \alpha_i)^+ < r_{k+1}.\end{aligned} \tag{60}$$

Using the DMT in (1), (60) is minimized as

$$\Delta_l^{ls}(\mathbf{r}) = \frac{b}{L}\bar{r}_1^l + d^*\left(r_{l+1}\right).$$

If $\frac{b}{L}\bar{r}_1^l \leq x$, we have that the minimum of (59) is achieved by $\beta^* = \left(x - \frac{b}{L}\bar{r}_1^l\right)^+$ if $\frac{b}{L}\bar{r}_1^l > (x-\beta)$ and is given by

$$\Delta_l^{ls}(\mathbf{r}) = x + d^*\left(r_{l+1}\right).$$

If $\frac{b}{L}\bar{r}_1^l \leq (x-\beta) < \frac{b}{L}\bar{r}_1^{l+1}$, the optimization problem in (59) is equivalent to

$$\begin{aligned}\Delta_l^{ls}(\mathbf{r}) = \inf (x-\beta)^+ + \beta + S_A(\boldsymbol{\alpha}) \\ \text{s.t. } \frac{b}{L}\sum_{i=1}^{M_*}(1-\alpha_i)^+ < \left(\frac{b}{L}r_1^{l+1} - (x-\beta)\right)^+, \\ \frac{b}{L}\bar{r}_1^l \leq (x-\beta) < \frac{b}{L}\bar{r}_1^{l+1}.\end{aligned} \tag{61}$$

The infimum of (61) is achieved by the largest β, since increasing β enlarges the range of $\boldsymbol{\alpha}$. Then, $\beta^* = (x - \frac{b}{L}\bar{r}_1^l)^+$, and we have,

$$\Delta_l^{ls}(\mathbf{r}) = x + d^*\left(r_{l+1}\right).$$

Finally, if $\frac{b}{L}\bar{r}_1^{l+1} \leq (x-\beta)$, there are no feasible solutions for (59). Therefore, putting all together we have

$$\Delta_l^{ls}(\mathbf{r}) = \inf \max\left\{\frac{b}{L}\bar{r}_1^l, x\right\} + d^*(r_{l+1}).$$

Similarly, at layer L, the infimum is achieved by $\boldsymbol{\alpha}^* = 0$ and $\beta^* = 0$ and is given by

$$\Delta_L^{ls}(\mathbf{r}) = \max\left\{\frac{b}{L}\bar{r}_1^L, x\right\}, \quad \text{for } r_L \leq M_*.$$

Note that the condition on r_L always holds.

C. Solution of the distortion exponent

Assume that for a given layer $\hat{l}$ we have $\bar{r}_1^{\hat{l}-1}\frac{b}{L} \leq x \leq \bar{r}_1^{\hat{l}}\frac{b}{L}$. Then, $\Delta_l^{ls}(\mathbf{r}) = x + d(r_{l+1})$ for $l = 0, ..., \hat{l}-1$. Using the KKT conditions, the maximim distortion exponent is obtained when all the distortion exponents are equal.

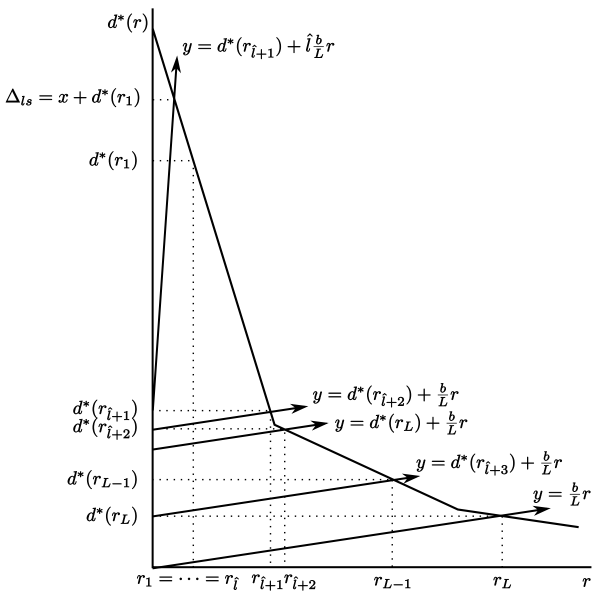
From $\Delta_0^{ls}(\mathbf{r}) = \cdots = \Delta_{\hat{l}-1}^{ls}(\mathbf{r})$ we have $r_1 = \cdots = r_{\hat{l}}$, and thus, $\bar{r}_1^{\hat{l}} = \hat{l}r_1$. Then, the exponents are given by

$$\begin{aligned}
\Delta_0^{ls}(\mathbf{r}) &= x + d^*(r_1) \\
\Delta_{\hat{l}}^{ls}(\mathbf{r}) &= b\frac{\hat{l}}{L}r_1 + d^*(r_{\hat{l}+1}) \\
&\cdots \\
\Delta_{L-1}^{ls}(\mathbf{r}) &= b\frac{\hat{l}}{L}r_1 + b\frac{1}{L}\bar{r}_{\hat{l}+1}^{L-1} + d^*(r_L) \\
\Delta_L^{ls}(\mathbf{r}) &= b\frac{\hat{l}}{L}r_1 + b\frac{1}{L}\bar{r}_{\hat{l}+1}^{L}.
\end{aligned}$$

