On Compound Channels With Side Information at the Transmitter

Patrick Mitran, *Student Member, IEEE*,
Natasha Devroye, *Student Member, IEEE*, and
Vahid Tarokh, *Senior Member, IEEE*

***Abstract*—Costa has proved that for noncausally known Gaussian interference at a power constrained transmitter communicating over an additive white Gaussian noise channel there is no capacity loss when compared to a scenario where interference is not present. For the case of a transmitter communicating over a quasistatic (i.e., nonergodic) fading channel, his method does not apply. In this correspondence, we derive upper and lower bounds on the capacity of compound channels with side information at the transmitter, first for finite alphabet channels and then, based on this result, for channels on standard alphabets (this includes real alphabets). For the special case of a degenerate compound channel with only one possible realization, our bounds are equivalent to the well-known capacity with side-information formula of Gel'fand and Pinsker. For the quasistatic fading channel, when fading is Ricean, we suggest a scheme based on our lower bound for which the performance is found to be relatively good even for moderate K-factor. As $K \rightarrow \infty$, the uncertainty on the channel vanishes and our scheme obtains the performance of dirty paper coding, namely that the interference is perfectly mitigated. As $K \rightarrow 0$, the proposed scheme treats the interferer as additional noise. These results may be of importance for the emerging field of cognitive radios where one user may be aware of another user's intended message to a common receiver, but is unaware of the channel path gains.**

***Index Terms*—Cognitive radios, compound channels, dirty paper coding, fading channels, Gel'fand–Pinsker channel, side information.**

I. Introduction

Gel'fand and Pinsker first considered the issue of communicating over a channel $P_{Y|X}$ when side information S is noncausally available at the encoder [10]. By building on the work of Gel'fand and Pinsker, Costa showed for the independent and identically distributed (i.i.d.) case the surprising result that for additive Gaussian noise channels (with noise power N and input power constraint P) with a Gaussian interferer (with power Q) that the interference could be perfectly mitigated [3]—i.e., that regardless of the power of the interferer Q, the capacity is $\frac{1}{2}\log(1+P/N)$.

This result was obtained by starting with one of the results of Gel'fand and Pinsker in [10], namely that in the discrete alphabet case, for any auxiliary random variable U distributed jointly with X and S according to $P_{U,X|S}$, the rate

$$R = I(U;Y) - I(U;S) \tag{1}$$

is achievable. Costa then generalized this result to real alphabets and showed that if the joint distribution is chosen such that $U = X + \alpha S$ with $X \sim N(0,P)$ and $\alpha = P/(P+N)$ then $R = \frac{1}{2}\log(1+P/N)$, the capacity of the channel without interference.

That the capacity is the same as if the interferer were not present has been generalized to the case where the additive noise is a stationary Gaussian process and the interferer is any power limited ergodic process [2]. Other extensions may be found for which both S and N are Gaussian but not necessarily stationary or ergodic [23], the interfering sequence is chosen arbitrarily [24] or extension to various multi-user settings with Gaussian channels [16].

There has also been much renewed interest in Gel'fand–Pinsker channels and in particular, the application of Costa's surprising dirty paper coding result. This is in part due to its application to the derivation of the capacity of Gaussian Broadcast Channels [1], [15], [18]–[20], [22]. In all these works, it is assumed that the transmitter has perfect channel state information, i.e., the fading coefficients of the wireless channel are assumed known to the transmitter and the receiver(s).

In this work, we are interested in generalizing the results of Gel'fand and Pinsker to the case where the channel is parameterized continuously by β (belonging to a compact set $\mathcal{C}$) unknown to the transmitter, but assumed known at the decoder. This model reflects fading channels where the channel state information may be estimated at the receiver, but is unknown to the transmitter. For communication over a channel without side information, such channels are often called compound channels[1] [6], [21].

While an interesting problem in itself, our motivation for considering compound channels with side information at the transmitter is due to the study of cognitive radios [8], [17]. In such a framework, a cognitive radio is aware of its environment in real-time. In particular, consider the situation where another (possibly noncognitive) user transmits to a receiver in either a frequency or time slot. The cognitive radio may sense this transmission and, if sufficiently near the transmitting node, rapidly acquire the entire message. Had there been no uncertainty on the channel, the cognitive radio could then dirty paper code against the first user's transmission by treating it as known interference. Hence, a greedy cognitive radio could communicate in the *same* time or frequency slot at no extra penalty to itself, regardless of the rate at which the first user is *attempting* to communicate.

Had both users known each other's messages then collaboration would be equivalent to a 2×1 multiple-input–single-output (MISO) system. Had neither user known the other user's message, then sharing the channel is a multiple-access problem.

In a recent work [7], the 2×2 interference channel has been generalized to the cognitive case and achievable regions derived. In the 2×2 interference channel, two users (say X_1 and X_2) wish to communicate to two receivers (say Y_1 and Y_2), with the performance measured by the rate at which independent information can be sent from X_1 to Y_1 and X_2 to Y_2. In the genie-aided cognitive setting, one further assumes that the second user has noncausal knowledge of the first user's message. In [7], achievable regions were derived by combining dirty paper-like coding techniques with those of [14]. In particular, using dirty paper-like methods, the second user can mitigate any part of the first user's message. However, in the case of compound channels (which is a more realistic model for wireless interference channels), it is not clear how side information can be employed at a transmitter if the interference (and transmitted signal) are both subject to fading.

The scenarios we are interested in is when there is uncertainty on the channel. In wireless channels, it is common that only the receiver has channel state information. Hence we study this particular kind of compound channel with side information at the transmitter.

For the traditional discrete compound channel $P^{\beta}_{Y|X}$ where β denotes the unknown parameter at the transmitter, it is well known that the capacity [21] is given by

$$C = \sup_{P_X} \inf_{\beta} I^{\beta}(X;Y) \tag{2}$$

[1]In [21], the channel is said to be compound provided that β is unknown to either the transmitter, the receiver or both. Here, we are interested in the scenario where β is known to the receiver but not the transmitter.

Manuscript received February 7, 2005; revised December 21, 2005.

The authors are with the Division of Engineering and Applied Sciences, Harvard University, Cambridge, MA 02138 USA (e-mail: mitran@deas.harvard.edu; ndevroye@deas.harvard.edu; vahid@deas.harvard.edu).

Communicated by K. Kobayashi, Associate Editor for Shannon Theory.
Digital Object Identifier 10.1109/TIT.2006.871044

where $I^{\beta}(X;Y) := I(P_X, P^{\beta}_{Y|X})$ denotes the mutual information between X and Y given the realization of the channel parameter is β.

Our main result in the finite alphabet case for compound channels with side information at the transmitter is as follows:

Theorem 1: The capacity C of the discrete memoryless compound channel $P^{\beta}_{Y|X,S}$, $\beta \in \mathcal{C}$ ($\mathcal{C}$ compact), with side information S at the transmitter is bounded by $C_{\ell} \leq C \leq C_u$ where

$$C_{\ell} = \sup_{P_{U|X,S,W}, P_{X|S,W}, P_W} \inf_{\beta \in \mathcal{C}} [I^{\beta}(U;Y|W) - I(U;S|W)] \quad (3)$$

$$C_u = \sup_{\{P^{\beta}_{U|X,S,W}\}_{\beta}, P_{X|S,W}, P_W} \inf_{\beta \in \mathcal{C}} [I^{\beta}(U;Y|W) - I(U;S|W)] \quad (4)$$

and the suprema are over all finite alphabet auxiliary random variables U and finite alphabet time-sharing random variables W and $\{P^{\beta}_{U|X,S,W}\}_{\beta}$ denotes any family of distributions (i.e., in C_U a distribution $P_{U|X,S,W}$ is chosen for each β before the infimum over β is computed). □

We remark that since the joint distribution $P_{X,S,W}$ in C_u does not depend on β, the upper bound C_u is in general tighter than if a genie had revealed β to the transmitter. We note that for the degenerate case where $\mathcal{C}$ is a singleton, both bounds reduce to the well known result of Gel'fand and Pinsker.

Bounds on the capacity of a two-sided side-information channel [4] may be derived as a special case of Theorem 1. In a two-sided side-information scenario, not only does the transmitter have a side-information sequence S_1, but the receiver is presented with a correlated sequence S_2 and the channel is governed by $P^{\beta}_{Y|X,S_1,S_2}$. By augmenting the channel so that it yields an output $Y' = (Y, S_2)$ with $P^{\beta}_{Y'|X,S_1} = P_{S_2|S_1} P^{\beta}_{Y|X,S_1,S_2}$, the problem is reformulated so that Theorem 1 applies, from which it follows that the rate

$$C'_{\ell} = \sup_{P_{U,X|S_1,W}, P_W} \inf_{\beta} [I^{\beta}(U;Y,S_2|W) - I(U;S_1|W)]$$

is achievable and likewise, an upper bound may be determined.

When the input is standard (this includes Euclidean space), we generalize Theorem 1 (Theorem 2 in Section IV) with the addition that the auxiliary random variable must now be allowed to be standard (but the time-sharing random variable remains finite alphabet) and the side information S must be quantized. Under some suitable conditions on the compound channel and the choice of auxiliary random variable, the quantized side information may be replaced with the actual side information S. This will turn out to be the case for the proposed scheme for cognitive radio channels.

We also consider the case where the input is power constrained (Theorem 3 in Section IV). Our result there is that except for a mild technicality, the bounds are essentially the same as in Theorem 2 except that the suprema must now be limited to those distributions for which the average constraint is met (i.e., those distributions for which $E|X|^2 < P$).

Returning to the cognitive radio scenario, we consider the problem of encoding a message V with knowledge of a Gaussian interfering signal S of power Q. The encoder output X is also power constrained to $P = Q$ and the signal received at the decoder is $Y = \beta_1 X + \beta_2 S + Z$ where Z is independent Gaussian noise and the compound channel parameter is $\beta := (\beta_1, \beta_2)$.

Similar to Costa's scheme, we suggest $U = X + \alpha S$, where α is now chosen as a function of the second order statistics of β_1 and β_2. The scheme proposed in Section V selects

$$\alpha = \frac{\mu_1^* \mu_2 \mathrm{SNR}}{(|\mu_1|^2 + \sigma_1^2)\,\mathrm{SNR} + 1} \quad (5)$$

where μ_i and σ_i^2 are the mean and variance of β_i. We note the following three facts about this choice for Ricean fading channels where β_1 and β_2 have K-factors K_1 and K_2, respectively, as follows.

1) If $K_1, K_2 \to \infty$, then the scheme is identical to Costa's with $\alpha = P/(P+N)$ and the interference is perfectly mitigated.
2) If either $K_1 \to 0$ or $K_2 \to 0$, the scheme treats the interferer as noise.
3) The performance does not depend on the phase difference between μ_1 and μ_2 as this choice of α rotates the mean channels so that their phases are aligned.

Furthermore, with this selection of α, it is numerically found that performance in terms of achievable rates for given outage probabilities is good over a wide range of K-factors. This conclusion is drawn by comparing to the outage capacity of an interference-free scenario and a scenario where the interference is treated as noise.

