The Capacity of a Vector Gaussian Arbitrarily Varying Channel

BRIAN HUGHES, MEMBER, IEEE, AND PRAKASH NARAYAN, MEMBER, IEEE

Abstract—**The random coding capacity of a vector Gaussian arbitrarily varying channel (VGAVC) is determined, along with a simple general method for computing this capacity. The VGAVC is a discrete-time memoryless vector channel with an input power constraint and additive Gaussian noise that is further corrupted by an additive "jamming signal." The statistics of this jamming signal are unknown and may be arbitrary subject only to a power constraint.**

I. INTRODUCTION

THE AIM of this paper is to determine the maximum amount of information that can be reliably transmitted across a vector communication channel that is corrupted by additive Gaussian noise and an intelligent jammer. We impose no restrictions on the class of jamming signals considered, beyond the fundamental limitation of bounded power. The channel under investigation, called a *vector Gaussian arbitrarily varying channel* (VGAVC) is described as follows (cf. Fig. 1). Once each second the transmitter sends an m-dimensional random vector, say $\boldsymbol{u}_i^*$ at time i, representing the output of an information source of rate R (bits per channel use) to the receiver. The sequence $\{\boldsymbol{u}_i^*\}$ can be chosen arbitrarily subject only to a power constraint (to be specified later). The channel output is defined by

$$\boldsymbol{y}_i^* \equiv \boldsymbol{u}_i^* + \boldsymbol{\eta}_i^* + \boldsymbol{s}_i^*$$

where $\{\boldsymbol{\eta}_i^*\}$, called the *background noise*, is an independent and identically distributed (i.i.d.) sequence of zero-mean Gaussian m-vectors with covariance matrix Σ. The sequence $\{\boldsymbol{s}_i^*\}$ represents hostile jamming or other noise sources with unknown statistics. The only restriction we impose on this sequence is a power constraint (also to be specified later).

There is a rich literature on the discrete arbitrarily varying channel (AVC). Much of this is summarized in [1]. Early results relevant to jammed Gaussian channels can be found in [2]–[4]; however, the first investigation of a channel substantially like the GAVC was reported in [5] and [6], where bounds on the achievable rates of reliable transmission were reported. Başar and Wu [7] have investigated the use of essentially the same channel for a different source transmission problem in which the source is a discrete-time memoryless Gaussian source, and reliability is measured by mean-square distortion. Ahlswede [8] has established the capacity of the GAVC when $\{\boldsymbol{s}_i^*\}$ is a scalar Gaussian sequence with arbitrarily varying variance. In [9] we determined the random coding λ-capacity of the GAVC for a variety of transmitter and jamming power constraints. Here we extend some of these results to VGAVC's, i.e., to GAVCs with vector input and output alphabets.

We are interested in determining the relationship between achievable error probability and coding rates when random coding is used. The methods and results of Blackwell *et al.* [10] for discrete AVC's do not apply to the channels considered here due to the presence of *cost constraints* on the transmitter and jammer power. These constraints often lead to results for the GAVC which are unlike those of discrete AVC's. This is exemplified by the observation that unlike discrete AVC's, the GAVC with *ensemble*-averaged power constraints does not have a capacity in the usual sense [9]. Rather, the achievable error probability of the code is a continuously increasing function of its rate. We emphasize that this aberrant behavior is due to the imposition of cost constraints and not the continuous input and output alphabets of the GAVC; indeed, discrete AVC's with ensemble-averaged cost constraints on the transmitter and jammer also generally fail to have a capacity [11], [12].

Our results can be summarized as follows. A coding theorem and a strong converse are proved that characterize the capacity of the VGAVC over the class of power-limited codes along with a simple and general method for computing this capacity. We find that the capacity of the VGAVC has a "water-filling" interpretation, much like the capacity of the m-dimensional additive Gaussian noise channels that it generalizes [13, theorem 7.5.1].

The remainder of the paper is organized as follows. In Section II we define the problem and summarize our results. Section III contains the proofs of these results.

Manuscript received June 30, 1987; revised December 31, 1987. This work was supported in part by the Naval Research Laboratory and the Office of Naval Research under Grants N00014-83-G-0192 and N00014-84-G-0101. This paper was presented in part at the 1987 Conference on Information Sciences and Systems, Johns Hopkins University, Baltimore, MD, March 1987.

B. Hughes is with the Department of Electrical and Computer Engineering, Johns Hopkins University, Baltimore, MD 21218.

P. Narayan is with the Electrical Engineering Department, University of Maryland, College Park, MD 20742.

IEEE Log Number 8823617.

0018-9448/88/0900-0995$01.00

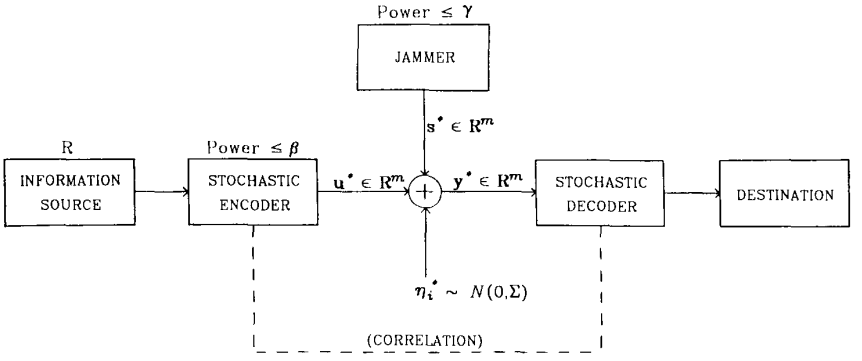

Fig. 1. Vector Gaussian arbitrarily varying channel (VGAVC).

II. Definitions and Summary of Results

For our purposes, a *codeword* of length n is a real $m \times n$ matrix, $U = \{u_{ij}\}_{i,j=1}^{m,n}$, selected by the transmitter. A *jamming sequence* of length n is an $m \times n$ matrix, $S = \{s_{ij}\}_{i,j=1}^{m,n}$, selected by the jammer. Given a codeword U and a jamming sequence S, the output of the VGAVC is defined by

$$Y^* = U + H^* + S$$

where $H^* = \{\eta_{ij}^*\}_{i,j=1}^{m,n}$ denotes an $m \times n$ random matrix whose columns are i.i.d $N(0,\Sigma)$, where 0 is the origin in $\boldsymbol{R}^m$ and Σ is a nonnegative-definite $m \times m$ matrix.[1]

An (n, M) *(deterministic) block code* is a system

$$C_n = \{(U_1, D_1), \cdots, (U_M, D_M)\}$$

where $\{U_k\}_{k=1}^{M}$ are codewords of length n, and the *decoding sets* $\{D_i\}_{i=1}^{M}$ are disjoint Borel subsets of $\boldsymbol{R}^{m \times n}$. We permit the use of random codes, denoted by

$$C_n^* = \{(U_1^*, D_1^*), \cdots, (U_M^*, D_M^*)\} \tag{2.1}$$

and also random jamming sequences, denoted by S^*. The code C_n^*, the jamming sequence S^*, and the background noise H^* are constrainted to be mutually independent.

