Joint Source and Channel Coding for MIMO Systems*

Tim Holliday, Andrea Goldsmith
Stanford University

Abstract

A significant amount of recent research has focused on characterizing the diversity-multiplexing tradeoff region in multiple antenna wireless systems. In this paper we focus on finding the point on this diversity-multiplexing region that minimizes end-to-end distortion. Our goal is to find the optimal balance between the increased data rate provided by multiplexing versus the error protection provided by diversity. We first present analytical results for the distortion achieved by concatenating a vector quantizer with a MIMO channel. We show that in the high SIR regime we can find a closed form expression for the end-to-end distortion as a function of the optimal point on the diversity-multiplexing tradeoff curve. We also show that this framework can be used to minimize end-to-end distortion for a broad class of source and channel codes. We demonstrate this with an example using progressive video encoding and space-time channel codes.

1 Introduction

Multiple antennas can significantly improve performance of wireless systems. For example, a data rate increase equal to the minimum number of transmit/receive antennas can be obtained by multiplexing data streams across the antennas. Alternatively, transmit or receive diversity can be utilized across multiple antennas to decrease the probability of error. Recently Zheng and Tse [5] demonstrated that both diversity and multiplexing can be accomplished simultaneously. However, there is a fundamental tradeoff between the two quantities: higher spatial multiplexing gain leads to lower diversity and vice versa. The main result in [5] is an explicit characterization of the diversity-multiplexing tradeoff region.

Our goal in this paper is to answer the following question: "Given the diversity-multiplexing region, where should one choose to operate?". In order to answer this question we require a performance metric from a higher layer. The metric of interest in this paper will be end-to-end distortion. Specifically, our system model consists of a source encoder concatenated with a MIMO channel encoder. Our goal is to determine the optimal point on the diversity-multiplexing region that minimizes the distortion due to both the source encoder and channel decoding errors.

This formulation differs from the traditional joint source-channel coding problem [1] in many ways. The traditional formulation determines the optimal fraction of bits to assign to the source encoder and channel encoder as the dimension of the source and

*This research is supported by NSF grants CCR-0325639 and CNS-0338807.

number of channel uses tend to infinity. The traditional result is that the source should be encoded at a rate arbitrarily close to the channel capacity. This rate can be achieved since we may code for the channel over arbitrarily long block lengths and drive the error probability to zero.

In this paper we consider a fundamentally different formulation. We assume that the dimension of our source and the block lengths over which we may code are both finite. This assumption is required for any system with delay constraints. However, as a result we will always have some probability of error in the channel, which induces the tradeoff between diversity and multiplexing. We determine the optimal tradeoff of diversity and multiplexing for minimizing end-to-end distortion.

The rest of this paper is organized as follows. In the next section we present the channel model and summarize the diversity-multiplexing tradeoff results from [5]. In Section 3 we present our source encoding framework and develop analytical bounds on the end-to-end distortion. Section 4 presents the optimization required to minimize end-to-end distortion as well as closed form expressions for the optimal operating point on the diversity-multiplexing curve. In Section 5 we present a similar formulation for optimizing diversity and multiplexing in progressive video transmission using space-time codes. We conclude in Section 6.

2 Channel Model

We will use the same channel model and notation from [5]. Consider a wireless link with M transmit antennas and N receive antennas. The fading coefficients h_{ij} that model the gain from transmit antenna i to receive antenna j are i.i.d. complex Gaussian with unit variance. The channel gain matrix H with elements $H(i,j) = (h_{ij} : i \in \{1, \ldots M\}, j \in \{1, \ldots, N\})$ is assumed to be known at the receiver and unknown at the transmitter. We assume that the channel remains constant over a block of T symbols, while each block is i.i.d. Therefore, in each block we can represent the channel as

$$Y = \sqrt{\frac{SNR}{M}} HX + W, \tag{1}$$

where $X \in \mathcal{C}^{M\text{x}T}$ and $Y \in \mathcal{C}^{N\text{x}T}$ are the transmitted and received signals, respectively. The additive noise vector W is i.i.d. complex Gaussian with unit variance.