This correspondence is structured as follows. In Section II, we introduce our notation more formally. In Section III, we define the problem for the finite alphabet case and outline the proof of our first main result, Theorem 1, which states bounds on the capacity of the finite alphabet compound channel with side information. The proof itself may be found in Appendix II. In Section IV, we generalize this result to standard alphabets and prove the rest of our main results, Theorems 2 and 3. In Section V, we apply these results to derive lower bounds on the outage capacity for our cognitive radio problem. This is achieved by proposing a distribution on U and X as a function of the second order statistics of the channel. Numerical evaluation of this lower bound is compared to two other scenarios: treating the interference as noise and an interference-free scenario. Our results show that for a wide range of K-factors, performance is much better than treating the interference as noise and for low SNR such as 5 dB, performance is close to the interference-free upper bound. In Section VI, we conclude this work. The proofs of some technical lemmas may be found in Appendix I.

II. Preliminaries and Notation

In the first part of this correspondence, all variables are discrete random variables with finite alphabets. We denote the realization of a random variable X by lowercase symbols such as x and the finite alphabet by $\mathcal{X}$. Sequences denoted by bold face $\boldsymbol{x} := x_1^n := (x_1, \ldots, x_n)$ are always of length n.

We denote a probability mass function (PMF) or distribution on X by P_X and the space of PMF's on X by $\mathcal{E}_X$ with metric

$$d\left(P_X^1, P_X^2\right) = \max_x \left|P_X^1[x] - P_X^2[x]\right|. \quad (6)$$

Likewise, a conditional distribution on Y given X is denoted by $P_{Y|X}$ and the space of all conditional distributions $\mathcal{E}_{Y|X}$ is given the metric

$$d\left(P_{Y|X}^1, P_{Y|X}^2\right) = \max_{y,x} \left|P_{Y|X}^1[y|x] - P_{Y|X}^2[y|x]\right|. \quad (7)$$

In this correspondence, we will employ the notion of strongly and conditionally typical sequences as defined in [6]. We denote by $N(a|\boldsymbol{x})$ the number of occurrences of the letter $a \in \mathcal{X}$ in the sequence $\boldsymbol{x}$. Then we have the following definitions.

Definition 1: For any distribution P_X on X, a sequence $\boldsymbol{x}$ is called (P_X, δ)-typical if

$$\left|\frac{1}{n} N(a|\boldsymbol{x}) - P_X[a]\right| \leq \delta \quad (8)$$

and no $a \in \mathcal{X}$ with $P_X[a] = 0$ occurs in $\boldsymbol{x}$. The set of such sequences is denoted by $T^n(P_X, \delta)$. □

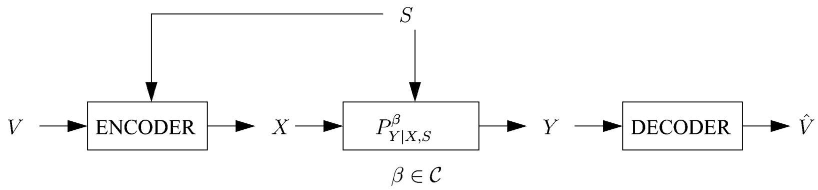

Fig. 1. Memoryless compound channel with side information at the transmitter.

Definition 2: For a stochastic matrix $P_{Y|X}$ a sequence $\boldsymbol{y}$ is $(P_{Y|X}, \delta)$-typical with respect to $\boldsymbol{x}$ if

$$\frac{1}{n}|N(a,b|\boldsymbol{x},\boldsymbol{y}) - N(a|\boldsymbol{x})P_{Y|X}[b|a]| \leq \delta \tag{9}$$

and $N(a,b|\boldsymbol{x},\boldsymbol{y}) = 0$ if $P_{Y|X}[b|a] = 0$. The set of such $\boldsymbol{y}$ sequences will be denoted by $T^n(P_{Y|X}, \delta)(\boldsymbol{x})$. □

In the Proof of Theorem 1, we will need the following two results. The first is a strengthening of the upper bound of [5, Lemma 13.6.2] such that ϵ_1 depends only on δ and both $|\mathcal{X}|$ and $|\mathcal{Y}|$, but not the underlying distributions. The second essentially follows easily from [6, Lemma 2.12]. The proofs may be found in Appendix I.

Lemma 1: Let $\boldsymbol{Y}$ be drawn i.i.d. according to P_Y, which is consistent with the joint distribution $P_{Y|X} \times P_X$. If there exists $(\boldsymbol{x}, \boldsymbol{y}) \in T^n(P_{Y|X} \times P_X, \delta)$ then

$$P[(\boldsymbol{x}, \boldsymbol{Y}) \in T^n(P_{Y|X} \times P_X, \delta)] \leq 2^{-n[I(P_X, P_{Y|X}) - \epsilon_1)]} \tag{10}$$

where $I(P_X, P_{Y|X})$ is the mutual information betwen X and Y under the joint distribution $P_{Y|X} \times P_X, \epsilon_1 \to 0$ as $\delta \to 0$ and $n \to \infty$ and ϵ_1 does not depend on the underlying distributions. □

Lemma 2: If $\boldsymbol{x} \in T^n(P_X, \delta)$ and $\boldsymbol{Y}$ is generated i.i.d. according to $P_{Y|X}$ then for any $\delta' > 0$

$$P[(\boldsymbol{x}, \boldsymbol{Y}) \notin T^n(P_{Y|X} \times P_X, \delta + \delta')] \leq \frac{K}{n(\delta')^2} \tag{11}$$

where $K = |\mathcal{X}||\mathcal{Y}|$. □

In the second part of this correspondence, we will remove the restriction on finite alphabets. If X is a random variable on a standard space [11], $(A_X, \mathcal{B}_X)$, where A_X is the alphabet and $\mathcal{B}_X$ denotes the Borel σ-field of events, then we denote a distribution on X by p_X (as opposed to P_X when we wish to emphasize that X is known to have a finite alphabet). We also define the following metric, known as the variational metric [12], between two distributions p_X^1 and p_X^2 on X,

$$d\left(p_X^1, p_X^2\right) = 2 \sup_{F \in \mathcal{B}_X} \left|p_X^1(F) - p_X^2(F)\right|. \tag{12}$$

In the case that X has a finite alphabet, it is clear that the metrics in (6) and (12) induce the same topology. Finally, we say that any family of conditional distributions $p^{\beta}_{Y|X}$ when A_X and A_Y are standard is continuously parametrized by β if $p^{\beta}_{XY} = p^{\beta}_{Y|X} \times p_X$ is continuous in β under (12) for all p_X.

III. The Discrete Case

A. Problem Formulation

Consider the discrete memoryless compound channel with side information at the transmitter as illustrated in Fig. 1. There, a sender wishes to encode a message V for transmission over a channel $P^{\beta}_{Y|X,S}$ where X denotes the input to the channel from the encoder, S denotes noncausal side information known to the encoder only and the channel is *continuously* parameterized by $\beta \in \mathcal{C}$ ($\mathcal{C}$ compact) which is unknown to the encoder but known to the receiver.

In particular, β is chosen from the set $\mathcal{C}$ and kept fixed for the duration of the frame. Prior to the encoding of the message, a sequence $\boldsymbol{s}$ with each letter generated independently according to the distribution $P_S[S_i = s_i]$ (and statistically independent of β) is revealed to the encoder. The encoder wishes to send a message $V \in \{1, \ldots, 2^{nR}\}$. Based on knowledge of $\boldsymbol{s}$ (but no knowledge of β), the encoder constructs a sequence $\boldsymbol{x}$ which is transmitted over the channel.

The receiver observes a stochastic sequence $\boldsymbol{Y}$ whose distribution is given by

$$P^{\beta}_{\boldsymbol{Y}|\boldsymbol{X},\boldsymbol{S}} = \prod_{i=1}^{n} P^{\beta}_{Y|X,S}[Y_i = y_i | X_i = x_i, S_i = s_i]. \tag{13}$$

Then, based only on this observation and knowledge of $\beta \in \mathcal{C}$, the decoder produces an estimate $\hat{V}$ of the message. As the actual channel realization $\beta \in \mathcal{C}$ is unknown at the encoder, we have the following performance measure for a code of block length n

$$\lambda_n = \sup_{\beta \in \mathcal{C}} P[\hat{V} \neq V | \beta] \tag{14}$$

where we assume that the distribution on the messages V is uniform.

We say that a rate R is achievable if there exists a sequence of codes of block length $n, n = 1, 2, 3, \ldots$ for which $\lambda_n \to 0$. We denote by C the supremum of achievable rates and call it the capacity of the discrete memoryless compound channel with side information at the transmitter.

B. General Approach

The result will be proved for time sharing random variables W with $|\mathcal{W}| = 1$. The extension to $|\mathcal{W}|$ finite is straightforward but notationally cumbersome. The proof combines the techniques of Gel'fand and Pinsker [10] as well as Wolfowitz [21]. Due to their careful interplay we present the entire proof in detail in Appendix II. Here, we present the intuition of the general approach.

As in [10], channel coding is to be performed on the fictitious auxiliary random variable U. In particular, if we view the output Y to be the result of the Markov chain $U \to (X, S) \to Y$, then the link $U \to Y$ may be viewed as a compound channel $P^{\beta}_{Y|U}$. For such a channel, 2^{nR_U} distinct codewords U with input distribution P_U can be transmitted reliably over the compound channel where $R_U = \inf_{\beta} I^{\beta}(U; Y)$.

The achieved information rate is not R_U however. The difficulty is that $\boldsymbol{s}$ is given and not the result of the chosen $\boldsymbol{u}$ codeword. In particular, it is only likely to be jointly typical with a randomly selected $\boldsymbol{u}$ codeword with probability $2^{-nI(U;S)}$. Hence, if with each message $v \in \{1, \ldots, 2^{nR}\}$ is associated a bin of $2^{n[I(U;S)+\epsilon]}$ auxiliary codewords, it is likely that one will be jointly typical with $\boldsymbol{s}$. (The set of bins is denoted $\mathcal{B}(\boldsymbol{U})$ and is called the codebook.) The cost of this approach is that the rate R achieved is then $R = R_U - I(U; S)$.

The general outline of the constructive proof of the lower bound is as follows. In the first part, we shall show that there is a sequence of codes of length n for which the maximal probability of error on a *simplified* compound channel $\mathcal{C}^{(n)}$ goes to zero. In particular, $\mathcal{C}^{(n)}$ will be a finite subset of $\mathcal{C}$ chosen to grow polynomially in n and such that under the metric in (7), for each $\beta \in \mathcal{C}$, there is a $\beta' \in \mathcal{C}^{(n)}$ such that $d(P^{\beta}_{Y|U,S}, P^{\beta'}_{Y|U,S}) \le 1/n^4$.

