Adversarial interference models for multiantenna cooperative systems

Anand D. Sarwate and Michael Gastpar
Department of Electrical Engineering and Computer Sciences
University of California, Berkeley
Berkeley, CA 94720, USA
Email: {asarwate, gastpar}@eecs.berkeley.edu

***Abstract*— According to the cognitive radio paradigm, a terminal or subsystem will opportunistically select a frequency band for data transmission. Because the spectrum is shared, a cognitive system may face interference that cannot be given a statistical description. An adversarial interference model may be appropriate for finding achievable rates under these conditions. In this work we investigate the benefit offered by using many antennas. In addition to adding spatial diversity and multiplexing gains, multiple antennas can allow a cognitive system to mitigate the effects of adversarial interference that is known to come from a system with fewer antennas.**

I. Introduction

Cognitive radio [1] refers to a class of wireless systems that use sensing to adapt their behavior in order to coexist with other systems. Cognitive radios have been proposed as a solution for exploiting unused spectrum by adopting flexible, frequency-agile devices. The spectrum is licensed to a *primary* user who shares the band with *secondary* or "cognitive" users. With the Federal Communication Commission's recent decision to allow spectrum reuse in the 700 MHz band [2], these systems are moving ever closer to reality and have raised a number of interesting theoretical questions. One popular information theoretic model for cognitive radio, due to Devroye, Mitran, and Tarokh [3], modifies an interference channel to give a "cognitive" user access to the other user's message noncausally. The cognitive user can then optimize its own rate subject to conditions guaranteeing minimal degradation to the other "primary" user. Gastpar [4] has looked at coexistence from the perspective of limiting the received power at a primary system. These optimistic perspectives on coexistence assume a high degree of coordination and cooperation between primary and secondary or cognitive systems.

In this work we do not address the coexistence conditions posed by the cognitive radio paradigm but instead focus on the issue of interference modeling. More specifically, we will study a model for multiple-input multiple-output (MIMO) Gaussian channels [5] in which additional interference comes from the primary system. An isolationist approach is to treat the interfering signal from the primary as additional noise. A game theoretic model for this problem, carried out most fully by Baker and Chao [6], uses the mutual information as a payoff between one player who can choose a transmit covariance matrix and another who can choose the noise covariance matrix. This approach relies on the convexity of the space of allowable noise matrices, which may not always hold. We will explicitly model the number of antennas used by an interfering system and demonstrate that using this knowledge leads to a significant rate increase when the receiver has more than one antenna.

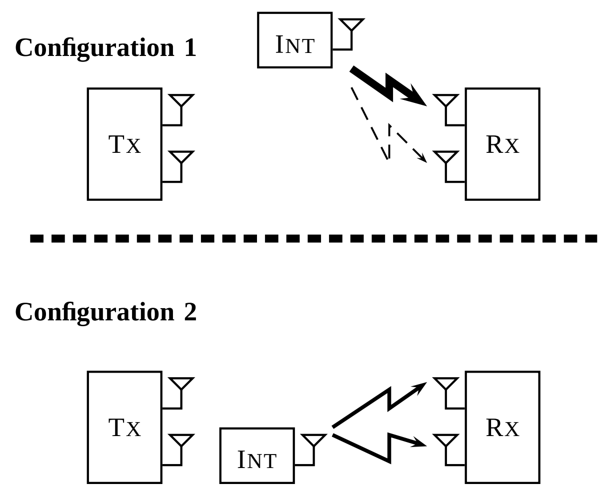

Fig. 1. Two different configurations for a system with a single-antenna interferer. Because the location of the interferer is unknown, the subspace in which the interference lies may be unknown prior to transmission.

To illustrate our problem, consider the two possible configurations shown in Figure 1. A 2×2 MIMO system is subject to unknown interference from a single antenna system. Because the location of the interferer is not known prior to transmission, the MIMO system must choose a rate and coding scheme that will work regardless of the interferer's location. The fact that the interferer has a single antenna means that the interference lies in a one-dimensional subspace of the received signal. One approach is to ignore this information and target a rate that only takes into account a power constraint on the interference. We will see that this is a suboptimal strategy, particularly in the limit of large interference power.

We will examine these issues through an information theoretic model known as the arbitrarily varying channel (AVC) [7], [8]. In the AVC the interference is modeled by a time varying state that is subject to an average power constraint but is otherwise arbitrary. We must guarantee reliable communication under all state realizations, so we can assume that the state is controlled by a malicious *jammer* whose objective is to maximize the probability of decoding error. The capacity expressions depend crucially on the error model and whether the transmitter and receiver are allowed to *randomize* their strategy. We will investigate two particular combinations in this work – the maximal-error capacity under randomized coding C_r and the average-error capacity under deterministic coding $\bar{C}_d$.