Both the transmitter and jammer are subject to limitations on transmitted power. For any $m \times n$ matrix, say $V = \{v_{ij}\}_{i,j=1}^{m,n}$, denote the (time-averaged) *power* of V by

$$P(V) \equiv \frac{1}{n} \sum_{i=1}^{m} \sum_{j=1}^{n} v_{ij}^2. \tag{2.2}$$

Then for fixed $\beta > 0$, we say that the random code (2.1) satisfies a time-averaged *input power constraint* almost surely (a.s.) if

$$P(U_k^*) \le \beta \text{ (a.s.)}, \qquad 1 \le k \le M. \tag{2.3}$$

A jamming sequence S^* satisfies a time-averaged *jamming power constraint* if

$$P(S^*) \le \gamma \quad \text{(a.s.)}. \tag{2.4}$$

[1]We use the following notation throughout this paper: $N(\mu,\Sigma)$ denotes a Gaussian distribution with mean vector μ and covariance matrix Σ. Asterisks are used as superscripts to denote random quantities and to distinguish them from deterministic quantities.

In the language of [9], (2.3) and (2.4) are *peak* power constraints.

Given an (n, M) random code C_n^* the (maximum) *error probability* is defined by

$$\lambda_\gamma(C_n^*) \equiv \sup_{S:\, P(S) \le \gamma} \max_{1 \le k \le M} \Pr\{U_k^* + H^* + S \in \bar{D}_k^*\} \tag{2.5}$$

where $\bar{D}_k^* \equiv \boldsymbol{R}^{m \times n} - D_k^*$. We say that an (n, M) random code C_n^* is an (n, M, λ) random code for some $0 < \lambda \le 1$, if it satisfies (2.3) and if

$$\lambda_\gamma(C_n^*) \le \lambda.$$

The random coding *capacity* of the VGAVC over the class of codes that satisfy (2.3), if it exists, is defined to be the largest nonnegative number C such that for any $\epsilon > 0$ and $0 < \lambda \le 1$, there exists an (n, M, λ) random code with

$$M \ge 2^{n(C-\epsilon)}$$

for all sufficiently large n, and for all $0 \le \lambda < 1$ there does not exist an (n, M, λ) random code with

$$M \ge 2^{n(C+\epsilon)}$$

for all sufficiently large n.

Before proceeding further, it is convenient to make a simplifying observation. For any $m \times m$ nonnegative-definite matrix Σ there exists an orthogonal transformation of $\boldsymbol{R}^m$, say Θ, and a diagonal matrix diag$(\boldsymbol{a})$, where $\boldsymbol{a} = (a_1, \cdots, a_m)$ and $\boldsymbol{a} \ge 0$, such that

$$\Theta \Sigma \Theta^T = \text{diag}(\boldsymbol{a})$$

where the superscript T denotes the matrix transpose [14]. The codeword power, jamming sequence power, and error probability of C_n^* are unchanged when Y^* is multiplied by Θ; therefore, it follows that an (n, M, λ) code exists for the VGAVC with noise covariance matrix Σ if and only if one exists for diag$(\boldsymbol{a})$. We can therefore assume, without loss of generality, that $\Sigma = \text{diag}(\boldsymbol{a})$ and further that $a_1 \le \cdots \le a_m$.

We now present several theorems that fix the value of C. First consider the special case $\gamma = 0$ (no jammer is present). The results here are well-known [13, theorem 7.5.1]. The

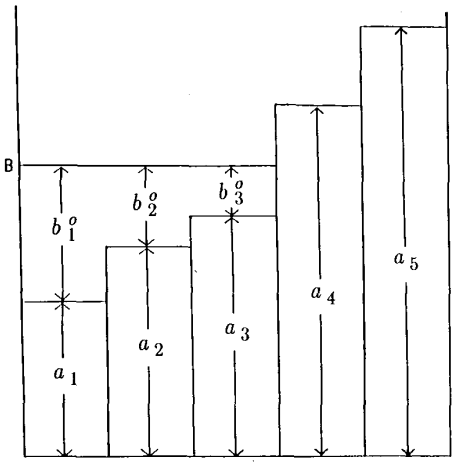

Fig. 2. Water-filling interpretation of parallel discrete-time Gaussian channels for $m = 5$ (after [13, p. 344]).

channel has capacity C_0 that is given by[2]

$$C_0 \equiv \max_{b_i \geq 0:\, \Sigma b_i \leq \beta} \sum_{i=1}^{m} \frac{1}{2} \log_2 \left(1 + \frac{b_i}{a_i}\right) \tag{2.6}$$

if $a_1 > 0$, and by $C_0 \equiv +\infty$ if $a_1 = 0$. The input power distribution $\boldsymbol{b}^0 = (b_1^0, \cdots, b_m^0)$ that attains the maximum in (2.6) has the form

$$\begin{aligned} b_i^0 + a_i &= B, \qquad \text{if } a_i < B \\ b_i^0 &= 0, \qquad \text{if } a_i \geq B \end{aligned} \tag{2.7}$$

where B is such that $\Sigma b_i^0 = \beta$. This distribution has the simple interpretation illustrated in Fig. 2. We can think of pouring a volume β of water into a container whose bottom consists of a series of plateaus whose heights are equal to the components of $\boldsymbol{a}$. The level to which this water settles is B, and b_i^0 gives the depth of the ith plateau below the water's surface. Consistent with this interpretation, we call $\boldsymbol{b}^0$ the *water-filling vector of power* β for $\boldsymbol{a}$ if it satisfies (2.7).

Suppose now that $\gamma > 0$ (jammer present). Consider the quantity

$$R(\beta, \gamma) \equiv \max_{\boldsymbol{b} \geq 0:\, \boldsymbol{e}\cdot\boldsymbol{b} \leq \beta} \min_{\boldsymbol{c} \geq 0:\, \boldsymbol{e}\cdot\boldsymbol{c} \leq \gamma} r(\boldsymbol{b}, \boldsymbol{c}) \tag{2.8}$$

where

$$r(\boldsymbol{b}, \boldsymbol{c}) \equiv \sum_{i=1}^{m} \frac{1}{2} \log_2 \left(1 + \frac{b_i}{a_i + c_i}\right) \tag{2.9}$$

and $\boldsymbol{e} \equiv (1, \cdots, 1)$. In Theorem 3 it will be shown that the value of the program (2.8) is unchanged if we switch the order of maximization and minimization, i.e.,

$$R(\beta, \gamma) = \min_{\boldsymbol{c} \geq 0:\, \boldsymbol{e}\cdot\boldsymbol{c} \leq \gamma} \max_{\boldsymbol{b} \geq 0:\, \boldsymbol{e}\cdot\boldsymbol{b} \leq \beta} r(\boldsymbol{b}, \boldsymbol{c}).$$

From (2.6), we see that $R(\beta, \gamma)$ can be interpreted as the smallest channel capacity that can be inflicted on the transmitter by a jammer whose signal is limited to *stationary memoryless Gaussian noise* of expected total power γ. The following theorems, whose proofs are deferred to Section III, establish that $R(\beta, \gamma)$ is the capacity of the VGAVC. The latter channel encompasses a far broader collection of jamming signals; viz., it includes all (possibly nonstationary non-Gaussian) signals of power at most γ (i.e., all signals that satisfy (2.4)).

