On the Capacity of Channels with Unknown Interference

MANJUNATH V. HEGDE, MEMBER, IEEE, WAYNE E. STARK, MEMBER, IEEE, AND DEMOSTHENIS TENEKETZIS, MEMBER, IEEE

Abstract—**We model the process of communicating in the presence of interference, which is unknown or hostile, as a two-person zero-sum game with the communicator and the jammer as the players. The objective function we consider is the rate of reliable communication. The communicator's strategies are encoders and distributions on a set of quantizers. The jammer's strategies are distributions on the noise power subject to certain constraints. We consider various conditions on the jammer's strategy set and on the communicator's knowledge. For the case where the decoder is uninformed of the actual quantizer chosen we show that, from the communicator's perspective, the worst-case jamming strategy is a distribution concentrated on a finite number of points, thereby converting a functional optimization problem into a nonlinear programming problem. Moreover we are able to characterize the worst-case distributions by means of necessary and sufficient conditions which are easy to verify. For the case where the decoder is informed of the actual quantizer chosen we are able to demonstrate the existence of saddle-point strategies. The analysis is also seen to be valid for a number of situations where the jammer is adaptive.**

I. INTRODUCTION

THE APPLICABILITY of game-theoretic models in jamming situations is by now well established [3], [7], [18], [19], [21]–[23]. In this paper we formulate fairly general models for a number of jamming situations as two-person zero-sum games between the communicator and the jammer. We allow the jammer the choice of one of a set of noise distributions satisfying peak and average power constraints. By way of countermeasure the communicator is allowed to randomize the input symbols as well as randomize the quantizer at the output side. We intend the analysis to be applicable to the performance of soft-decision decoding for jammed channels.

Before describing the channel model we will use, we provide the motivation for considering the problem. Typically, in a spread-spectrum channel the performance in additive white Gaussian noise is identical to the performance of nonspread systems; namely, the bit error probability decreases exponentially with the signal-to-noise ratio. However, when subject to worst-case partial-band or pulsed jamming (wherein power is concentrated in time or frequency to affect only a fraction of the symbols transmitted while allowing the remaining to be received "error-free") the bit error probability of a spread-spectrum system decreases only inverse linearly with the signal-to-noise ratio. This is a significant degradation, typically on the order of 30–40 dB compared to an additive white Gaussian noise channel for a bit error probability on the order of 10^{-5}.

To remedy this situation, most systems use some form of error-correction coding. As has been well-known in the communication field, hard-decision decoding requires roughly a 2-dB larger signal-to-noise ratio than soft-decision decoding for the same error probability. Thus considerable interest has focused on soft-decision decoding. One problem that has been observed is that if a (soft) decoding algorithm designed for a nonjammed channel is used for a jammed channel, then the performance is extremely poor when the jamming strategy is optimized. One method for "overcoming" this difficulty is to assume the jamming noise has one of two distributions (usually one having zero variance called the "off" state and the other called the "on" state) and that the decoder knows when the jammer is "on" and when the jammer is "off." Most systems analyses do not incorporate jamming strategies that affect the reliability of the side information (see, however, [24]).

Thus there is considerable interest in decoding algorithms that do not assume side information and do not do hard-decision decoding. However, most of these algorithms still assume the jammer pulses between one of two levels. In this paper we investigate the case of a decoder that processes symbols from a finite alphabet (i.e., multilevel quantization) and where the only constraints on the jammer are average and peak power. We formulate the problem as a game with two players. The jammer, whose strategy set consists of distributions on the power of the jamming noise, and the communicator, whose strategy set consists of encoders and distributions on the set of quantizers. The objective function is the rate of reliable communication, with the communicator wishing to maximize the rate and the jammer seeking to minimize the rate. We first

Manuscript received August 19, 1987; revised November 9, 1988. This work was supported in part by the Office of Naval Research under Contract N00014-85-K0545, by the National Science Foundation under Grant ECS-8517708, and by a Rackman Research Grant of the University of Michigan. This paper was presented in part at the 25th Annual Allerton Conference on Communication, Control, and Computing, University of Illinois, Urbana-Champaign, October 1987.

M. V. Hegde was with the Electrical Engineering and Computer Science Department, University of Michigan, Ann Arbor. He is now with the Electrical and Computer Engineering Department, Louisiana State University, Baton Rouge, LA 70803.

W. E. Stark and D. Teneketzis are with the Electrical Engineering and Computer Science Department, University of Michigan, Ann Arbor, MI 48109-2122.

IEEE Log Number 8929035.

0018-9448/89/0700-0770$01.00

show that this game is equivalent to a game with mutual information as the objective function and the communicator's strategies replaced by distributions on the input to the channel and distributions on the quantizer selected. We look for worst-case jamming strategies and investigate when the game admits a saddle point. Other work done on an information-theoretic modeling of spread-spectrum systems subject to jamming can be found in [6], [11], [15], [18], [21]–[23]. These papers, however, do not consider multilevel jamming and soft-decision decoding, both of which are considered in this paper.

We now describe the basic setup of our problem and assumptions. After the model is described we will explain how the model applies to a frequency-hopped spread-spectrum communication system. We consider a modulator that transmits one out of M symbols. This transmitted symbol is denoted by the random variable X. The received signal which has been corrupted by the jammer in some fashion is demodulated and quantized into one of L values. To forbid the jammer from using knowledge of the quantizer in designing his worst-case strategy, we allow randomization of the quantizer over some given set of quantizers. Clearly, such randomization increases the size of the communicator's strategy set. Thus we view this situation as a game with two players: the jammer and the communicator. The jammer selects the noise power in the channel, and the communicator chooses the encoder, the decoder, and the quantizer. The jammer can be thought of as modulating a generic noise variable by varying the power according to some distribution. The strategy set for the jammer is the set of all distributions on the power of the jamming noise subject to the given constraints on the peak and average power.

We assume that the jamming strategy, while fixed for a whole codeword, is to choose independently the noise power in the channel from symbol to symbol. There are several reasons for using this model. First, since we are examining the performance of very long codes, we will not, for example, let the jammer pulse on for a whole codeword and then off for a whole codeword or equivalently jam the whole frequency band for a whole codeword. Second, a strategy that is used in many coded systems is interleaving. This, in effect, makes each of the encoders/decoders see a memoryless channel. Third, but not of lesser importance, since the point of the paper is to examine the multilevel jamming strategies and multilevel quantization strategies, we do not complicate the problem by including a jammer with memory.

The strategy set for the communicator is the set of (block) encoders and decoders and distributions on quantizers. Let us denote by E a particular choice of encoder, decoder, and quantizer distribution, and let X denote the input of the channel. Furthermore, let P denote a distribution on the input alphabet, G a distribution on the set of quantizers, F a distribution on the noise power chosen by the jammer, Y a random variable denoting the output of the quantizer, and $I(G, P; F)$ the mutual information, $I(X; Y)$, between X and Y under the choice of F, P, and G. The payoff we are interested in analyzing is the rate of reliable communication (R say) in this situation. The communicator wants to maximize it, and the jammer wants to minimize it. Thus the lower and the upper value of this game would be $\max_E \min_F R(E, F)$, and $\min_F \max_E R(E, F)$, respectively.

Consider the upper value of the game, $\min_F \max_E R(E,F)$. From the channel coding theorem [8, theorem 1.5, p. 104] we see that for each choice of F, $\max_E R(E, F)$ is $\max_{P,G} I(G, P; F)$, and so the upper value of the game is $\min_F \max_{G,P} I(G, P; F)$.

Now consider the lower value of the game, $\max_E \min_F R(E, F)$. From the compound channel coding theorem [8, corollary 5.10, p. 173] we see that this lower value is $\max_{P,G} \min_F I(G, P; F)$.

As a consequence of these observations, we recognize that we may equivalently view the situation as a two-person zero-sum game with the communicator and jammer as players, with the jammer's strategy set being the set of distributions F (subject to some constraint), the communicator's strategy set being the set of distributions (P, G), and with the mutual information $I(G, P; F)$ being the payoff or objective function.

Our basic model can be easily seen to fit a frequency-hop communication system in which the modulation uses an M-ary signal set, using say D dimensions where $D \leq M$ (see the example in Section II). The spread-spectrum bandwidth is divided into a large number of frequency slots. There are several ways that one can hop the modulated signal. One possibility is to have all of the M possible signals use the same pseudorandom hopping pattern. In this case the particular frequency slot used is independent of the data transmitted. Another possibility is to have M frequency hopping patterns, one for each data symbol. In this case the frequency slot used depends on which of the M data symbols is transmitted. The jammer can distribute his total power in any fashion over the whole set of frequency slots. However, the distribution the jammer chooses remains the same for the duration of the codeword. In the first type of hopping system the jammer may be able to add noise in either all or none of the signal dimensions. In the second case the appropriate model is for the noise added in each dimension to be independent. We will say more about these two cases when the model is described mathematically in Section II.

We now summarize the results obtained in this paper. For the general setup just described we show that the worst-case jamming strategy from the communicator's perspective is to pulse between a finite number of power levels. We also consider the case of random quantizing strategies where the demodulator output is quantized into a finite number of outputs by a randomized quantizer, i.e., the quantization thresholds are random. For the case of randomized quantizer thresholds we show that the optimal randomized quantizer can perform better than the nonrandomized quantizer and that from the jammer's point of view the worst-case distribution of the quantizer thresholds is concentrated on a finite number of points.

The remainder of the paper is organized as follows. In Section II we define the models we will be considering and give examples for which our models apply. In Sections III and IV we derive results concerning the worst-case jamming strategy and the optimal quantizer strategy for the cases where the decoder is uninformed about the actual quantizer chosen and informed about the actual quantizer chosen, respectively. Finally, in Section V we discuss our results and state our conclusions and extensions.

II. Channel Models

In this section we describe the models we use in the subsequent analysis. In all cases we consider a modulator that transmits one out of M signals in D dimensions ($D \le M$). This transmitted signal is denoted by the random variable X. The received signal which is corrupted by the jammer in some fashion is demodulated and quantized into one of L values. The received signal is denoted by the random variable Y.