The class of channel models we will consider are multiple-input multiple-output (MIMO) arbitrarily varying channels. We will call a channel with M_t transmitter antennas, M_r receiver antennas and M_j jammer a (M_t, M_r, M_k) MIMO AVC. For the $(M, 1, 1)$ channel we can use existing results to show how extra transmit antennas improve the capacity. For the (M, M, M) channel, capacities have been found by Hughes and Narayan [9] and by Csiszár [10]. The solution is given by a "mutual waterfilling" procedure. The jammer allocates its power via waterfilling [11] over the noise spectrum, and then the transmitter allocate power by waterfilling over the combined noise-plus-interference spectrum. The main example that we will consider is the $(2, 2, 1)$ AVC, for which we will show that the transmitter and receiver can exploit the fact that the jammer only has a single antenna to increase the capacity above the $(2, 2, 2)$ case.

II. THE (M,1,1) CHANNEL : EXTRA TRANSMIT ANTENNAS

As a warm-up, we turn first to the $(M, 1, 1)$ channel model in which we augment a simple point-to-point single-input single-output (SISO) channel by adding more antennas at the transmitter. By applying standard results from the AVC literature, we can characterize the benefits of adding extra antennas in both the randomized and deterministic coding settings.

Consider a channel with multiple transmit antennas and a single receive antenna:

$$Y(t) = \mathbf{h}^T \mathbf{X}(t) + W(t) + S(t) \ . \tag{1}$$

The input $\mathbf{X} \in \mathbb{R}^M$ and interference S satisfy total power constraints:

$$\sum_{t=1}^{N} \|\mathbf{X}(t)\|^2 \leq n\Gamma \tag{2}$$

$$\sum_{t=1}^{N} S(t)^2 \leq n\Lambda \ . \tag{3}$$

The noise $W(t)$ is assumed to be iid with Gaussian distribution $\mathcal{N}(0, \sigma^2)$. We will assume that the vector of channel gains $\mathbf{h}$ is known to both the transmitter and receiver. In the arbitrarily varying channel, we do not make a distributional assumption on the interference S, and in fact we assume that the jammer's input S could be chosen arbitrarily.

Let $[N] = \{1, 2 \ldots, N\}$. An (n, N) deterministic code for this channel is a pair of maps (ϕ, ψ), where the encoder is $\phi : [N] \to \mathbb{R}^{n \times M}$ and the decoder is $\psi : \mathbb{R}^n \to [N]$. We require $\|\phi(i)\|^2 \leq n\Gamma$ for all i in order to satisfy the input power constraint. An (n, N, K) randomized code is a pair of random variables (Φ, Ψ) taking values in the set of deterministic codes. We will assume that the pair (Φ, Ψ) is uniformly distributed on a set $\{(\phi_k, \psi_k) : k \in [K]\}$ of $K = K(n)$ random codes. We will define the key size for a randomized code to be $K(n)$. The maximal probability of error for a randomized code is given by

$$\varepsilon = \max_{\mathbf{s}:\|\mathbf{s}\|^2 \leq n\Lambda} \max_{i \in [N]} \frac{1}{K} \sum_{k=1}^{K} \mathbb{P}\left(\psi_k(\mathbf{Y}) \mid \phi_k(i), \mathbf{s}\right) \ . \tag{4}$$

The average error for a deterministic code is given by

$$\bar{\varepsilon} = \max_{\mathbf{s}:\|\mathbf{s}\|^2 \leq n\Lambda} \frac{1}{N} \sum_{i=1}^{N} \mathbb{P}\left(\psi(\mathbf{Y}) \mid \phi(i), \mathbf{s}\right) \ . \tag{5}$$

A rate R is said to be achievable under maximal error using randomized coding if there exists a sequence of (n, N, K) randomized codes with $\varepsilon \to 0$ as $n \to \infty$. A rate R is said to be achievable under average error using deterministic coding if there exists a sequence of (n, N) deterministic codes with $\bar{\varepsilon} \to 0$ as $n \to \infty$.

In the case where $M = 1$, it is well-known that if the encoder and decoder are permitted to use joint randomization that the capacity is equal to the additive white Gaussian noise (AWGN) channel:

$$C_r(\Gamma, \Lambda) = \frac{1}{2} \log \left(1 + \frac{h^2 \Gamma}{\sigma^2 + \Lambda}\right) \tag{6}$$

However, if common randomness is not available, the capacity is zero unless the received signal power exceeds that of the interference:

$$\bar{C}_d(\Gamma, \Lambda) = \begin{cases} C_r(h^2\Gamma, \Lambda) & h^2\Gamma > \Lambda \\ 0 & h^2\Gamma \leq \Lambda \end{cases} \tag{7}$$

These results extend straightforwardly to the case with multiple transmit antennas.

A. Randomized coding

In [12] we showed that the amount of common randomness needed to achieve the randomized coding capacity grows at most logarithmically in the blocklength n of the code. Thus the amount of overhead to achieve $C_r(h^2\Gamma, \Lambda)$ may be quite modest.