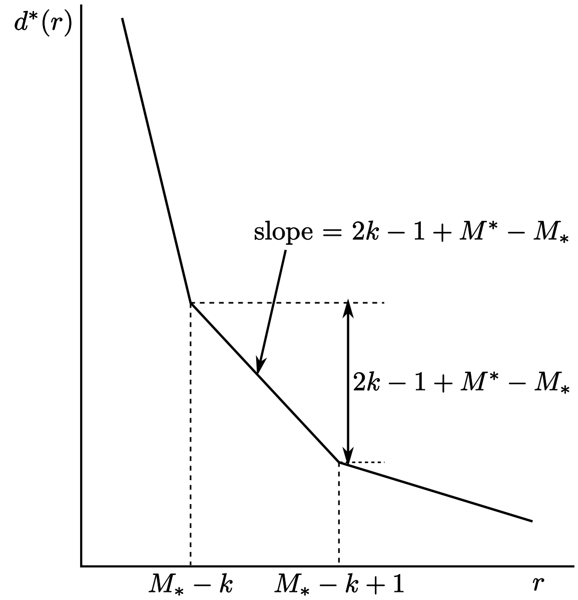
Equating all these exponents, we have

$$\begin{aligned}
b\frac{1}{L}r_L &= d^*(r_L) \\
b\frac{1}{L}r_{L-1} + d(r_L) &= d^*(r_{L-1}) \\
&\cdots \\
b\frac{1}{L}r_{\hat{l}+1} + d^*(r_{\hat{l}+2}) &= d^*(r_{\hat{l}+1}) \\
b\frac{l}{L}r_1 + d^*(r_{\hat{l}+1}) &= d^*(r_1) + x.
\end{aligned}$$

See Figure 10(a) for a geometric interpretation of the rate allocation for LS-JDS satisfying the above



(a) Rate allocation for the source layers of LS-JDS illustrated on DMT curve of the MIMO channel.



(b) The DMT curve of an $M_t \times M_r$ MIMO system is composed of M_* line segments, of which the i-th one is shown in the figure.

Fig. 10. Geometrical interpretation of the LS-JDS rate allocation and the DMT curve.

equalities: we have $L - \hat{l}$ straight lines of slope b/L and each line intersects in the y axis at a point with the same ordinate as the intersection of the previous line with the DMT curve. The more layers we have the higher the distortion exponent of LS-JDS can climb. The remaining $\hat{l}$ layers allow a final climb of slope $\hat{l}b/L$. Note that the higher $\hat{l}$, the higher the slope but the lower the starting point $d^*(r_{\hat{l}+1})$.

Next, we adapt Lemma 3 from [9] to our setup. Let q be a line with equation $y = -\alpha(t - M)$ for some $\alpha > 0$ and $M > 0$ and let $q_i = 1, ..., L$ be the set of lines defined recursively from L to 1 as $y = (b/L)t + d_{i+1}$, where $b > 0$, $d_{L+1} \triangleq 0$, and d_i is the y component of the intersection of q_i with q. Then, sequentially solving the intersection points for $i = \hat{l} + 1, ..., L$ we have:

$$d_i - d_{i+1} = M\frac{b}{L}\left(\frac{\alpha}{\alpha + b/L}\right)^{L-i+1}.$$

Summing all the terms for $i = \hat{l}+1, ..., L$ we obtain

$$d_i = M\alpha \left[1 - \left(\frac{\alpha}{\alpha + b/L}\right)^{L-i+1}\right].$$

In the following we consider a continuum of layers, i.e., we let $L \to \infty$. Let $\hat{l} = \kappa L$ be the numbers of layers needed so that $b\hat{l}/Lr_1 = b\kappa r_1 = x$, that is, from $l = 1$ to $l = \kappa L$.

When $M_* = 1$, the DMT curve is composed of a single line with $\alpha = M^*$ and $M = 1$. In that case, with layers from $\kappa L + 1$ to L the distortion exponent increases up to

$$d^*(r_{L\kappa+1}) = M\alpha \left[1 - \left(\frac{\alpha}{\alpha + b/L}\right)^{L(1-\kappa)}\right].$$

In the limit of infinite layers, we obtain

$$\lim_{L\to\infty} d^*(r_{L\kappa+1}) = M\alpha \left(1 - e^{-\frac{b(1-\kappa)}{\alpha}}\right).$$

We still need to determine the distortion achieved due to the climb with layers from $l = 1$ to $l = \kappa L$ by determining r_1, which is found as the solution to $\Delta_0^{ls}(\mathbf{r}) = \Delta_{L\kappa}^{ls}(\mathbf{r})$, i.e.,

$$b\kappa r_1 + d^*(r_{L\kappa+1}) = x - \alpha(r_1 - M), \tag{62}$$

Since $x = b\kappa r_1$, $r_1 = x/b\kappa$, and from (62) we get to

$$d^*(r_{L\kappa+1}) = -\alpha\left(\frac{x}{b\kappa} - M\right),$$

which, in the limit of infinite layers, solves for

$$\kappa^* = \frac{M^*}{b}\mathcal{W}\left(\frac{e^{\frac{b}{M^*}x}}{M^*}\right),$$

where $\mathcal{W}(z)$ is the Lambert W function, which gives the principal solution for w in $z = we^w$. The distortion exponent in the MISO/SIMO case is then found as

$$\Delta_{ls}^*(b, x) = x + M^*\left(1 - e^{-\frac{b(1-\kappa^*)}{M^*}}\right).$$

For MIMO channels, the DMT curve is formed by M_* linear pieces, each between $M_* - k$ and $M_* - k + 1$ for $k = 1, ..., M_*$. From the value of the DMT at $M_* - k$ to the value at $M_* - k + 1$, there is a gap of $M^* - M_* + 2k - 1$ in the y abscise. Each piece of the curve can be characterized by $y = -\alpha(t - M)$, where for the k-th interval we have $\alpha = \phi_k$ and $M = M_k$ as in (20). See Figure 10(b) for an illustration.