Theorem 1 (Coding Theorem for the VGAVC): Let λ and ϵ be arbitrary positive numbers with $\lambda \leq 1$. When n is sufficiently large, there exists an (n, M, λ) random code so that

$$M > 2^{n(R(\beta,\gamma)-\epsilon)}.$$

Theorem 2 (Strong Converse for the VGAVC): Let λ and ϵ be arbitrary numbers with $0 \leq \lambda < 1$ and $\epsilon > 0$. Suppose that an (n, M, λ) random code exists. If n is sufficiently large, then

$$M < 2^{n(R(\beta,\gamma)+\epsilon)}.$$

We now address the problem of computing $R(\beta, \gamma)$. As stated earlier, for $\gamma = 0$ the optimizing power distribution in (2.6) is easily calculated from (2.7). The following theorem shows that a simple procedure also exists for computing the power distribution $\boldsymbol{b}^0$ and $\boldsymbol{c}^0$, that optimize $r(\boldsymbol{b}, \boldsymbol{c})$ when $\gamma > 0$.

Theorem 3: Consider the following two-player zero-sum game (cf. [15]):

$$\text{Program I:} \quad \max_{\boldsymbol{b} \geq 0:\, \boldsymbol{e}\cdot\boldsymbol{b} \leq \beta} \min_{\boldsymbol{c} \geq 0:\, \boldsymbol{e}\cdot\boldsymbol{c} \leq \gamma} r(\boldsymbol{b}, \boldsymbol{c})$$

$$\text{Program II:} \quad \min_{\boldsymbol{c} \geq 0:\, \boldsymbol{e}\cdot\boldsymbol{c} \leq \gamma} \max_{\boldsymbol{b} \geq 0:\, \boldsymbol{e}\cdot\boldsymbol{b} \leq \beta} r(\boldsymbol{b}, \boldsymbol{c}).$$

Let $\boldsymbol{c}^0$ be the water-filling vector of power γ for the background noise power vector $\boldsymbol{a}$, and let $\boldsymbol{b}^0$ be the water-filling vector of power β for $\boldsymbol{a} + \boldsymbol{c}^0$. Then $\boldsymbol{c}^0$ and $\boldsymbol{b}^0$ are saddlepoint strategies for the game defined above; i.e., for any other nonnegative sequences $\boldsymbol{c} \geq 0$ and $\boldsymbol{b} \geq 0$ such that $\boldsymbol{e}\cdot\boldsymbol{b} \leq \beta$ and $\boldsymbol{e}\cdot\boldsymbol{c} \leq \gamma$ where $\boldsymbol{e} = (1, \cdots, 1)$, the following double inequality holds:

$$r(\boldsymbol{b}, \boldsymbol{c}^0) \leq r(\boldsymbol{b}^0, \boldsymbol{c}^0) \leq r(\boldsymbol{b}^0, \boldsymbol{c}). \tag{2.10}$$

The optimizing power distributions of Theorem 3 also have a simple water-filling interpretation, illustrated in Fig. 3. A volume γ of water is poured into the container of Fig. 2. The depth of the ith plateau below the water's surface is c_i^0. We then pour an *additional* volume β of water into the container. The amount by which the new water level rises above the old over the ith plateau is b_i^0. It is interesting to note that $\boldsymbol{c}^0$ and $\boldsymbol{b}^0$ are "mutually water filling" in the sense that $\boldsymbol{c}^0$ is also the water-filling vector of power γ for the sequence $\boldsymbol{a} + \boldsymbol{b}^0$.

III. Proofs of Theorems 1–3

We will prove Theorem 3 first, because its conclusions are required in the proofs of Theorems 1 and 2.

Proof of Theorem 3: It suffices to prove the following

[2] Only the weak converse is established in [13, theorem 7.5.1]; however, Theorem 2 confirms that the strong converse holds.

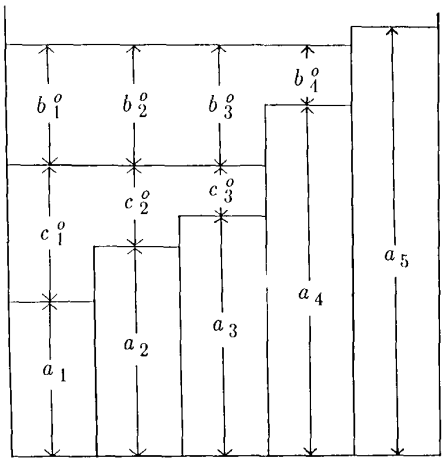

Fig. 3. Water-filling interpretation of optimizing power distributions from Theorem 3 ($m = 5$).

two statements:

a) $$\max_{\boldsymbol{b} \geq 0:\ \boldsymbol{e}\cdot\boldsymbol{b} \leq \beta} r(\boldsymbol{b}, \boldsymbol{c}^0) = r(\boldsymbol{b}^0, \boldsymbol{c}^0)$$

b) $$\min_{\boldsymbol{c} \geq 0:\ \boldsymbol{e}\cdot\boldsymbol{c} \leq \gamma} r(\boldsymbol{b}^0, \boldsymbol{c}) = r(\boldsymbol{b}^0, \boldsymbol{c}^0)$$

where $r(\boldsymbol{b}, \boldsymbol{c})$ is defined in (2.9) and $\boldsymbol{b}^0$ and $\boldsymbol{c}^0$ are as defined in Theorem 3. Statement a) is well-known [13, theorem 7.5.1]; therefore, it only remains to show b). The function $r(\boldsymbol{b}^0, \cdot)$ is convex over $\{\boldsymbol{c}:\ \boldsymbol{c} \geq 0\}$; thus necessary and sufficient conditions [13, theorem 4.4.1] for a given sequence $\boldsymbol{c}'$ to minimize $r(\boldsymbol{b}^0, \cdot)$ are

$$\frac{\partial r(\boldsymbol{b}^0, \boldsymbol{c}')}{\partial c_i} = \rho, \qquad \text{for all } i\text{: } c_i' > 0$$

$$\frac{\partial r(\boldsymbol{b}^0, \boldsymbol{c}')}{\partial c_i} \geq \rho, \qquad \text{for all } i\text{: } c_i' = 0 \tag{3.1}$$

for some real ρ.

We now show that $\boldsymbol{c}^0$ satisfies these conditions. Note that the water-filling sequences have the following properties (see Fig. 3): $c_i^0 > 0$ implies $c_i^0 + a_i = c_1^0 + a_1$ and $b_i^0 = b_1^0$; $c_i^0 = 0$ implies $a_i \geq c_1^0 + a_1$, $b_i^0 + a_i \geq c_1^0 + a_1 + b_1^0$, and $b_i^0 \leq b_1^0$. Therefore, if $c_i^0 > 0$, then

$$\frac{\partial r(\boldsymbol{b}^0, \boldsymbol{c}^0)}{\partial c_i} = \frac{-b_i^0}{2(a_i + c_i^0)(a_i + c_i^0 + b_i^0)} = \frac{-b_1^0}{2(a_1 + c_1^0)(a_1 + c_1^0 + b_1^0)}.$$

Otherwise, if $c_i^0 = 0$, then

$$\frac{\partial r(\boldsymbol{b}^0, \boldsymbol{c}^0)}{\partial c_i} = \frac{-b_i^0}{2a_i(a_i + b_i^0)} \geq \frac{-b_1^0}{2(a_1 + c_1^0)(a_1 + c_1^0 + b_1^0)}.$$

Thus $(c_1^0, \cdots, c_m^0)$ satisfies (3.1) with

$$\rho \equiv \frac{-b_1^0}{2(a_1 + c_1^0)(a_1 + c_1^0 + b_1^0)},$$

thereby proving statement b). This completes the proof of Theorem 3.