Proposition 1: For the Gaussian AVC with M transmit antennas and a single receive antenna, full channel state information, and power constraint Γ, the randomized coding capacity under maximal error is given by

$$C_r(\Gamma, \Lambda) = \frac{1}{2} \log \left(1 + \frac{\|\mathbf{h}\|^2 \Gamma}{\sigma^2 + \Lambda}\right) \ . \tag{8}$$

Proof: This result simply follows from the point-to-point Gaussian AVC under randomized coding. To find the capacity, we must find the optimal power allocation to the antennas. This is simply a matter of maximizing the received power

$$\left(\sum_{m=1}^{M} h_m \sqrt{\Gamma_m}\right)^2 \tag{9}$$

subject to the constraint that $\sum_m \Gamma_m \leq \Gamma$. The optimal allocation assigns $\Gamma_m = (h_m^2/\|\mathbf{h}\|^2)\Gamma$, which makes the received power $\|\mathbf{h}\|^2 \Gamma$. ∎

B. Deterministic coding

When the transmitter and receiver can jointly randomize, they can induce a random distribution on the jammer's input to make it similar in distribution to additional Gaussian noise. However, in certain situations secret key agreement prior to transmission may not be possible, so it is of interest to see the benefits that multiple antennas can have in reducing the threshold behavior of the deterministic coding capacity. We begin with the trivial observation that extra antennas reduces the threshold on the transmitter's power.

Proposition 2: For the Gaussian AVC with M transmit antennas and a single receive antenna, full CSI, and power constraint Γ, the deterministic coding capacity under average error is given by

$$\bar{C}_d(\Gamma, \Lambda) = \frac{1}{2} \log \left(1 + \frac{\|\mathbf{h}\|^2 \Gamma}{\sigma^2 + \Lambda} \right) \tag{10}$$

if $\Gamma > \Lambda / \|\mathbf{h}\|^2$, and 0 otherwise.

Proof: Because the threshold for the Gaussian AVC depends on the received power [8], we can simply find the threshold in the randomized coding capacity given by Proposition 1. ∎

The assumption of full CSI implies that the decoder could feed back some information to the transmitter. Complete information about the channel gains $\mathbf{h}$ could require an amount of feedback commensurate with $O(\log n)$, which could also enable randomized communication. Feedback of only M bits is sufficient to inform the transmitter of the signs of h_m for each m. It can then use an equal power allocation to make the received signal power equal to

$$\frac{\Gamma}{M} \left(\sum_{m=1}^{M} |h_m| \right)^2 . \tag{11}$$

We can extend this line of reasoning to channels with inputs and outputs in $\mathbb{C}$, subject to phase fading, by looking at the tradeoff between the amount of feedback and the reduction in the power threshold for each M.

Consider the channel in (1) with inputs, outputs, and interference taking complex values and noise $W(t) \sim \mathcal{CN}(0, \sigma^2)$. For simplicity, let $\mathbf{h}$ be a vector of phase shifts, so $h_m = \exp(j2\pi\phi_m)$. The decoder knows $\mathbf{h}$ and can quantize the phases and send them to the transmitter. Suppose that the receiver uses kM bits to uniformly quantize the M phases. The quantization error δ_m for each phase is at most 2^{-k}.

Proposition 3: Let

$$P(M) = \frac{\Gamma}{M} \left(\frac{M+1}{2} \right)^2 \cos^2(2\pi 2^{-k}) + \frac{\Gamma}{M} \sin^2(2\pi 2^{-k}) \ , \tag{12}$$

for M odd, and

$$P(M) = \frac{\Gamma}{M} \left(\frac{M}{2} \right)^2 \cos^2(2\pi 2^{-k}) \tag{13}$$

for M even. Then the following rate is achievable on the complex M-antenna MISO AVC with phase fading and k-bits of quantized phase information per antenna:

$$R = \begin{cases} \log\left(1 + \frac{P(M)}{\Lambda + \sigma^2}\right) & P(M) > \Lambda \\ 0 & P(M) \leq \Lambda \end{cases} \tag{14}$$

Proof: If the transmitter does a uniform power allocation to each of the antennas, the received signal power can be lower bounded by $P(M)$. ∎

III. The (2,2,1) channel : rank-limited jamming

We saw in the $(M, 1, 1)$ case that under randomized coding, the benefits of adding extra antennas at the transmitter were the same as in the standard Gaussian MISO channel. In particular, the single-antenna constraint on the jammer could not be exploited because the receiver was also limited to a single antenna. If we add a second antenna at the receiver, the story changes considerably. In this section we will give a characterization for the easiest non-trivial MIMO channel, the $(2, 2, 1)$ MIMO AVC.