Hence, we will have a sequence of codes that performs well on increasingly better approximations $\mathcal{C}^{(n)}$ to $\mathcal{C}$. It will then be shown that this sequence of codes also performs well on the entire compound channel $\mathcal{C}$ in the sense that the maximal probability of error on $\mathcal{C}$ is at most greater than that on $\mathcal{C}^{(n)}$ by an amount that is only a function of n and the alphabet sizes and which goes to 0 as $n \to \infty$.

The proof of the existence of a sequence of codes which is good on $\mathcal{C}^{(n)}$ will employ random coding arguments. We will show that averaged over all randomly chosen bins $\mathcal{B}(\boldsymbol{U})$

$$\begin{aligned} E_{\mathcal{B}(\boldsymbol{U})} \max_{\beta' \in \mathcal{C}^{(n)}} P[\hat{V} \neq V | \beta'] &\le E_{\mathcal{B}(\boldsymbol{U})} P[\mathsf{E}_e] + E_{\mathcal{B}(\boldsymbol{U})|\bar{\mathsf{E}}_e} \max_{\beta' \in \mathcal{C}^{(n)}} \Big\{ P[\mathsf{E}_{d1} | \bar{\mathsf{E}}_e, \beta'] \\ &\quad + P[\mathsf{E}_{d2} | \bar{\mathsf{E}}_e, \bar{\mathsf{E}}_{d1}, \beta'] \Big\} \end{aligned} \tag{15}$$

goes to 0 as $n \to \infty$, where E_e denotes an encoding error event and together E_{d1} and E_{d2} cover all possible decoding error events and by symmetry the right side may be evaluated for a fixed $V = v$, say $v = 1$. (E_{d1} is the event that $\boldsymbol{Y}$ is not jointly typical with $\boldsymbol{u}$ and E_{d2} is the event that there is another $\boldsymbol{u}$ which is jointly typical with $\boldsymbol{Y}$.) Conditioned on $\bar{\mathsf{E}}_e$, we will see that $P[\mathsf{E}_{d1} | \bar{\mathsf{E}}_e, \beta']$ may be bounded in a fashion independent of the chosen codebook or β'. If we denote this bound by $P[\mathsf{E}_{d1} | \bar{\mathsf{E}}_e]$, then it will suffice to demonstrate that

$$\begin{aligned} E_{\mathcal{B}(\boldsymbol{U})} P[\mathsf{E}_e] &+ P[\mathsf{E}_{d1} | \bar{\mathsf{E}}_e] \\ &+ \sum_{\beta' \in \mathcal{C}^{(n)}} E_{\mathcal{B}(\boldsymbol{U})|\bar{\mathsf{E}}_e} P[\mathsf{E}_{d2} | \bar{\mathsf{E}}_e, \bar{\mathsf{E}}_{d1}, \beta'] \to 0 \end{aligned} \tag{16}$$

as $n \to \infty$ to prove the existence of a sequence of good codes on $\mathcal{C}^{(n)}$. The details may be found in Appendix II.

IV. Continuous Alphabets

In this section, we generalize the discrete memoryless compound channel with side information to the case where the input and output alphabets are not necessarily drawn from a finite alphabet. In particular, in this section, we assume that all variables X, Y, S, and U are standard [11]. This includes most cases of practical relevance such as Euclidean space $\mathbb{R}^n$ (and all finite alphabet variables as well).

The usual approach to generalize to continuous alphabet channels results which are proved for the finite alphabet case centers around invoking quantization arguments [3], [9]. The traditional argument for a noncompound continuous-alphabet channel $p_{Y|X}$ is essentially summarized as follows. Consider any input distribution p_X for the continuous channel. By choosing a sufficiently fine but finite alphabet approximation to p_X and sufficiently finely quantizing the continuous output Y, the mutual information between the quantized inputs and outputs may be made arbitrarily close to $I(X;Y)|_{p(X)}$.

We shall take the same approach. However, for the case of compound channels, such arguments require careful justification on two points. First, one must choose a sequence of successively finer quantizers at the input and at the output in such a way that the mutual information between the quantized inputs and quantized outputs approaches that of the desired continuous inputs and outputs for all possible channel realizations $p^{\beta}_{Y|X}$, i.e., we may not choose the quantizer as a function of the channel. For the case of noncompound channels, this is not an issue as there is only one possible channel realization. Second, as capacity formulas for compound channels typically involve minimizing over the possible realization β *before* maximizing over the input distribution, one must not only show that convergence is achieved for all β, but that convergence is uniform in β.

The first issue will be resolved by restricting the alphabets to be standard and will be elaborated on in Section IV-A. The second issue will be dealt with by invoking a result from analysis. This result is essentially a variation of Dini's theorem, which provides sufficient conditions for a sequence of functions on a compact space to converge uniformly when it is known to converge pointwise. This technical lemma is also discussed in Section IV-A.

Our main result in this section is stated as follows.

Theorem 2: Let $p^{\beta}_{Y|X,S}$ be a compound channel continuously parameterized in $\beta \in \mathcal{C}$ ($\mathcal{C}$ compact) where the input and output alphabets are standard. Then, the capacity of the compound channel with side information at the transmitter, C, is bounded by $C_\ell \le C \le C_u$ where

$$\begin{aligned} C_\ell = \sup_{p_{U|X,q(S),W}, p_{X|q(S),W}, P_W} \inf_{\beta \in \mathcal{C}} [I^{\beta}(U;Y|W) \\ - I(U;q(S)|W)] \end{aligned} \tag{17}$$

$$\begin{aligned} C_u = \sup_{\{p^{\beta}_{U|X,S,W}\}_\beta, p_{X|S,W}, P_W} \inf_{\beta \in \mathcal{C}} [I^{\beta}(U;Y|W) \\ - I(U;S|W)] \end{aligned} \tag{18}$$

and the suprema are over all distributions on standard alphabets U, all distributions on finite alphabet random variables W and all quantizers $q(\cdot)$ for S. □

For some cases of interest, if the compound channel and the auxiliary random variable are sufficiently well-behaved, then $q(S)$ may be replaced with S in the lower bound C_ℓ. This will be the case for the proposed distribution in Section V (where the desired joint distribution on U, X, S and Y is then Gaussian).

Examples of channels to which this theorem applies includes multiple-input–multiple-output (MIMO) channels with additive white Gaussian noise (AWGN) when the channel gain matrix H is continuously parameterized by β. For finite alphabet channels, Theorem 2 is seen to also apply.

For continuous input channels, it is often desired to impose constraints on the input to the channel, such as peak or average power constraints. For example, if the input X is peak constrained to a maximal amplitude of M, then this is equivalent to restricting $\mathcal{X} = [-M, M]$ for channels with real inputs and Theorem 2 directly applies.

For average power constrained channels (such that the input codeword must have power less than M say), it is sufficient to limit the supremum for C_ℓ in Theorem 2 to those distributions for which $EX^2 < M$ and EX^4 exists. This is because for any $p_{U,X|S,W}, P_W$ for which this is true, the input codeword will satisfy the required power constraint with high probability (provided the frame length n is large enough) by the weak law of large numbers.[2] It is equally true that in the upperbound of Theorem 1, any induced distribution by the encoder must satisfy the power constraint. Hence, we may add the constraint that $EX^2 < K$ to the upperbound C_u as well.

Theorem 3: Consider the compound channel described in Theorem 2. Furthermore, assume there is a power constraint $\frac{1}{n}\sum_i X_i^2 < M$ on the input to the channel. Then the constrained capacity is bounded as in Theorem 2 with the additional conditions that all distributions considered must satisfy $EX^2 < M$ and, for the lowerbound only, EX^4 is finite. □

[2]That the weak law of large numbers applies to X^2 follows from the fact that the fourth moment of X exists.

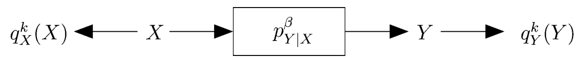

Fig. 2. Quantization of input and output variables of a channel.

A. Some Supporting Results

In this section, we briefly review some basic facts about standard spaces and state the analytical lemma. For a definition of standard spaces and a survey of important results of note, the reader is referred to the books by Gray [11], [12].

We define a quantizer q_X on a measurable space $(A_X, \mathcal{B}_X)$ (recall that A_X denotes the alphabet and $\mathcal{B}_X$ the Borel set) to be a measurable mapping $q_X : A_X \to D_X$ where $|D_X|$ is finite. We say that a quantizer q'_X is finer than a quantizer q_X if the atoms $Q_X := \{q_X^{-1}(a); a \in A_X\}$ are unions of atoms of Q'_X.

The following is a slightly stronger restatement in [12, Lemma 5.5.5] which follows from the discussion following [12, Lemma 5.5.1].

Lemma 3: Suppose that X and Y are random variables with standard alphabets defined on a common probability space. Suppose that $q_X^k, k = 1, 2, \ldots$ is a sequence of progressively finer quantizers such that the partitions asymptotically generate $\mathcal{B}_X$ and define a sequence of quantizers q_Y^k similarly. Then, for any distribution p_{XY},

$$I(X;Y) = \lim_{k\to\infty} I\left(q_X^k(X); q_Y^k(Y)\right). \tag{19}$$

That is, the same sequence of quantizers works for all distributions p_{XY}. □

Fig. 2 illustrates the idea here. The input to the channel is always X and the immediate output Y, no matter the quantizer employed. The mutual information terms on the right of (19) are between the quantized value of input X and the quantized value of output Y. An alternate interpretation is possible. We note that the arrow $q(X) \leftarrow X$ may be inverted. Then the interpretation is that the quantized random variable $q(X)$ is generated first. X is obtained from $q(X)$ by dithering the atoms Q_X. For example, if X is a uniform random variable in $[0,1]$, and the quantizer evaluates the most significant bit of the binary representation of X, the alternate interpretation is that the most significant bit is selected first and then the remaining bits are "added." In general, the distribution of the dithering depends on the realization of $q^k(X)$.

We close this subsection with an analytical result whose proof is very similar to that of Dini's Theorem and may be found in Appendix I.

Lemma 4: Let X be a compact space and consider a sequence of continuous functions $f_k(x) : X \to \mathbb{R}$. Suppose further that $f_k(x)$ converges pointwise and nondecreasingly to a (possibly noncontinuous) function $f(x)$. Furthermore, if $g(x)$ is a continuous function such that $f(x) \geq g(x)$, then for every $\epsilon > 0$, there exists an $N(\epsilon)$ such that when $k > N(\epsilon)$, $f_k(x) > g(x) - \epsilon$ uniformly for all x. □

B. Proof of Theorem 2

In this section we shall present the Proof of Theorem 2. First, a key supporting lemma is shown.

Lemma 5: Consider any family of distributions $p^\beta_{Y|X}$ continuously parametrized by β taking values in a compact set $\mathcal{C}$. Then, for each p_X, there exists a sequence of successively finer quantizers q_X^k, q_Y^k such that for each $\epsilon > 0$, there is an $N(\epsilon)$ such that for every $k > N(\epsilon)$

$$I^\beta\left(q_X^k(X); q_Y^k(Y)\right) \geq \inf_\beta I^\beta(X;Y) - \epsilon. \tag{20}$$

□

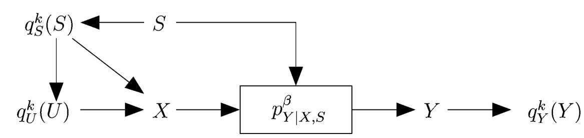

Fig. 3. Quantization of input and output variables of the compound channel with side information.