We will again consider a continuum of layers, i.e., we let $L \to \infty$, and we let $l = L\kappa$ be the number of lines required to have $b\kappa r_1 = x$. Then, for the remaining lines from $l + 1$ to L, let $L(1-\kappa)\kappa_k$ be the number of lines with slope b/L required to climb up the whole interval k. Since the gap in the y abscise from the value at $M_* - k$ to the value at $M_* - k + 1$, is $M^* - M_* + 2k - 1$, climbing the whole k-th interval with $L(1-\kappa)\kappa_k$ lines requires

$$d_{L-L(1-\kappa)\kappa_k} = M^* - M_* + 2k - 1,$$

where

$$d_{L-L(1-\kappa)\kappa_k} = M\alpha \left[1 - \left(\frac{\alpha}{\alpha + b/L}\right)^{L(1-\kappa)\kappa_k+1}\right].$$

In the limit we have

$$\lim_{L\to\infty} d_{L-L(1-\kappa)\kappa_k} = M\alpha \left[1 - e^{-\frac{b(1-\kappa)\kappa_k}{\alpha}}\right].$$

Then, each required portion, κ_k, is found as

$$\kappa_k = \frac{M^* - M_* + 2k - 1}{b(1-\kappa)} \ln\left(\frac{M_* - k + 1}{M_* - k}\right).$$

This gives the portion of lines required to climb up the k -th segment of the DMT curve. In the MIMO case, to be able to go up exactly to the k-th segment with lines from $l + 1$ to L we need to have $\sum_{j=1}^{k-1} \kappa_j < 1 \leq \sum_{j=1}^{k} \kappa_j$. This is equivalent to the requirement $c_{k-1} < b(1-\kappa) \leq c_k$ using c_i as defined in Theorem 6. To climb up each line segment we need $\kappa_k(1-\kappa)L$ lines (layers) for $k = 1, ..., M_* - 1$, and for the last segment climbed we have $(1 - \sum_{j=1}^{k-1} \kappa_j)L$ lines remaining, which gives an extra ascent of

$$M\alpha\left(1-e^{-\frac{b(1-\kappa)(1-\sum_{j=1}^{k-1}\kappa_j)}{\alpha}}\right).$$

Then, we have climbed up to the value

$$\begin{aligned}d_{L\kappa+1}=&\sum_{i=1}^{k-1}(M^*-M_*+2i-1)\\&+(M_*-k+1)(M^*-M_*+2k-1)\left(1-e^{-\frac{b(1-\kappa)(1-\sum_{j=1}^{k-1}\kappa_j)}{M^*-M_*+2k-1}}\right).\end{aligned}$$

With the remaining lines, i.e., from $l=1$ to $l=\kappa L$, the extra climb is given by solving $\Delta_0^{ls}(\mathbf{r})=\Delta_{\kappa L}^{ls}(\mathbf{r})$, i.e.,

$$x+d^*(r_1)=b\kappa r_1+d_{L\kappa+1}.$$

The diversity gain $d^*(r_1)$ at segment k is given by

$$d^*(r_1)=-\alpha(r_1-M)+\sum_{i=1}^{k-1}(M^*-M_*+2i-1).$$

Since we have $b\kappa r_1=x$, this equation simplifies to

$$d^*\left(\frac{x}{b\kappa}\right)=d_{L\kappa+1}.$$

Therefore, using $c_{k-1}\triangleq b(1-\kappa)\sum_{j=1}^{k-1}\kappa_j$, we solve κ from

$$-\alpha\left(\frac{x}{b\kappa}-M\right)=M\alpha\left(1-e^{-\frac{b(1-\kappa)-c_{k-1}}{\alpha}}\right),$$

and find

$$\kappa^*=\frac{\alpha}{b}\mathcal{W}\left(\frac{e^{\frac{b-c_{k-1}}{\alpha}}x}{M\alpha}\right).$$

The range of validity for each k is given by $c_{k-1}<b(1-\kappa)\leq c_k$. Since for a given c, the solution

to $c = b(1-\kappa^*)$ is found as

$$b = \frac{xe^{c_{k-1}-c}}{M} + c,$$

when $c = c_{k-1}$, we have

$$b > \frac{x}{M} + c_{k-1} = c_{k-1} + \frac{x}{M_* - k + 1}.$$

When $c = c_k$, since $c_{k-1} - c_k = \alpha \ln(M/(M_* - k))$, we have

$$b \leq \frac{xe^{c_{k-1}-c_k}}{M} + c_k = c_k + \frac{x}{M_* - k}.$$

Putting all together, we obtain the condition of the theorem and the corresponding distortion exponent.

APPENDIX VI
PROOF OF THEOREM 7

We consider the usual change of variables, $\lambda_i = \rho^{-\alpha_i}$ and $\gamma = \rho^{-\beta}$. Let r_l be the multiplexing gain of the l-th layer and $\mathbf{r} \triangleq [r_1, ..., r_L]$, such that $R_i = r_i \log \rho$, and define $\bar{r}_1^l \triangleq \sum_{i=1}^{l} r_i$.