For simplicity, we will treat our MIMO channel as a vector Gaussian channel:

$$\mathbf{Y}(t) = \mathbf{X}(t) + \mathbf{g}S(t) + \mathbf{W}(t) \ , \tag{15}$$

where $\mathbf{X}$, $\mathbf{Y}$, and $\mathbf{W}$ taking values in $\mathbb{R}^2$, $\mathbf{g}$ is an arbitrary unit vector in $\mathbb{R}^2$, the interference $S(t)$ is subject to the same average power constraint $\sum_{t=1}^{n} S(t) \leq n\Lambda$, and the noise $\mathbf{W}(t) \sim \mathcal{N}(0, \Sigma_W)$ and is iid over time. The transmitter is also subject to a power constraint $\sum_{t=1}^{n} \|\mathbf{X}(t)\|^2 \leq n\Gamma$. We can, without loss of generality, take the noise covariance matrix to be diagonal, so $\Sigma_W = \mathrm{diag}(\sigma_1^2, \sigma_2^2)$. The interference is constrained to a rank-1 subspace, albeit an unknown one. We must therefore design a coding scheme that works for all values of $\mathbf{g}$.

In the case without the rank constraint on the interference, the jammer can also allocate power to all the degrees of freedom in this channel. This channel is equivalent to a vector Gaussian AVC [9] and the capacity for general M under randomized coding is known to be given by a "mutual waterfilling" strategy. Both the transmitter and jammer choose diagonal covariance matrices. The jammer chooses a covariance $\mathrm{diag}(\Lambda_1, \Lambda_2, \ldots, \Lambda_M)$ by waterfilling over the noise spectrum:

$$\lambda^* = \max \left\{ \lambda : \left(\lambda - \sigma_m^2\right)^+ \leq \Lambda \right\} \tag{16}$$

$$\Lambda_m = \left(\lambda^* - \sigma_m^2\right)^+ \ . \tag{17}$$

The transmitter then chooses a covariance $\mathrm{diag}(\Gamma_1, \Gamma_2, \ldots, \Gamma_M)$ based on this worst jamming strategy:

$$\gamma^* = \max \left\{ \gamma : \left(\gamma - \sigma_m^2 - \Lambda_m\right)^+ \leq \Gamma \right\} \tag{18}$$

$$\Gamma_m = \left(\gamma^* - \sigma_m^2 - \Lambda_m\right)^+ \ . \tag{19}$$

Hughes and Narayan [9] showed that this allocation is a saddle point for the mutual information and is achievable for the Gaussian AVC with randomized coding. Later, Csiszár [10] showed that the capacity for deterministic codes is also given by this allocation if $\Gamma > \Lambda$.

In what follows we will focus on randomized coding for these channels. The mutual information saddle point for the vector Gaussian AVC can be expressed in the following way:

$$\max_{\Sigma_X : \mathrm{tr}(\Sigma_X) \le \Gamma} \min_{\Sigma_S : \mathrm{tr}(\Sigma_S) \le \Lambda} \frac{1}{2} \log \frac{|\Sigma_X + \Sigma_S + \Sigma_W|}{|\Sigma_S + \Sigma_W|} = \min_{\Sigma_S : \mathrm{tr}(\Sigma_S) \le \Lambda} \max_{\Sigma_X : \mathrm{tr}(\Sigma_X) \le \Gamma} \frac{1}{2} \log \frac{|\Sigma_X + \Sigma_S + \Sigma_W|}{|\Sigma_S + \Sigma_W|} . \quad (20)$$

Proposition 4: For the $(M, M, 1)$ MIMO AVC, the following rate is achievable using randomized coding:

$$R_{\text{wfill}} = \sum_{m=1}^{M} \frac{1}{2} \log \left(1 + \frac{\Gamma_m}{\Lambda_m + \sigma_m^2} \right) , \quad (21)$$

where $\{\Gamma_m\}$ and $\{\Lambda_m\}$ are given by the waterfilling solutions in (16)–(19).

Proof: By relaxing the rank constraint on the jammer, we arrive at the standard vector Gaussian AVC channel

$$\mathbf{Y} = \mathbf{X} + \mathbf{S} + \mathbf{W} . \quad (22)$$

Since any coding scheme for this channel must be robust to a rank-limited jammer, all rates achievable for this channel are also achievable on the rank-constrained jamming channel. ■

A. Optimizing for the rank-constrained jammer

However, the rank constraint on the jammer should admit rates higher than R_{wfill}, since in many cases the jammer's waterfilling strategy does not satisfy its rank constraint. If the transmitter fixes a covariance matrix Σ_X first, we can achieve a rate (using the results of [9]):

$$\max_{\Sigma_X : \mathrm{tr}(\Sigma_X) \le \Gamma} \min_{\mathbf{g} : \|\mathbf{g}\| = 1} \frac{1}{2} \log \frac{|\Sigma_X + \Sigma_W + \Lambda \mathbf{g}\mathbf{g}^T|}{|\Sigma_W + \Lambda \mathbf{g}\mathbf{g}^T|} . \quad (23)$$

Unfortunately, even the inner minimization is not convex in general, so standard optimization techniques are difficult to apply. In what follows we will characterize the above quantity for the case $M = 2$ and show that rates higher than R_{wfill} are achievable.