Proof of Theorem 1: Let ϵ and λ be arbitrary positive numbers with $\lambda \leq 1$. We now show that an (n, M, λ) code exists with $M > 2^{n(R(\beta,\gamma)-\epsilon)}$ for all sufficiently large n. We suspect that the *best* transmitter strategy and the *worst* jammer strategy are both asymptotically Gaussian (as in [9]). Therefore, define $p(Y|U)$ and $p(U)$ to be the conditional density and unconditional density, respectively, of Y^* when U^* has i.i.d. $N(0, \operatorname{diag}(\boldsymbol{b}^0))$ columns, S^* has i.i.d. $N(0, \operatorname{diag}(\boldsymbol{c}))$ columns, and $Y^* = U^* + H^* + S^*$; i.e.,

$$p(Y|U) = \prod_{j=1}^{n} \left[2\pi(c_i + a_i)\right]^{-m/2} \exp\left(- \sum_{i=1}^{m} \frac{(y_{ij} - u_{ij})^2}{2(c_i + a_i)} \right)$$

$$p(Y) = \prod_{j=1}^{n} \left[2\pi(b_i^0 + c_i + a_i)\right]^{-m/2} \cdot \exp\left(- \sum_{i=1}^{m} \frac{y_{ij}^2}{2(b_i^0 + c_i + a_i)} \right)$$

for all $Y = \{y_{ij}\}_{i,j=1}^{m,n}$ and $U = \{u_{ij}\}_{i,j=1}^{m,n}$, where $\boldsymbol{b}^0$ is the optimal transmitter power vector defined in Theorem 3 and $\boldsymbol{c} \geq 0$ is an arbitrary jamming power vector. For any given $m \times n$ matrices U and Y and any $\boldsymbol{c} \geq 0$, define the *mutual information* between Y and U as [13, p. 29]

$$I_c(U; Y) \equiv \log_2 \left[\frac{p(Y|U)}{p(Y)} \right] = nr(\boldsymbol{b}^0, \boldsymbol{c}) + \log_2 e \sum_{i,j=1}^{m,n} \frac{y_{ij}^2}{2(b_i^0 + c_i + a_i)} - \frac{(y_{ij} - u_{ij})^2}{2(c_i + a_i)}.$$

Finally, let G_n consist of real m-vectors $\boldsymbol{c} \geq 0$ such that $n\boldsymbol{c}/\gamma$ has only integer components and such that $\boldsymbol{e}\cdot\boldsymbol{c} \leq \gamma$; clearly, it follows that $|G_n| \leq (n+1)^m$.[3]

To prove Theorem 1, we require the following two lemmas whose proofs are contained in the Appendix.

Lemma 1: Let $U^* = \{u_{ij}^*\}_{i,j=1}^{m,n}$ be a random matrix with independent elements so that u_{ij}^* has distribution $N(0, b_i^0)$, and let Y be any $m \times n$ matrix. Then

$$\Pr\left\{ \frac{1}{n} I_c(U^*; Y) > \alpha \right\} \leq 2^{-n\alpha} \tag{3.2}$$

holds for any real α and $\boldsymbol{c} \geq 0$.

Lemma 2: Let U^* be as in Lemma 1. Then for any $m \times n$ matrix S that satisfies $P(S) \leq \gamma$, there is a jamming

[3] $|A|$ denotes the cardinality of the set A.

power vector $c^S \in G_n$ so that

$$\Pr\left\{\frac{1}{n} I_{c^S}(U^*; U^* + H^* + S) \le r(b^0, c^S) - \delta\right\} \le 2^{-n\delta^2/47m} \tag{3.3}$$

holds for all $0 < \delta \le 5m$ and $n > (1 + 2/\delta \ln 2)m^2$.

Remark: Observe that the right side of (3.3) does not depend on S and tends to zero for all $\delta > 0$.

We now proceed with the proof of Theorem 1. Fix $\epsilon > 0$, and to avoid trivialities, assume that $\epsilon \le 10m$. Let $M \equiv 2^{\lfloor n(R(\beta,\gamma)-\epsilon)\rfloor}$. For any $m \times n$ matrices U and Y we say that Y is G_n-*typical* of U if and only if

$$\frac{1}{n} I_c(U; Y) > r(b^0, c) - \epsilon/2$$

holds for *some* $c \in G_n$. Consider an $(n, 4M)$ random code whose codewords $\{U_1^*, \cdots, U_{4M}^*\}$ are independent replicas of $U^* = \{u_{ij}^*\}_{i,j=1}^{m,n}$, where the components u_{ij}^* are independent and $N(0, b_i^0)$ distributed. We decode a particular received sequence Y as message i if and only if Y is G_n-typical of U_i^* and is not G_n-typical of U_j^* for any $j \ne i$; otherwise, we declare an error. Let D_i^* be the set of received sequences that are decoded as U_i^*. It is easily seen that

$$\{U_i^* + H^* + S \in \bar{D}_i^*\} = B_i \cup \left(\bigcup_{j \ne i} B_j\right)$$

where

$$B_i \equiv \bigcap_{c \in G_n} \left\{\frac{1}{n} I_c(U_i^*; U_i^* + H^* + S) \le r(b^0, c) - \epsilon/2\right\}$$

and

$$B_j \equiv \bigcup_{c \in G_n} \left\{\frac{1}{n} I_c(U_j^*; U_i^* + H^* + S) > r(b^0, c) - \epsilon/2\right\}.$$

Therefore, for each S satisfying $P(S) \le \gamma$,

$$\begin{aligned}
&\Pr\{U_i^* + H^* + S \in \bar{D}_i^*\} \\
&\quad \le \Pr(B_i) + \sum_{j \ne i} \Pr(B_j) \\
&\quad \overset{a)}{\le} 2^{-n\epsilon^2/188m} + \sum_{j \ne i} \sum_{c \in G_n} \\
&\qquad \cdot \Pr\left\{\frac{1}{n} I_c(U_j^*; U_i^* + H^* + S) > r(b^0, c) - \epsilon/2\right\} \\
&\quad \overset{b)}{\le} 2^{-n\epsilon^2/188m} + (4M - 1)|G_n| 2^{-n(r(b^0,c)-\epsilon/2)} \\
&\quad \overset{c)}{\le} 2^{-n\epsilon^2/188m} + 8(n+1)^m 2^{-n\epsilon/2}
\end{aligned} \tag{3.4}$$

holds for all $n > (1 + 4/\epsilon \ln 2)m^2$, $1 \le i \le 4M$. The justification of these steps is as follows. Step a) follows by observing that $\epsilon \le 10m$ and

$$B_i \subset \left\{\frac{1}{n} I_{c^S}(U_i^*; U_i^* + H^* + S) < r(b^0, c^S) - \epsilon/2\right\}$$

and applying Lemma 2 with $\delta = \epsilon/2$. Step b) follows from applying Lemma 1 to the right-most term in a), and step c) follows from $M \le 2^{n(R(\beta,\gamma)-\epsilon)}$, $|G_n| \le (n+1)^m$ and (2.10).