First, we derive the outage set $\mathcal{O}_l^{bs}$ for each layer in the high SNR regime, which we denote by $\mathcal{L}_l$. For the power allocation $\rho_l = \rho^{\xi_{l-1}} - \rho^{\xi_l}$, the l.h.s. of the inequality in the definition of $\mathcal{O}_l^{bs}$ in (21) is given by

$$\begin{aligned} I(\mathbf{U}_l; \mathbf{V}|\mathbf{U}_1^{l-1}) &= I(\mathbf{U}_l^L; \mathbf{V}|\mathbf{U}_1^{l-1}) - I(\mathbf{U}_{l+1}^L; \mathbf{V}|\mathbf{U}_1^{l-1}) \\ &= \log \frac{\det\left(\mathbf{I} + \frac{\rho^{\xi_{l-1}}}{M_*}\mathbf{H}\mathbf{H}^H\right)}{\det\left(\mathbf{I} + \frac{\rho^{\xi_l}}{M_*}\mathbf{H}\mathbf{H}^H\right)} \\ &= \log \prod_{i=1}^{M_*} \frac{1 + \frac{\rho^{\xi_{l-1}}}{M_*}\lambda_i}{1 + \frac{\rho^{\xi_l}}{M_*}\lambda_i} \\ &\doteq \rho^{\sum_{i=1}^{M_*}(\xi_{l-1}-\alpha_i)^+ - (\xi_l - \alpha_i)^+}. \end{aligned} \tag{63}$$

The r.h.s. of the inequality in the definition of $\mathcal{O}_l^{bs}$ in (21) can be calculated as in (54). Then, from (63) and (54), $\mathcal{L}_l$ follows as:

$$\mathcal{L}_l \triangleq \Bigg\{(\boldsymbol{\alpha},\beta) : b\sum_{i=1}^{M_*}[(\xi_{l-1}-\alpha_i)^+ - (\xi_l-\alpha_i)^+] < \left(\sum_{i=1}^{l} br_i - (x-\beta)\right)^+ - \left(\sum_{i=1}^{l-1} br_i - (x-\beta)\right)^+\Bigg\}.$$

Since $\mathcal{O}_l^{bs}$ are mutually exclusive, in the high SNR we have

$$\begin{aligned} ED_{bs}(\mathbf{R},\boldsymbol{\xi}) &= \sum_{l=0}^{L}\int_{\mathcal{O}_{l+1}^{bs}} D_d\left(\sum_{i=0}^{l} b/2R_i,\gamma\right)p_h(\mathbf{H})p_\Gamma(\gamma)d\mathbf{H}d\gamma \\ &\doteq \sum_{l=0}^{L}\int_{\mathcal{L}_{l+1}} \rho^{-(\max\{\sum_{i=0}^{l} br_i,(x-\beta)^+\}+\beta+S_A(\boldsymbol{\alpha}))}d\boldsymbol{\alpha}d\beta \\ &\doteq \sum_{l=0}^{L}\rho^{-\Delta_l(\mathbf{r},\boldsymbol{\xi})} \\ &\doteq \rho^{-\Delta_{bs}^{L}(\mathbf{r},\boldsymbol{\xi})}, \end{aligned} \tag{64}$$

where, from Varadhan's lemma, the exponent for each integral term is given by

$$\Delta_l^{bs}(\mathbf{r},\boldsymbol{\xi}) = \inf_{\mathcal{L}_{l+1}} \max\left\{b\bar{r}_0^l,(x-\beta)^+\right\}+\beta+S_A(\boldsymbol{\alpha}). \tag{65}$$

Then, the distortion exponent is found as

$$\Delta_{bs}^{L}(b,x) = \max_{\mathbf{r},\boldsymbol{\xi}}\min_{l=0,\dots,L}\left\{\Delta_l^{bs}(\mathbf{r},\boldsymbol{\xi})\right\}. \tag{66}$$

Similarly to the DMT, we consider the *successive decoding diversity gain*, defined in [9], as the solution to the probability of outage with successive decoding of each layer, given by

$$\begin{aligned} d_{ds}(r_l,\xi_{l-1},\xi_l) &\triangleq \inf_{\boldsymbol{\alpha}^+} S_A(\boldsymbol{\alpha}) \\ &\quad \text{s.t. } r_l > \sum_{i=1}^{M_*}[(\xi_{l-1}-\alpha_i)^+ - (\xi_l-\alpha_i)^+]. \end{aligned} \tag{67}$$

Without loss of generality, consider the multiplexing gain r_l given by $r_l = k(\xi_{l-1}-\xi_l)+\delta_l$, where

$k \in [0, 1, ..., M_* - 1]$ and $0 \leq \delta_l < \xi_{l-1} - \xi_l$. Then, the infimum for (67) is found as

$$d_{ds}(r_l, \xi_{l-1}, \xi_l) = \Phi_k \xi_{l-1} - \Upsilon_k \delta_l, \tag{68}$$

with

$$\alpha_i^* = \begin{cases} \xi_{l-1}, & 1 \leq i < M_* - k, \\ \xi_{l-1} - \delta_l, & i = M_* - k, \\ 0, & M_* - k < i \leq M_*. \end{cases}$$

Now, we solve (65), using (68) for each layer, as a function of the power allocation ξ_{l-1} and ξ_l, and the rate r_l.