The general philosophy that we will use is that of game theory with the players being the jammer and the communicator. The jamming strategies are distributions dF on D random variables, $Z_1, Z_2, \cdots, Z_D$. These random variables represent the power of the jammer in each of the signal dimensions and are modeled as modulating a generic noise variable present in the channel. For example, if $D=1$ and N is a zero-mean unit-variance Gaussian random variable, then the jammer's noise may be of the form $Z_1 N$. We note here that the distribution of the generic random variable N is not important (except for the constraints on the mean and variance), and all the results hold for any such random variable. The jammer has an average-power constraint and a peak-power constraint. More generally, the jammer is constrained by

$$\int f(z_1, z_2, \cdots, z_D)\, dF(z_1, z_2, \cdots, z_D) \le K_J \qquad (1)$$

and

$$0 \le Z_j \le b_j, \qquad j = 1, \cdots, D \qquad (2)$$

where b_j is the peak-power constraint and $f(z_1, \cdots, z_D)$ is some continuous functional of $(z_1, \cdots, z_D)$. For average power constrained channels with no peak constraint we let b_j become very large. The output of the demodulator is quantized into one of L values, say $0, 1, \cdots, L-1$. The output of the quantizer, Y, is also the output of the channel for coding.

Before proceeding, we illustrate this model with an example. Consider a frequency-hop communication system. The modulated signal is one of two orthogonal tones, i.e., binary frequency shift keying ($D = M = 2$). $X = 0$ corresponds to transmitting a tone in the first and $X = 1$ to transmitting a tone in the second dimension. Before transmission the modulated signal is hopped over a set of q distinct frequencies. The signal is affected by a jammer with constraints on the total power and peak power. The jammer may distribute the total available power in any manner over the set of q frequency slots (subject to the constraints to be mentioned later). Let $W_{i,j}$ be the (random) amount of jamming power in the ith frequency slot and jth signal dimension $i = 1, \cdots, q$ and $j = 1, 2$. The actual noise in the ith frequency slot and jth dimension is $N_j W_{i,j}$ where N_j is the generic (unit variance) noise random variable in dimension j. The received signal is the sum of the transmitted signal and the jamming signal. The frequency dehopper (which is synchronized to the transmitted hopping pattern) dehops the received signal, i.e., selects the appropriate hopping frequency slot for demodulation. Thus the output of the frequency dehopper is the modulated signal plus the jamming noise at the frequency slot chosen by the hopping pattern. Since the frequency hopper chooses each of the q frequency slots with probability $1/q$, the noise power in dimension j at the input to the demodulator is $W_{i,j}$ with probability $1/q$ for $i = 1, \cdots, q$ and $j = 1, 2$. Thus $Z_j = W_{i,j}$ with probability $1/q$. In this example $f(z_1, z_2) = (z_1^2 + z_2^2)/2$, $K_J = 1$, and b_1 and b_2 are arbitrary constants greater than 1. The demodulator is a noncoherent matched filter which basically measures the energy in each of the $D = 2$ signal dimensions and produces a vector (R_1, R_2). The conditional probability distribution of R_j given $Z_j = z_j$ depends on z_j and on the distribution of N_j. The output of the demodulator is quantized by a quantizer from the set Q of possible quantizers with, in this example, four outputs. With Y denoting the output of the quantizer we write

$$Y = q(R_1, R_2) = \begin{cases} 0, & r \le \theta \\ 1, & \theta < r \le 1 \\ 2, & 1 < r \le 1/\theta \\ 3, & 1/\theta < r \end{cases}$$

where $r = R_2^2 / R_1^2$ and θ is a number between 0 and 1. Thus by integrating the conditional distribution of the random variables R_1 and R_2 over the regions just defined we can determine the conditional probability transition matrix $[p(y|x, \theta, z)]$ for every $z = (z_1, z_2)$ and θ. The interpretation of the quantizer is the following. $Y = 0$ represents a transmitted symbol 0 received with high quality, whereas $Y = 1$ represents a transmitted symbol 0 with low quality, etc. The quantizer is parameterized by θ which is between 0 and 1 (see Viterbi [26]). Examples for other types of quantizers and modulators are easy to find.

The strategies for the communicator are to choose a distribution $dG(\theta)$ on Θ, the random quantization thresholds and a distribution, $dP(x)$ on the input alphabet. We will let $\boldsymbol{Q}$ be the parameter space for the quantizers and assume $\boldsymbol{Q}$ is some compact subset of $\boldsymbol{R}$. For each $(z_1, \cdots, z_D)$ and $\theta \in \boldsymbol{Q}$ there is a probability distribution on the output of the channel given the input of the channel:

$$\Pr\{Y = y | X = x, \Theta = \theta, Z_1 = z_1, Z_2 = z_2, \cdots, Z_D = z_D\} = p(y|x, \theta, z_1, z_2, \cdots, z_D). \qquad (3)$$

The foregoing model describes the input/output relation

of the channel for a particular symbol. In addition, we model the channel as being memoryless.

We now introduce some notation. Let

A	$= \{0,1,\cdots, M-1\}$, input alphabet,
B	$= \{0,1,\cdots, L-1\}$, output alphabet,
$\boldsymbol{Q}$	quantizer parameter space (some compact subset of $\boldsymbol{R}$),
Z	$= (Z_1,\cdots, Z_D)$, $0 \le Z_i \le b_i$,
$p(y\|x,\theta,z)$	transition probability from x to y given θ, z,
$P_{y\|x}(\theta, z)$	corresponding stochastic matrix, $P_{y\|x}(\theta,z) = [p(y\|x,\theta,z)]$,
$P_{y\|x}(\theta)$	$= \int_K p(y\|x,\theta,z)\, dF(z)$.

We assume that

1) $p(y|x,\theta,z)$ is continuous in z for all θ, x;
2) $p(y|x,\theta,z)$ is continuous in θ for all x, z.

Let $\boldsymbol{S}$ denote the set of all probability distributions on the Borel sets of $K \triangleq \{\bar{z} = (z_1,\cdots, z_D): 0 \le z_i \le b_i\}$. The mutual information between X and Y when they are related by the stochastic matrix $\bar{P}_{y|x}(G,F)$ is

$$I(G,P;F) = I\big(\bar{P}_{y|x}(G,F)\big) = \sum_{x,y} \bar{P}_{y|x}(G,F)P(x)\log \frac{\bar{P}_{y|x}(G,F)}{\sum_{x'} \bar{P}_{y|x'}(G,F)P(x')} \tag{4}$$

where

$$\bar{P}_{y|x}(G,F) = \int_K \int_Q P_{y|x}(\theta,z)\, dG(\theta)\, dF(z).$$

The performance measure we are interested in is the largest rate such that nearly error-free communication can be achieved, i.e., channel capacity. Another performance measure of interest is the channel cut-off rate R_0 (many researchers [15] believe this to be a practical limit to the set of rates for which reliable communication is possible). Similar results to those in this paper can be derived with R_0 as the performance measure (see [13]). We consider two different information structures for the communicator:

I) The decoder is unaware of the actual quantizer chosen but only knows the distribution $dG(\theta)$ on the set of quantizers. The jammer knows only the set of quantizers but not the distribution $dG(\theta)$ chosen by the communicator. He is also aware of the fact that the decoder does not know the actual quantizer chosen.

II) The decoder knows the actual quantizer chosen. Again, the jammer knows only the set of quantizers. He also knows that the decoder is aware of the actual quantizer chosen.

Case I is seen to apply to situations where, possibly for implementation reasons, the decoding is fixed and not altered with the specific quantizer chosen. It may also be viewed as worst case in the sense that the decoder's knowledge of the specific quantizer and the utilization of such knowledge can only improve the communicator's performance. When there is no randomization of the quantizer, i.e., the quantizer is fixed, Cases I and II are the same and our results for both cases apply.

Several special jamming strategies are of interest because of their correspondence to physical problems. We will classify the cases as follows:

A) arbitrary joint distribution on $Z_1, Z_2,\cdots, Z_D$;
B) $Z_1 = Z_2 = \cdots = Z_D = Z$;
C) one-dimensional jamming, i.e., at most one of the random variables $Z_i \neq 0$;
D) independent jamming, i.e., $Z_1, Z_2,\cdots, Z_D$ are independent.

Case B corresponds to the physical situation where the jammer is not able to place different amounts of power in different dimensions of the signal space of each slot but can place different amounts of power in different frequency slots. Case C corresponds to the case where only one of the dimensions of a slot can be jammed at once. Case D corresponds to a frequency-hop communication system with independent hopping for the different symbols. The standard game-theoretic description is given next.

Communicator's Perspective

The communicator is interested in the maximum rate at which information can be reliably transmitted no matter what strategy the jammer employs. The communicator designs his system assuming the jammer will somehow find out the strategy he is using and then choose the worst possible distribution on the power levels. The largest such rate is

$$\max_{G,P} \min_{F} I(G,P;F)$$

where $I(G,P;F) \triangleq I(X;Y)$ and (dG, dP) is chosen by the communicator and dF is chosen by the jammer. That this is the maximum rate of reliable transmission is well-known since what we are dealing with is a compound channel with a finite input alphabet and a finite output alphabet [8, pp. 172–173].

Jammer's Perspective

The jammer is interested in finding the minimum value of the rate so that information cannot be reliably transmitted at any higher rate no matter what strategy the communicator employs. The jammer designs his system assuming the communicator will somehow find out the strategy he is using and then design the optimal communication system. The jammer attempts to minimize the rate above which reliable communication cannot occur. The smallest such rate is

$$\min_{dF} \max_{dG,dP} I(G,P;F).$$

That this is the smallest rate the jammer can guarantee is obvious because for each F the rate above which reliable communication is impossible is $\max_{dG,dP} I(G,P;F)$.

In Case I, no simplification of the mutual information occurs. However, in Case II the appropriate mutual information can be written as an expectation of the mutual information for a fixed θ:

$$I(G, P; F) = E_G(I(\Theta, P; F))$$

where E_G refers to taking expectations with respect to dG and $I(\Theta, P; F) \triangleq I(X; Y|\Theta)$.