The last line of (3.4), which is independent of S and i, tends to zero for all $\epsilon > 0$, as desired. The code defined earlier, however, does not satisfy (2.3) since the codewords are Gaussian. This can be remedied by selecting a subset of M codewords as follows. Let $A^* = \{i\colon P(U_i^*) > \beta, 1 \le i \le 4M\}$. If $|A^*| \le 3M$, select in any way a subset of M codewords that satisfy $P(U_i^*) \le \beta$ (together with the corresponding decoding sets) and call the resulting code C_n^*; otherwise, if $|A^*| > 3M$, declare an error. It follows that

$$\lambda_\gamma(C_n^*) \le 2^{-n\epsilon^2/188m} + 8(n+1)^m 2^{-n\epsilon/2} + \Pr\{|A^*| > 3M\} \tag{3.5}$$

for all $n > (1 + 4/\epsilon \ln 2)m^2$. Note that $|A^*|$ is binomial $(4M, p)$, where $p \equiv \Pr\{P(U^*) > \beta\}$, and that $P(U^*)$ is gamma distributed with $EP(U^*) = \beta$. Since the median of the gamma distribution is less than or equal to the mean, it follows that $p \le 1/2$. By Chebyshev's inequality [16, p. 190] we have

$$\Pr\{|A^*| > 3M\} \le \frac{4Mp(1-p)}{(3M - 4Mp)^2} \le \frac{1}{M} \le 2^{-n(R(\beta,\gamma)-\epsilon)}.$$

Substituting this last result into (3.5), we find that the right side of (3.5) can be made less than any $\lambda > 0$ for n chosen sufficiently large. This completes the proof of Theorem 1.

Proof of Theorem 2: For all $\epsilon > 0$ and $0 \le \lambda < 1$, we now prove that an $(n, 2^{n(R(\beta,\gamma)+\epsilon)}, \lambda)$ random code does not exist for all sufficiently large n. Let

$$C_n = \{(U_1, D_1), \cdots, (U_M, D_M)\}$$

be any (n, M, λ) *deterministic* code. Since $R(\beta, \cdot)$ is continuous and decreasing, we can choose $\bar{\gamma} < \gamma$ so that

$$R(\beta, \bar{\gamma}) \le R(\beta, \gamma) + \frac{\epsilon}{2}. \tag{3.6}$$

Let $\bar{c}^0$ be the water-filling vector of power $\bar{\gamma}$ for a. Let $S^* = \{s_{ij}^*\}_{i,j=1}^{m,n}$ be an $m \times n$ random matrix with independent elements so that s_{ij}^* has distribution $N(0, \bar{c}_i^0)$. As defined, S^* does not satisfy (2.4). Therefore, define $\hat{S}^* \equiv S^*$ if $P(S^*) \le \gamma$; otherwise, $\hat{S}^* \equiv 0$. Clearly, $\hat{S}^*$ satisfies (2.4), and by the law of large numbers [16, p. 363]

$$\epsilon_n \equiv \Pr\{\hat{S}^* \ne S^*\} \to 0$$

as $n \to \infty$. It further follows that

$$\begin{aligned}
\lambda_\gamma(C_n) &\ge \max_{1 \le k \le M} \Pr\{U_k + H^* + \hat{S}^* \in \bar{D}_k^*\} \\
&\ge \max_{1 \le k \le M} \Pr\{U_k + H^* + S^* \in \bar{D}_k^*\} - \epsilon_n.
\end{aligned}$$

Let $\bar{b}^0$ be the water-filling vector of power β for $a + \bar{c}^0$, and for $0 < \delta < 1$ define the following mn-dimensional

ellipsoidal regions:

$$E_\delta^1 \equiv \left\{ Y \left| \sum_{i,j=1}^{m,n} \frac{y_{ij}^2}{mn(1+\delta)\left(\bar{b}_i^0 + \bar{c}_i^0 + a_i\right)} \le 1 \right. \right\}$$

$$E_\delta^2 \equiv \left\{ Y \left| \sum_{i,j=1}^{m,n} \frac{y_{ij}^2}{mn(1-\delta)\left(\bar{c}_i^0 + a_i\right)} \le 1 \right. \right\}.$$

We require the following two lemmas whose proofs are contained in the Appendix.

Lemma 3: Let $U = \{u_{ij}\}_{i,j=1}^{m,n}$ be any $m \times n$ matrix such that $P(U) \le \beta$ (cf. (2.2)). Then for any $\delta > 0$,

$$\Pr\left\{ U + H^* + S^* \in \bar{E}_\delta^1 \right\} < \frac{4}{mn\delta^2}. \tag{3.7}$$

Lemma 4: Let $\nu(A)$ denote the volume of any $A \subset \boldsymbol{R}^{m\times n}$, and define $p \equiv \Pr(H^* + S^* \in A)$. Then

$$\nu(A) \ge \nu\left(E_{g^{-1}(p^{2/mn})}^2\right) \tag{3.8}$$

where $g(\delta) \equiv (1-\delta)e^{\delta}$, for $0 \le \delta < 1$.

To prove Theorem 2, we proceed as follows. Fix $1 > \delta > 0$ (to be chosen later). Choose $n_0 = n_0(\lambda, \delta, m, \gamma, \bar{\gamma}) \ge 2$ (which does not depend on C_n) so that

$$\epsilon_n \le \frac{1-\lambda}{4} \qquad \frac{4}{mn\delta^2} \le \frac{1-\lambda}{4} \qquad \left(\frac{1-\lambda}{2}\right)^{2/mn} \ge g(\delta) \tag{3.9}$$

for all $n \ge n_0$; this is possible since $g(\delta) < 1$ for all $1 > \delta > 0$. Define $D_i' \equiv D_i \cap E_\delta^1$. From Lemma 3 and (3.9) it follows that

$$\Pr\left\{ U_k + H^* + S^* \in D_k' \right\} \ge \frac{1-\lambda}{2}.$$

From (3.9) and Lemma 4 it follows that

$$\nu(D_k') \ge \nu\left(E_\delta^2\right). \tag{3.10}$$

The volume of the unit sphere in mn dimensions is

$$\frac{\pi^{mn/2}}{\Gamma\left(\frac{mn+2}{2}\right)}.$$

Thus, by an elementary change of variables, the volume of an mn-dimensional ellipsoid with axes $\{r_{ij}\}_{i,j=1}^{m,n}$ is

$$\frac{\pi^{mn/2}}{\Gamma\left(\frac{mn+2}{2}\right)} \prod_{i,j=1}^{m,n} r_{ij}.$$

