Communication and Interference Coordination

Ricardo Blasco-Serrano, Ragnar Thobaben, and Mikael Skoglund

KTH Royal Institute of Technology and ACCESS Linnaeus Centre

SE-100 44, Stockholm, Sweden

E-mail: {ricardo.blasco, ragnar.thobaben, mikael.skoglund}@ee.kth.se

Abstract—We study the problem of controlling the interference created to an external observer by a communication processes. We model the interference in terms of its type (empirical distribution), and we analyze the consequences of placing constraints on the admissible type. Considering a single interfering link, we characterize the communication-interference capacity region. Then, we look at a scenario where the interference is jointly created by two users allowed to coordinate their actions prior to transmission. In this case, the trade-off involves communication and interference as well as coordination. We establish an achievable communication-interference region and show that efficiency is significantly improved by coordination.

I. Introduction

Communication is subject to undesirable and often unavoidable interference that degrades the performance of neighboring transceivers and impairs the operation of nearby electronic devices. From an information-theoretic point of view, interference has traditionally been studied using the interference channel, which models the mutual effects between two user pairs that communicate simultaneously. This channel abstraction captures the fundamental tradeoff between the communication rates of the two pairs. In spite of decades of efforts, our understanding of this tradeoff is only partial or restricted to some special cases (see [1, Chapter 6] for a basic summary). In addition, the model is less appropriate for the cases where the impairment is created to a different type of device that is not necessarily communicating. An alternative view of interference that goes beyond communication-impairment effects was proposed in [2]. The authors modeled the communication-induced disturbances in terms of the undesired information rate and investigated the limits on the communication rate imposed by a constraint on the disturbance. They characterized explicitly the rate-disturbance region for the single disturbance case and gave partial results for other cases.

In this work, we take a similar approach although our model for the interference is quite different. Instead of endowing the interference with an informational meaning, we characterize it in terms of its type (i.e., empirical distribution). Thus, we study which communication rates are compatible with constraints placed on the type of the interference created by the communication process. Our results are therefore related to the study of channels with constraints on the channel inputs (e.g., see [1, Sec. 3.3] and references therein) and on the channel outputs [3, Sec. 29]. Our motivation is similar to that in [4], where output constraints were used as a model for the external power restrictions encountered, for example, in cognitive radio systems. As we shall see, our results for the single user can be interpreted as a generalization of those in [4] for discrete channels. Moreover, our work is also connected to [5], which studies the empirical distributions of capacity-achieving codes, although our codes are characterized both by communication properties (i.e., vanishing error probabilities) *and* interference constraints (i.e., convergence of the interference type in an appropriate sense).

We also consider a multiuser set-up in which the transmitters are allowed to coordinate their actions to mitigate the joint effect of their interference and improve the overall efficiency. This is closely related to the problem of coordination in networks, which was studied in [6]. Most relevant to our work, the authors characterized (empirical) coordination in terms of the type of the sequences of actions and established the fundamental limits for a variety of network topologies. We show that this framework for coordination is very useful when different transmitters are subject to a common interference constraint.

In the remainder of this section we introduce the basic mathematical concepts and establish the notation. We consider the single user case in Section II and a multiple user case in Section III. Finally, we conclude our work in Section IV.

A. Preliminaries

We consider exclusively random variables with finite alphabets. We denote them and their realizations using upper case and lower case letters, respectively (e.g., X and x). We use bold face for vectors and specify their lengths using superindices (e.g., $\boldsymbol{x}^n$). We use calligraphic letters (e.g., $\mathcal{T}$ or $\mathcal{T}$) to denote sets. Given a set $\mathcal{T}$, we denote its complement by $\mathcal{T}^c$.