We construct a family of codes for this channel $\{C(SNR)\}$ of block length T for each SNR level. Define $P_e(SNR)$ as the average probability of error and $R(SNR)$ as the number of bits per symbol for the codebook. A channel code scheme $\{C(SNR)\}$ is said to achieve multiplexing gain r and diversity gain d if

$$\lim_{\log SNR \to \infty} \frac{R(SNR)}{\log SNR} = r, \tag{2}$$

and

$$\lim_{\log SNR \to \infty} \frac{\log P_e(SNR)}{\log SNR} = -d. \tag{3}$$

For each r we define the optimal diversity gain $d^*(r)$ as the supremum of the diversity gain achieved by any scheme. The main result from [5] that we will require in the next section is summarized in the following statement.

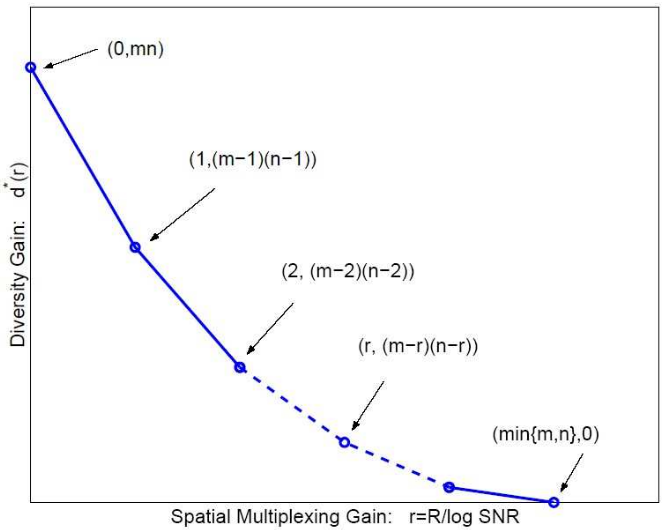

Figure 1: The optimal diversity-multiplexing tradeoff for $T \geq M + N - 1$.

Diversity-Multiplexing Tradeoff [5]: Assume the block length $T \geq M+N-1$. Then the optimal tradeoff between diversity gain and multiplexing gain is $d^*(r) = (M-r)(N-R)$, for $0 \leq r \leq \min(M, N)$. This function $d^*(r)$ is plotted in Figure 1.

In this framework the rate of the codebook $\{C(SNR)\}$ must scale with $\log SNR$, otherwise the multiplexing gain will go to zero. Hence, in the following sections we will assume, without loss of generality, that the rate of the codebook is $Tr \log SNR$ for any choice of $0 \leq r \leq \min(M, N)$ and block length T. We also assume that the codebook achieves the optimal diversity gain $d^*(r)$ for any choice of r.

3 End-to-End Distortion Model

This section presents our system model for the end-to-end transmission of source data. The source coding setup and notation follow [3]. We assume the original source data u is a random variable with probability density $f(u)$, which has support on a closed, bounded subset of $\Re^k$ with non-empty interior. A s-bit quantizer is applied to u via the following transformation:

$$Q(u) = \sum_{i=1}^{2^s} v_i I_{A_i}, \tag{4}$$

where I_{A_i} is the standard indicator function, and $\{A_i\}_{i=1}^{2^s}$ is a partition of $\Re^k$ into disjoint regions. Each region A_i is represented by a single codevector v_i.

We assume the encoder/decoder pair achieves the high-resolution noiseless distortion [1]

$$D_s(Q) = 2^{-ps/k+O(1)}, \tag{5}$$

as $s \to \infty$, where

$$D_s(Q) = \sum_{i=1}^{2^s} \int_{A_i} ||u - v_i|| f(u) du, \tag{6}$$

and $||u - v_i||^p$ is the pth power of the Euclidian norm. The high-resolution asymptotic regime is often used in source coding theory to achieve analytical results. As noted in

Section 1, in this paper we consider the high SIR regime. However, we also require the rate of our channel codebook $\{C(SNR)\}$ to scale as $r \log SNR$. Hence, the source coder will receive an increasing number of bits as $SNR \to \infty$, placing us in the high-resolution regime.