In all of our analysis we assume that the jammer and the decoder/quantizer have complete information about the set of strategies available to each one of these so that no secret information is considered. As mentioned previously, the performance measure we consider is the largest rate such that reliable communication (in the sense of arbitrarily small error probability) is possible.

We are now ready to state the results. In brief, our results show that when the decoder is informed of the quantization rule, then (under a compatibility assumption) there is a saddle-point in Cases A and B, i.e., the jammer's rate and the communicator's rate are equal (Theorem 5). However, when the decoder is not informed of the quantization rule, then the jammer's rate and the communicator's rate may differ. The optimal distributions F from the communicator's point of view and the G from the jammer's point of view are finite dimensional (in all the Cases A, B, C, and D) (Theorem 1). This converts a functional optimization problem into a finite-dimensional nonlinear programming problem.

III. Case AI: Decoder Uninformed

The communicator has to determine the distributions $(dG(\theta), dP(x))$ that maximize the amount of information $I(G, P; F)$ transmitted. The jammer has to find the noise distribution $dF(z)$ to minimize the information received by the decoder. Thus the communicator's goal is to achieve

$$\max_{dG(\theta), dP(z)} \min_{dF(z)} I(G, P; F)$$

whereas the jammer wants to achieve

$$\min_{dF(z)} \max_{dG(\theta), dP(z)} I(G, P; F).$$

In this section we show that for any choice of strategy by either player there is a simple characterization of the optimal reaction strategy of his opponent.

Theorem 1: a) The jammer can achieve the minimum in $\max_{dG(\theta), dP(x)} \min_{dF(z)} I(G, P; F)$ with a distribution concentrated at at most $M(L-1)+2$ points.

b) The communicator can achieve the maximum in $\min_{dF(z)} \max_{dG(\theta), dP(x)} I(G, P; F)$ with a distribution concentrated at at most $M(L-1)+1$ points.

Discussion: Theorem 1a) says that the communicator in trying to achieve $\max_{dG(\theta), dP(x)} \min_{dF(z)} I(G, P; F)$ has to consider only reaction strategies of the jammer that have a finite number of points of support, i.e., for each $(dG(\theta), dP(x))$ chosen by the communicator the worst-case jammer distribution may be assumed to be concentrated at a finite number of points and this number is bounded uniformly (in $(dG(\theta), dP(x))$) by $M(L-1)+2$. It follows that for a fixed quantizer (i.e., no randomization of the quantization) the worst-case jammer is one who chooses such a finite-dimensional distribution. Similarly, Theorem 1b) says that the jammer may, in trying to achieve $\min_{dF(z)} \max_{dG(\theta), dP(x)} I(G, P; F)$, consider only finite-dimensional reaction strategies on the communicator's part.

To prove these results, we use the following facts: 1) the convexity and concavity properties of the mutual information function (it is convex in the channel transition matrix and concave in the input distribution), 2) the equivalence of weak convergence with Levy convergence in our situation [13], a fact which we use to show the continuity of our objective function in the strategies as well as compactness of our strategy sets (this allows us to conclude that there is a worst-case jamming strategy and a best-case communicator strategy), and 3) Dubins' theorem to demonstrate that the optimal reaction strategies are described by distributions concentrated on a finite number of points. Dubins' theorem allows the extreme points of certain convex sets to be written as finite linear combinations of extreme points of larger convex sets. (For an introduction to the use of Dubins' theorem in information theory, see [25]. Some results concerning the Levy metric are contained in Appendix III.)

Proof of Theorem 1: We prove part a) in detail. The modifications required to obtain part b) are straightforward. We start by first proving two intermediate results, Lemmas 1 and 2.

Lemma 1: $I(G, P; F)$ is a Levy-continuous functional of $dF(z)$ for any fixed $(dG(\theta), dP(x))$.

Proof: First we note that for every $(dG(\theta), dP(x))$, $I(\overline{P}_{y|x})$ is a convex function of $\overline{P}_{y|x}$ [8, p. 50], i.e.,

$$I\left(\alpha \overline{P}^1_{y|x} + (1-\alpha)\overline{P}^2_{y|x}\right) \le \alpha I\left(\overline{P}^1_{y|x}\right) + (1-\alpha) I\left(\overline{P^2_{y|x}}\right), \qquad 0 \le \alpha \le 1$$

and

$$p(y|x, z) = \int_Q p(y|x, \theta, z)\, dG(\theta)$$

is a continuous function of z (since $p(y|x, \theta, z)$ is continuous in z and $p(y|x, \theta, z) \le 1$, this follows from the dominated convergence theorem). Also

$$p(y|x) = \int_K \int_Q p(y|x, \theta, z)\, dG(\theta)\, dF(z)$$

$$= \int_K p(y|x, z)\, dF(z).$$

Hence $p(y|x)$ is a Levy-continuous functional of $dF(z)$, and therefore $\overline{P}_{y|x}$ is a Levy-continuous functional of $dF(z)$.

Now $I(G, P; F)$ is a convex function of $\overline{P}_{y|x}$, and hence it is continuous in the interior of the finite-dimensional set $\boldsymbol{W}$ of all stochastic matrices. (Thus $I(G, P; F)$ is continuous at any point $\overline{P}_{y|x}$ such that at least one row of $\overline{P}_{y|x}$ is

not a one point distribution, i.e., $\bar{P}_{y|x}$ is not deterministic.) Hence $I(G, P; F)$ is a Levy-continuous function of $dF(z)$ for any fixed $(dG(\theta), dP(x))$.

Let $\boldsymbol{S} \triangleq$ set of all probability distributions on the Borel subsets of K, and

$$\boldsymbol{S}^1 \triangleq \left\{ dF(z) \in \boldsymbol{S}: \int f(z)\, dF(z) = K_J \right\} \tag{5}$$

be a hyperplane in $\boldsymbol{S}$.

Lemma 2: $I(G, P; F)$ achieves its maximum (minimum) in $\boldsymbol{S}^1$.

Proof: We note that $\boldsymbol{S}$ is compact in the Levy topology [13, appendix C]. Also $\boldsymbol{S}^1$ is a hyperplane in $\boldsymbol{S}$ which is closed (since $dF(z) \to \int_K f(z)\, dF(z)$ is Levy-continuous) in the Levy topology. Hence $\boldsymbol{S}^1$, being a closed subset of a compact set, is itself (Levy) compact.

Thus Lemma 1 asserts that for fixed $(dG(\theta), dP(x))$, $I(G, P; F)$ is a Levy-continuous functional on the compact set $\boldsymbol{S}^1$. Hence it achieves its minimum (maximum) at some point $dF^*(z) \in \boldsymbol{S}^1$.

The lemmas are now used to complete the proof of Theorem 1. From Lemma 2 we know that $I(G, P; F)$ achieves its minimum in $\boldsymbol{S}^1$. Denote the corresponding $\bar{P}_{y|x}$ by $\bar{P}^*_{y|x} = [p^*(y|x)]$, i.e.,

$$\bar{P}^*_{y|x} = \int_K \int_Q p(y|x, \theta, z)\, dG(\theta)\, dF^*(z). \tag{6}$$

Now consider the set

$$\Lambda = \left\{ dF(z) \in \boldsymbol{S}^1: \int_K \int_Q p(y|x, z, \theta)\, dG(\theta)\, dF(z) = p^*(y|x),\ x \in A,\ y \in B^1 \right\} \tag{7}$$

where $B^1 = \{0, 1, \cdots, L-2\}$. The set Λ is the intersection of $\boldsymbol{S}$ with $M(L-1)+1$ hyperplanes viz. $\boldsymbol{S}^1$ and the $M(L-1)$ hyperplanes

$$h_{yx} = \left\{ dF(z) \in \boldsymbol{S}: \int_K \int_Q p(y|x, z, \theta) \cdot dG(\theta)\, dF(z) = p^*(y|x) \right\}. \tag{8}$$

Furthermore, $\boldsymbol{S}$ is convex; $\boldsymbol{S}$ is linearly bounded ($\boldsymbol{S}$ being compact in a metric space is bounded, and hence its intersection with any line is bounded), and $\boldsymbol{S}$ being a compact subset of a metric space is closed and any line l in the metric space is closed. Thus $\boldsymbol{S}$ is also linearly closed. Hence we have that $\boldsymbol{S}$ is a convex, linearly closed, and linearly bounded set. By Dubins' theorem [10] we can conclude that since Λ is the intersection of $\boldsymbol{S}$ with $M(L-1)+1$ hyperplanes, every extreme point of Λ is a convex combination of $M(L-1)+2$ or fewer points of $\boldsymbol{S}$.

From our construction of Λ we know that $I(G, P; F)$ is constant on Λ. Hence for fixed $(dG(\theta), dP(x))$, $I(G, P; F)$ assumes its minimum value at an extreme point of Λ also. Hence $I(G, P; F)$ assumes its minimum value at some point $dF(z)$ which is a convex combination of $M(L-1)+2$ or fewer extreme points of $\boldsymbol{S}$.

Since the extreme points of $\boldsymbol{S}$ are the one-point distributions, we can finally assert that for each $(dG(\theta), dP(x))$ the jammer can achieve the minimum in

$$\max_{dG(\theta),\, dP(x)} \min_{dF(z)} I(G, P; F)$$

with a distribution concentrated at $M(L-1)+2$ points. This concludes the proof of a).

For channels that are symmetric for each θ and z, i.e., $p(y|x_1, z, \theta)$ is some permutation of $p(y|x_i, z, \theta)$, we see that the set Λ is actually the intersection of $\boldsymbol{S}$ with $(L-1)+1$ hyperplanes only and hence part a) of the theorem holds with $(L-1)+2 = L+1$ instead of $M(L-1)+2$. For M-ary symmetric channels, i.e., channels with M inputs and M outputs and such that for each θ and z, $p(y_i|x_i, z, \theta) = 1-\epsilon$ and $p(y_i|x_j, z, \theta) = \epsilon/(M-1)$, $i \neq j$, the bound on the number of points of support reduces to 3.