Proof: By the continuity of $p^\beta_{Y|X}$ in β it follows that $I^\beta(q_X^k(X); q_Y^k(Y))$ is also continuous in β. By Lemma 3, this sequence of functions converges pointwise to the function $I^\beta(X;Y) \geq I^*$ where I^* is the constant $\inf_\beta I^\beta(X;Y)$. Since the quantizers are successively finer, we have that

$$I^\beta(q_X^{k+1}(X); q_Y^{k+1}(Y)) \geq I^\beta(q_X^k(X); q_Y^k(Y)).$$

Hence, by Lemma 4, it follows that such an $N(\epsilon)$ can be found. ■

Proof of Theorem 2: We first note that the proof of the upper bound of Theorem 1 requires no particular assumption about the nature of the random variables X, S, U and Y except for (58), where if the only assumptions on Y and S are that they are standard, then U must be assumed standard.

As for the direct part, again the result will be proved for the case that $|\mathcal{W}| = 1$. For any finite $|\mathcal{W}|$, the generalization is straightforward. Consider a given $p_{U,X|q_S(S)}$. We will show that by employing sufficiently fine quantizers for all random variables but S, for any $\epsilon > 0$, the rate

$$R := \inf_{\beta\in\mathcal{C}}[I^\beta(U;Y) - I(U; q_S(S))] - 2\epsilon$$

may be achieved.

Let

$$p^\beta_{U,X,q_S(S),S,Y} = p_{U,X|q_S(S)} \times p_{q_S(S)|S} \times p_S \times p^\beta_{Y|X,S}$$

and $p^\beta_{q_U^k(U), q_X^k(X), q_S(S), S, q_Y^k(Y)}$ denote the distribution induced on the quantized $q_U^k(U), q_X^k(X), q_Y^k(Y)$ as well as S and $q_S(S)$ by $p^\beta_{U,X,q_S(S),S,Y}$.

The general outline is illustrated in Fig. 3. The encoder is given the nonquantized S. It first proceeds to quantize it to obtain $q_S(S)$. From this, encoding is performed on the fictitious finite alphabet auxiliary random variable $q_U^k(U)$ as in Theorem 1. The encoder then generates the finite-alphabet input $q_X^k(X)$ from $q_S(S)$ and $q_U^k(U)$. The actual input X to the channel is generated stochastically according to $p_{X|q_X^k(X), q_U^k(U), q_S(S)}$. The effect of the latter is to dither the atoms of q_X^k such that the distribution $p_{X, q_S(S), q_U^k(U)}$ induces the original quantized $P_{q_X^k(X), q_S(S), q_U^k(U)}$. The output of the channel is Y which is quantized to $q_Y^k(Y)$ and decoding proceeds as in Theorem 1. Note that due to the dithering of the atoms of $q_X^k(X)$, $p^\beta_{q_U^k(U), q_X^k(X), q_S(S), S, q_Y^k(Y)}$ is also the joint distribution on $q_U^k(U), q_X^k(X), q_S(S), S$, and $q_Y^k(Y)$ induced by the finite-alphabet encoder.

For each k, the rate achieved by the encoder and decoder operating on the quantized alphabets is then

$$R_k = \min_{\beta\in\mathcal{C}}\left[I^\beta\left(q_U^k(U); q_Y^k(Y)\right) - I\left(q_U^k(U); q_S(S)\right)\right]. \tag{21}$$

Now, by Lemma 5, one can find an $N(\epsilon)$ such that for all $k > N_1(\epsilon)$

$$I^\beta(q_U^k(U); q_Y^k(Y)) > \inf_\beta I^\beta(U;Y) - \epsilon.$$

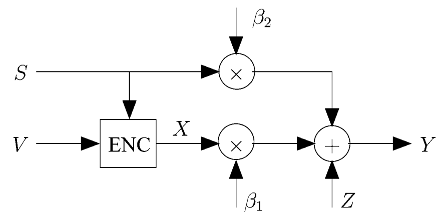

Fig. 4. Communications over a fading channel with a fading interferer whose signal, but not fading coefficient, is known at the transmitter.

Also, by Lemma 3, one can find an N_2 such that for $k > N_2$

$$|I(q_U^k(U); q_S(S)) - I(U; q_S(S))| < \epsilon.$$

Then $R_k > R - 2\epsilon$. ■

We remark that the reason that finely quantizing S does not allow us to replace $q_S(S)$ with S is that while S is a dithering of the atoms of $q_S(S)$, it is a dithering that is not correlated with $q_U(U)$ nor $q_X(X)$. However, the amount of dithering becomes rather small for fine quantizers and if the compound channel and desired $p_{U|X,S}$ are well behaved, this effect may become negligible, in which case $q(S)$ may be replaced with S. A case of this for a particular choice of $p_{U,X|S}$ will be observed in the next section.

V. Application to Fading Channels

In this section, we are interested in studying achievable outage probabilities for communicating over a fading channel with a fading interferer whose signal, but not fading coefficient, is known at the transmitter. This scenario is motivated by cognitive radio channels [7], [17]. In the simplest cognitive framework, two nearby users share a common channel to a single receiver. As the users are nearby, it is reasonable to assume that one user, say the second, may quickly acquire the message of the first when the first transmits to the receiver. If all the channel parameters are known, then a greedy second user may communicate simultaneously using a dirty paper coding scheme with no loss to capacity for itself, regardless of the rate at which the first user is attempting to communicate (this may degrade the performance of the first user). In the case that the channel parameters are unknown at the transmitter, the channel is compound and a traditional dirty paper coding scheme cannot be employed.

The model we consider is illustrated in Fig. 4. There, a message V is encoded with knowledge of a Gaussian interfering signal S of power Q. The encoder output X is power constrained to P. The received signal at the decoder is $Y = \beta_1 X + \beta_2 S + Z$ where the compound parameter at the encoder is $\beta := (\beta_1, \beta_2)$ and the noise power of Z is N.

We shall now assume a probability distribution on the parameters β. In particular, we assume that β_1 and β_2 are independent Ricean distributed with parameters K_1 and K_2, i.e., $\beta_1 = \mu_1 + Z_1$ where Z_1 is a zero mean circularly symmetric complex Gaussian random variable with variance σ_1^2 such that $|\mu_1|^2 + \sigma_1^2 = 1$ and $|\mu_1|^2/\sigma_1^2 = K_1$ and likewise for β_2.

For a given $p_{U,X|q(S),W}$ and P_W, we say that a rate R has outage probability P_{out} if $P_{\text{out}} = P[R > I^\beta(U;Y|W) - I(U;q(S)|W)]$. This definition is justified since for all β such that $|\beta| < L$ for some constant L and $R \leq I^\beta(U;Y|W) - I(U;q(S)|W)$, Theorem 3 assures us that a rate R is achievable. If L is taken sufficiently large, then the probability of observing a realization $|\beta| \leq L$ can be made arbitrarily small and our definition takes the operational significance that there is a code of rate R for which communication is successful on a set of β whose probability is $1 - P_{\text{out}}$.

We seek to determine a lower bound on achievable rates for a given outage probability. This will be accomplished by suggesting a distribution $p_{U,X|S}$ and numerically evaluating (R, P_{out}) pairs. (For the rest of this section, we assume that $W = \phi$). Similar to Costa's dirty paper coding scheme, the desired joint distribution on U, X and S is specified by $U = X + \alpha S$, where the parameter α should depend only on the second order statistics μ_1, μ_2, σ_1 and σ_2 while X is zero mean Gaussian with variance P.

However, as required by Theorem 2, we first consider $p_{U_k,X|q_k(S)}$ for some appropriate sequence of successively finer quantizers, i.e., we consider $U_k = X + \alpha q_k(S) + W_k$ where W_k is independent uniform additive dithering noise. Then, it can be shown that $\inf_\beta I^\beta(U_k;Y) \to \inf_\beta I^\beta(U;Y)$ on any bounded subset of β as $k \to \infty$ (see Appendix III).

We may therefore choose $U = X + \alpha S$. With such a choice of auxiliary random variable, the covariance matrix of (U, Y) is

$$\text{COV}(U,Y) = \begin{pmatrix} P + |\alpha|^2 Q & \beta_1^* P + \alpha \beta_2^* Q \\ \beta_1 P + \alpha^* \beta_2 Q & |\beta_1|^2 P + |\beta_2|^2 Q + N \end{pmatrix}. \quad (22)$$

Hence,

$$\begin{aligned} I(U;Y) = &\log((|\beta_1|^2 P + |\beta_2|^2 Q + N)(P + |\alpha|^2 Q)) \\ &- \log((|\beta_1|^2 P + |\beta_2|^2 Q + N)(P + |\alpha|^2 Q) \\ &- |\beta_1 P + \alpha^* \beta_2 Q|^2) \quad (23) \\ I(U;S) = &\log((P + |\alpha|^2 Q)/P). \quad (24) \end{aligned}$$

Then, if we define $I(\beta_1, \beta_2) := I(U;Y) - I(U;S)$ and make the simplifications $P = Q$ and $\text{SNR} = P/N = Q/N$ (if $P < Q$ say, then the extra interference power could be absorbed into the μ_2 and σ_2 statistics), we obtain

$$\begin{aligned} I(\beta_1, \beta_2) = &\log((|\beta_1|^2 + |\beta_2|^2)\text{SNR} + 1) \\ &- \log(\text{SNR}|\beta_2 - \alpha\beta_1|^2 + 1 + |\alpha|^2). \quad (25) \end{aligned}$$

It is not clear what choice of α will minimize outage probability for a given rate, but a reasonable choice of α can be obtained by selecting it to minimize the expected value of $\text{SNR}|\beta_2 - \alpha\beta_1|^2 + 1 + |\alpha|^2$. If β_1 and β_2 have small variance, then this choice will approximately maximize $EI(\beta_1, \beta_2)$. Carrying out the expectation, α is found to be

$$\alpha = \frac{\mu_1^* \mu_2 \text{SNR}}{(|\mu_1|^2 + \sigma_1^2)\,\text{SNR} + 1} \quad (26)$$

which is the desired choice of α in terms of the second-order statistics of the fading coefficients. We note the following three facts about this choice for Ricean-fading channels:

1) if $K_1, K_2 \to \infty$, then the scheme is identical to Costa's with $\alpha = P/(P+N)$ and the interference is perfectly mitigated;
2) if either $K_1 \to 0$ or $K_2 \to 0$, the scheme treats the interferer as noise;
3) the performance does not depend on the phase difference between μ_1 and μ_2 as this choice of α rotates the mean channels so that their phases are aligned.