Proposition 5: For the $(2, 2, 1)$ MIMO AVC, the optimal input distribution is diagonal. If $\Sigma_X = \mathrm{diag}(\Gamma_1, \Gamma_2)$, then the $\mathbf{g}$ minimizing the mutual information between $\mathbf{X}$ and $\mathbf{Y}$ is equal to $(1, 0)^T$ if

$$\frac{\Gamma_1 / \sigma_1^2}{\Gamma_2 / \sigma_2^2} > \frac{(\Gamma_1 + \sigma_1^2 + \Lambda)}{(\Gamma_2 + \sigma_2^2 + \Lambda)} , \quad (24)$$

and is equal to $(0, 1)^T$ if the reverse inequality holds. If equality holds in (24), then all values of $\mathbf{g}$ yield the same mutual information.

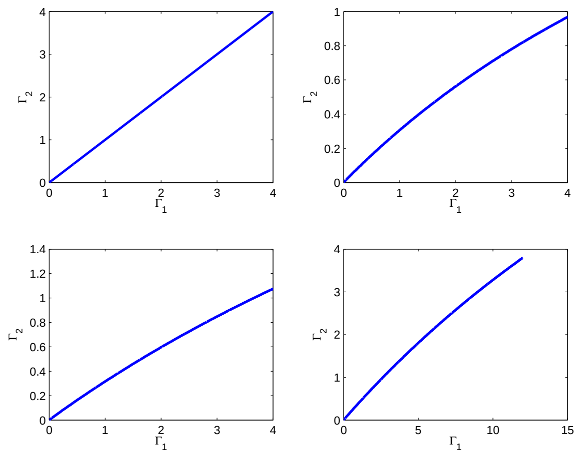

Fig. 2. Four examples of the threshold for the jammer's strategy. The upper left plot is for $(\sigma_1^2, \sigma_2^2, \Lambda) = (2, 2, 4)$, the upper right for $(\sigma_1^2, \sigma_2^2, \Lambda) = (3, 1, 4)$, the lower left for $(\sigma_1^2, \sigma_2^2, \Lambda) = (3, 1, 8)$, and the lower right for $(\sigma_1^2, \sigma_2^2, \Lambda) = (5, 2, 4)$.

Proof: We will prove the second statement first. Suppose that the input covariance is diagonal and consider the problem of minimizing

$$F(\mathbf{g}) = \frac{1}{2} \log \frac{|\Sigma_X + \Sigma_W + \Lambda \mathbf{g}\mathbf{g}^T|}{|\Sigma_W + \Lambda \mathbf{g}\mathbf{g}^T|} , \quad (25)$$

over all unit vectors $\mathbf{g} = (g, \sqrt{1 - g^2})^T$. Differentiating F and some algebra gives the following:

$$\frac{dF}{dg} = g\Lambda \left(\frac{(\Gamma_2 - \Gamma_1) + (\sigma_2^2 - \sigma_1^2)}{|\Sigma_X + \Sigma_W + \Lambda \mathbf{g}\mathbf{g}^T|} - \frac{(\sigma_2^2 - \sigma_1^2)}{|\Sigma_W + \Lambda \mathbf{g}\mathbf{g}^T|} \right) = g\Lambda \left(\frac{\Gamma_2 \sigma_1^2 (\Gamma_1 + \sigma_1^2 + \Lambda) - \Gamma_1 \sigma_2^2 (\Gamma_2 + \sigma_2^2 + \Lambda)}{|\Sigma_X + \Sigma_W + \Lambda \mathbf{g}\mathbf{g}^T| \cdot |\Sigma_W + \Lambda \mathbf{g}\mathbf{g}^T|} \right) . \quad (26)$$

Note that the only point where the derivative is 0 is when $g = 0$. However, this point may be a maximum or a minimum, depending on the sign in the numerator. This yields the threshold in (24).

Suppose that the transmitter chooses a non-diagonal Σ_X. Regardless of the actual value of the transmit covariance, $\mathbf{g} = (1, 0)^T$ and $\mathbf{g} = (0, 1)^T$ are possible channel realizations. For these two choices of $\mathbf{g}$, the input covariance Σ_X' created by zeroing the off-diagonal elements of Σ_X yields a larger mutual information. Therefore the max-min in (23) is maximized by a diagonal Σ_X. ■

The previous proposition says that for a given diagonal covariance matrix, the jammer's optimal strategy is to jam one of the subchannels. Figure 2 shows the the boundary given by the threshold for different values of the parameters σ_1^2, σ_2^2, Γ_1, Γ_2, and Λ. The region above each curve corresponds to $\mathbf{g} = (1, 0)^T$, and below to $\mathbf{g} = (0, 1)^T$. Within each region, the transmitter covariance can be optimized to maximize the mutual information, as in the following proposition.