For b) we note that the jammer wants to achieve

$$\min_{dF(z)} \max_{dG(\theta),\, dP(x)} I(G, P; F).$$

This may be written as

$$\min_{dF(z)} \max_{dG(\theta)} C(G, F)$$

where $C(G, F) \triangleq \max_{dP(x)} I(G, P; F)$.

We note that, as in Lemma 1, for any fixed $dF(z)$, $C(G, F)$ is a continuous functional of $dG(\theta)$. (Simply note that $C(G, F)$, being the maximum of functions convex in $\bar{P}_{y|x}$, is also convex in $\bar{P}_{y|x}$ and proceed as before.) Using our hypothesis that $p(y|x, \theta, z)$ is continuous in θ, we can show that $\min_{dF(z)} \max_{dG(\theta)} C(G, F)$ can be achieved for any $dF(z)$ by the decoder/quantizer with a distribution $dG(\theta)$ that is concentrated at no more than $M(L-1)+1$ points.

Again, for symmetric channels we note that part b) of the theorem holds with L instead of $M(L-1)+1$. For M-ary symmetric channels this number is 2. The number of points of support is one less than Case A as we have not imposed any constraints on the distributions $dG(\theta)$ chosen by the quantizer.

A. Necessary and Sufficient Conditions

We now characterize the aforementioned finite-dimensional distributions by means of necessary and sufficient conditions. We first briefly introduce the appropriate definitions and results from optimization theory and then specialize them to our cases.

Let Ω be a convex set and f a function from Ω into $\boldsymbol{R}$. For some fixed x_0, if for all x

$$\lim_{\alpha \downarrow 0} \frac{f((1-\alpha)x_0 + \alpha x) - f(x_0)}{\alpha} \tag{9}$$

exists, f is said to be weakly differentiable at x_0 and the foregoing limit is denoted by $f'_{x_0}(x)$, the weak derivative at

x_0. If f is weakly differentiable in Ω at x_0 for all x_0 in Ω, f is said to be weakly differentiable in Ω. We now state an optimization theorem that follows from [14, p. 178].

Optimization Theorem: Let f be a continuous weakly differentiable concave map from a compact convex set to $\boldsymbol{R}$. Let

$$C \triangleq \sup_{x \in \Omega} f(x). \tag{10}$$

Then 1) $C = \max f(x) = f(x_0)$ for some $x_0 \in \Omega$; 2) a necessary and sufficient condition for $f(x_0) = C$ is $f'_{x_0}(x) \le 0$ for all $x \in \Omega$.

Constrained Optimization Theorem [14, p. 217]: Let Ω be a convex subset of a linear vector space and f and g concave functionals on Ω to $\boldsymbol{R}$. Assume there is an $x_1 \in \Omega$ such that $g(x_1) < 0$, and let

$$C' \triangleq \sup_{\substack{x \in \Omega \\ g(x) \le 0}} f(x). \tag{11}$$

If C' is finite, then there exists a constant $\lambda \ge 0$ such that

$$C' = \sup_{x \in \Omega} [f(x) - \lambda g(x)]. \tag{12}$$

Furthermore, if the supremum in the first equation is achieved by $x_0 \in \Omega$ and $g(x_0) \le 0$, it is achieved by x_0 in the second equation and $\lambda g(x_0) = 0$ [14, p. 217].

Now given any $dG(\theta)$ and the power constraint we define

$$U_c(K_J, G) \triangleq \sup_{\substack{F \in \boldsymbol{S} \\ h_F \le K_J}} -I(G, P; F) \tag{13}$$

where $h_F \triangleq \int_K f(z)\, dF(z)$. To simplify notation, we define D: $\boldsymbol{S} \to \boldsymbol{R}$ by $D(F) = \int_K f(z)\, dF(z) - K_J$. Using the constrained optimization theorem we will infer in Theorem 2 that a nonnegative constant $\lambda = \lambda(G, K_J)$ exists for $D(F) \le 0$ such that

$$U_c(G, K_J) = \sup_{F \in \boldsymbol{S}} [-I(G, P; F) - \lambda D(F)]. \tag{14}$$

We now formulate necessary and sufficient conditions for the characterization of the optimal distributions of Theorem 1 in the following two theorems.

Theorem 2: $U_c(G, K_J)$ is achieved by a distribution $F_0 \in \boldsymbol{S}$ satisfying $D(F) \le 0$ and a necessary and sufficient condition for $U_c(G, K_J) = -I(G, P; F_0)$ is that for some constant $\lambda \ge 0$

$$\int_K [-i(z; G, F_0) - \lambda f(z)]\, dF(z) \le -I(G, P; F_0) - \lambda K_J \tag{15}$$

for all $F \in \boldsymbol{S}$ where

$$i(z; G, F_0) \triangleq \sum_{x, y} p(x) p(y|x, z) \cdot \log \left(\frac{\int p(y|x, z)\, dF_0(z)}{\sum_x p(x) \int p(y|x, z)\, dF_0(z)} \right).$$

Proof: $D: \boldsymbol{S} \to \boldsymbol{R}$ is clearly linear, bounded, concave, continuous, and weakly differentiable in $\boldsymbol{S}$ with $D'_{F_1}(F_2) = D(F_2) - D(F_1)$. By choosing F_1 as a distribution with unit mass appropriately, we can infer that $D(F_1) < 0$. Next we show that $I(G, P; F)$ is convex in F:

$$\begin{aligned} &I(G, P; \alpha F_1 + (1-\alpha) F_2) \\ &\quad = I\left(\bar{P}_{y|x}(G, \alpha F_1 + (1-\alpha) F_2)\right) \\ &\quad = I\left(\int_K \int_Q p(y|x, \theta, z)\, dG(\theta) (\alpha dF_1 + (1-\alpha)\, dF_2) \right) \\ &\quad = I\left(\alpha \bar{P}_{y|x}(G; F_1) + (1-\alpha) \bar{P}_{y|x}(G; F_2)\right) \\ &\quad = I\left(\alpha \bar{P}^1_{y|x} + (1-\alpha) \bar{P}^2_{y|x}\right) \\ &\quad \le \alpha I\left(\bar{P}^1_{y|x}\right) + (1-\alpha) I\left(\bar{P}^2_{y|x}\right) \\ &\qquad \text{(by the convexity of } I(.) \text{ with respect to } P_{y|x}) \\ &\quad = \alpha I(G, p; F_1) + (1-\alpha) I(G, P; F_2). \end{aligned} \tag{16}$$

Then since $U_c(G, K_J)$ is finite, we can infer from the constrained optimization theorem that there exists some constant $\lambda \ge 0$ such that $U_c = \sup_{F \in \boldsymbol{S}} [-I(G, P; F) - \lambda D(F)]$.

We now show that $I(G, P; F)$ is weakly differentiable at all $F \in \boldsymbol{S}$. Let $L(\alpha) = I(G, P; \alpha F_1 + (1-\alpha) F_2)$. Since $I(G, P; F)$ is convex in F, $L(\alpha)$ is convex in α. Therefore, $(L(\alpha) - L(0))/\alpha$ is nondecreasing in α and bounded from below and thus $\lim_{\alpha \downarrow 0} (L(\alpha) - L(0))/\alpha$ exists. Furthermore, we have the following.

Lemma 3:

$$I'_{F_1}(G, P; F_2) = \int i(z; G, F_1)\, dF_2(z) - I(G, P; F_1).$$

Proof of Lemma 3: See Appendix I.

We now have that $-I(G, P; F) - \lambda D(F)$ is concave, continuous, and weakly differentiable in F. Thus by the optimization theorem there is a distribution function $F_0 \in \boldsymbol{S}$ such that $U_c(G, K_J) = -I(G, P; F_0) - \lambda D(F_0)$. The necessary and sufficient condition becomes

$$-I'_{F_0}(G, P; F) - \lambda D'_{F_0}(F) \le 0 \qquad \text{for all } F \in \boldsymbol{S} \tag{17}$$

or

$$\int_K [-i(z; G, F_0) - \lambda f(z)]\, dF(z) \le -I(G, P; F_0) - \lambda h_{F_0}. \tag{18}$$

If $h_{F_0} < K_J$, the power constraint is trivial and the constant λ is zero, i.e., $D(F_0) < 0$ but $\lambda D(F_0) = 0$. Thus the necessary and sufficient condition is established.

From Theorem 1 we know that it is possible to find F_0 from the set of distributions with a finite number of points of support. Finding such an F_0 entails determining the set of points of increase as well as the amounts of increase of F_0 at those points. Let E_0 denote the set of points of increase of F_0. We now show the following.

Theorem 3: Let F_0 be a probability distribution satisfying the power constraint. Then F_0 achieves $U_c(G, K_J)$ if

and only if for some $\lambda \geq 0$,

C1)
$$-i(z; G, F_0) \leq -I(G, P; F_0) + \lambda(f(z) - K_J), \quad \text{for all } z \in K$$

C2)
$$-i(z; G, F_0) = -I(G, P; F_0) + \lambda(f(z) - K_J), \quad \text{for all } z \in E_0.$$

Proof: Sufficiency is clear because if both conditions 1) and 2) hold, then the conditions of Theorem 2 hold. We show necessity.

Assume that F_0 is "optimal," but C1 is not true. Then there must exist some $z_1 \in K$ such that $-i(z; G, F_0) > -I(G, P; F_0) + \lambda(f(z) - K_J)$. Let $F_1(z)$ be a probability distribution with a unit increase at such a point $z_1 \in K$. Then

$$\int_K [-i(z; G, F_0) - \lambda f(z)]\, dF_1(z) > -I(G, P; F_0) - \lambda K_J \tag{19}$$

which contradicts Theorem 2. Hence C1 must be true.