Assume that the rate of the channel codebook $\mathcal{C}\{SNR\}$ is matched to the rate of the quantizer (i.e. $s = Tr \log SNR$). Each codevector from the quantizer $v_1, \ldots, v_{2^s}$ is mapped into a codeword from $\mathcal{C}\{SNR\}$ through a permutation mapping π. We assume the mapping π is chosen equally likely at random from the $2^s!$ possibilities. The codeword $\pi(i)$ is transmitted over the channel in (1) and decoded at the receiver. Let $q(\pi(j)|\pi(i))$ be the probability that codeword $\pi(j)$ is decoded at the receiver given that $\pi(i)$ was transmitted. In general, the probability $q(\cdot|\cdot)$ will depend on the SNR, the quantizer Q, and the permutation mapping π. Hence, we can write the end-to-end distortion as follows,

$$D_T(Q, SNR, \pi) = \sum_{i=1}^{2^s} \sum_{j=1}^{2^s} q(\pi(j)|\pi(i)) \int_{A_j} ||u - v_j||^p f(u) du. \tag{7}$$

It is shown in [3] that (7) can be bounded by

$$D_T(Q, SNR, \pi) \leq D_s(Q) + O(1) \sum_{i=1}^{2^s} P(A_i) P_{e|\pi(i)}, \tag{8}$$

where $P_{e|\pi(i)}$ is the probability of codeword error given that codeword $\pi(i)$ was transmitted. One arrives at this bound by splitting (7) into two pieces; one corresponding to correctly received codewords and the other corresponding to erroneous decoding. The term corresponding to correct transmission is bounded by the noiseless distortion $D_s(Q)$ while the term corresponding to errors is bounded by a constant multiplied by the average block error probability.

By construction, the rate of our channel codebook (and hence the source encoder) is $s = Tr \log SNR$, therefore

$$D_s(Q) = 2^{-ps/k + O(1)} = 2^{-\frac{pTr}{k} \log SNR + O(1)}, \tag{9}$$

as $s \to \infty$, or as $\log SNR \to \infty$. Moreover, from the diversity/multiplexing results summarized in Section 1 we have the following equivalence for the probability of codeword error

$$P_e = 2^{-d^*(r) \log SNR + o(\log SNR)} = 2^{(N-r)(M-r) \log SNR + o(\log SNR)}, \tag{10}$$

as $\log SNR \to \infty$. We may then write the bound on total distortion as

$$D_T(Q, SNR, \pi) \leq 2^{-\frac{pTr}{k} \log SNR + O(1)} + 2^{(N-r)(M-r) \log SNR + o(\log SNR)}, \tag{11}$$

as $\log SNR \to \infty$. The terms in (11) provide us with an explicit characterization of the diversity-multiplexing tradeoff and its impact on end-to-end distortion. The first term in (11), corresponding to the noiseless encoder distortion, is strictly decreasing in the multiplexing rate r. The second term, corresponding to channel error probabilities, is strictly increasing with r. Hence, it is clear that there will be an optimal choice of $0 \leq r \leq \min(N, M)$.

4 Minimizing Total Distortion

In order to achieve analytic results for the minimum distortion bound we consider the asymptotic regime of $\log SNR \to \infty$. In general, we minimize an exponential sum of the form $2^{f(r)\log SNR} + 2^{g(r)\log SNR}$ by choosing the exponents $f(r)$ and $g(r)$ to be within $O(1)$ of each other. (Note that if the exponents were not of the same order then one term in the sum would dominate the other as $\log SNR \to \infty$.) In our particular case this means we should choose r such that

$$d^*(r) \quad = \quad (N-r)(M-r) = \frac{pTr}{k} + o(1). \tag{12}$$

Solving for r in (12) yields the optimal multiplexing rate of the codebook. Note however that we may not always have a solution to this equation. If the block length T is substantially larger than the dimension of the source vector k then we may not have enough antennas to create the diversity required to solve the equation. Likewise, if k is much larger than T then we will not have enough antennas to generate the multiplexing rate required to match the exponents. For the case where k and T are of the same order we have the following result.