Definition 1 (Total Variation). Let $P_{X,Y}$ and $Q_{X,Y}$ be two probability distributions defined on $\mathcal{X} \times \mathcal{Y}$. The total variation between them is defined as

$$\|P_{X,Y} - Q_{X,Y}\|_{\text{TV}} \triangleq \frac{1}{2} \sum_{x,y} |P_{X,Y}(x,y) - Q_{X,Y}(x,y)|.$$

$\diamond$

Definition 2 (Type). Let $\boldsymbol{x}^n \in \mathcal{X}^n$ and $\boldsymbol{y}^n \in \mathcal{Y}^n$. The type of the tuple $(\boldsymbol{x}^n, \boldsymbol{y}^n)$ is defined as

$$T_{\boldsymbol{x}^n, \boldsymbol{y}^n}(x,y) \triangleq \frac{1}{n} \sum_{i=1}^{n} \mathbb{1}\left\{(x_i, y_i) = (x,y)\right\}$$

for all $(x,y) \in \mathcal{X} \times \mathcal{Y}$, where $\mathbb{1}\{\cdot\}$ is the indicator function.

$\diamond$

Definition 3 (Typical sequence). Let $x^n \in \mathcal{X}^n$ and $\epsilon > 0$. We say that the sequence x^n is (ϵ-)typical with respect to a distribution P_X if $\|T_{x^n} - P_X\|_{\text{TV}} < \epsilon$. We denote by $\mathcal{T}_\epsilon^{(n)}(P_X)$ the set of all such sequences. $\diamond$

Most of our results involve the following notion of convergence of sequences of probability distributions. Consider a sequence (indexed by n) of random vectors X^n with $X^n \sim P_{X^n}$ for some sequence of distributions P_{X^n}, and the corresponding sequence of types T_{X^n}. Consider also a sequence of deterministic distributions $G^{(n)}$. We say that T_{X^n} converges in probability in total variation to $G^{(n)}$ if

$$\lim_{n \to \infty} \Pr(\|T_{X^n} - G^{(n)}\|_{\text{TV}} \geq \epsilon) = 0$$

for all $\epsilon > 0$. We denote this using the shorthand notation

$$\|T_{X^n} - G^{(n)}\|_{\text{TV}} \to 0 \text{ in probability.}$$

(The specialization of this notion of convergence to the case of fixed G or to deterministic sequences is straightforward.)

II. SINGLE USER

Consider the scenario depicted in Figure 1. This corresponds to a discrete memoryless channel (DMC) with one input X and two outputs Y and Z. The output Y is the observation at the intended receiver, while Z corresponds to an undesired interference created to an external observer. The channel is governed by a conditional probability mass function (pmf) $P_{Y,Z|X}$. The encoder-decoder pair can use the channel for communicating a random message M as long as the interference z^n has a certain shape, measured in terms of its type $T_{z^n}(z)$. For this purpose, they use a code.

Definition 4 (Code). An $(n, 2^{nR})$-code for the scenario in Figure 1 consists of:
- a message set $\mathcal{M} \triangleq \{1, \ldots, \lceil 2^{nR} \rceil\}$,
- an encoding function $x^n : \mathcal{M} \to \mathcal{X}^n$,
- a decoding function $\hat{m} : \mathcal{Y}^n \to \mathcal{M} \cup \{e\}$.

$\diamond$

We assume that the message is uniformly distributed over the message set.

Definition 5 (Achievability). We say that the communication rate R is achievable with interference type G_Z if there exists a sequence of $(n, 2^{nR})$-codes such that

$$\lim_{n \to \infty} \Pr(\hat{M} \neq M) = 0, \tag{1}$$

$$\|T_{Z^n} - G_Z\|_{\text{TV}} \to 0 \text{ in probability} \tag{2}$$

under the distribution induced by the codes. $\diamond$

The *communication-interference capacity region* $\mathcal{C}$ of the DMC $P_{Y,Z|X}$ is the closure of the set of all rate-interference type tuples (R, G_Z) that are achievable.

Our main result for the channel model in Figure 1 is a complete characterization of the communication-interference capacity region (Theorem 6). This region is convex and depends only on the marginals $P_{Y|X}$ and $P_{Z|X}$. Convexity
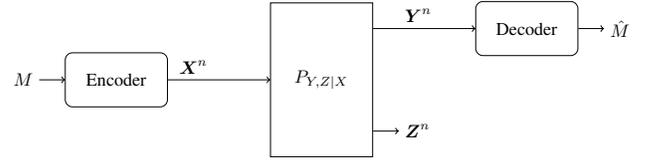

Fig. 1. Scenario for single-user communication with interference constraint.

is easily proven using standard time-sharing arguments. The dependency on the marginals also follows from well-known arguments (see e.g., [1, Lemma 5.1]).