Now assume that F_0 is "optimal," but C2 is not true. Then since C1 is true, $-i(z; G, F_0) < -I(G, P; F_0) + \lambda(f(z) - K_J)$ for all x in E', where E' is some subset of E_0 with positive measure, i.e.,

$$\int_{E'} dF_0(z) = c > 0. \tag{20}$$

Because $\int_{E_0 - E'} dF_0(z) = 1 - c$ and on $E_0 - E'$

$$i(z; G, F_0) = I(G, P; F_0) + \lambda(f(z) - K_J) \tag{21}$$

and

$$\begin{aligned}\int_K [i(z; G, F_0) - \lambda F(z)]\, dF_0(z) &= \int_{E'} [i(z; G, F_0) - \lambda f(z)]\, dF_0(z) \\ &\quad + \int_{E_0 - E'} [i(z; G, F_0) - \lambda f(z)]\, dF_0(z) \\ &\quad \cdot \int_{K - E_0} [i(z; G, F_0) - \lambda f(z)]\, dF_0(z),\end{aligned}$$

we have

$$-I(G, P; F_0) - \lambda K_J < -I(G, P; F_0) - \lambda K_J \tag{22}$$

i.e., a contradiction. Hence C2 must be true.

Theorems 1 and 3 reduce the calculation of the distributions describing the reaction strategies to finite-dimensional nonlinear programming problems. They can be used to simplify the search for conservative strategies which are optimal for either player. In Theorem 4 we assert the existence of conservative strategies for each player.

Theorem 4: For the game described in Case AI, there exists a conservative strategy $(d\bar{G}(\theta), d\bar{P}(x))$ for the communicator and a conservative strategy $d\bar{F}(z)$ for the jammer, i.e. strategies such that

a) $$\min_{dF(z)} I(\bar{G}, \bar{P}; F) = \max_{dP(x), dG(\theta)} \min_{dF(z)} I(G, P; F) \tag{23}$$

and

b) $$\max_{dP(x), dG(\theta)} I(G, P; \bar{F}) = \min_{dF(z)} \max_{dP(x), dG(\theta)} I(G, P; F). \tag{24}$$

Proof: From Lemmas 1 and 2 we note that a) $I(G, P; F)$ is lower semicontinuous in $dF(z)$ for each $(dG(\theta), dP(x))$, and b) there exists $(dG(\theta), dP(x))$ such that $I(G, P; F)$ is lower semicompact in $dF(z)$. Theorem 4a) now follows from a fundamental existence theorem [2, p. 209, th. 1]. Theorem 4b) follows in a similar way.

B. The Remaining Cases

Case BI: With $F(z)$ now recognized as a one-dimensional distribution, Theorems 1 and 2 are easily seen to be true.

Case CI: We redefine $\boldsymbol{S}$ as follows: $\boldsymbol{S} = \bigcup_{i=1}^{M} L_i$, where L_i is the space of product distributions such that

$$\Pr(Z_i \geq 0) \geq 0$$
$$\Pr(Z_j = 0) = 1, \qquad j \neq i.$$

By our previous arguments each L_i is Levy compact, and hence so is $\boldsymbol{S}$. Now the proofs of Theorem 1 and Theorem 2 follow as before.

Case DI: We perform the analysis by fixing $D-1$ of the D distributions $dF_1, \cdots, dF_D$. By minor modifications in the proof of Lemma 1 we see that $I(X; Y)$ is a Levy continuous functional of $dF_i(z)$ for each i. Defining $\boldsymbol{S}$ and $\boldsymbol{S}^1$ similarly, except that now both are spaces of distributions of $dF_i(z_i)$ instead of $dF(z)$, we see that for each $(dG(\theta), dP(x))$ the jammer can achieve the minimum in

$$\max_{(dG(\theta), dP(x))} \min_{dF(z) = dF_1(z_1), dF_2(z_2), \cdots, dF_D(z_D)} I(G, P; F) \tag{25}$$

with a distribution dF_i concentrated at no more than $M(L-1)+2$ points.

Since i is arbitrary, we can assert that the jammer can achieve the minimum in (16) with distributions dF_i, $i = 1, \cdots, D$, each of which is concentrated at no more than $M(L-1)+2$ points. Part b) of Theorem 1 and Theorem 2 are easily seen to be true as stated.

IV. Case AII: Decoder Informed

We have an arbitrary joint distribution on $Z_1, \cdots, Z_D$. The jammer chooses $dF(z)$ and knows that the decoder knows θ. The communicator chooses $dG(\theta)$ and, further, the decoder knows θ.

In this case we make a "compatibility" assumption, that is, for every θ and $dF(z)$ the capacity-achieving input distribution $dP(x)$ remains the same.

While "compatibility" certainly restricts our model applicability, we show by example that it is often a worst-case

assumption. For instance, we know [9] that if $M = L$ and if the jammer's strategy set is restricted so that for each distribution $dF(z)$ and quantizer θ, $\Pr\{\text{error}|x\} \le \epsilon$ for every x, then the saddle-point strategy for the jammer is to choose a distribution such that

$$p(y|x) = \frac{1}{M}, \quad \text{for all } y, x, \qquad \text{if } \epsilon > 1 - \frac{1}{M}$$

and

$$p(y|x) = \begin{cases} \dfrac{\epsilon}{M-1}, & y \neq x, \\ 1-\epsilon, & y = x, \end{cases} \qquad \text{if } \epsilon \le 1 - \frac{1}{M}$$

and the saddle-point strategy for the communicator is to choose a uniform distribution on the input alphabet. In our model this corresponds to choosing the canonical noise variables so that $p(y|x, \theta)$ is a symmetric channel for each θ. Such symmetry (and thereby "compatibility") is obtained in a number of other situations as a saddle-point strategy. Under certain conditions, when we have convex constraints in the M noise variables affecting the M inputs of the channel which are invariant under any permutation of the M variables (i.e., a "symmetric" constraint), then the choice of a uniform distribution on the input and the choice of a symmetric channel are saddle-point strategies for the communicator and the jammer, respectively (see Appendix II). To describe one more example, if we have M inputs and M outputs,

$$y_i = n_i, \qquad i = 1, \cdots, M, i \neq j$$
$$y_j = A + n_j, \qquad i = j$$

where the n_i are $N(0, v_i)$, $i = 1, \cdots, M$ independent random variables with the constraint $\sum_{i=1}^{M} v_i = c$, then from arguments similar to those in Appendix II it can be seen that the saddle-point strategy is to choose $v_i = c/M$ and a uniform distribution on the input.

Utilization of the "compatibility" assumption allows us to write the problem for the communicator and jammer as

$$\min_{dF(z)} \max_{dG(\theta)} E_G(C(\theta, F))$$

and

$$\max_{dG(\theta)} \min_{dF(z)} E_G(C(\theta, F))$$

where $C(\theta, F) = \max_{dP(x)} I(\theta; F)$ and $I(\theta; F) = I(X; Y|\theta)$.

In this section we prove the existence of a saddle-point. The main result is stated in the following theorem.

Theorem 5: There exists a pair of distributions $(dG^*(\theta), dF^*(z))$ such that

$$E_G(C(\theta, F^*)) \le E_{G^*}(C(\theta, F^*)) \le E_{G^*}(C(\theta, F))$$

for all feasible $dG(\theta), dF(z)$, i.e., $(dG^*(\theta), dF^*(z))$ is a saddle-point for the game case AII.

Proof: The set of all feasible dF's, i.e.,

$$\left\{ dF(z) : \int_K f(z)\, dF(z) \le K_J \right\}, \qquad 0 \le z_i \le b_i$$

is clearly convex and compact. The set of all dG's is also convex and compact.

We note that for any fixed $dF(z), C(\theta, F)$ is a continuous function of θ:

$$p(y|x, \theta) = \int_K p(y|x, \theta, z)\, dF(z)$$

is by our earlier arguments a continuous function of θ. Hence $P_{y|x}(\theta)$ is a continuous function of θ. Also $C(\theta, F) = C(P_{y|x}(\theta))$, and we know that $C(P_{y|x}(\theta))$ is convex in $P_{y|x}(\theta)$. Therefore, for every $\theta \in Q$ such that $P_{y|x}(\theta)$ is not deterministic, $C(P_{y|x}(\theta))$ is a continuous function of $P_{y|x}(\theta)$. Hence for fixed $dF(z), C(\theta, F) = C(P_{y|x}(\theta))$ is a continuous function of θ and so

$$E_G(C(\theta, F)) = \int_Q C(\theta, F)\, dG(\theta) \tag{26}$$

is a Levy continuous functional of $dG(\theta)$.

Since $E_G(C(\theta, F))$ is linear in $dG(\theta)$, it is also a concave function of $dG(\theta)$. Next we note that $C(\theta, F)$ is convex in $dF(z)$ for each θ since $C(\theta, F) = C(P_{y|x}(\theta))$. Hence

$$C(\theta, \alpha F^1 + (1-\alpha)F^2) \le \alpha C(\theta, F^1) + (1-\alpha) C(\theta, F^2) \qquad 0 \le \alpha \le 1.$$

Taking expectations with respect to G,

$$\int_Q C(\theta, \alpha F^1 + (1-\alpha)F^2)\, dG(\theta) \le \int_Q \left(\alpha C(\theta, F^1) + (1-\alpha) C(\theta, F^2)\right) dG(\theta).$$

Therefore,

$$E_G\left(C(\theta, \alpha F^1 + (1-\alpha)F^2)\right) \le \alpha E_G(C(\theta, F^2)) + (1-\alpha) E_G(C(\theta, F^2)).$$

Consequently, $E_G(C(\theta, F))$ is a convex function in $dF(z)$.