When no layer is successfully decoded, i.e., $l = 0$, we have

$$\begin{aligned} \Delta_0^{bs}(\mathbf{r}, \boldsymbol{\xi}) &= \inf (x - \beta)^+ + \beta + S_A(\boldsymbol{\alpha}) \\ \text{s.t. } & b \sum_{i=1}^{M_*} \left[(\xi_0 - \alpha_i)^+ - (\xi_1 - \alpha_i)^+ \right] < (br_1 - (x - \beta))^+. \end{aligned}$$

The infimum is achieved by $\beta^* = x$ and using (67), we have

$$\Delta_0^{bs}(\mathbf{r}, \boldsymbol{\xi}) = x + d_{ds}\left(r_1, \xi_0, \xi_1\right).$$

At layer l, the distortion exponent is given by the solution of the following optim

$$\begin{aligned} \Delta_l^{bs}(\mathbf{r}, \boldsymbol{\xi}) \;=\; & \inf \max\{b\bar{r}_1^l, (x - \beta)^+\} + \beta + S_A(\boldsymbol{\alpha}) \\ & \text{s.t. } b \sum_{i=1}^{M_*} \left[(\xi_l - \alpha_i)^+ - (\xi_{l+1} - \alpha_i)^+ \right] < (b\bar{r}_1^{l+1} - x + \beta)^+ - (b\bar{r}_1^l - x + \beta)^+. \end{aligned}$$

If $b\bar{r}_1^l \geq x$, the infimum is obtained for $\beta^* = 0$ and solving

$$\begin{aligned} \Delta_l^{bs}(\mathbf{r}, \boldsymbol{\xi}) \;=\; & \inf \max\{b\bar{r}_1^l, x\} + S_A(\boldsymbol{\alpha}) \\ & \text{s.t. } \sum_{i^1}^{M_*} \left[(\xi_l - \alpha_i)^+ - (\xi_{l+1} - \alpha_i)^+ \right] < r_{k+1}. \end{aligned}$$

Using (67), we obtain the solutio as

$$\Delta_l^{bs}(\mathbf{r},\boldsymbol{\xi}) = \max\{x, b\bar{r}_1^l\} + d_{ds}\left(r_{l+1}, \xi_l, \xi_{l+1}\right).$$

If $b\bar{r}_1^l \leq x$, the infimum is given by $\beta^* = (x - b\bar{r}_1^l)^+$, and again, we have a version of (67) with the distortion exponent

$$\Delta_l^{bs}(\mathbf{r},\boldsymbol{\xi}) = x + d_{ds}\left(r_{l+1}, \xi_l, \xi_{l+1}\right).$$

At layer L, the distortion exponent is the solution to the optimization problem

$$\begin{aligned}\Delta_L^{bs}(\mathbf{r},\boldsymbol{\xi}) = \inf \max &\left\{b\bar{r}_1^L, (x-\beta)^+\right\} + \beta + S_A(\boldsymbol{\alpha}) \\ \text{s.t.}\, b\sum_{i=1}^{M_*} &[(\xi_{L-1}-\alpha_i)^+ - (\xi_L - \alpha_i)^+] \geq \left(b\bar{r}_1^L - (x-\beta)\right)^+ - \left(b\bar{r}_1^{L-1} - (x-\beta)\right)^+.\end{aligned}$$

The infimum is achieved by $\boldsymbol{\alpha}^* = 0$ and $\beta^* = 0$, and is given by

$$\Delta_L^{bs}(\mathbf{r},\boldsymbol{\xi}) = \max\left\{b\bar{r}_1^L, x\right\}, \quad \text{for } r_L \leq M_*(\xi_{L-1} - \xi_L).$$

Note that the condition on r_L always holds.

Gathering all the results, the distortion exponent problem in (66) is solved as the minimum of the exponent of each layer, $\Delta_l^{bs}(\mathbf{r},\boldsymbol{\xi})$, which can be formulated as

$$\begin{aligned}\Delta_{bs}^L(b,x) = &\max_{\mathbf{r},\boldsymbol{\xi}}\, t \\ \text{s.t.}\quad & t \leq x + d_{sd}\left(r_1, \xi_0, \xi_1\right), \\ & t \leq \max\{b\bar{r}_1^l, x\} + d_{sd}\left(r_{l+1}, \xi_l, \xi_{l+1}\right), \quad \text{for } l = 1, \ldots, L-1, \\ & t \leq \max\{b\bar{r}_1^L, x\}.\end{aligned} \tag{69}$$

If $x \geq b\bar{r}_1^L$, then $\max\{x, b\bar{r}_1^l\} = x$ for all l, and the minimum distortion exponent is given by $\Delta_L^{bs}(\mathbf{r},\boldsymbol{\xi}) = x$, which implies $\Delta_{mj}^L(b,x) = x$. If $x \leq br_1$, then $\max\{x, b\bar{r}_1^l\} = b\bar{r}_1^l$ for all l. In general, if $b\bar{r}_1^q < x \leq b\bar{r}_1^{q+1}$, $q = 0, ..., L$, and $\bar{r}_1^0 \triangleq 0$, $\bar{r}_1^{L+1} \triangleq \infty$, then (69) can be formulated, using $r_l =$