Fig. 5 shows the achievable rates for given outage probabilities of $10^{-1}, 10^{-2}$ and 10^{-3} at $\text{SNR} = 5$ dB over a wide range of K-factors when both β_1 and β_2 have the same K-factor. This is reasonable if we are to assume that the interferer and user are nearby, as would be the case in cognitive radio channels. Also, for comparison, we illustrate the performance of a scenario when there is no interference (which provides an upper bound on the outage capacity) and when the interference is treated as noise.

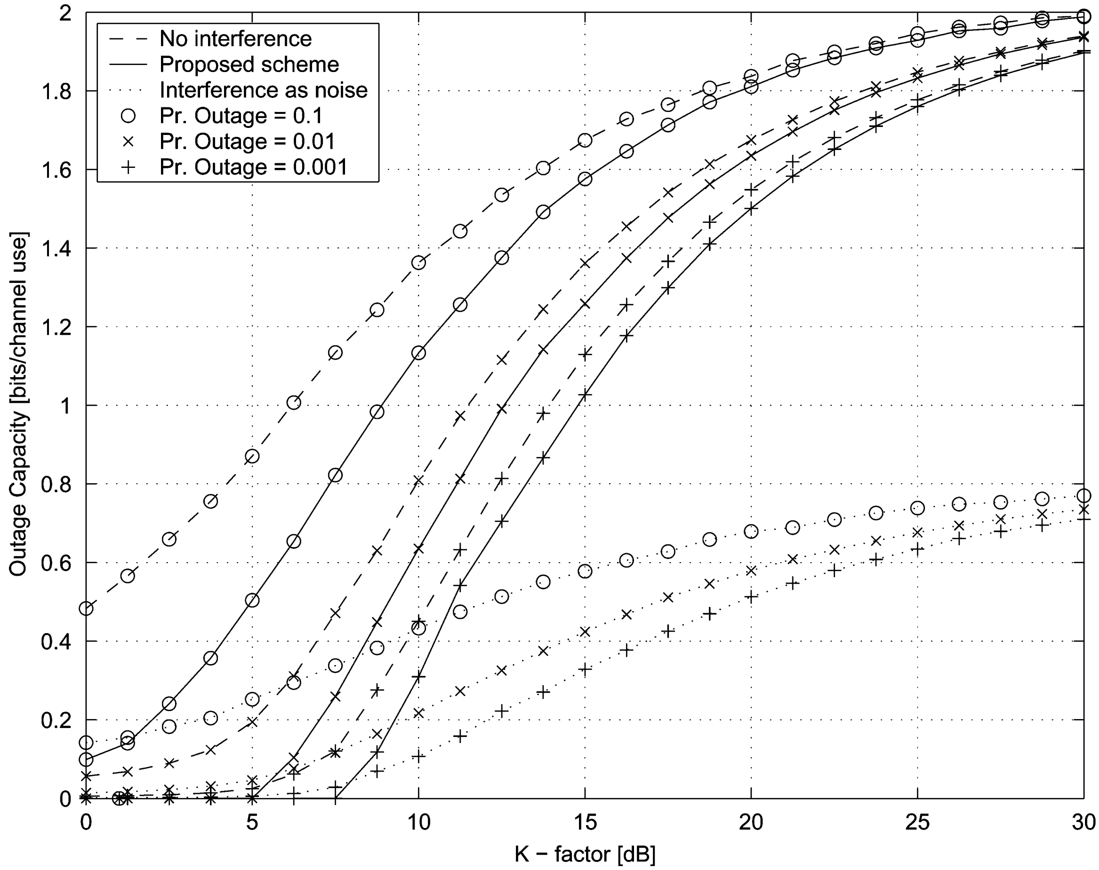

Fig. 5. Communications over a fading channel with a fading interferer whose signal, but not fading coefficient, is known at the transmitter for SNR $= 5$ dB with $P = Q = 1$.

From the figure, it is seen that for K-factors of 10 dB or greater, significant gains are achieved over treating the interference as noise. (It has been found experimentally that in the New Jersey area, 70% of wireless channels have a K-factor of at least 5 dB and 55% have K factor of 10 dB or more [13]). Furthermore, for very high reliability such as an outage probability of 10^{-2} or 10^{-3}, the achievable rate is very near the interference-free upper bound. In all cases, for large K-factor, the scheme achieves the performance of the upperbound (asymptotically). This is not surprising as in the large K regime, there is little or no uncertainty on the channel and the proposed scheme mimics Costa's dirty paper coding, i.e., the interference is perfectly mitigated.

Figs. 6 and 7 show similar plots for 10 and 20 dB SNR, respectively. Again, over a wide range of K-factors, the scheme outperforms treating the interferer as noise by a wide margin and for high K-factors, completely mitigates the interference. In the low K regime, the scheme is not as near the upper bound as it was for SNR $= 5$ dB. However, as our scheme is only a lower bound on performance, the exact outage capacity for this problem remains open.

VI. Conclusion

We have derived bounds on the capacity of compound channels with side information at the transmitter, first for the finite alphabet case and then, based on these results, for standard alphabets. Bounds for two-sided state-information problems (i.e., a side information sequence at the transmitter and another at the receiver) can also be derived as a special case of this result. While the exact capacity remains an open problem, we suspect that the lower bound is in fact tight. The apparent looseness in the upper bound is due to the fact that the empirical distribution on U that is derived in the converse is itself obtained from the distribution on Y induced by the encoder. As the latter is clearly a function of the channel realization, so is the distribution on U. Intuitively, this could be plausible—the auxiliary codebook is fictitious and one may perhaps be able to generate a different auxilliary codebook for each possible channel realization. Despite this observation, we believe that this is not the case and that the lower bound is tight.

We have also considered the problem of a user communicating over a fading channel with a fading interferer whose signal, but not fading coefficient, is known at the transmitter. This scenario is of interest for the emerging field of cognitive radios. By proposing an input distribution, we have derived achievable rates for given outage probabilities. We have also compared these rates to the outage capacities of two schemes: communication over an interference-free fading channel and treating the interference as noise. Our results indicate significant gains are possible over the latter while in the low SNR regime, performance is near the upper bound of the interference-free scenario.

Appendix I
Proof of Lemmas

We will need the following Lemma, which slightly extends [6, Lemma 2.10].

Lemma 6:

1) If $\boldsymbol{x} \in T^n(P_X, \delta)$ and $\boldsymbol{y} \in T^n(P_{Y|X}, \delta')(\boldsymbol{x})$ then $(\boldsymbol{x}, \boldsymbol{y}) \in T^n(P_{Y|X} \times P_X, \delta + \delta')$.
2) If $(\boldsymbol{x}, \boldsymbol{y}) \in T^n(P_{Y|X} \times P_X, \delta)$ then $\boldsymbol{x} \in T^n(P_X, \delta|\mathcal{Y}|)$ and $\boldsymbol{y} \in T^n(P_{Y|X}, 2\delta|\mathcal{Y}|)(\boldsymbol{x})$. □

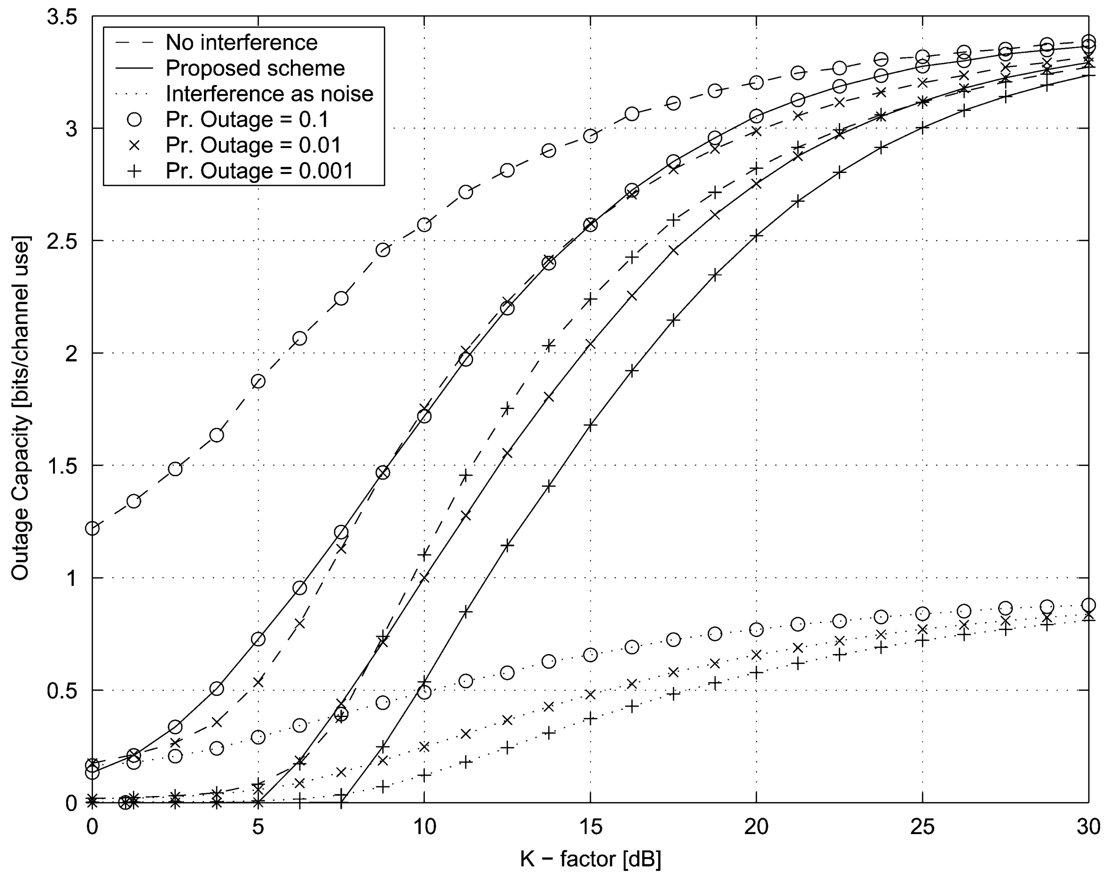

Fig. 6. Communications over a fading channel with a fading interferer whose signal, but not fading coefficient, is known at the transmitter for SNR $= 10$ dB with $P = Q = 1$.