Also $E_G(C(\theta, F))$ is Levy continuous in $dF(z)$. To prove this, it suffices to show that for any sequence F_n converging to F in the Levy metric

$$E_G(C(\theta, F_n)) \to E_G(C(\theta, F)).$$

Since convergence in the Levy metric is in our case equivalent to weak convergence [13, appendix C], it suffices to show this for $F_n \xrightarrow{w} F$. However,

$$\begin{aligned} \lim_n E_G(C(\theta, F_n)) &= \lim_n \int_Q C(\theta, F_n)\, dG \\ &= \int_Q \lim_n C(\theta, F_n)\, dG \\ &\quad \text{(by the dominated convergence theorem)} \\ &= \int_Q C(\theta, F)\, dG \\ &\quad \text{(since } C(\theta, F) \text{ is Levy continuous in } F) \\ &= E_G(C(\theta, F)) \end{aligned}$$

which proves Levy continuity in $dF(z)$. From these properties of the objective function and the convexity and compactness of the feasible strategy sets we recognize that the hypotheses of the Sion minimax theorem of game theory are satisfied [2, theorem 7, p. 218]. This concludes the proof of Theorem 3.

We note that these saddle-point distributions need not have finite support. However, in this case we have an equilibrium, and with no further knowledge of each other's choice of strategy, the jammer and the quantizer should be content utilizing $dG^*(\theta)$ and $dF^*(z)$.

Using the optimization theorem and the constrained optimization theorem, we can derive necessary and sufficient conditions at these saddle points. Given any $dG(\theta)$ and the power constraint, we define

$$\bar{U}_c(K_J, G) \triangleq \sup_{\substack{F \in \boldsymbol{S} \\ h_F \le K_J}} -E_G(C(\theta, F)) \tag{27}$$

and given any $dF(z)$ we define

$$\bar{V}_c(F) \triangleq \sup_{G \in \mathcal{G}} E_G(C(\theta, F)) \tag{28}$$

where $\mathcal{G}$ is the space of distributions on Q. Then we have the following.

Theorem 6: The saddle-point strategies dF^*, dG^* satisfy the following inequalities:

$$E_{G^*}\left(\int(-\bar{i}(z;\theta,F^*)-\lambda f(z))\,dF(z)\right) \le E_{G^*}(-C(\theta,F^*))-\lambda K_J \tag{29}$$

for some $\lambda \ge 0$, for all F where

$$\bar{i}(z;\theta,F) \triangleq \sum_{x,y} P(x)p(y|x,z,\theta)\cdot\log\frac{\int p(y|x,z,\theta)\,dF(z)}{\sum_x P(x)\int p(y|x,z,\theta)\,dF(z)}.$$

Also

$$E_G(C(\theta,F^*)) \le E_{G^*}(C(\theta,F^*)) \tag{30}$$

for all G.

Proof: For any F, denote by $D_{G_0}(E_G(C(\theta,F)))$ the weak derivative of $E_G(C(\theta,F))$ at G_0, and for any G denote by $D_{F_0}(E_G(C(\theta,F)))$ the weak derivative of $E_G(C(\theta,F))$ at F_0. Using Lemma 3 and the dominated convergence theorem, we have

$$D_{F_1}(E_G(-C(\theta,F_2))) = E_G\left(-\int\bar{i}(z;\theta,F_1)\,dF_2\right)+E_G(C(\theta,F_1)) \tag{31}$$

for any F_1, F_2.

Also

$$D_{G_1}(E_{G_2}(C(\theta,F))) = E_{G_2}(C(\theta,F))-E_{G_1}(C(\theta,F)). \tag{32}$$

Now letting $F_1 = F^*, G_1 = G^*$ in (32) and using the constrained optimization theorem and the optimization theorem and the properties of $E_G(C(\theta,F))$ as in Theorem 2, we have that a necessary and sufficient condition for F^* to achieve $\bar{U}_c(K_J, G^*)$ is

$$E_{G^*}\left(-\int(\bar{i}(z;\theta,F^*)-\lambda f(z))\,dF(z)\right) \le E_{G^*}(-C(\theta,F^*))-\lambda K_J \tag{33}$$

for some $\lambda \ge 0$, for all F.

Letting $F_1 = F^*, G_1 = G^*$ in the second equation gives us similarly that a necessary and sufficient condition to achieve $\bar{V}_c(F^*)$ is

$$E_G(C(\theta,F^*)) \le E_{G^*}(C(\theta,F^*)) \tag{34}$$

for all G. Since at a saddle-point $\bar{U}_c(K_J, G^*)$ and $\bar{V}_c(F^*)$ are simultaneously achieved, the theorem follows.

A. The Remaining Cases

Case BII: Theorem 3 holds with $F(z)$ as a one-dimensional distribution.

Case CII: Although $\boldsymbol{S}$ is compact, it is not convex and so we cannot demonstrate that there is a saddle-point strategy.

Case DII: Again, we have that $E_G(C(\theta,F))$ is a Levy-continuous functional of $dG(\theta)$ and is concave in $dG(\theta)$. Also $E_G(C(\theta,F))$ is Levy continuous in $(dF_1(z),\cdots, dF_D(z))$. However, $E_G(C(\theta,F_1,\cdots,F_D))$ is not convex in $(F_1,\cdots,F_D)$. Hence we cannot assert the existence of a saddle point in this case.

B. Fixed Quantizer

Before concluding this section we also point out that if we did not have randomized quantization, then without "compatibility" the game would have a saddle point where the jammer's saddle-point distribution need be concentrated at at most $M(L-1)+2$ points. We summarize this in Theorem 7.

Theorem 7: For any quantizer θ, there exists a pair of distributions $dP^*(x), dF^*(z)$ such that

$$I(\theta,P,F^*) \le I(\theta,P^*,F^*) \le I(\theta,P^*,F) \tag{35}$$

for all feasible dP, dF. Moreover, $dF^*(z)$ can be chosen to be concentrated at at most $M(L-1)+2$ points, and necessary and sufficient conditions for $dF^*(z)$ and $dP^*(x)$ are that for some $\lambda_1, \lambda_2 \ge 0$,

$$-i(z;\theta,F^*) \le -I(\theta,P^*,F^*)+\lambda_1(f(z)-K_J) \tag{36}$$

for all $z \in K$ and

$$-i(z;\theta,F^*) = -I(\theta,P^*,F^*)+\lambda_1(f(z)-K_J) \tag{37}$$

for all $z \in E_0$, where $i(\cdot;\cdot,\cdot)$ is as defined in Theorem 2 with G concentrated on θ. Also

$$I_x(\theta,P^*,F^*) = \lambda_2 \tag{38}$$

for all $x \ni P^*(x) > 0$ and

$$I_x(\theta,P^*,F^*) \le \lambda_2 \tag{39}$$

for all $x \ni P^*(x) = 0$, where

$$I_x(\theta, P^*, F^*) \triangleq \sum_y p(y|x,\theta) \log \frac{p(y|x,\theta)}{\sum_x P^*(x) p(y|x,\theta)}.$$

Proof: From the proof of Theorem 5 we know that all we need to show is that $I(\theta, P, F)$ is (Levy) continuous in $dP(x)$. We show this by considering any sequence $dP_n(x) \xrightarrow{w} dP(x)$ and showing $I(\theta, P_n, F) \to I(\theta, P, F)$. Since x belongs to the finite set A, weak convergence is equivalent to convergence in any finite-dimensional metric.

Now

$$\begin{aligned}
&|I(\theta, P_n, F) - I(\theta, P, F)| \\
&= \left| \sum_{x,y} P_n(x) p(y|x,\theta) \log \frac{p(y|x,\theta)}{\sum_x P_n(x) p(y|x,\theta)} \right. \\
&\quad \left. - \sum_{x,y} P(x) p(y|x,\theta) \log \frac{p(y|x,\theta)}{\sum_x P(x) p(y|x,\theta)} \right| \\
&\le \left| \sum_{x,y} P_n(x) p(y|x,\theta) \log \frac{p(y|x,\theta)}{\sum_x P_n(x) p(y|x,\theta)} \right. \\
&\quad \left. - \sum_{x,y} P_n(x) p(y|x,\theta) \log \frac{p(y|x,\theta)}{\sum_x P(x) p(y|x,\theta)} \right| \\
&\quad + \left| \sum_{x,y} P_n(x) p(y|x,\theta) \log \frac{p(y|x,\theta)}{\sum_x P(x) p(y|x,\theta)} \right. \\
&\quad \left. - \sum_{x,y} P(x) p(y|x,\theta) \log \frac{p(y|x,\theta)}{\sum_x P(x) p(y|x,\theta)} \right| \\
&\le \left| \sum_{x,y} P_n(x) p(y|x,\theta) \right| \left| \log \frac{\sum_x P_n(x) p(y|x,\theta)}{\sum_x P(x) p(y|x,\theta)} \right| \\
&\quad + \sum_x D |P_n(x) - P(x)| \qquad (40)
\end{aligned}$$

where

$$\begin{aligned}
D &= \max_{x,y} p(y|x,\theta) \log \frac{p(y|x,\theta)}{\sum_x p(y|x,\theta)} \\
&\le LD \left| \log \frac{\sum_x P_n(x) p(y|x,\theta)}{\sum_x P(x) p(y|x,\theta)} \right| \\
&\quad + \sum_x D |P_n(x) - P(x)|. \qquad (41)
\end{aligned}$$

Again, since A is finite, we can say that for all $\delta > 0$ there exists an N such that for all $n > N$

$$\begin{aligned}
1 - \delta \le \frac{P_n(x)}{P(x)} &\le 1 + \delta, && \text{for all } x \in A \\
1 - \delta \le \frac{P_n(x) p(y|x,\theta)}{P(x) p(y|x,\theta)} &\le 1 + \delta, && \text{for all } x \in A \\
1 - \delta \le \frac{\sum_x P_n(x) p(y|x,\theta)}{\sum_x P(x) p(y|x,\theta)} &\le 1 + \delta, && \text{for all } x \in A. \qquad (42)
\end{aligned}$$

By the continuity of the log function we can say that for all $\epsilon > 0$ there exists a $\delta > 0$ such that

$$-\epsilon \le \left| \log \frac{\sum_x P_n(x) p(y|x,\theta)}{\sum_x P(x) p(y|x,\theta)} \right| \le \epsilon.$$

The second term in (41) can also clearly be made $\le \epsilon$ for sufficiently large n. Thus the continuity of $I(\theta, P, F)$ with respect to P is confirmed, and the first part of the theorem follows. The bound on the number of points of support of dF^* follows from Theorem 1a). The necessary and sufficient conditions are derived as before from Theorem 3 and well-known results on channel capacity [13, p. 91].