Theorem 6. *The communication-interference capacity region $\mathcal{C}$ of the DMC $P_{Y,Z|X}$ is the set of rate-interference type tuples (R, G_Z) such that*

$$R \leq \max_{P_X \in \mathcal{P}} I(X; Y)$$

where

$$\mathcal{P} \triangleq \left\{ P_X : \sum_x P_X P_{Z|X} = G_Z \right\}. \tag{3}$$

$\square$

Observe that this result agrees with our basic understanding of communication and coordination. In particular, the capacity expression is reminiscent of that for the point-to-point channel but the maximization is over the restricted set $\mathcal{P}$ of input distributions P_X that induce the desired interference type G_Z. We will refer to the set $\mathcal{P}$ defined in (3) as the *pre-image* of G_Z. It is simple to show that the pre-image of a given G_Z is a closed and convex set.

The result in Theorem 6 is different from those involving constraints on the channel output in [3, Sec. 29] and [4]. For example, satisfying an interference power constraint does not directly imply convergence of the type of the interference in the sense defined above. In contrast, convergence of the type ensures that the power constraint is satisfied. However, our characterization of the interference in terms of its type does not extend to continuous alphabets.

In the remainder of this section we will prove Theorem 6. For this purpose, we first introduce the following auxiliary results (Lemmas 7-10).

Lemma 7. *The interference type T_{Z^n} induced by a sequence of $(n, 2^{nR})$-codes can only converge in probability to distributions G_Z with non-empty pre-image, that is, $\mathcal{P} \neq \emptyset$.* $\square$

Proof: First, observe that convergence in probability

$$\|T_{Z^n} - G_Z\|_{\text{TV}} \to 0$$

implies that

$$\mathbb{E}\{\|T_{Z^n} - G_Z\|_{\text{TV}}\} \to 0$$

because the total variation is bounded. In turn, this means that

$$\mathbb{E}\{T_{Z^n}\} \to G_Z$$

by a simple application of Jensen's inequality. Now, note that

$$
\begin{aligned}
\mathbb{E}\{T_{\boldsymbol{Z}^n}\} &= \sum_x \mathbb{E}\{T_{\boldsymbol{X}^n, \boldsymbol{Z}^n}\} \\
&= \sum_x \mathbb{E}\{T_{\boldsymbol{X}^n}\} P_{Z|X} \\
&= f(\mathbb{E}\{T_{\boldsymbol{X}^n}\}),
\end{aligned}
$$

where $f : \mathcal{X} \to \mathcal{Z}$ is a continuous function and $\mathbb{E}\{T_{\boldsymbol{X}^n}\}$ is a bounded sequence of probability distributions on $\mathcal{X}$. Thus, by the Bolzano-Weierstrass theorem [7, Theorem 3.6], the sequence $\mathbb{E}\{T_{\boldsymbol{X}^n}\}$ has a convergent subsequence, which we denote by $\bar{P}_X^{(n)}$. That is,

$$
\bar{P}_X^{(n)} \to \hat{P}_X,
$$

where $\hat{P}_X$ is the corresponding limit (i.e., a probability distribution on $\mathcal{X}$). By convergence $\mathbb{E}\{T_{\boldsymbol{Z}^n}\} \to G_Z$ and by continuity of the function f, we establish that

$$
\begin{aligned}
\lim_{n \to \infty} f(\mathbb{E}\{T_{\boldsymbol{X}^n}\}) &= \lim_{n \to \infty} f(\bar{P}_X^{(n)}) \\
&= f(\hat{P}_X) \\
&= G_Z.
\end{aligned}
$$

This means that $\hat{P}_X(x) \in \mathcal{P}$. Therefore, $\mathcal{P} \neq \emptyset$. ∎