Proposition 6: Suppose $\sigma_1^2 > \sigma_2^2$. Let β be the value of Γ_1

for which equality holds in (24) with $\Gamma_2 = \Gamma - \Gamma_1$, and let

$$\alpha = \frac{1}{2}(\Gamma + (\sigma_2^2 - \sigma_1^2) + \Lambda(1 - 2g^2)) \ . \tag{27}$$

Then the transmitter can maximize the mutual information by choosing

$$\Gamma_1 = \min\{\alpha, \beta\} \ . \tag{28}$$

Furthermore, for this power allocation the the worst jammer allocation is $\mathbf{g} = (0, 1)^T$.

Proof: Let $\mathbf{e}_1 = (1, 0)^T$ and $\mathbf{e}_2 = (0, 1)^T$. If we set $\Gamma_2 = \Gamma - \Gamma_1$ we can rewrite the threshold in (24) as:

$$\sigma_2^2 \left(1 + \frac{\sigma_2^2 + \Lambda}{\Gamma - \Gamma_1}\right) - \sigma_1^2 \left(2 + \frac{\sigma_1^2 + \Lambda}{\Gamma_1}\right) > 0 \ . \tag{29}$$

Differentiating the left side with respect to Γ_1 we obtain

$$\sigma_2^2 \left(\frac{\sigma_2^2 + \Lambda}{(\Gamma - \Gamma_1)^2}\right) + \sigma_1^2 \left(\frac{\sigma_1^2 + \Lambda}{\Gamma_1^2}\right) \ , \tag{30}$$

which is strictly positive, so the left side of the threshold is an increasing function of Γ_1. Thus for small Γ_1 the worst jammer direction is $\mathbf{g} = \mathbf{e}_2$ and for large Γ_1 the worst jammer direction is $\mathbf{g} = \mathbf{e}_1$.

Suppose that $\sigma_1^2 > \sigma_2^2$ and look at the point $\Gamma_1 = \Gamma_2 = \Gamma/2$. The threshold (24) is clearly not satisfied, so the worst $\mathbf{g}$ is equal to $\mathbf{e}_1$ only for $\Gamma_1 > \Gamma/2$. Let $\Sigma_X = \mathrm{diag}(\Gamma_1, \Gamma - \Gamma_1)$ and

$$F_i(\Gamma_1) = \frac{1}{2} \log \frac{|\Sigma_X + \Sigma_W + \Lambda \mathbf{e}_i \mathbf{e}_i^T|}{|\Sigma_W + \Lambda \mathbf{e}_i \mathbf{e}_i^T|} \ , \tag{31}$$

for $i = 1, 2$. We claim that $F_1(\Gamma_1)$ is a decreasing function of Γ_1. Differentiating with respect to Γ_1 gives:

$$\frac{dF_1}{d\Gamma_1} = \frac{\Gamma - 2\Gamma_1 + \sigma_2^2 - \sigma_1^2 - \Lambda}{2 \cdot (\Gamma_1 + \sigma_1^2 + \Lambda)(\Gamma - \Gamma_1 + \sigma_2^2)} \tag{32}$$

Since the worst $\mathbf{g}$ is $\mathbf{e}_1$ for $\Gamma_1 > \Gamma/2$, the derivative is negative, which shows that $F_1(\Gamma_1)$ decreases from the threshold point of (24). Thus the transmitter will choose a covariance such that the worst $\mathbf{g} = \mathbf{e}_2$ to maximize the mutual information.

Turning to $F_2(\Gamma_1)$, we can again differentiate:

$$\frac{dF_2}{d\Gamma_1} = \frac{\Gamma - 2\Gamma_1 + \sigma_2^2 - \sigma_1^2 + \Lambda}{2 \cdot (\Gamma_1 + \sigma_1^2)(\Gamma - \Gamma_1 + \sigma_2^2 + \Lambda)} \tag{33}$$

The maximum is at

$$\Gamma_1 = \frac{1}{2}\left(\Gamma + \sigma_2^2 - \sigma_1^2 + \Lambda\right) \ , \tag{34}$$

unless this point exceeds the threshold in (24). In this case, we choose Γ_1 such that equality holds in (24). ∎

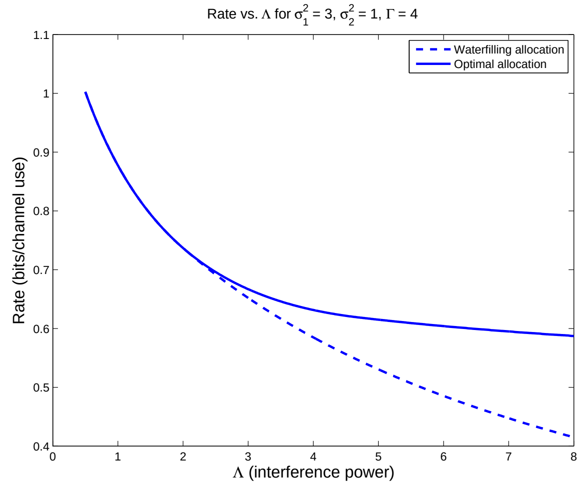

Fig. 3. An example of the capacity region versus the interference power Λ for $\sigma_1^2 = 3$, $\sigma_2^2 = 1$, $\Gamma = 4$.