V. Conclusion

We have constructed fairly general channel models which are capable of representing a number of jamming situations. The jammers we have considered have all been nonadaptive, and by using results from the compound channel, we were able to give operational significance to our minimax performance measures, i.e., we asserted the existence of encoders and decoders that can perform at arbitrarily low probabilities of error at rates close to our performance measures. Our analysis is clearly also applicable to many restrictions on the jammer's strategy set other than the ones we have considered.

In the case where the decoder is uninformed (Case I) we have shown that the worst-case jammer strategy (as well as best communicator strategy) need only be one of the class of distributions with finite support. We have a bound on the number of these points of support in terms of the sizes of the input and the output alphabet. Thus we have reduced the computation of the worst-case jamming strategies to a finite-dimensional nonlinear programming problem. Moreover we can characterize these distributions by necessary and sufficient conditions that are fairly easy to test.

In cases where the decoder is informed, we reduce the communicator's strategy set (either by using the "compatibility" assumption or by fixing a quantizer). In such instances, when we have convexity with respect to the jammer's strategy (as in cases AII and BII), we were able to demonstrate the existence of a saddle-point strategy. For the case of nonrandomized quantization we were further able to characterize these saddle-point strategies.

We reiterate that all the above presupposes nonadaptive jamming. The compound channel model which we use indirectly by our choice of objective function is appropriate in this case. We can allow for more sophisticated jammers if we incorporate the cases where the jammer's strategies are allowed to depend on the previous (and present) channel inputs. The appropriate channel model to use then is that of the arbitrarily "star" varying channel (A^*VC) [8, p. 233]. This model generalizes the arbitrarily varying channel (AVC) and includes it as a special case. It is known that the m-capacity (i.e., capacity with maximum probability of error over all the codewords) of the A^*VC is the same as that of the corresponding AVC [8, p. 232]. This capacity is known for the case of binary output alphabet (and finite input alphabet) and equals $\max_{dP(x)} \min_{W \in \bar{\bar{\mathscr{W}}}} I(X;Y)$ where X and Y are the input and the output, respectively, W is any channel chosen from the set of channels $\mathscr{W}$, and $\bar{\bar{\mathscr{W}}}$ is the row-convex closure of $\mathscr{W}$ [8]. In our case the jammer's strategy set is already row-convex closed and hence the appropriate programs would be a) for the communicator

$$\max_{(dG(\theta),\, dP(x))} \min_{dF(z)} I(G,F),$$

and b) for the jammer

$$\min_{dF(z)} \max_{(dG(\theta),\, dP(x))} I(G,F)$$

which is the same objective function as the one we have used. Similarly, in the case where the decoder is informed we would obtain the same objective functions. Thus all the results derived in the previous chapter for the case of mutual information can be extended to the case of the A^*VC channel with binary output. This model may be viewed as a worst-case representation of adaptive jamming. Unfortunately, the m-capacity of the AVC is as yet unknown for output sizes greater than 2. On the other hand, the a-capacity of the AVC (i.e., the capacity with average probability of error) is known to be either 0 or else $\max_{dP(x)} \min_{W \in \bar{\mathscr{W}}} I(X;Y)$ where $\bar{\mathscr{W}}$ is the convex closure of the set $\mathscr{W}$ to which W belongs [8, p. 214]. (In [9] a necessary and sufficient computable condition is given for determining if the capacity is positive.) Since in our model the set of channels is convex as well as row-convex, the a-capacity is known to be greater than 0 if and only if the m-capacity is greater than 0 [1]. Thus with average probability of error, whenever the jammer's strategy set is such that he cannot force the capacity to be 0, then all the results of the preceding chapter extend to the case of the A^*VC channel.

Appendix I

Lemma 3: We have

$$I'_{F_1}(G;F_2) = \int i(z;G,F_1)\, dF_2(z) - I(G;F_1)$$

where

$$i(z;G,F_1) = \sum_{x,y} p(x)\, p(y|x,z) \log \left(\frac{\int p(y|x,z)\, dF_1}{\sum_x p(x) \int p(y|x,z)\, dF_1} \right).$$

Proof: It follows that

$$\begin{aligned} I'_{F_1}(G;F_2) = \lim_{\alpha \downarrow 0} \frac{1}{\alpha} \Bigg\{ & \sum_{x,y} p(x) \Bigg(\int\int p(y|x,z,\theta) \cdot [(1-\alpha)\, dF_1 + \alpha dF_2]\, dG(\theta) \Bigg) \\ & \cdot \log \frac{\left(\int\int p(y|x,z,\theta)[(1-\alpha)\, dF_1 + \alpha dF_2]\, dG(\theta) \right)}{\sum_x p(x) \left(\int\int p(y|x,z,\theta)[(1-\alpha)\, dF_1 + \alpha dF_2]\, dG(\theta) \right)} \\ & - \sum_{x,y} p(x) \left(\int\int p(y|x,z,\theta)\, dF_1 dG(\theta) \right) \\ & \cdot \log \frac{\left(\int\int p(y|x,z,\theta)\, dF_1 dG(\theta) \right)}{\sum_x p(x) \left(\int\int p(y|x,z,\theta)\, dF_1 dG(\theta) \right)} \Bigg\}. \end{aligned} \tag{1}$$

Denoting $\int p(y|x,z,\theta)\, dG(\theta)$ by $p(y|x,z)$,

$$\begin{aligned} I'_{F_1}(G;F_2) = \lim_{\alpha \downarrow 0} \frac{1}{\alpha} \Bigg\{ & \sum_{x,y} p(x) \int p(y|x,z)[(1-\alpha)\, dF_1 + \alpha dF_2] \\ & \cdot \log \frac{\int p(y|x,z)[(1-\alpha)\, dF_1 + \alpha dF_2]}{\sum_x p(x) \int p(y|x,z)[(1-\alpha)\, dF_1 + \alpha dF_2]} \\ & - \int p(y|x,z)\, dF_1 \log \frac{\int p(y|x,z)\, dF_1}{\sum_x p(x) \int p(y|x,z)\, dF_1} \Bigg\} \\ = \lim_{\alpha \downarrow 0} \frac{1}{\alpha} \Bigg\{ & \sum_{x,y} p(x) \Bigg[\int p(y|x,z)\, \alpha dF_2 \\ & \cdot \log \left(\frac{p(y|x,z)[(1-\alpha)\, dF_1 + \alpha dF_2]}{\sum_x p(x) \int p(y|x,z)[(1-\alpha)\, dF_1 + \alpha dF_2]} \right) \\ & - \int p(y|x,z)\, \alpha dF_1 \\ & \cdot \log \left(\frac{\int p(y|x,z)[(1-\alpha)\, dF_1 + \alpha dF_2]}{\sum_x p(x) \int p(y|x,z)[(1-\alpha)\, dF_1 + \alpha dF_2]} \right) \Bigg] \Bigg\} \\ & + \lim_{\alpha \downarrow 0} \frac{1}{\alpha} \sum_{x,y} p(x) \Bigg[\int p(y|x,z)\, dF_1 \\ & \cdot \log \left(\frac{\int p(y|x,z)[(1-\alpha)\, dF_1 + \alpha dF_2]}{\sum_x p(x) \int p(y|x,z)[(1-\alpha)\, dF_1 + \alpha dF_2]} \right) \\ & - \int p(y|x,z)\, dF_1 \log \left(\frac{\int p(y|x,z)\, dF_1}{\sum_x p(x) \int p(y|x,z)\, dF_1} \right) \Bigg] \\ = a + b\,(\text{say}). \end{aligned}$$

By choosing a sequence $\alpha_n \downarrow 0$ and using weak convergence of $(1-\alpha_n)dF_1 + \alpha_n dF_2$ to dF_1

$$a = \int i(z; G, F_1)\, dF_2 - I(G; F_1)$$

$$b = \frac{d}{d\alpha}\left[\sum_{x,y} p(x)\int p(y|x,z)\, dF_1 \cdot \log\left(\frac{\int p(y|x,z)[(1-\alpha)\, dF_1 + \alpha dF_2]}{\sum_x p(x)\int p(y|x,z)[(1-\alpha)\, dF_1 + \alpha dF_2]}\right)\right]_{\alpha=0},$$

at $\alpha = 0$.

Taking the derivative

$$b = \sum_{x',y} p(x')\int p(y|x',z)\, dF_1 \cdot \left\{\frac{\sum_x p(x)\int p(y|x',z)[(1-\alpha)\, dF_1 + \alpha dF_2]}{\int p(y|x,z)[(1-\alpha)\, dF_1 + \alpha dF_2]}\right. \cdot \frac{1}{d^2}\left[\left(\sum_x p(x)\int p(y|x,z)[(1-\alpha)\, dF_1 + \alpha dF_2]\right) \cdot \int p(y|x',z)\,(dF_2 - dF_1) - \int p(y|x',z)[(1-\alpha)\, dF_1 + \alpha dF_2] \cdot \left.\left(\sum_x p(x)\int p(y|x,z)\,(dF_2 - dF_1)\right)\right]\right\}$$

where

$$d \triangleq \sum_x p(x)\int p(y|x,z)[(1-\alpha)\, dF_1 + \alpha dF_2].$$

After some algebraic manipulation it can be shown that $b \to 0$ as $\alpha \downarrow 0$.