Lemma 8. *Let G_Z be given and have pre-image $\mathcal{P}$ such that $\mathcal{P} \neq \emptyset$ and $\mathcal{P}^c \neq \emptyset$. Consider the sets*

$$
\tilde{\mathcal{P}}_\epsilon \triangleq \{\tilde{P}_X : \|\tilde{P}_X - P_X\|_{TV} \geq \epsilon \text{ for all } P_X \in \mathcal{P}\},
$$

$$
\tilde{\mathcal{G}}_\epsilon \triangleq \left\{\tilde{G}_Z : \sum_x P_{Z|X} \tilde{P}_X = \tilde{G}_Z \text{ for some } \tilde{P}_X \in \tilde{\mathcal{P}}_\epsilon\right\},
$$

defined for any fixed $\epsilon > 0$ such that $\tilde{\mathcal{P}}_\epsilon \neq \emptyset$. Let

$$
d^\star = \inf_{\tilde{G}_Z \in \tilde{\mathcal{G}}_\epsilon} \|G_Z - \tilde{G}_Z\|_{TV}.
$$

Then, we have that $d^\star > 0$. □

Proof: Assume that $d^\star = 0$. Note that $\tilde{\mathcal{P}}_\epsilon$ is a compact set and that $\tilde{G}_Z$ is a continuous function of $\tilde{P}_X$. Therefore, $\tilde{\mathcal{G}}_\epsilon$ is a compact set, too. Note also that $\|G_Z - \tilde{G}_Z\|_{TV}$ is a continuous function of $\tilde{G}_Z$. Thus, by Weierstrass' extreme value theorem [7, Theorem 4.16], there must exist some $\tilde{G}_Z \in \tilde{\mathcal{G}}_\epsilon$ (and hence some $\tilde{P}_X \in \tilde{\mathcal{P}}_\epsilon$) such that

$$
\|G_Z - \tilde{G}_Z\|_{TV} = 0.
$$

That is, $G_Z = \tilde{G}_Z$. However, this would imply that $\tilde{P}_X \in \mathcal{P}$, which is a contradiction. Thus, we must have $d^\star > 0$. ∎

Lemma 9. *Let $\epsilon > 0$ and consider two arbitrary pmfs Q_Z and $\tilde{Q}_Z$ defined on $\mathcal{Z}$ with typical sets $\mathcal{T}_\epsilon^{(n)}(Q_Z)$ and $\mathcal{T}_\epsilon^{(n)}(\tilde{Q}_Z)$, respectively. If the total variation between the pmfs satisfies $\|Q_Z - \tilde{Q}_Z\|_{TV} > 2\epsilon$ then the two typical sets are disjoint. That is, $\mathcal{T}_\epsilon^{(n)}(Q_Z) \cap \mathcal{T}_\epsilon^{(n)}(\tilde{Q}_Z) = \emptyset$.* □

Proof: Let $\boldsymbol{z}^n \in \mathcal{T}_\epsilon^{(n)}(Q_Z)$, that is,

$$
\|Q_Z - T_{\boldsymbol{z}^n}\|_{TV} < \epsilon.
$$

Then

$$
\begin{aligned}
\|\tilde{Q}_Z - T_{\boldsymbol{z}^n}\|_{TV} &= \|\tilde{Q}_Z - Q_Z + Q_Z - T_{\boldsymbol{z}^n}\|_{TV} \\
&\geq \|\tilde{Q}_Z - Q_Z\|_{TV} - \|Q_Z - T_{\boldsymbol{z}^n}\|_{TV} \\
&> 2\epsilon - \epsilon.
\end{aligned}
$$

Thus $\boldsymbol{z}^n \notin \mathcal{T}_\epsilon^{(n)}(\tilde{Q}_Z)$ and $\mathcal{T}_\epsilon^{(n)}(Q_Z) \cap \mathcal{T}_\epsilon^{(n)}(\tilde{Q}_Z) = \emptyset$. ∎

Lemma 10. *Let G_Z be fixed and have pre-image $\mathcal{P}$. If a sequence of $(n, 2^{nR})$-codes induces an interference type $T_{\boldsymbol{Z}^n}$ such that*

$$
\|T_{\boldsymbol{Z}^n} - G_Z\|_{TV} \to 0 \quad \text{in probability,} \tag{4}
$$

then the expectation of the type of the codewords $\mathbb{E}\{T_{\boldsymbol{X}^n}\}$ satisfies