B. Examples and comparison

We will now discuss some examples comparing the optimal strategies for the rank deficient adversarial model to the waterfilling strategy and show that in some cases the rank constraint increases the achievable rates over those given by waterfilling. Figure 3 shows the waterfilling and optimal rates as the interference power Λ is varied. The curves are equal until the point where $\Lambda = \sigma_1^2 - \sigma_2^2$, at which point the jammer's waterfilling strategy cannot be realized by allocating all power to a single channel. For large interference powers, the rank constraint allows the transmitter and receiver to communicate at rates strictly superior to those guaranteed by the relaxed waterfilling allocation.

As a second example, we can examine the asymptotic behavior of the capacity as $\Lambda \to \infty$. The optimal jamming strategy is still to jam the less noisy channel, so the noise-plus-interference spectrum becomes more and more unbalanced. Clearly the subchannel with noise $\sigma_2^2 + \Lambda$ contributes no rate to the capacity in the limit. However, any power in the first subchannel will still contribute. As $\Lambda \to \infty$ the expression in (34) increases without bound, so the limiting behavior is given by the threshold (24). The right side of (24) goes to 1, so the optimal power allocation reduces to

$$\frac{\Gamma_1}{\Gamma_2} = \frac{\sigma_1^2}{\sigma_2^2} \ . \tag{35}$$

Corollary 1: For the $(2, 2, 1)$ MIMO AVC with $\sigma_1^2 > \sigma_2^2$, the randomized coding capacity in the limit as $\Lambda \to \infty$ is given by

$$C_r(\Gamma) = \frac{1}{2} \log \left(1 + \frac{\Gamma}{\sigma_1^2 + \sigma_2^2}\right) \ . \tag{36}$$

We can also take the limit as both Γ and Λ go to ∞ while keeping the ratio $\rho = \Gamma/\Lambda$ fixed.

Corollary 2: For the $(2, 2, 1)$ MIMO AVC with $\sigma_1^2 > \sigma_2^2$, the randomized coding capacity in the limit as $\Gamma, \Lambda \to \infty$ with fixed $\rho = \Gamma/\Lambda$ scales according to

$$C_r(\rho, \Gamma) = O(\log \Gamma) + \frac{1}{2} \log \left(1 + \frac{\rho}{2}\right) \ . \tag{37}$$

Proof: Equation (34) shows that the optimal Γ_1 goes to $\Gamma/2$, which gives:

$$C_r(\Gamma, \Lambda) = \frac{1}{2} \log \left(1 + \frac{\Gamma}{2\sigma_1^2}\right) + \frac{1}{2} \log \left(1 + \frac{\Gamma/2}{\sigma_2^2 + \Lambda}\right) \ . \tag{38}$$

Taking the limit yields the result. ■

For the MIMO AVC, the previous two corollaries show that the signal-to-interference ratio is not a good measure in the case of rank-deficient interference.

IV. The (M,M,1) channel

We now turn to the more general channel

$$\mathbf{Y}(t) = \mathbf{X}(t) + \mathbf{g}S(t) + \mathbf{W}(t) \ , \tag{39}$$

where all vectors are in $\mathbb{R}^M$, $\mathbf{W}(t)$ is iid with distribution $\mathcal{N}(0, \Sigma_W)$ where $\Sigma_W = \mathrm{diag}(\sigma_1^2, \sigma_2^2, \ldots, \sigma_M^2)$. The coding definitions extend from the earlier definitions. We will again consider randomized coding for these channels and so we can focus on the mutual information

$$\max_{\Sigma_X : \mathrm{tr}(\Sigma_X) \leq \Gamma} \ \min_{\mathbf{g} : \|\mathbf{g}\| = 1} \frac{1}{2} \log \frac{|\Sigma_X + \Sigma_W + \Lambda \mathbf{g}\mathbf{g}^T|}{|\Sigma_W + \Lambda \mathbf{g}\mathbf{g}^T|} \ . \tag{40}$$

As in the $(2, 2, 1)$ case, the inner minimization for a fixed Σ_X is not convex.