Therefore, from (3.10) the volume of $D_1' \cup \cdots \cup D_m'$ is at least

$$M \frac{\pi^{mn/2}}{\Gamma\left(\frac{mn+2}{2}\right)} \left(mn(1-\delta)\right)^{mn/2} \prod_{i=1}^{m} \left(a_i + \bar{c}_i^0\right)^{n/2}.$$

On the other hand the volume of $D_1' \cup \cdots \cup D_m'$ is no more than the volume of E_δ^1 which is

$$\frac{\pi^{mn/2}}{\Gamma\left(\frac{mn+2}{2}\right)} \left(mn(1+\delta)\right)^{mn/2} \prod_{i=1}^{m} \left(\bar{b}_i^0 + a_i + \bar{c}_i^0\right)^{n/2}.$$

Therefore,

$$\begin{aligned} \frac{\log_2 M}{n} &\le \sum_{i=1}^{m} \frac{1}{2} \log_2 \left(1 + \frac{\bar{b}_i^0}{a_i + \bar{c}_i^0}\right) + \frac{m}{2} \log_2 \left(\frac{1+\delta}{1-\delta}\right) \\ &= R(\beta, \bar{\gamma}) + \frac{m}{2} \log_2 \left(\frac{1+\delta}{1-\delta}\right) \\ &< R(\beta, \gamma) + \frac{\epsilon}{2} + \frac{m}{2} \log_2 \left(\frac{1+\delta}{1-\delta}\right). \end{aligned}$$

Choosing $1 > \delta > 0$ small enough to ensure that

$$m \log_2 \left(\frac{1+\delta}{1-\delta}\right) < \epsilon,$$

we conclude that $\lambda < 1$ implies that $M < 2^{n(R(\beta,\gamma)+\epsilon)}$ for all $n > n_0$, thereby establishing a bound of the form given in Theorem 2 for all (n, M, λ) *deterministic* codes. However, since the bounds obtained were *uniform* (i.e., n_0 does not depend on C_n) over the class of deterministic codes, they hold for random codes as well. This completes the proof of Theorem 2.

IV. Concluding Remarks

We have established the random coding capacity of the VGAVC when the transmitter and jammer are subject to time-averaged power constraints. Although the proof of this result is complicated, the final capacity formula (as given by Theorem 3) has a simple interpretation: C is identical to the capacity of the vector additive Gaussian noise channel that would be formed if the jammer transmitted a sequence of i.i.d $N(0, \mathrm{diag}(\boldsymbol{c}^0))$ random vectors (cf. (2.6)). Thus although the definition of error probability (2.5) presumes that an intelligent jammer will exploit knowledge of the statistics of the transmitter's random code to inflict the largest possible error probability, we find that the jammer, regardless of how he distributes his power, can do no more harm (in the sense of limiting achievable rates of reliable transmission) then memoryless Gaussian noise with the water-filling power distribution.

The results of this paper generalize [9, theorem 1] to vector channels. In [9] it was shown that for time-averaged power constraints on the transmitter and the jammer the capacity of the scalar GAVC is the same as the Gaussian channel that results when the jammer transmits memoryless Gaussian noise at the maximum allowable power. We might also consider imposing *ensemble*-averaged power constraints, as in [9]. Under such constraints it is likely that the VGAVC, like the scalar GAVC's of [9], will have no capacity in the usual sense. The exact form of the region of achievable rates and error probabilities, however, is not presently known.

The capacity of the VGAVC for *deterministic* codes, if it exists, remains an open problem, as it is in the scalar case [9]. The presence of a cost structure on the set of transmitter and jamming symbols causes the results for the GAVC to be qualitatively unlike those of the discrete AVC. It is important to note that many methods that have proved useful in the study of discrete AVC's, notably the Ahlswede [17] elimination technique, cannot be applied when cost constraints are considered [9], [19]. It is likely that new techniques will need to be developed to deal with the special features of these channels.

Acknowledgment

The authors thank A. Ephremides for useful discussions and an anonymous reviewer for a careful reading of this paper and for many helpful comments. The first author would also like to acknowledge the support of the Information Technology Division of the Naval Research Laboratory, Washington, DC.

Appendix

Proof of Lemma 1

Using the inequality $u(t) \le 2^t$ for all real t (where $u(t)$ is the unit step function), we derive the following upper bound:

$$\begin{aligned}
&\Pr\left\{\frac{1}{n} I_c(U^*; Y) > \alpha\right\} \\
&\quad \le \boldsymbol{E}\{2^{(I_c(U^*;Y)-n\alpha)}\} \\
&\quad = 2^{n(r(\boldsymbol{b}^0,\boldsymbol{c})-\alpha)} \prod_{i,j=1}^{m,n} \boldsymbol{E}\left\{\exp\left(\frac{y_{ij}^2}{2(b_i^0+c_i+a_i)} - \frac{(y_{ij}-u_{ij}^*)^2}{2(c_i+a_i)}\right)\right\} \\
&\quad \overset{a)}{=} 2^{n(r(\boldsymbol{b}^0,\boldsymbol{c})-\alpha)} \prod_{i,j=1}^{m,n} \frac{1}{\sqrt{1+b_i^0/(c_i+a_i)}} \\
&\quad = 2^{n(r(\boldsymbol{b}^0,\boldsymbol{c})-\alpha)} \times 2^{-nr(\boldsymbol{b}^0,\boldsymbol{c})} = 2^{-n\alpha}.
\end{aligned}$$

Step a) follows by observing that if X is $N(\mu, \sigma^2)$ and $b < (2\sigma^2)^{-1}$, then [18, appendix 7C]

$$\boldsymbol{E}\{e^{bX^2}\} = \frac{1}{\sqrt{1-2b\sigma^2}} \exp\left\{\frac{b\mu^2}{1-2b\sigma^2}\right\}. \tag{A.1}$$

This completes the proof of Lemma 1.

Proof of Lemma 2

Using the inequality $u(t) \le 2^{\rho t}$ for all real t and $\rho \ge 0$, we obtain the following upper bound for all $\boldsymbol{c}$:

$$\begin{aligned}
&\Pr\left\{\frac{1}{n} I_c(U^*; U^*+H^*+S) \le r(\boldsymbol{b}^0,\boldsymbol{c}) - \delta\right\} \\
&\quad = \Pr\left\{\log_2 e \sum_{i,j=1}^{m,n} \frac{(\eta_{ij}^*+s_{ij})^2}{2(c_i+a_i)} - \frac{(u_{ij}^*+\eta_{ij}^*+s_{ij})^2}{2(b_i^0+c_i+a_i)} \ge n\delta\right\} \\
&\quad \le 2^{-n\rho\delta} \prod_{i,j=1}^{m,n} \boldsymbol{E}\left\{\exp\left(\frac{\rho(\eta_{ij}^*+s_{ij})^2}{2(c_i+a_i)} - \frac{\rho(u_{ij}^*+\eta_{ij}^*+s_{ij})^2}{2(b_i^0+c_i+a_i)}\right)\right\}.
\end{aligned} \tag{A.2}$$