Theorem 1: If $\frac{1}{\min(M,N)-1} \le \frac{pT}{k} \le (M-1)(N-1)$, then choosing r^* to satisfy (12) will minimize the bound (11) on the end-to-end distortion $D_T(Q, SNR, \pi)$, as $\log SNR \to \infty$.

If we can match the exponents in (11) we also have the following bound on the rate at which total distortion tends to zero with SNR.

Theorem 2: If $d^*(r^*) = \frac{pTr}{k}$ then

$$\lim_{\log SNR \to \infty} \frac{D_T(Q, SNR, \pi)}{\log SNR} \le -d^*(r^*). \tag{13}$$

Proof: We have

$$\begin{aligned} D_T(Q, SIR, \pi) \quad &\le \quad 2^{-\frac{pTr}{k}\log SNR + O(1)} + 2^{(N-r)(M-r)\log SIR + o(\log SIR)} \\ &= \quad 2^{-d^*(r^*)\log SNR}\, 2^{O(1)+o(\log SNR)}. \end{aligned}$$

Then as $\log SNR \to \infty$

$$\begin{aligned} \lim_{\log SNR \to \infty} \frac{\log\left[2^{-d^*(r^*)\log SNR}\, 2^{O(1)+o(\log SNR)}\right]}{\log SIR} \quad &= \quad -d^*(r^*) \lim_{\log SNR \to \infty} \frac{\log\left[2^{O(1)+o(\log SNR)}\right]}{\log SIR} \\ &= \quad -d^*(r^*). \quad \blacksquare \end{aligned}$$

4.1 Non-asymptotic Bounds

We may also consider the behavior of our distortion bound and the corresponding choice of r^* for large, but finite, SNR. In this case we must solve the following convex optimization to find the optimal diversity-multiplexing tradeoff.

$$\begin{aligned} \min_{r} \quad & 2^{-\frac{pT}{k} r \log SIR} + 2^{-(M-r)(N-r)\log SNR} \\ \text{s.t.} \quad & 0 \le r \le \min(M, N). \end{aligned} \tag{14}$$

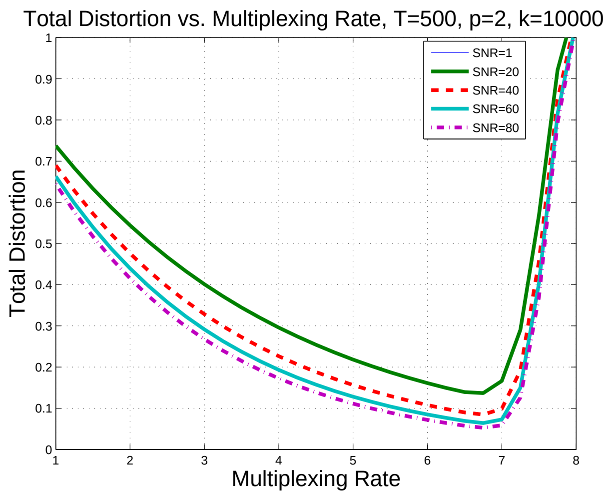

Figure 2: Total distortion vs. number of antennas assigned to multiplexing for differing levels of SIR. T is substantially smaller than k.

Figures 2-4 contain plots that compare the total end-to-end distortion versus the number of antennas assigned to multiplexing. Each plot contains four curves that represent different SNR levels. The difference between the three plots is the ratio of the block length T to source vector dimension k. Notice that for T much smaller than k we will use almost all of our antennas for multiplexing. For k much smaller than T we will use nearly all of our antennas for diversity. For k of the same order as T we will choose an intermediate number of antennas for multiplexing and diversity.