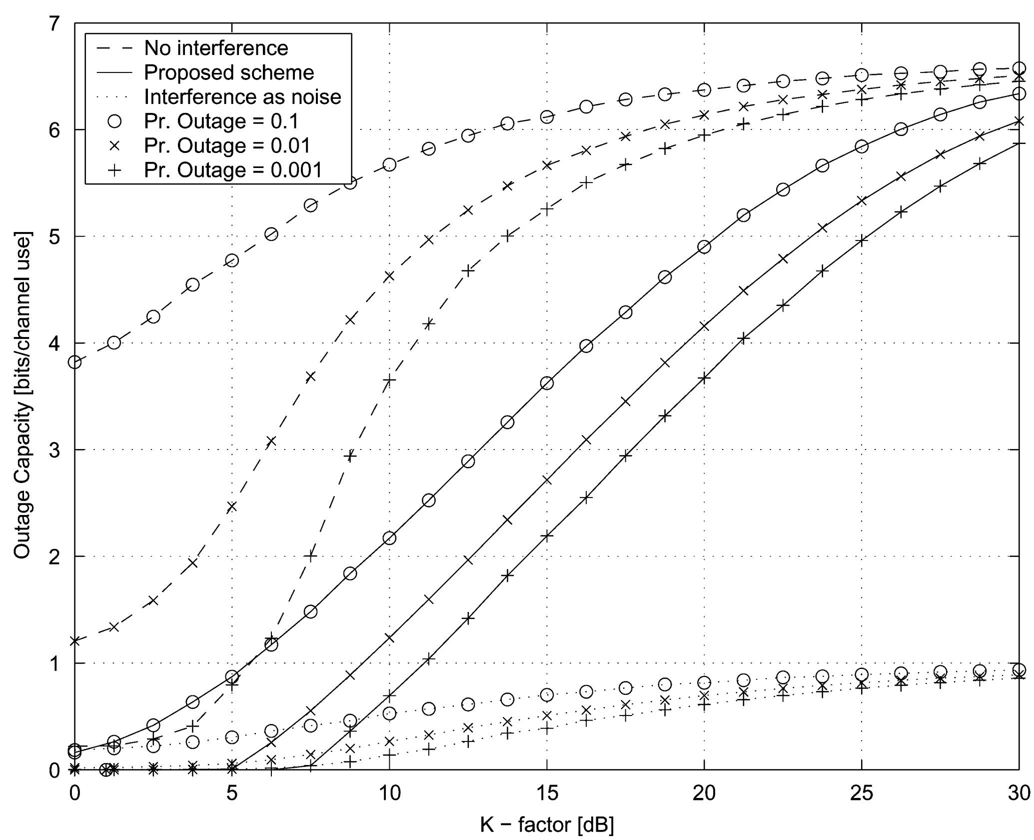

Fig. 7. Communications over a fading channel with a fading interferer whose signal, but not fading coefficient, is known at the transmitter for SNR $= 20$ dB with $P = Q = 1$.

Proof: Part 1) is proved in [6] while part 2) follows from Definitions 1 and 2 and the triangle inequality. ■

Proof of Lemma 1: The left of (10) may be upper-bounded by $\sum P_Y[\boldsymbol{y}]$ where the sum is over all $\boldsymbol{y}$ sequences such that $(\boldsymbol{x}, \boldsymbol{y}) \in T^n(P_{Y|X} \times P_X, \delta)$. If Q_Y denotes the type (empirical distribution) of $\boldsymbol{y}$, then $P_Y[\boldsymbol{y}] \le 2^{-nH(Q_Y)}$ (by [6, eq. (2.5)]).

Since the $\boldsymbol{y}$ sequences we are summing over satisfy $\boldsymbol{y} \in T^n(P_Y, \delta|\mathcal{X}|)$, we have by the uniform continuity of the entropy function $H(\cdot)$ that $H(Q_Y) \ge H(P_Y) - \epsilon_2$ and $\epsilon_2 \to 0$ as $\delta \to 0$ and does not depend on P_Y. We may thus bound the left side of (10) by

$$\le \sum 2^{-n[H(P_Y)-\epsilon_2]} \tag{27}$$

where the sum is over all $\boldsymbol{y}$ sequences such that $(\boldsymbol{x}, \boldsymbol{y}) \in T^n(P_{Y|X} \times P_X, \delta)$.

By Lemma 6, we may further upperbound the desired expression by extending the sum to all $\boldsymbol{y} \in T^n(P_{Y|X}, 2\delta|\mathcal{Y}|)(\boldsymbol{x})$. Since $\boldsymbol{x} \in T^n(P_X, \delta|\mathcal{Y}|)$, by [6, Lemma 2.13], there are at most $2^{n[H(P_{Y|X}|P_X)+\epsilon_3]}$ such $\boldsymbol{y}$ sequences where ϵ_3 does not depend on $P_{Y|X}$ or P_X and $\epsilon_3 \to 0$ as $\delta \to 0$ and $n \to \infty$. The result then follows since $I(P_X, P_{Y|X}) = H(P_Y) - H(P_{Y|X}|P_X)$. ■

Proof of Lemma 2: By Lemma 6,

$$\begin{aligned} &P[(\boldsymbol{x}, \boldsymbol{Y}) \notin T^n(P_{Y|X} \times P_X, \delta + \delta')] \\ &\quad \le P[\boldsymbol{Y} \notin T^n(P_{Y|X}, \delta')(\boldsymbol{x})]. \end{aligned} \tag{28}$$

The result then follows directly from [6, Lemma 2.12] where it is shown that the right side of (28) is bounded by $|\mathcal{X}||\mathcal{Y}|/n(\delta')^2$. ■

Proof of Lemma 4: For each x, one can find an $N(x)$ such that for all $k > N(x), f_k(x) > g(x) - \epsilon$. By continuity of $f_k(x)$ and $g(x)$, it is also true that for some open neighborhood of x, denoted $B(x)$, for each $y \in B(x), f_k(y) > g(y)$ whenever $k > N(x)$. Since X is compact, the open cover $B(x)$ can be reduced to a finite cover $B(x^{(1)}), \ldots, B(x^{(M)})$. The theorem is satisfied if we take $N(\epsilon) = \max_{1 \le i \le M} N(x^{(i)})$. ■

Appendix II
Proof of Theorem 1

A. Lower Bound C_ℓ

For any $P_{U,X|S} \in \mathcal{E}_{U,X|S}, \epsilon > 0$, we show that the rate $R = \inf_{\beta \in \mathcal{B}}[I^\beta(U;Y) - I(U;S)] - 2\epsilon$ is achievable.

Codebook Generation: From $P_{U,X|S}$ and P_S, the encoder computes the marginalization P_U and proceeds to generate $2^{n(I^*(U;Y)-\epsilon)}$ codewords $\boldsymbol{u}$, where $I^*(U;Y) = \inf_{\beta \in \mathcal{C}} I^\beta(U;Y)$ and the ith letter of $\boldsymbol{u}$ is distributed according to $P_U[U = u_i]$.

These codewords are then evenly distributed into $2^{n(R-2\epsilon)}$ bins in a uniformly random manner, i.e., there are $2^{n(I(U;S)+\epsilon)}$ codewords in each bin. The jth codeword in the vth bin is denoted $\boldsymbol{u}(v, j)$.

Codeword Selection: The encoder wishes to transmit message $V \in \{1, \ldots, 2^{n(R-\epsilon)}\}$. To that end, it looks in the Vth bin for a codeword $\boldsymbol{u}$ that is $(P_{U|S} \times P_S, \delta)$-jointly typical with $\boldsymbol{s}$.

From [5, Lemma 13.6.2], the probability of any generated $\boldsymbol{u}$ sequence being thus jointly typical is at least $2^{-n(I(U;S)-\epsilon_1)}$ with $\epsilon_1 \to 0$ as $\delta \to 0$ and $n \to \infty$. The probability that there is no jointly typical $\boldsymbol{u}$ sequence in the Vth bin is then well known to be bounded by

$$E_{\mathcal{B}(\boldsymbol{U})}P[\mathsf{E}_e] \le \exp(-2^{n(I(U;S)+\epsilon)}2^{-n(I(U;S)-\epsilon_1)}) \tag{29}$$

which decays doubly exponentially provided $\epsilon_1 < \epsilon$ which itself can be assured by taking δ sufficiently small and n sufficiently large.

The encoder then creates an $\boldsymbol{x}$ sequence stochastically by choosing each letter x_i according to $P_{X|U,S}[X = x_i|U = u_i, S = s_i]$.

Decoding on $\mathcal{C}^{(n)}$: With the choice of $P_{U,X|S}$, we may compute $P_{X,S|U}$ through Baye's rule. We note that by the selection of a $\boldsymbol{u}$ as a jointly typical sequence with $\boldsymbol{s}$, the output $\boldsymbol{y}$ appears to be the result of the Markov chain $U \to (X,S) \to Y$ with $P^\beta_{Y\,|\,U,S} = \sum_x P^\beta_{Y|X,S}[Y = y|X = x, S = s]P[X = x, S = s|U = u]$. We denote the set of such $P^\beta_{Y\,|\,U,S}$ as the causal equivalent compound channel.

The remainder of the proof is structured as follows. First, we show that for a sequence of appropriately chosen subsets $\mathcal{C}^{(n)}$ of $\mathcal{C}$, a sequence of codes of length n whose average probability of error goes to zero on the subsets $\mathcal{C}^{(n)}$ may be found. In the second part, we show that this code is in fact good on the entire compound channel $\mathcal{C}$.

First, consider subsets $\mathcal{C}^{(n)} \subset \mathcal{C}$ such that for each $\beta \in \mathcal{C}$, there is a $\beta' \in \mathcal{C}^{(n)}$ with

$$d\left(P^\beta_{Y\,|\,U,S}, P^{\beta'}_{Y\,|\,U,S}\right) \le 1/n^4 \tag{30}$$

where $d(\cdot,\cdot)$ is the metric defined in (7). Since the space of all possible distributions $\mathcal{E}_{Y\,|\,U,S}$ is bounded, the cardinality of $\mathcal{C}^{(n)}$ is at most $O(n^{4|\mathcal{U}||\mathcal{Y}||\mathcal{S}|})$, i.e., polynomial.

Now, define the following decoding sets D,

$$\begin{aligned} D_{\beta,v,j} = \{\boldsymbol{y} : (\boldsymbol{y}, \boldsymbol{u}(v,j)) \in T^n(P^\beta_{Y,U}, \delta_2), \nexists (v' \ne v, j') \text{ s.t.,} \\ (\boldsymbol{y}, \boldsymbol{u}(v', j')) \in T^n(P^\beta_{Y,U}, \delta_2)\} \end{aligned} \tag{31}$$

where $\delta_2 = |\mathcal{S}|(\delta + \delta')$ for some $\delta' > 0$.

The decoder receives $\boldsymbol{y}$ and knows $\beta \in \mathcal{C}^{(n)}$. The decoder searches through the sequence of sets $D_{\beta,v,j}$ until it finds a $\hat{v}$ and $\hat{j}$ such that $\boldsymbol{y} \in D_{\beta,\hat{v},\hat{j}}$. These sets were chosen to be disjoint (over v). Hence, while $\boldsymbol{y}$ may belong to more than one D-set, they all have the same $\hat{v}$. The decoder then declares the bin $\hat{v}$ to be its estimate of the encoded message V.