The factors on the right side in (A.2) can be evaluated as follows:

$$\begin{aligned}
&\boldsymbol{E}\left\{\exp\left(\frac{\rho(\eta_{ij}^*+s_{ij})^2}{2(c_i+a_i)} - \frac{\rho(u_{ij}^*+\eta_{ij}^*+s_{ij})^2}{2(b_i^0+c_i+a_i)}\right)\right\} \\
&\quad = \boldsymbol{E}\left\{\exp\left(\frac{\rho(\eta_{ij}^*+s_{ij})^2}{2(c_i+a_i)}\right)\right. \\
&\qquad \left.\times \boldsymbol{E}\left\{\exp\left(-\frac{\rho(u_{ij}^*+\eta_{ij}^*+s_{ij})^2}{2(b_i^0+c_i+a_i)}\right)\middle|\eta_{ij}^*\right\}\right\} \\
&\quad \overset{a)}{=} \sqrt{\frac{(b_i^0+c_i+a_i)}{[(1+\rho)b_i^0+c_i+a_i]}} \times \boldsymbol{E}\left\{\exp\left(\pi_i(\rho)(\eta_{ij}^*+s_{ij})^2\right)\right\} \\
&\quad \overset{b)}{=} \sqrt{\frac{(b_i^0+c_i+a_i)}{[(1+\rho)b_i^0+c_i+a_i]}} \\
&\qquad \times \frac{1}{\sqrt{1-2\pi_i(\rho)a_i}} \times \exp\left(\frac{\pi_i(\rho)s_{ij}^2}{1-2\pi_i(\rho)a_i}\right)
\end{aligned}$$

where

$$\pi_i(\rho) \equiv \frac{\rho(1+\rho)b_i^0}{2(c_i+a_i)\left((1+\rho)b_i^0+c_i+a_i\right)}.$$

Steps a) and b) above follow from (1). Substituting b) into (A.2), we find that

$$\Pr\left\{\frac{1}{n} I_c(U^*; U^*+H^*+S) \le r(\boldsymbol{b}^0,\boldsymbol{c}) - \delta\right\} \le e^{-nE(\rho,\boldsymbol{c},S)} \tag{A.3}$$

for all $n \ge 1$, $\rho \ge 0$, and $\boldsymbol{c} \ge 0$, where

$$\begin{aligned}
E(\rho,\boldsymbol{c},S) \equiv \rho\delta\ln 2 + \sum_{i=1}^{m}\left[\frac{1}{2}\ln(1-2\pi_i(\rho)a_i)\right. \\
\left. + \frac{1}{2}\ln\left(1+\frac{\rho b_i^0}{b_i^0+c_i+a_i}\right) - \frac{\pi_i(\rho)c_i^J}{1-2\pi_i(\rho)a_i}\right]
\end{aligned} \tag{A.4}$$

where $c_i^J \equiv (1/n)\sum_{j=1}^n s_{ij}^2$. We can assume that $\sum c_i^J = \gamma$, since otherwise we could bound (A.4) below by multiplying the constants c_i^J by $\gamma/\sum c_i^J$. Under this assumption we have a simple lower bound for $\rho < 1$:

$$\begin{aligned}
E(\rho,\boldsymbol{c},S) &\overset{a)}{\ge} \rho\delta\ln 2 \\
&\quad + \sum_{i=1}^{m}\left[\frac{\rho b_i^0}{2((1+\rho)b_i^0+c_i+a_i)} - \frac{\pi_i(\rho)(c_i^J+a_i)}{1-2\pi_i(\rho)a_i}\right] \\
&\overset{b)}{\ge} \rho\delta\ln 2 + \left[\frac{\rho}{1+\rho} - \frac{\rho\theta_n}{1-\rho}\right]\sum_{i=1}^{m}\frac{b_i^0}{2(b_i^0+c_i+a_i)}
\end{aligned}$$

where

$$\theta_n \equiv \max_{1\le i\le m}\left(\frac{c_i^J+a_i}{c_i+a_i}\right).$$

These steps are justified in the following way: a) follows by applying the inequality $\ln(1+x) \ge x/(1+x)$ for all $x > -1$;

b) follows from $\rho < 1$ and

$$\begin{aligned}
&\frac{\pi_i(\rho)\left(c_i^J + a_i\right)}{1 - 2\pi_i(\rho)\, a_i} \\
&\quad = \frac{\rho(1+\rho)\, b_i^0\left(c_i^J + a_i\right)}{2\left(c_i + a_i\right)\left(\left(1-\rho^2\right) b_i^0 + c_i + a_i\right) + 2\rho(1+\rho)\, b_i^0 c_i} \\
&\quad \le \left(\frac{c_i^J + a_i}{c_i + a_i}\right) \times \frac{\rho(1+\rho)\, b_i^0}{2\left(\left(1-\rho^2\right) b_i^0 + c_i + a_i\right)} \\
&\quad \le \frac{\rho\theta_n}{1-\rho} \times \frac{b_i^0}{2\left(b_i^0 + c_i + a_i\right)}.
\end{aligned}$$

We now choose $\boldsymbol{c}^S \in G_n$ to give the desired result. Let $\bar{i}$ maximize c_i^J, and define[4] $c_i^S \equiv \gamma \lfloor n c_i^J/\gamma \rfloor / n$ for $i \neq \bar{i}$, and $c_{\bar{i}}^S = \gamma - \sum_{i \neq \bar{i}} c_i^S$. It follows that

$$\left(\frac{c_i^J + a_i}{c_i^S + a_i}\right) \le 1 \le \left(\frac{c_i^J + a_i}{c_i^S + a_i}\right)$$

for $i \neq \bar{i}$; hence $\theta_n = (c_i^J + a_i)/(c_i^S + a_i)$. By definition $c_i^J - c_i^S \le m\gamma/n$ and $c_i^J \ge \gamma/m$, so that $c_i^S \ge \gamma/m - m\gamma/n$. If we restrict $n > m^2$, it follows that

$$\theta_n \le \frac{c_i^J}{c_i^S} \le 1 + \frac{m}{n - m^2}. \tag{A.5}$$

It follows from (A.5) and b) that for the $\boldsymbol{c}^S$ chosen earlier,

$$\begin{aligned}
E\left(\rho, \boldsymbol{c}^S, S\right) &\ge \rho\delta \ln 2 + \left[\frac{\rho}{1+\rho} - \frac{\rho}{1-\rho} - \frac{\rho}{1-\rho}\frac{m}{n-m^2}\right] \\
&\quad \cdot \sum_{i=1}^{m} \frac{b_i^0}{2\left(b_i^0 + c_i^S + a_i\right)} \\
&\ge \rho\delta \ln 2 - \left[\frac{2\rho^2}{1-\rho^2} - \frac{\rho}{1-\rho}\frac{m}{n-m^2}\right]\frac{m}{2} \\
&\ge \rho\left(\delta \ln 2 - \tau_n\right) - 2\rho^2 m, \qquad \text{for } \rho < 1/2
\end{aligned}$$

where $\tau_n \equiv m^2/(n - m^2)$. Let $\rho^0 = (\delta \ln 2 - \tau_n)/4m$. Restricting $n > (1 + 2/\delta \ln 2)m^2$ and $\delta \le 5m$ ensures $\tau_n < \delta \ln 2/2$ and $0 \le \rho^0 < 1/2$. Therefore,

$$E\left(\rho^0, \boldsymbol{c}^S, S\right) \ge \frac{(\delta \ln 2)^2}{32m} \ge \frac{\delta^2 \ln 2}{47m}. \tag{A.6}$$

Combining (A.6) with (A.3), it follows that for all $0 < \delta \le 5m$ and $n > (1 + 2/\delta \ln 2)m^2$

$$\Pr\left\{\frac{1}{n} I_c\left(U^*; U^* + H^* + S\right) \le r\left(\boldsymbol{b}^0, \boldsymbol{c}\right) - \delta\right\} \le 2^{-n\delta^2/47m},$$

as desired. This completes the proof of Lemma 2.