Let $\mathbf{e}_j$ denote the j-th elementary vector, i.e. the unit vector with 1 in the j-th entry and 0 elsewhere. These vectors correspond to the jammer choosing to jam one of the subchannels defined by Σ_W. We believe that the behavior seen in the $(2, 2, 1)$ case extends to the $(M, M, 1)$ case in the following sense:

1) If the input covariance matrix Σ_X is diagonal, then the $\mathbf{g}$ minimizing the mutual information is equal to $\mathbf{e}_j$ for some j. The optimal input covariance is diagonal.
2) If $\sigma_1^2 > \sigma_2^2 > \cdots > \sigma_M^2$, then the optimal input covariance Σ_X forces the minimizing $\mathbf{g}$ to be $\mathbf{e}_M$.

V. Distributive and cooperative implementations

Multiple antennas can lead to significant gains in capacity (see e.g. Telatar [5]). This, in turn, has fueled the development of cooperative techniques where multiple separate terminals jointly appear as a virtual MIMO array than can partially capitalize on these gains [13]. Our results show that MIMO can also lead to significant gains in robustness, particularly when the transmitter and receiver have additional knowledge of *how* the interfering signals are generated. In the context of cognitive radio, this interference could come from a primary system, legacy system, or other cognitive systems in the same space. Our results say that the presence of extra antennas at the receiver can lead to an increase in rates over a rank-deficient but adversarial interferer. Future work will determine the exact degree to which our gains carry over to distributive and cooperative implementations.

For example, the hierarchical MIMO approach of Özgür, Lévêque and Tse [13] uses MIMO cooperation on a local level to communicate on a long-haul link in an large ad-hoc network with many nodes. If this network coexists in an environment with a few powerful interferers comprising another system, a naive "sum power" approach to channel modeling may result in a pessimistic estimate of the MIMO link's capacity. By explicitly accounting for the density of the interfering system, additional gains may be possible.

One weakness of our model is that it assumes the gains from the jammer to the receiver are fixed over time. If we adopt a fast-fading model, the rank constraints will no longer become operative and the approach via mutual information games may be more appropriate. However, for quasistatic channels our results point to an important and hitherto unexploited aspect of interference modeling. In future work we will clarify the investigate the more general (M, M, J) case for which we hope to find analogous results to the $(2, 2, 1)$ case.

Acknowledgment

This work was supported in part by the National Science Foundation under award CCF-0347298. A.D. Sarwate would like to acknowledge useful discussions with P. Venkitasubramaniam and H. Palaiyanur.

References

[1] J. Mitola and G. Maguire, "Cognitive radio: Making software radios more personal," *IEEE Personal Communications*, vol. 6, no. 4, pp. 13–18, August 1999.

[2] Federal Communications Commission, "Auction of 700 mhz band licenses scheduled for january 24, 2008; notice and filing requirements, minimum opening bids, reserve prices, upfront payments, and other procedures for auctions 73 and 76," Public Notice (DA 07-4171), October 5th 2007.

[3] N. Devroye, P. Mitran, and V. Tarokh, "Achievable rates in cognitive radio channels," *IEEE Trans. Inf. Theory*, vol. 52, no. 5, pp. 1813–1827, May 2006.

[4] M. Gastpar, "On capacity under receive and spatial spectrum-sharing constraints," *IEEE Transactions on Information Theory*, vol. 53, no. 2, pp. 471–487, February 2007.

[5] I. Telatar, "Capacity of multi-antenna gaussian channels," *European Transactions on Telecommunication*, vol. 10, no. 6, pp. 585–595, 1999.

[6] C. Baker and I. Chao, "Information capacity of channels with partially unknown noise. I. finite-dimensional channels," *SIAM Journal of Applied Mathematics*, vol. 56, no. 3, pp. 946–963, June 1996.

[7] B. Hughes and P. Narayan, "Gaussian arbitrarily varying channels," *IEEE Transactions on Information Theory*, vol. 33, no. 2, pp. 267–284, 1987.

[8] I. Csiszár and P. Narayan, "Capacity of the Gaussian arbitrarily varying channel," *IEEE Transactions on Information Theory*, vol. 37, no. 1, pp. 18–26, 1991.

[9] B. Hughes and P. Narayan, "The Capacity of a Vector Gaussian Arbitrarily Varying Channel," *IEEE Transactions on Information Theory*, vol. 34, no. 5, pp. 995–1003, 1988.

[10] I. Csiszár, "Arbitrarily varying channels with general alphabets and states," *IEEE Transactions on Information Theory*, vol. 38, no. 6, pp. 1725–1742, 1992.

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

[12] A. Sarwate and M. Gastpar, "Randomization Bounds on Gaussian arbitrarily varying channels," in *Proceedings of the 2006 International Symposium on Information Theory*, Seattle, WA, 2006.

[13] A. Özgür, O. Lévêque, and D. Tse, "Hierarchical cooperation achieves optimal capacity scaling in ad hoc networks," *IEEE Transactions on Information Theory*, vol. 53, no. 10, pp. 3549–3572, October 2007.