$k(\xi_{l-1}-\xi_l)+\delta_l$, $\boldsymbol{\delta} \triangleq [\delta_1, \cdots, \delta_L]$ and $\boldsymbol{\xi}$, as the following linear optimization program:

$$\begin{aligned}
\Delta_{mj}^L(b,x) = & \min_{\substack{1\leq q\leq L,\\ 0\leq k\leq M_*-1.}} \min_{\boldsymbol{\delta},\boldsymbol{\xi}} -t \\
\text{s.t.} \quad & t \leq x+\Phi_k\xi_0-\Upsilon_k\delta_1, \\
& t \leq x+\Phi_k\xi_l-\Upsilon_k\delta_{l+1}, \quad \text{for} \ \ l=1,\dots,q, \\
& t \leq b\sum_{i=1}^{l}[k(\xi_{i-1}-\xi_i)+\delta_i]+\Phi_k\xi_l-\Upsilon_k\delta_{l+1}, \\
& \quad \text{for} \ \ l=q,\dots,L-1, \\
& t \leq b\sum_{i=1}^{L}[k(\xi_{i-1}-\xi_i)+\delta_i], \\
& 0\leq \delta_l < \xi_{l-1}-\xi_l, \quad \text{for} \ \ l=1,\dots,L, \\
& 0\leq \xi_L \leq ... \leq \xi_1 \leq \xi_0 = 1, \\
& \sum_{l=1}^{l'}[bk(\xi_{l-1}-\xi_l)+\delta_l] < x.
\end{aligned} \tag{70}$$

The linear program (70) can be efficiently solved using numerical methods. In Figure 4, the numerical solution is shown. However, in the following we provide a suboptimal yet more compact analytical solution by fixing the multiplexing gains $\mathbf{r}$. We fix the multiplexing gains as $\hat{r}_l = [(k+1)(\xi_{l-1}-\xi_l)-\epsilon_1]$, $\epsilon_1 > 0$ for $k=0,...,M_*-1$, and $\delta_l \triangleq (\xi_{l-1}-\xi_l)-\epsilon_1$, when the bandwidth ratio satisfies

$$b \in \left[\frac{\Phi_{k+1}+x}{k+1}, \frac{\Phi_k+x}{k}\right). \tag{71}$$

Assume $br_1 \geq x$. Then, each distortion exponent is found as

$$\begin{aligned}
\hat{\Delta}_0^{bs}(\mathbf{r},\boldsymbol{\xi}) &= x+\Phi_k\xi_l-\Upsilon_k\delta_{l+1}, \\
\hat{\Delta}_l^{bs}(\mathbf{r},\boldsymbol{\xi}) &= b\bar{r}_1^l+\Phi_k\xi_l-\Upsilon_k\delta_{l+1}, \quad \text{for } l=1,...,L-1, \\
\hat{\Delta}_L^{bs}(\mathbf{r},\boldsymbol{\xi}) &= b\bar{r}_1^L.
\end{aligned} \tag{72}$$

Similarly to the other schemes, for which the distortion exponent is maximized by equating the exponents, we look for the power allocation $\boldsymbol{\xi}$, such that all distortion exponent terms $\Delta_l^{bs}(\hat{\mathbf{r}},\boldsymbol{\xi})$ in

(66) are equal.

Equating all distortion exponents $\hat{\Delta}_l^{bs}(\hat{\mathbf{r}}, \boldsymbol{\xi})$ for $l = 2, ..., L-1$, i.e., $\hat{\Delta}_{l-1}^{bs}(\hat{\mathbf{r}}, \boldsymbol{\xi}) = \hat{\Delta}_l^{bs}(\hat{\mathbf{r}}, \boldsymbol{\xi})$, we have

$$d_{sd}\left(\hat{r}_l, \xi_{l-1}, \xi_l\right) = br_l + d_{sd}\left(\hat{r}_{l+1}, \xi_l, \xi_{l+1}\right). \tag{73}$$

Since $\hat{r}_l = [(k+1)(\xi_{l-1} - \xi_l) - \epsilon_1]$, we have

$$d_{sd}\left(\hat{r}_l, \xi_{l-1}, \xi_l\right) = \Phi_k \xi_{l-1} - \Upsilon_k(\xi_{l-1} - \xi_l - \epsilon_1).$$

Substituting in (73), we find that the power allocations for $l \geq 2$ need to satisfy,

$$(\xi_l - \xi_{l+1}) = \eta_k(\xi_{l-1} - \xi_l) + \mathcal{O}(\epsilon_1),$$

where η_k is defined in (22) and $\mathcal{O}(\epsilon_1)$ denotes a term that tends to 0 as $\epsilon_1 \to 0$. Then, for $l = 2, ..., L-1$ we obtain

$$\xi_l - \xi_{l+1} = \eta_k^{l-1}(\xi_1 - \xi_2) + \mathcal{O}(\epsilon_1), \tag{74}$$

and ξ_l can be found as

$$\begin{aligned}
1 - \xi_l &= (1 - \xi_1) + \sum_{i=1}^{l-1}(\xi_i - \xi_{i+1}) + \mathcal{O}(\epsilon_1) \\
&= (1 - \xi_1) + \sum_{i=1}^{l-1}\eta_k^{i-1}(\xi_1 - \xi_2) + \mathcal{O}(\epsilon_1) \\
&= (1 - \xi_1) + (\xi_1 - \xi_2)\frac{1 - \eta_k^{l-1}}{1 - \eta_k} + \mathcal{O}(\epsilon_1).
\end{aligned}$$

Then, for $l = 2, ..., L$, we have

$$\xi_l = \xi_1 - (\xi_1 - \xi_2)\frac{1 - \eta_k^{l-1}}{1 - \eta_k} + \mathcal{O}(\epsilon_1). \tag{75}$$