Proof of Lemma 3

Let $b_i^U \equiv (1/n)\sum_{j=1}^{n} u_{ij}^2$ and define

$$Z \equiv \sum_{i,j=1}^{m,n} \frac{\left(u_{ij} + \eta_{ij}^* + s_{ij}^*\right)^2}{mn(1+\delta)\left(\bar{b}_i^0 + \bar{c}_i^0 + a_i\right)}.$$

[4] $\lfloor x \rfloor$ represents the unique integer l such that $x \le l \le x + 1$.

It is easy to show that

$$\max_{b_i \ge 0:\, \sum b_i \le \beta} \sum_{i=1}^{m} \frac{b_i + \bar{c}_i^0 + a_i}{\bar{b}_i^0 + \bar{c}_i^0 + a_i} = m.$$

Using this, we obtain the following bounds:

$$\begin{aligned}
\boldsymbol{E}(Z) &= \sum_{i=1}^{m} \frac{\left(b_i + \bar{c}_i^0 + a_i\right)}{m(1+\delta)\left(\bar{b}_i^0 + \bar{c}_i^0 + a_i\right)} \le \frac{1}{(1+\delta)} \\
\operatorname{var}(Z) &= \sum_{i=1}^{m} \frac{2\left(\bar{c}_i^0 + a_i\right)^2 + 4\left(\bar{c}_i + a_i\right) b_i}{m^2 n(1+\delta)^2\left(\bar{b}_i^0 + \bar{c}_i^0 + a_i\right)^2} \\
&< \frac{4}{m^2 n(1+\delta)^2} \sum_{i}^{m} \frac{b_i + \bar{c}_i^0 + a_i}{\bar{b}_i^0 + \bar{c}_i^o + a_i} \le \frac{4}{mn(1+\delta)^2}.
\end{aligned}$$

We now apply Chebyshev's inequality [16, p. 190] to obtain

$$\Pr(Z > 1) \le \frac{\operatorname{var}(Z)}{(1 - \boldsymbol{E}(Z))^2} < \frac{4}{mn\delta^2}.$$

This completes the proof of Lemma 3.

Proof of Lemma 4

For any set $B \subset \boldsymbol{R}^{m \times n}$, let $P(B) \equiv \Pr(H^* + S^* \in B)$. Observe that $H^* + S^*$ has a probability density function that is invariant on the boundary of E_δ^2; further, it is a strictly decreasing function of $0 \le \delta < 1$. It follows that $P(E_\delta^2) \le P(A)$ implies $\nu(A) \ge \nu(E_\delta^2)$. To see this, note that $P(A - E_\delta^2) \ge P(E_\delta^2 - A)$ implies $\nu(A - E_\delta^2) \ge \nu(E_\delta^2 - A)$.

Application of the Chernoff bound [18, p. 97] to $P(E_\delta^2)$ yields

$$P\left(E_\delta^2\right) \le \left[(1-\delta)\, e^{\delta}\right]^{mn/2} = \left[g(\delta)\right]^{mn/2}$$

where $g(\cdot)$ is strictly decreasing from 1 to 0 on $0 \le \delta < 1$. Thus for $\delta = g^{-1}(P(A)^{2/mn})$, we have $P(A) \ge P(E_\delta^2)$, as desired. This completes the proof of Lemma 4.

References

[1] I. Csiszár and J. Körner, *Information Theory: Coding Theorems for Discrete Memoryless Systems*. New York: Academic, 1981.

[2] C. E. Shannon, "A mathematical theory of communication," *Bell Syst. Tech. J.*, vol. 27, pp. 379–423 and 623–656, July and Oct. 1948.

[3] N. M. Blachman, "Communication as a game," in *WESCON Conf. Rec.*, 1957, pp. 61–66.

[4] R. L. Dobrushin, "Optimal information transmission over a channel with unknown parameters," *Radiotekh. Electron.*, vol. 4, no. 12, pp. 1951–1956, 1959.

[5] N. M. Blachman, "On the capacity of a band-limited channel perturbed by statistically dependent interference," *IRE Trans. Inform. Theory*, vol. IT-8, pp. 48–55, Jan. 1962.

[6] ——, "The effect of statistically dependent interference upon channel capacity," *IRE Trans. Inform. Theory*, vol. IT-8, pp. 553–557, Sept. 1962.

[7] T. Başar and Y. Wu, "Solutions to a class of minimax decision problems arising in communications systems," in *Proc. 23rd Conf. Decision and Control*, Dec. 1984, pp. 1182–1187.

[8] R. Ahlswede, "The capacity of a channel with arbitrarily varying Gaussian channel probability functions," in *Trans. 6th Prague Conf. Information Theory, Statistical Decision Functions, and Random Processes*, Sept. 1971, pp. 13–21.

[9] B. Hughes and P. Narayan, "Gaussian arbitrarily varying channels," *IEEE Trans. Inform. Theory*, vol. IT-33, pp. 267–284, Mar. 1987.

[10] D. Blackwell, L. Breiman, and A. J. Thomasian, "The capacities of

certain channel classes under random coding," *Ann. Math. Statist.*, vol. 31, pp. 558–567, 1960.
[11] B. Hughes and P. Narayan, "Interleaving and channels with unknown memory," in *Proc. 19th Ann. Conf. Information Sciences and Systems*, Mar. 1985.
[12] I. Csiszár and P. Narayan, "Arbitrarily varying channels with constrained inputs and states," *IEEE Trans. Inform. Theory*, vol. 34, pp. 27–34, Jan. 1988.
[13] R. G. Gallager, *Information Theory and Reliable Communication*. New York: Wiley, 1968.
[14] G. E. Shilov, *Linear Algebra*. Englewood Cliffs, NJ: Prentice-Hall, 1984.
[15] D. Blackwell and M. A. Girshick, *Theory of Games and Statistical Decisions*. New York: Wiley, 1954.
[16] A. N. Shirayayev, *Probability*. New York: Springer-Verlag, 1984.
[17] R. Ahlswede, "Elimination of correlation in random codes for arbitrarily varying channels," *Z. Wahrscheinlichkeitstheorie Verw. Gebiete*, vol. 44, pp. 159–175, 1978.
[18] J. Wozencraft and I. M. Jacobs, *Principles of Communication Engineering*. New York: Wiley, 1965.
[19] I. Csiszár and P. Narayan, "The capacity of the arbitrarily varying channel revisited: Positivity and constraints," *IEEE Trans. Inform. Theory*, vol. 34, pp. 181–193, Mar. 1988.