It is interesting to note that we will never choose full multiplexing or full diversity for any SNR in this optimization since either choice would cause a term in (14) to go to one. This result runs contrary to intuition and requires careful interpretation. In the diversity and multiplexing gain definitions in Section 1, a selected diversity gain of 0 does not necessarily correspond to a probability of error equal to 1. Rather, $d^*(r) = 0$ tells us that the probability of error is $O(1)$ and does not tend to zero as $SNR \to \infty$. Hence, if we examine the bound in (11) and choose $d^*(r) = 0$; then in the high-SNR regime we will have the first term tend to zero with the second term remaining strictly greater than zero. Clearly this will result in sub-optimal end-to-end distortion at high SNR, and therefore we will always choose to have at least some diversity in our system. A similar converse argument can be made to show that we will always choose to have some multiplexing gain in the system as well.

5 Progressive Video Encoding

While the results in the previous section lead to an interesting analytical result, they only apply to a very narrow class of source/channel encoders and distortion metrics. However, the basic optimization problem (14) can be applied to a broad class of encoders and distortion measures. Furthermore, this optimization can be applied in non-asymptotic settings, thereby allowing us to study the diversity-multiplexing tradeoff in current and future operational wireless systems. In this section we present an example of end-to-end distortion optimization, via the diversity-multiplexing tradeoff, for source/channel

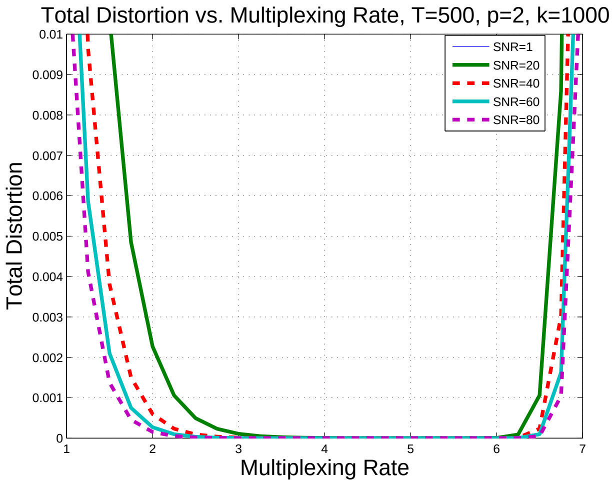

Figure 3: Total distortion vs. number of antennas assigned to multiplexing for differing levels of SIR. T is on the same order as k.

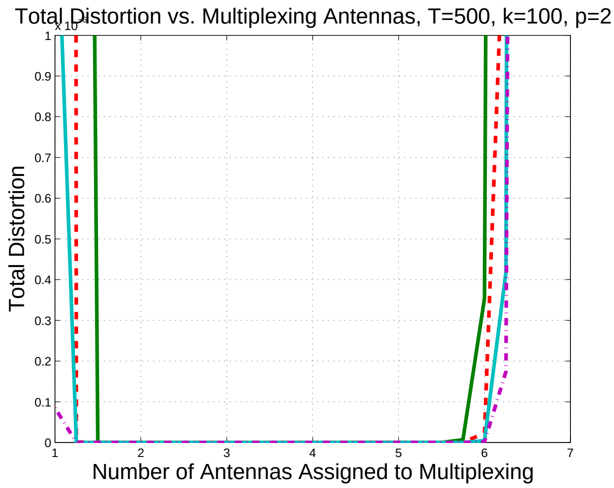

Figure 4: Total distortion vs. number of antennas assigned to multiplexing for differing levels of SIR. T is substantially larger than k.

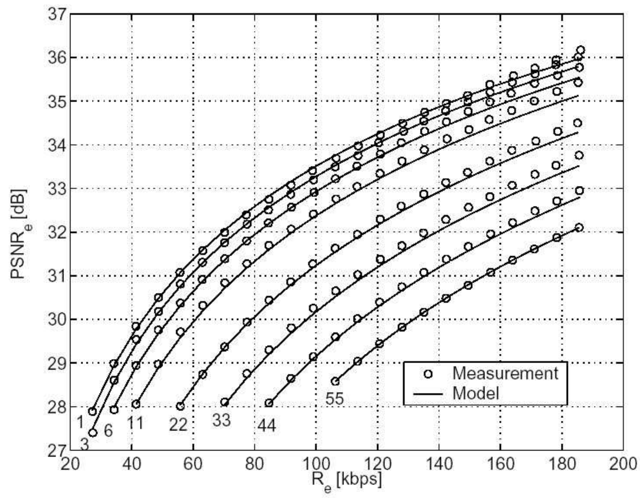

Figure 5: PSNR vs. Rate for different encoding options of the source encoder presented in [2]. We use the topmost curve corresponding to $\beta = .01$.

distortion models that are fitted to real video streams and wireless channels.