With this choice of decoding sets, for each $\beta \in \mathcal{C}^{(n)}$, we mimic the random coding argument. In particular, the probability of a decoding error conditioned on successful encoding, $P[\mathsf{E}_d|\bar{\mathsf{E}}_e, \beta]$, is bounded by

$$P[\mathsf{E}_d|\bar{\mathsf{E}}_e, \beta] \le P[\mathsf{E}_{d1}|\bar{\mathsf{E}}_e, \beta] + P[\mathsf{E}_{d2}|\bar{\mathsf{E}}_{d1}, \bar{\mathsf{E}}_e, \beta] \tag{32}$$

where E_{d1} denotes the event that $\boldsymbol{Y}$ is not $(P^\beta_{Y|U} \times P_U, \delta_2)$-jointly typical with $\boldsymbol{u}(v,j)$ and E_{d2} denotes the event that there is a $(v' \ne v, j')$ such that $(\boldsymbol{u}(v', j'), \boldsymbol{Y})$ is $(P^\beta_{U,Y}, \delta_2)$-typical.

Now, since encoding was successful, $(\boldsymbol{u}(v,j), \boldsymbol{s})$ are $(P_{U,S}, \delta)$-typical. Since by Lemma 2, for any jointly typical $(\boldsymbol{u}, \boldsymbol{s})$, the probability that $(\boldsymbol{u}, \boldsymbol{s}, \boldsymbol{y})$ is not $(P^\beta_{U,S,Y}, \delta + \delta')$-typical is less than $|\mathcal{U}||\mathcal{S}||\mathcal{Y}|/n(\delta')^2$, the first term of the right of (32) is bounded by

$$\begin{aligned} P[\mathsf{E}_{d1}|\bar{\mathsf{E}}_e, \beta] &\le |\mathcal{U}||\mathcal{S}||\mathcal{Y}|/n(\delta')^2 && (33) \\ &:= P[\mathsf{E}_{d1}|\bar{\mathsf{E}}_e] && (34) \end{aligned}$$

where $P[\mathsf{E}_{d1}|\bar{\mathsf{E}}_e]$ is independent of β and the codebook conditioned on $\bar{\mathsf{E}}_e$. Furthermore, $P[\mathsf{E}_{d1}|\bar{\mathsf{E}}_e] \to 0$ when $\delta'(n) = n^{1/4}$ and $n \to \infty$.

It therefore remains to show that

$$\sum_{\beta \in \mathcal{C}^{(n)}} E_{\mathcal{B}(\boldsymbol{U})|\bar{\mathsf{E}}_e} P[\mathsf{E}_{d2}|\bar{\mathsf{E}}_e, \bar{\mathsf{E}}_{d1}, \beta] \to 0. \tag{35}$$

Since $(\boldsymbol{u}(v,j), \boldsymbol{y})$ are $(P_{U,Y}, \delta_2)$-typical, by Lemma 1 the probability of E_{d2} averaged over all randomly chosen codewords in bins other than bin v is no more than

$$\begin{aligned} &E_{\mathcal{B}(\boldsymbol{U})|\bar{\mathsf{E}}_e} P[\mathsf{E}_{d2}|\bar{\mathsf{E}}_e, \bar{\mathsf{E}}_{d1}, \beta] \\ &\quad \le 2^{n(I^*(U;Y)-\epsilon)}2^{-n(I^\beta(U;Y)-\epsilon_1)} && (36) \\ &\quad \le 2^{n(\epsilon_1 - \epsilon)} && (37) \end{aligned}$$

where the last inequality follows from our definition of $I^*(U;Y) = \inf_\beta I^\beta(U;Y)$. We note that since ϵ_1 does not depend on the underlying distribution, neither of these bounds depends on β. By letting $n \to \infty$ as $\delta \to 0$, then $\epsilon_1 < \epsilon$ and the bound in (37) decays exponentially, uniformly for all β. Since $|\mathcal{C}^{(n)}|$ is polynomial, the sum in (35) asymptotically vanishes.

Hence, one can find a sequence of good codebooks for which

$$P\left[\mathsf{E}|\mathcal{C}^{(n)}\right] := P[\mathsf{E}_e] + \max_{\beta' \in \mathcal{C}^{(n)}} P[\mathsf{E}_d|\bar{\mathsf{E}}_e, \beta'] \tag{38}$$

goes to zero as $n \to \infty$.

Decoding on $\mathcal{C}$: We wish to show to show that for the same sequence of codebooks,

$$P[\mathsf{E}|\mathcal{C}] := P[\mathsf{E}_e] + \sup_{\beta \in \mathcal{C}} P[\mathsf{E}_d|\bar{\mathsf{E}}_e, \beta] \tag{39}$$

also goes to zero as $n \to \infty$. It will suffice to show that

$$\sup_{\beta \in \mathcal{C}} P[\mathsf{E}_d|\bar{\mathsf{E}}_e, \beta] \to 0. \tag{40}$$

If we define $D_{\beta',v} = \cup_j D_{\beta',v,j}$, we note that the probability of correct decoding for a $\beta' \in \mathcal{C}^{(n)}$ may be expressed as

$$\begin{aligned} &P[\bar{\mathsf{E}}_d \,|\, \bar{\mathsf{E}}_e, \beta'] \\ &= \sum_{v,j,\boldsymbol{s}} P[\boldsymbol{Y} \in D_{\beta',v}|\boldsymbol{u}(v,j), \boldsymbol{s}] P[v,j,\boldsymbol{s}|\bar{\mathsf{E}}_e] \qquad (41)\\ &\geq 1 - \lambda_n^{(n)} \qquad (42)\end{aligned}$$

with $\lambda_n^{(n)} \to 0$ as $n \to \infty$ and is not a function of $\beta' \in \mathcal{C}^{(n)}$.

We will now show that if given any actual channel realization $\beta \in \mathcal{C}$, decoding is performed with the decoding sets for β', β' chosen as a function of β according to (30), then

$$\begin{aligned} &P[\boldsymbol{Y} \in D_{\beta',v}|\boldsymbol{u}(v,j), \boldsymbol{s}, \beta] \\ &\geq \frac{(n^2-1)}{n^2}\left(P[\boldsymbol{Y} \in D_{\beta',v} \,|\, \boldsymbol{u}(v,j), \boldsymbol{s}, \beta'] - \frac{|\mathcal{U}||\mathcal{S}||\mathcal{Y}|}{n}\right). \qquad (43)\end{aligned}$$

Hence, for any $\beta \in \mathcal{C}$

$$P[\bar{\mathsf{E}}_d \,|\, \bar{\mathsf{E}}_e, \beta] \geq \frac{(n^2-1)}{n^2}\left(1 - \lambda_n^{(n)} - \frac{|\mathcal{U}||\mathcal{S}||\mathcal{Y}|}{n}\right) \tag{44}$$

which implies (40). Hence, the decoder may choose $D_{\beta,v,j} = D_{\beta',v,j}$ and the code is still good.

So it remains to show (43). Now, for any $\boldsymbol{u}$ and $\boldsymbol{s}$, consider partitioning $D_{\beta',v}$ into $D^1_{\beta',v}$ and $D^2_{\beta',v}$, where

$$D^1_{\beta',v} = \left\{\boldsymbol{y} \in D_{\beta',v} : \exists i \text{s.t.} P^{\beta'}_{Y\,|\,U,S}[y_i \,|\, u_i, s_i] < 1/n^2\right\} \tag{45}$$

$$D^2_{\beta',v} = \left\{\boldsymbol{y} \in D_{\beta',v} : \forall i P^{\beta'}_{Y\,|\,U,S}[y_i \,|\, u_i, s_i] \geq 1/n^2\right\}. \tag{46}$$

Then

$$P^{\beta'}_{Y\,|\,U,S}[D^1_{\beta',k}|\boldsymbol{u}, \boldsymbol{s}] \leq n(1/n^2)|\mathcal{U}||\mathcal{Y}||\mathcal{S}| \tag{47}$$

$$P^{\beta}_{Y\,|\,U,S}[D^1_{\beta',k}|\boldsymbol{u}, \boldsymbol{s}] \leq n(2/n^2)|\mathcal{U}||\mathcal{Y}||\mathcal{S}| \tag{48}$$

where the last inequality follows from (30). Also, from (30) and (45), we have that for $\boldsymbol{y} \in D^2_{\beta',v}$

$$\frac{P^{\beta}[\boldsymbol{y} \,|\, \boldsymbol{u}, \boldsymbol{s}]}{P^{\beta'}[\boldsymbol{y} \,|\, \boldsymbol{u}, \boldsymbol{s}]} \geq \left(\frac{1/n^2 - 1/n^4}{1/n^2}\right)^n. \tag{49}$$

Combining these results

$$\begin{aligned} &P[\boldsymbol{Y} \in D_{\beta',v} \,|\, \boldsymbol{u}(v,j), \boldsymbol{s}, \beta] \\ &\geq \frac{P\left[\boldsymbol{Y} \in D^2_{\beta',v} \,|\, \boldsymbol{u}(v,j), \boldsymbol{s}, \beta\right]}{P\left[\boldsymbol{Y} \in D^2_{\beta',v} \,|\, \boldsymbol{u}(v,j), \boldsymbol{s}, \beta'\right]} \\ &\quad \times P\left[\boldsymbol{Y} \in D^2_{\beta',v} \,|\, \boldsymbol{u}(v,j), \boldsymbol{s}, \beta'\right] \qquad (50)\\ &\geq \left(\frac{n^2-1}{n^2}\right) P\left[\boldsymbol{Y} \in D^2_{\beta',v} \,|\, \boldsymbol{u}(v,j), \boldsymbol{s}, \beta'\right] \qquad (51)\\ &\geq \left(\frac{n^2-1}{n^2}\right)\left(P[\boldsymbol{Y} \in D_{\beta',v} \,|\, \boldsymbol{u}(v,j), \boldsymbol{s}, \beta'] - \frac{2|\mathcal{U}||\mathcal{Y}||\mathcal{S}|}{n}\right). \qquad (52)\end{aligned}$$