From $\hat{\Delta}_L^{bs}(\hat{\mathbf{r}},\boldsymbol{\xi})=b\bar{r}_1^L = b\sum_{i=1}^{L}(k+1)(\xi_{i-1}-\xi_i)$, we have

$$\begin{aligned}\hat{\Delta}_L^{bs}(\hat{\mathbf{r}},\boldsymbol{\xi}) &= b(k+1)(\xi_0-\xi_1)+b(k+1)(\xi_2-\xi_1)\sum_{i=1}^{L}\eta_k^{i-1}+\mathcal{O}(\epsilon_1)\\ &= b(k+1)\left[(\xi_0-\xi_1)+(\xi_2-\xi_1)\frac{1-\eta_k^{L-1}}{1-\eta_k}\right]+\mathcal{O}(\epsilon_1).\end{aligned} \tag{76}$$

Putting all together, from (72) we obtain

$$\begin{aligned}&\hat{\Delta}_0^{bs}(\hat{\mathbf{r}},\boldsymbol{\xi}) = x+\Phi_k\xi_0-\Upsilon_k(\xi_0-\xi_1-\epsilon_1),\\ &\hat{\Delta}_l^{bs}(\hat{\mathbf{r}},\boldsymbol{\xi}) = b(k+1)(\xi_0-\xi_1)+\Phi_k\xi_1-\Upsilon_k(\xi_1-\xi_2+\epsilon_1), \quad \text{for } l=1,...,L-1,\\ &\hat{\Delta}_L^{bs}(\hat{\mathbf{r}},\boldsymbol{\xi}) = b(k+1)[(\xi_0-\xi_1)+(\xi_2-\xi_1)\Gamma_k]+\mathcal{O}(\epsilon_1).\end{aligned} \tag{77}$$

By solving $\hat{\Delta}_{bs}^{L}(b,x)=\hat{\Delta}_0^{bs}(\hat{\mathbf{r}},\boldsymbol{\xi})=\hat{\Delta}_1^{bs}(\hat{\mathbf{r}},\boldsymbol{\xi})=\hat{\Delta}_L^{bs}(\hat{\mathbf{r}},\boldsymbol{\xi})$, and letting $\epsilon_1\to 0$, we obtain (23), and

$$\begin{aligned}\xi_1 \qquad &= \frac{(\Upsilon_k+\Phi_k\Gamma_k)(\Upsilon_k+b(k+1)-\Phi_k-x)}{(\Upsilon_k+b(1+k))(\Upsilon_k+b(1+k)\Gamma_k)-b(k+1)\Phi_k\Gamma_k},\\ \xi_1-\xi_2 &= \frac{\Phi_k(\Upsilon_k+b(k+1)-\Phi_k-x)}{(\Upsilon_k+b(1+k))(\Upsilon_k+b(1+k)\Gamma_k)-b(k+1)\Phi_k\Gamma_k}.\end{aligned} \tag{78}$$

For this solution to be feasible, the power allocation sequence has to satisfy $1\geq\xi_1\geq...\xi_L\geq 0$, i.e., $\xi_l-\xi_{l+1}\geq 0$. From (74) we need $\eta_k\geq 0$ and $\xi_1-\xi_2\geq 0$. We have $\eta_k\geq 0$ if $b\geq\frac{\Phi_{k+1}}{k+1}$, which holds in the regime characterized by (71). Then, $\xi_1-\xi_2\geq 0$ holds if $\Upsilon_k+b(k+1)-\Phi_k-x\geq 0$ and $(\Upsilon_k+b(1+k))(\Upsilon_k+b(1+k)\Gamma_k)-b(k+1)\Phi_k\Gamma_k\geq 0$. It can be shown that $(\Upsilon_k+b(1+k))(\Upsilon_k+b(1+k)\Gamma_k)-b(k+1)\Phi_k\Gamma_k$ is monotonically increasing in $b\geq 0$, and positive for $k=0,...,M_*-1$. Therefore, we need to check if $\Upsilon_k+b(k+1)-\Phi_k-x\geq 0$. This holds since this condition is equivalent to

$$b\geq\frac{\Phi_{k+1}+x}{k+1}.$$

Note that, in this regime, we have $\xi_1\geq 0$. In addition, $\xi_l=\xi_1+(\xi_1-\xi_2)\Gamma_k\geq 0$. Therefore, for each k the power allocation is feasible in the regime characterized by (71). It can also be checked that $br_1>x$ is satisfied. This completes the proof.

A. *Convergence for $L \to \infty$.*

In the limit of infinite layers, i.e., $L \to \infty$, this scheme converges to $\hat{\Delta}_{bs}^{\infty}(b,x) = \max\{x, b(k+1)\}$ when $0 \leq \eta_k < 1$, i.e.,

$$b \in \left[\frac{\Phi_{k+1}+x}{k+1}, \frac{\Phi_k}{k+1}\right),$$

and it converges to

$$\hat{\Delta}_{bs}^{\infty}(b,x) = \Phi_k + x\left(\frac{b(k+1)-\Phi_k}{b(k+1)-\Phi_{k+1}}\right),$$

when $\eta_k \geq 1$, that is, for

$$b \in \left[\frac{\Phi_k}{k+1}, \frac{\Phi_k + x}{k}\right).$$

References

[1] A. Wyner, "The rate-distortion function for source coding with side information at the decoder," *Information and Control*, vol. 38, no. 1, pp. 60–80, Jan. 1978.

[2] S. Shamai, S. Verdú, and R. Zamir, "Systematic lossy source-channel coding," *IEEE Trans. Information Theory*, vol. 44, no. 2, pp. 564–579, Mar. 1998.