We use the progressive video encoder model developed in [2]. The overall mean-square distortion is evaluated as

$$D_T = D_e + D_c, \tag{15}$$

where D_e is the distortion induced by the source encoder and D_c is the distortion created by errors in the channel. Although the total distortion is represented by two separate components, each component shares some common terms so we will still have a tradeoff between diversity and multiplexing.

The model for D_e consists of a six parameter analytic formula that is fitted to a particular traffic stream. Due to space constraints we will not present the full model here. Figure 5 contains a plot of the PSNR vs. rate for this source encoder. Each curve in Figure 5 represents a different encoding method parameterized by a scalar β. We will use the topmost curve ($\beta = .01$) for our example as it provides the best PSNR for any given rate. This source encoder setting also provides the highest sensitivity to channel errors, which allows us to highlight the tradeoff between multiplexing and diversity in our optimization.

The model for the channel distortion D_c is fitted to the following equation,

$$D_c = \sigma^2 P_e(N_u) \left[\frac{\gamma + \beta}{\gamma} \ln \left(1 + \frac{\gamma}{\beta} \right) - \frac{1}{\gamma} + \frac{1}{2} \right], \tag{16}$$

where the parameters σ^2, γ, and β must be estimated for a particular source encoder and traffic stream. N_u is the number of antennas utilized for multiplexing and $P_e(N_u)$ is the probability of codeword error as a function of N_u.

Our channel transmission scheme follows the setup in [4]. We utilize 8 transmit and 8 receive antennas with a set of linear space-time codes that can trade off multiplexing for diversity. The plot in Figure 6 shows the probability of error for these codes as a function of E_b/N_0. The curve corresponding to $N_u = 8$ shows the error probability of the full multiplexing scheme (i.e. the highest error probability). The curves for $N_u = 4$, $N_u = 2$, and $N_u = 1$ show the decrease in error probability as fewer antennas are assigned to multiplexing and utilized to increase diversity.

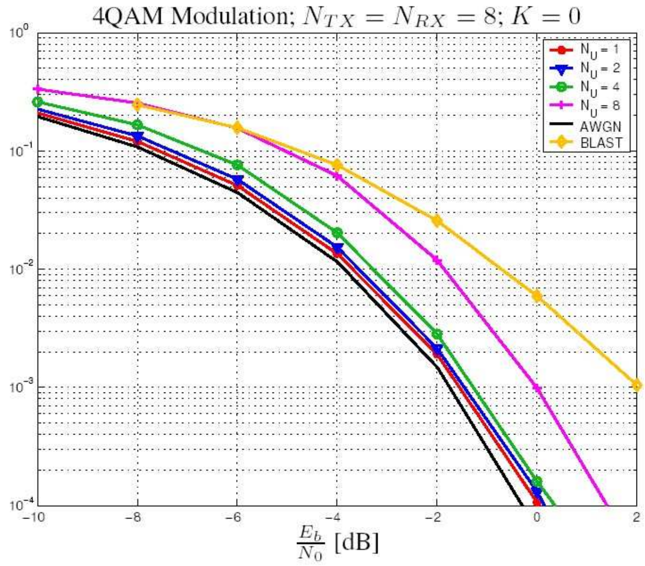

Figure 6: Probability of error vs. E_b/N_0 for the set of linear space-time codes used in our example. The parameter N_u is the number of antennas assigned to multiplexing. The error performance of BLAST and the full diversity AWGN channel are provided for comparison.