B. Upper Bound C_u

The encoder produces an output $\boldsymbol{x} = f(v, \boldsymbol{s})$ where f may be a stochastic function. When V is uniform in $\{1, \ldots, 2^{nR}\}$, this induces a distribution on the tuple $(V, \boldsymbol{S}, \boldsymbol{X}, \boldsymbol{Y})$. It is a simple extension of the basic Fano inequality [5] to show that for the induced distribution

$$\begin{aligned} H_\beta(V \,|\, \boldsymbol{Y}) &\leq h(P_e(\beta)) + P_e(\beta)\log(2^{nR} - 1) \qquad (53)\\ &\leq n[h(P_e(\beta)) + P_e(\beta)R] \qquad (54)\\ &:= n\delta(P_e(\beta)) \qquad (55)\end{aligned}$$

holds simultaneously for all $\beta \in \mathcal{C}$ where $P_e(\beta) = P[V \neq \hat{V}|\beta]$ and $h(\cdot)$ is the binary entropy function. Then, since V is independent of the side information

$$\begin{aligned} I^\beta(V; \boldsymbol{Y}) - I(V; \boldsymbol{S}) &= H(V) - H^\beta(V \,|\, \boldsymbol{Y}) \qquad (56)\\ &\geq n(R - \delta(P_e(\beta))) \qquad (57)\end{aligned}$$

holds simultaneously for all $\beta \in \mathcal{C}$. Furthermore, from [10, Lemma 4], we have that if we define the n random variables

$$U(i) = \left(V, Y_1^{i-1}, S_{i+1}^n\right) \tag{58}$$

then the induced distribution on the tuples $(U(i), X_i, S_i)$ by the encoder $f(v, \boldsymbol{s})$ is such that

$$\begin{aligned} \sum_{i=1}^n I^\beta(U(i); Y_i) - I(U(i); S_i) &\geq I^\beta(V; \boldsymbol{Y}) - I(V; \boldsymbol{S}) \qquad (59)\\ &\geq n(R - \delta(P_e(\beta))) \qquad (60)\end{aligned}$$

which again hold simultaneously for all $\beta \in \mathcal{C}$. Also, the induced distribution on the tuple may be factored as $P_{X_i,S_i} \times P^\beta_{U_i\,|\,X_i,S_i}$. Furthermore, if S and Y are drawn from finite alphabets, so is $U(i)$. Defining W to be a random variable uniformly distributed in $\{1, \ldots, n\}$ and $U = U(i), S = S_i, X = X_i$ when $W = i$

$$I^\beta(U; Y|W) - I(U; S|W) \geq R - \delta(P_e(\beta)). \tag{61}$$

If we minimize over all $\beta \in \mathcal{C}$ and rearrange terms

$$\sup_{\beta \in \mathcal{C}} \delta(P_e(\beta)) \geq R - \inf_{\beta \in \mathcal{B}} I^\beta(U; Y \,|\, W) - I(U; S \,|\, W) \tag{62}$$

which proves the converse for a $P_{X\,|\,S,W}$ and any family $\{P^\beta_{U\,|\,X,S,W}\}_\beta$. ■

Appendix III
Proposed Scheme

In this section, we show the following result.

Lemma 7: Let X, S, and Z be jointly Gaussian random variables whose covariance matrix is nonsingular and continuously parameterized by $\beta \in \mathcal{C}$ ($\mathcal{C}$ compact). Then there exists a sequence of quantizers, $q_k(S)$, for S such that $\inf_\beta I^\beta(X + q_k(S) + W_k; Z) \to \inf_\beta I^\beta(X + S; Z)$.

Proof: Let W_k be uniform additive noise in $[-L/k, L/k]$ and q_k be a uniform quantizer on $[-L, L]$ with k levels (if $|y| > L$ then $q_k(\cdot)$ is overloaded and the output is the constant $\text{sign}(y)(1 + 1/k)L$).

Denote by $f^\beta_{X,Y,Z}$ the joint probability density function (pdf) on X, Y, and Z when $Y = S$ and by $\tilde{f}^\beta_{X,Y,Z}$ the joint pdf on X, Y and Z when $Y = q_k(S) + W_k$. Likewise, denote by $f^\beta_{U,V,Z}$ and $\tilde{f}^\beta_{U,V,Z}$ the joint pdf on U, V, and Z when $U = X + Y$ and $V = X - Y$. We will not explicitly state the subscripts when they are clear from context.

By Lemma 5, for any ϵ_1, we can find quantizers for U and Z such that

$$\inf_\beta I^\beta(q_U(U); q_Z(Z))\big|_{f^\beta_{U,Z}} > \inf_\beta I^\beta(U; Z)\big|_{f^\beta_{U,Z}} - \epsilon_1.$$

We will show that by choosing k large enough

$$|I^\beta(q_U(U);q_Z(Z))|_{f^\beta_{U,Z}} - I^\beta(q_U(U);q_Z(Z))|_{\tilde{f}^\beta_{U,Z}}| < \epsilon_1 \quad (63)$$

which will prove the result. Since

$$\begin{aligned}
&\int_{x,y,z} |\tilde{f}^\beta(x,y,z) - f^\beta(x,y,z)|dxdydz \\
&\quad = \int_{u,v,z} |\tilde{f}^\beta(u,v,z) - f^\beta(u,v,z)|dudvdz && (64)\\
&\quad \geq \int_{u,z} |\int_v \tilde{f}^\beta(u,v,z) - f^\beta(u,v,z)dv|dudz && (65)\\
&\quad = \int_{u,z} |\tilde{f}^\beta(u,z) - f^\beta(u,z)|dudz && (66)\\
&\quad \geq d(\tilde{P}^\beta_{q_U(U),q_Z(Z)}, P^\beta_{q_U(U),q_Z(Z)}) && (67)
\end{aligned}$$

for any quantizers $q_U(\cdot)$ and $q_Z(\cdot)$, where $P^\beta_{q_U(U),q_Z(Z)}$ is the PMF induced by $f^\beta(u,z)$ and the quantizers $q_U(\cdot)$ and $q_Z(\cdot)$. By the uniform continuity of mutual information for finite-alphabet random variables, it will suffice to show that the left of (64) can be made uniformly small for all β.

Now, for any $\epsilon > 0$, choose L such that

$$\int_{\mathbb{R}^3 \setminus \{|x|\leq L, |y|\leq L, |z|\leq L\}} f^\beta(x,y,z) < \epsilon, \quad (68)$$

for all $\beta \in \mathcal{C}$. Also, let $K_L = \max_{|x|\leq L, |y|\leq L, |z|\leq L, \beta} f^\beta(x,z)$. Since $|\frac{d}{dy} f^\beta(y|x,z)|$ has a maximum on $|x| \leq L, |y| \leq L, |z| \leq L, \beta \in \mathcal{C}$, by choosing k sufficiently large, we can restrict

$$|\tilde{f}^\beta(y\,|\,x,z) - f^\beta(y\,|\,x,z)| < \delta \quad (69)$$

for any $\delta > 0$. Then, for $|x| \leq L, |y| \leq L, |z| \leq L$ and $\beta \in \mathcal{C}$

$$|\tilde{f}^\beta(x,y,z) - f^\beta(x,y,z)| < K_L\delta. \quad (70)$$

Then,

$$\begin{aligned}
&\int_{x,y,z} |\tilde{f}^\beta(x,y,z) - f^\beta(x,y,z)|dxdydz \\
&\quad \leq \int_{|x|\leq L, |y|\leq L, |z|\leq L} |\tilde{f}^\beta(x,y,z) f^\beta(x,y,z)|dxdydz && (71)\\
&\qquad + \int_{\mathbb{R}^3 \setminus \{|x|\leq L, |y|\leq L, |z|\leq L\}} \tilde{f}^\beta(x,y,z) + f^\beta(x,y,z)dxdydz \\
&\quad \leq L^3_\epsilon K_{L_\epsilon}\delta + 2\epsilon. && (72)
\end{aligned}$$

■

Acknowledgment

The authors would like to thank Daniel Yamins for insightful discussions.

References

[1] G. Caire and S. Shamai, "On the achievable throughput of a multiantenna Gaussian broadcast channel," *IEEE Trans. Inf. Theory*, vol. 49, no. 7, pp. 1691–1706, Jul. 2003.

[2] A. S. Cohen and A. Lapidoth, "The Gaussian watermarking game," *IEEE Trans. Inf. Theory*, vol. 48, no. 6, pp. 1639–1667, Jun. 2002.

[3] M. H. M. Costa, "Writing on dirty paper," *IEEE Trans. Inf. Theory*, vol. IT-29, no. 3, pp. 439–441, May 1983.

[4] T. M. Cover and M. Chiang, "Duality between channel capacity and rate distortion with two-sided state information," *IEEE Trans. Inf. Theory*, vol. 48, no. 6, pp. 1629–1639, Jun. 2002.

[5] T. M. Cover and J. A. Thomas, *Elements of Information Theory*. New York: Wiley, 1991.

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

[7] N. Devroye, P. Mitran, and V. Tarokh, "Achievable rates in cognitive radio channels," *IEEE Trans. Inf. Theory*, vol. 50, no. 11, Nov. 2004.

[8] ——, "Achievable rates in cognitive radio channels," in *Proc. Conf. Information Science and Systems*, Baltimore, MD, Jun. 2005, p. 302.

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

[10] S. I. Gel'fand and M. S. Pinsker, "Coding for channels with random parameters," *Probl. Contr. and Inf. Theory*, vol. 9, no. 1, pp. 19–31, 1980.

[11] R. M. Gray, *Probability, Random Processes, and Ergodic Properties*. New York: Springer-Verlag, 1988.

[12] ——, *Entropy and Information Theory*. New York: Springer-Verlag, 1990.

[13] L. J. Greenstein, V. Erceg, and D. G. Michelson, "Modeling diversity reception over narrowband fixed wireless channels," *Electron. Lett.*, vol. 34, no. 11, pp. 1146–1147, May 1998.

[14] T. S. Han and K. Kobayashi, "A new achievable rate region for the interference channel," *IEEE Trans. Inf. Theory*, vol. IT-27, no. 1, pp. 49–60, Jan. 1981.

[15] N. Jindal, S. Vishwanath, and A. Goldsmith, "On the duality of the Gaussian multiple-access and broadcast channels," *IEEE Trans. Inf. Theory*, vol. 50, no. 5, pp. 768–783, May 2004.

[16] Y.-H. Kim, A. Sutivong, and S. Sigurjonsson, "Multiple user writing on dirty paper," in *Proc. IEEE Int. Symp. Information Theory*, Chicago, IL, Jun./Jul. 2004, p. 534.

[17] J. Mitola, "Cognitive Radio," Ph.D. dissertation, Royal Institute of Technology (KTH), Stockholm, Sweden, 2000.

[18] S. Vishwanath, N. Jindal, and A. Goldsmith, "Duality, achievable rates, and sum-rate capacity of Gaussian MIMO broadcast channels," *IEEE Trans. Inf. Theory*, vol. 49, no. 10, pp. 2648–2657, Oct. 2003.

[19] P. Viswanath and D. N. C. Tse, "Sum capacity of the vector Gaussian broadcast channel and uplink-downlink duality," *IEEE Trans. Inf. Theory*, vol. 49, no. 8, pp. 1912–1921, Aug. 2003.

[20] H. Weingarten, Y. Steinberg, and S. Shamai (Shitz), "The capacity region of the Gaussian MIMO broadcast channel," *IEEE Trans. Inf. Theory*, submitted for publication.

[21] J. Wolfowitz, *Coding Theorems of Information Theory*. New York: Springer-Verlag, 1978.

[22] W. Yu and J. M. Cioffi, "Sum capacity of Gaussian vector broadcast channels," *IEEE Trans. Inf. Theory*, vol. 50, no. 9, pp. 1875–1892, Sep. 2004.

[23] W. Yu, A. Sutivong, D. Julian, T. M. Cover, and M. Chiang, "Writing on colored paper," in *Proc. IEEE Int. Symp. Information Theory*, Washington, DC, Jun. 2001, p. 302.

[24] R. Zamir, S. Shamai (Shitz), and U. Erez, "Nested linear/lattice codes for structured multiterminal binning," *IEEE Trans. Inf. Theory*, vol. 48, no. 6, pp. 1250–1276, Jun. 2002.