[3] C. Ng, D. Gündüz, A. Goldsmith, and E. Erkip, "Distortion minimization in Gaussian layered broadcast coding with successive refinement," *IEEE Trans. on Information Theory*, vol. 55, no. 11, pp. 5074–5086, Nov. 2009.

[4] F. Etemadi and H. Jafarkhani, "Optimal layered transmission over quasi-static fading channels," in *Proc. IEEE Int. Symp. Information Theory (ISIT)*, Jul. 2006.

[5] M. Shaqfeh, W. Mesbah, and H. Alnuweiri, "Utility maximization for layered transmission using the broadcast approach," *IEEE Trans. on Wireless Communications*, vol. 11, no. 3, pp. 1228–1238, Mar. 2012.

[6] J. N. Laneman, D. N. C. Tse, and G. W. Wornell, "Cooperative diversity in wireless networks: Efficient protocols and outage behavior," *IEEE Trans. on Information Theory*, vol. 50, no. 12, pp. 3062–3080, Dec. 2004.

[7] D. Gündüz and E. Erkip, "Distortion exponent of parallel fading channels," in *Proc. IEEE Int'l Symposium on Information Theory Proceedings (ISIT)*, Jul. 2006, pp. 694–698.

[8] ——, "Source and channel coding for cooperative relaying," *IEEE Trans. on Information Theory*, vol. 53, no. 10, pp. 3454–3475, Oct. 2007.

[9] ——, "Joint source–channel codes for MIMO block-fading channels," *IEEE Trans. on Information Theory*, vol. 54, no. 1, pp. 116–134, Jan. 2008.

[10] D. Gündüz, A. Goldsmith, and H. Poor, "Distortion exponent in mimo channels with feedback," in *Proc. IEEE Information Theory Workshop (ITW)*, Oct. 2009, pp. 293–297.

[11] J. Wang and J. Liang, "Distortion exponents of two-way relay networks," *IEEE Trans. on Signal Processing*, vol. 59, no. 9, pp. 4424–4437, Sep. 2011.

[12] S. Zhao and D. Tuninetti, "On the distortion exponent of block-fading gaussian interference channels," in *Proc. 48th Annual Allerton Conference on Communication, Control, and Computing (Allerton).*, Oct. 2010, pp. 1266–1272.

[13] S. Zhao, R. Timo, T. Chan, A. Grant, and D. Tuninetti, "The impact of side-information on gaussian source transmission over block-fading channels," in *Proc. IEEE Int'l Conference on Communications (ICC)*, May 2010, pp. 1–5.

[14] G. Caire and K. Narayanan, "On the distortion SNR exponent of hybrid digital–analog space–time coding," *IEEE Trans. on Information Theory*, vol. 53, no. 8, pp. 2867–2878, Aug. 2007.

[15] K. Bhattad, K. Narayanan, and G. Caire, "On the distortion SNR exponent of some layered transmission schemes," *IEEE Trans. on Information Theory*, vol. 54, no. 7, pp. 2943–2958, 2008.

[16] C. Ng, C. Tian, A. Goldsmith, and S. Shamai, "Minimum expected distortion in Gaussian source coding with uncertain side information," in *Proc. IEEE Information Theory Workshop (ITW)*, Sep. 2007, pp. 454–459.

[17] I. Estella and D. Gündüz, "Joint Source-Channel Coding with Time-Varying Channel and Side-Information," *Submitted to IEEE on Information Theory*, 2013. [Online]. Available: http://arxiv.org/abs/1312.0932

[18] ——, "Expected distortion with fading channel and side information quality," in *Proc. IEEE Int'l Conference on Communications (ICC)*, Jun. 2011, pp. 1–6.

[19] ——, "Distortion exponent in fading MIMO channels with time-varying side information," in *Proc. IEEE Int. Symposium on Information Theory*, St. Petersburg, Russia, Aug. 2011, pp. 548 –552.

[20] ——, "Distortion exponent with side-information diversity," in *IEEE Global Conference on Signal and Information Processing (GlobalSIP)*, Dec. 2013, pp. 941–944.

[21] L. Zheng and D. Tse, "Diversity and multiplexing: A fundamental tradeoff in multiple-antenna channels," *IEEE Trans. on Information Theory*, vol. 49, no. 5, pp. 1073–1096, May. 2003.

[22] M. Wilson, K. Narayanan, and G. Caire, "Joint source channel coding with side information using hybrid digital analog codes," *IEEE Trans. on Information Theory*, vol. 56, no. 10, pp. 4922–4940, Oct. 2010.

[23] W. Equitz and T. Cover, "Successive refinement of information," *IEEE Trans. on Information Theory*, vol. 37, no. 2, pp. 269–275, Mar. 1991.

[24] Y. Steinberg and N. Merhav, "On successive refinement for the Wyner-Ziv problem," *IEEE Trans. on Information Theory*, vol. 50, no. 8, pp. 1636–1654, Aug. 2004.

[25] K. Bhattad, K. Narayanan, and G. Caire, "On the distortion snr exponent of some layered transmission schemes," *IEEE Trans. on Information Theory*, vol. 54, no. 7, pp. 2943–2958, Jul. 2008.

[26] M. Abramowitz and I. A. Stegun, *Handbook of mathematical functions: with formulas, graphs, and mathematical tables*. Courier Dover Publications, 2012.

[27] A. Dembo and O. Zeitouni, *Large Deviations Techniques and Applications*, 2nd ed. Springer, New York, 1998.