Since this channel coding scheme does not permit us to assign fractions of antennas we must solve the following integer program for the optimal distortion and number of multiplexing antennas (rather than a convex program),

$$\begin{aligned} \min_{N_u} \quad & D_e + D_c \\ s.t. \quad & N_u \in \{1, 2, 4, 8\}. \end{aligned} \tag{17}$$

Figure 7 contains a set of curves that show the total distortion achieved as a function of the number of antennas assigned to multiplexing. The uppermost curve corresponds to the lowest SNR and the bottom curve corresponds to the highest SNR. We can see that we have an explicit tradeoff here that depends on SNR. At low SNR the total distortion is minimized by assigning few antennas to multiplexing in order to improve error performance. As SNR increases we assign more antennas to multiplexing to take advantage of the improved channel. One significant difference between this plot and the asymptotic results in Section 3 is that here we do assign our antennas to full multiplexing as the SNR becomes large. The reason we observe this behavior is that the rate of our codebook in this example does not scale with SNR. We have a finite number of space-time coding schemes and as the SNR becomes large we eventually reach a point where we would prefer to move to a higher rate code that is not available. Hence, the optimal choice in this case is to eventually move to full multiplexing. The implication of this result is that in order to fully take advantage of a MIMO system we require a sufficiently robust collection of channel codes that take advantage of all available SNRs.

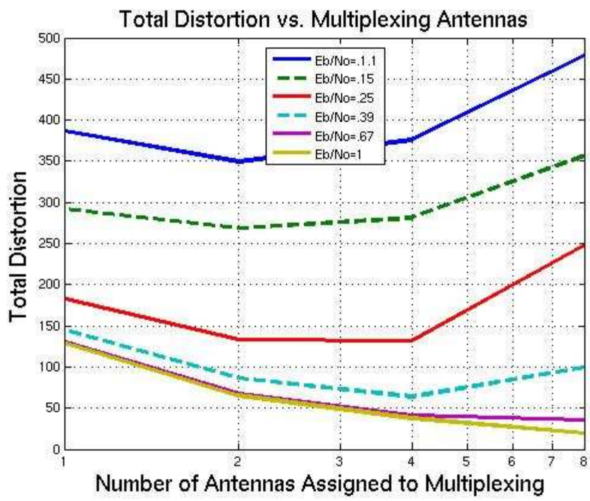

Figure 7: Total distortion vs. number of antennas assigned to multiplexing for differing levels of SIR.

6 Conclusion

We investigate the optimal diversity-multiplexing tradeoff in terms of minimizing end-to-end distortion. We derive an analytical formula that gives the optimal multiplexing for this minimization. We also show the exponential rate of decrease for the distortion is bounded by the diversity gain associated with the optimal multiplexing rate. We show that the framework is applicable to a broad class of source and channel codes. We apply the framework to progressive video encoding with space-time codes and obtain numerical results indicating end-to-end distortion and the optimal use of antennas for diversity and multiplexing.

References

[1] A. Gersho, "Asymptotically Optimal block quantization", *IEEE Trans. Inform. Theory* , Vol. 25, pp. 373-380, July 1979.

[2] B. Girod, K. Stuhlmuller, N. Farber, "Trade-Off Between Source and Channel Coding for Video Transmission", *Proc. of ICIP'2000*, Vol. 1, pp. 399-402, Sept. 2000.

[3] B. Hochwald, K. Zeger, "Tradeoff Between Source and Channel Coding", *IEEE Trans. Inform. Theory* , Vol. 43 , pp. 1412 - 1424, Sept. 1997.

[4] M. Kuhn, I. Hammerstroem, A. Wittneben, "Linear Scalable Space-Time Codes: Tradeoff Between Spatial Multiplexing and Transmit Diversity", *Proc. of SPAWC 2003*, June 2003.

[5] L. Zheng, D. Tse, "Diversity and Multiplexing: the Optimal Tradeoff in Multiple Antenna Channels", *IEEE Trans. Inform. Theory* , Vol. 49, pp. 1073 - 1096, May 2003.