Appendix II

Here we consider a communication game with two players, player A who chooses an input distribution r on the M-ary input alphabet, and player B who chooses the $M \times L$ transition probability matrix. Let X and Y denote the input and output random variables, respectively, and let n_i denote the distribution of the random variable associated with the conditional density $p(y|x_i)$. Let the set of all feasible $\bar{n}$'s $(= (n_1, \cdots, n_M))$ be compact. The channel $p(y|x)$ is a function of $\bar{n}(= (n_1, \cdots, n_M))$. Assume that function is linear and that for a choice of $n_i = n$, $i = 1, \cdots, M$ the channel chosen is symmetric. Let $I(r, \bar{n}) \triangleq I(X; Y)$ when A's choice is r and B's choice is $\bar{n}$. Let $n_1, \cdots, n_M$ be constrained by $f_i(n_1, \cdots, n_M) \le c_i$, $i = 1, \cdots, c$, where f_i is a convex symmetric function of $n_1, \cdots, n_M$, i.e., f_i is invariant under any permutation of $n_1, \cdots, n_M$. Then a saddle-point strategy exists for both players. For player A it is to choose a uniform distribution on the input. For player B it is to choose all the components of $\bar{n}$ equal; that is, there exists $\bar{n}^*$ with all its components equal such that

$$I(r, \bar{n}^*) \le I(r^*, \bar{n}^*) \le I(r^*, \bar{n})$$

where r^* corresponds to the uniform input distribution.

Proof: Step 1: $I(r, \bar{n}^*) \le I(r^*, \bar{n}^*)$. This follows from the fact that the mutual information between the input and the output of a symmetric channel is maximized by the uniform distribution.

Step 2: $I(r^*, \bar{n}^*) \le I(r^*, \bar{n})$. Since $I(X; Y)$ is a convex function of $p(y|x)$, which is linear in $\bar{n}$, $I(r, \bar{n})$ is convex in $\bar{n}$. Moreover, given the form of the constraints, the set of feasible $\bar{n}$'s is a convex set.

Now for any $\epsilon > 0$, let $\inf I(r^*, \bar{n}) + \epsilon$ be achieved at some $\bar{n}_1 \neq \bar{n}^*$. Then we show $I(r^*, \bar{n}^*) \le I(r^*, \bar{n}_1)$, proving that the minimum is also achieved at $\bar{n}^*$. The use of a uniform distribution on the input and the symmetry of the constraints implies that for any permutation of $\bar{n}_1 (\bar{n}_1^\alpha$ say) we have a new channel $p^\alpha(y|x)$ which involves just a relabeling of the inputs of the original channel. The mutual information $I(r^*, \bar{n}_1)$ is equal to $I(r^*, \bar{n}_1^\alpha)$. Now consider all the $M!$ permutations of $\bar{n}_1 = \bar{n}_1^\alpha$: $\alpha \in T$ (not all the permutations are distinct, but this does not matter). Take the convex combination $1/M! \sum_{\alpha \in T} \bar{n}_1^\alpha = \bar{n}_e$ (say). Every component of $\bar{n}_e$ is equal to $1/M! \sum_{i1}^{M} n_{1i}$. Also from the convexity of $I(r^*, \bar{n})$ w.r.t. $\bar{n}$ we know that

$$I\left(r^*, \frac{1}{M!}\sum_{\alpha \in T} n_1^\alpha\right) \le \frac{1}{M!}\sum_{\alpha \in t} I(r^*, n_1^\alpha) = I(r^*, n_1).$$

Therefore,

$$I(r^*, \bar{n}_e) \le I(r^*, \bar{n}_1)$$

and hence $\inf I(r^*, \bar{n}) + \epsilon$ is achieved at $\bar{n}_e$ too. The result follows from the observation that $I(r^*, \bar{n})$ is concave in r.

Appendix III

We append here a collection of results (without proof) on the Levy metric and topology which are utilized in various parts of the paper. The proofs may all be found in [13, appendices A, B, C].

Definition: 1: The Levy metric on the space of all D-dimensional distributions of K is defined as

$$d(F, G) = \inf\{h: F(x_1 - h, x_2 - h, \cdots, x_D - h) - h \le G(x_1, \cdots, x_D) \le F(x_1 + h, \cdots, x_D + h) + h, \text{ for all } (x_1, \cdots, x_D)\}$$

where F and G are any D-dimensional distributions on K and $(x_1, \cdots, x_D) \in K^D$. It is easy to verify that $d(F, G)$ satisfies the three properties of a metric:

1) $d(F, G) \ge 0$ and $= 0$, if and only if $F = G$;
2) $(F, G) = d(G, F)$;
3) $d(F, H) \le d(F, G) + d(G, H)$ for any D-dimensional distributions F, G, and H.

Definition: 2: A sequence of distribution functions F_n on $\boldsymbol{R}^D$ is said to converge weakly to F if and only if for any bounded continuous function $f(\bar{x})$ defined on $\boldsymbol{R}^D$ (where $\bar{x}$ is $(x_1, \cdots, x_D)$)

$$\int_{\boldsymbol{R}^D} f(\bar{x})\, dF_n(x) \int_{\boldsymbol{R}^D} f(x)\, dF(x).$$

This kind of convergence is written $F_n \xrightarrow{w} F$.

Theorem: With F and $F_1, F_2, \cdots$ denoting distribution functions of the random vector $\bar{X} = (X_1, X_2, \cdots, X_d)$ such that $t_i \leq X_i \leq u_i$, the following are equivalent:

1) $F_n \to F$ at every point $\bar{x}$ which is a continuity point of the distribution $F(\bar{x})$;
2) $d(F_n, F) \to 0$;
3) $F_n \xrightarrow{w} F$.

This theorem demonstrates the equivalence (in our situation) of weak convergence with Levy convergence, i.e., convergence in the Levy metric. We utilize this in showing the continuity of our objective functions in the strategies as well as in showing the compactness of our strategy sets.

Theorem: The set S of distribution functions of random variables $\bar{X} = (X_1, \cdots, X_D)$ such that $0 \leq X_i \leq b_i$ is compact in the space of distribution functions on $\bar{X}$.

This theorem demonstrates the compactness of our two strategy sets, allowing us to infer that there is a worst-case jamming strategy and a best-case communicator strategy.

References

[1] R. Ahlswede, "Elimination of correlation in random codes for arbitrarily varying channels," *Zeit. Wahrscheinlichkeitstheorie*, no. 33, pp. 159–175, 1978.

[2] J. P. Aubin, *Mathematical Methods of Game and Economic Theory*. New York: North-Holland, 1982.

[3] N. M. Blachman, "Communication as a game," in *Wescon 1957 Conf. Rec.*, 1957.

[4] D. Blackwell, L. Breiman, and A. J. Thomasian, "The capacity of a class of channels," *Ann. Math. Statist.*, vol. 30, pp. 1229–1241, 1959.

[5] ____, "The capacities of certain channel classes under random coding," *Ann. Math. Statist.*, vol. 31, pp. 558–567, 1960.

[6] J. M. Borden, D. J. Mason, and R. J. McEliece, "Some information theoretic saddlepoints," *SIAM. Control. Opt.*, vol. 23, no. 1, Jan. 1985.

[7] L. F. Chang, "An information-theoretic study of ratio-threshold antijam techniques," Ph.D. dissertation, University of Illinois, Urbana-Champaign, 1985.

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

[9] I. Csiszár and P. Narayan, "The capacity of the arbitrarily varying channel revisited: Positivity, constraints," *IEEE Trans. Inform. Theory*, vol. IT-34, no. 2, pp. 181–193, Mar. 1988.

[10] R. L. Dobrushin, "Optimum information transmission through a channel with unknown parameters," *Radio Eng. Electron.*, vol. 4, no. 12, 1959.

[11] L. E. Dubins, "On extreme points of convex sets," *J. Math. Anal. Appl.*, vol. 5, pp. 237–244, 1962.

[12] T. Ericson, "The arbitrarily varying channel and the jamming problem," *Acta Electron. Sinica*, vol. 14, no. 4, pp. 21–35, July 1986.

[13] R. G. Gallager, *Information Theory and Reliable Communication*. New York: Wiley, 1968.

[14] M. V. Hegde, "Performance analysis of coded, frequency-hopped spread-spectrum systems," Ph.D. dissertation, University of Michigan, Ann Arbor, Aug. 1987.

[15] D. G. Luenberger, *Optimization by Vector Space Methods*. New York: Wiley, 1969.

[16] J. L. Massey, "Coding and modulation in digital communications," in *Proc. Int. Zurich Sem. Digital Communications*, March 1974.

[17] R. J. McEliece and W. E. Stark, "Channels with block interference," *IEEE Trans. Inform. Theory*, vol. IT-30, pp. 44–53, Jan. 1984.

[18] R. J. McEliece and E. R. Rodemich, "A study of optimal abstract jamming strategies vs. noncoherent MFSK," in *Military Commun. Conf. Rec.*, 1983, pp. 1.1.1–1.1.6.

[19] R. J. McEliece, "Communication in the presence of jamming—An information theoretic approach," in *Secure Digital Communications*. New York: Springer-Verlag, 1983, pp. 127–166.

[20] R. J. McEliece and W. E. Stark, "The optimal code rate vs. a partial band jammer," in *Milcom Rec. 1982*, 1982, pp. 45.3.1–45.3.5.

[21] W. C. Peng, "Some communication jamming games," Ph.D. dissertation, University of Southern California, Los Angeles, Jan. 1986.

[22] W. L. Root, "Communication through unspecified additive noise," *Inform. Contr.*, vol. 4, pp. 15–29, 1961.

[23] W. E. Stark, "Coding for frequency-hopped spread-spectrum channels with partial-band interference," Ph.D. dissertation, University of Illinois, Urbana-Champaign, 1982.

[24] ____, "Coding for frequency-hopped spread-spectrum communication with partial-band interference—Part 1: Capacity and cutoff rate," *IEEE Trans. Commun.*, vol. COM-33, no. 10, Oct. 1986.

[25] ____, "Coding for frequency-hopped spread-spectrum communication with partial-band interference—Part 2: Coded performance," *IEEE Trans. Commun.*, vol. COM-33, no. 10, Oct. 1986.

[26] A. J. Viterbi, "A robust ratio threshold technique to mitigate tone and partial-band jamming in coded MFSK systems," in *Proc. 1982 IEEE Military Communication Conf.*, Oct. 1982, pp. 22.4.1–22.4.5.

[27] H. S. Witsenhausen, "Some aspects of convexity useful in information theory," *IEEE Trans. Inform. Theory*, vol. IT-26, pp. 265–271, May 1980.