$$
\|\mathbb{E}\{T_{\boldsymbol{X}^n}\} - P_X^{(n)}\|_{TV} \to 0 \tag{5}
$$

for some sequence $P_X^{(n)}$ with $P_X^{(n)} \in \mathcal{P}$ for all n. □

Proof: First, note that $\mathcal{P} \neq \emptyset$ by virtue of Lemma 7. Moreover, if $\mathcal{P}$ is equal to the whole simplex of probability distributions on $\mathcal{X}$ (i.e., $\mathcal{P}^c = \emptyset$) the proof is trivial. We prove the lemma for the case $\mathcal{P} \neq \emptyset, \mathcal{P}^c \neq \emptyset$ in two steps. i) First, we show that (4) implies that $\lim_{n \to \infty} \Pr(\boldsymbol{X}^n \notin \mathcal{T}_\epsilon^{(n)}(\mathcal{P})) = 0$ for any $\epsilon > 0$, where

$$
\mathcal{T}_\epsilon^{(n)}(\mathcal{P}) \triangleq \{\boldsymbol{x}^n : \|T_{\boldsymbol{x}^n} - P_X\|_{TV} < \epsilon \text{ for some } P_X \in \mathcal{P}\}.
$$

(The set $\mathcal{T}_\epsilon^{(n)}$ is a straightforward generalization of the typical set $\mathcal{T}_\epsilon^{(n)}$.) ii) Then, we show that this implies (5).

i) We prove the first step by contradiction. Assume that (4) is satisfied by some sequence of $(n, 2^{nR})$-codes with distribution $P_{\boldsymbol{X}^n}$ for which there exist $\delta > 0$ and $\epsilon_x > 0$ such that

$$
\delta \leq \limsup_{n \to \infty} \Pr(\boldsymbol{X}^n \notin \mathcal{T}_{\epsilon_x}^{(n)}(\mathcal{P})).
$$

Note that for every ϵ_x' such that $0 < \epsilon_x' < \epsilon_x$ we have that $\tilde{P}_{\epsilon_x} \subseteq \tilde{P}_{\epsilon_x'}$ and this implies that $\Pr(\boldsymbol{X}^n \notin \mathcal{T}_{\epsilon_x}^{(n)}(\mathcal{P})) \leq \Pr(\boldsymbol{X}^n \notin \mathcal{T}_{\epsilon_x'}^{(n)}(\mathcal{P}))$. For our purposes, it will be more convenient to write our expressions in terms of

$$
\tilde{\mathcal{P}}_{\epsilon_x} \triangleq \{\tilde{P}_X : \|\tilde{P}_X - P_X\|_{TV} \geq \epsilon_x \text{ for all } P_X \in \mathcal{P}\}.
$$

With this notation, the set $\{\boldsymbol{x}^n \notin \mathcal{T}_{\epsilon_x}^{(n)}(\mathcal{P})\}$ is equivalent to $\{\boldsymbol{x}^n : T_{\boldsymbol{x}^n} \in \tilde{\mathcal{P}}_{\epsilon_x}\}$. Observe that $\tilde{\mathcal{P}}_{\epsilon_x} \neq \emptyset$ for sufficiently small ϵ_x because $\tilde{\mathcal{P}}_{\epsilon_x} \subseteq \mathcal{P}^c$ and $\mathcal{P}^c$ is a set with non-empty interior. Thus, without loss of generality, we assume that $\tilde{\mathcal{P}}_{\epsilon_x} \neq \emptyset$.

Now, we define the following finite cover $\mathcal{Q}_{\epsilon_c}$ of the set $\tilde{\mathcal{P}}_{\epsilon_x}$. Given ϵ_c such that $0 < \epsilon_c < \epsilon_x$, the set $\mathcal{Q}_{\epsilon_c}$ is a *finite* set of distributions on $\mathcal{X}$ such that for every $\tilde{P}_X \in \tilde{\mathcal{P}}_{\epsilon_x}$ there exists some $P_X \in \mathcal{Q}_{\epsilon_c}$ with

$$
\|P_X - \tilde{P}_X\|_{TV} < \epsilon_c.
$$

Such a cover exists because the set $\tilde{\mathcal{P}}_{\epsilon_x}$ is compact. In fact, there exist more than one set with these properties. For convenience, we choose one (any) such set with the smallest possible cardinality. Thus, any distribution in $\tilde{\mathcal{P}}_{\epsilon_x}$ can be approximated by an element in the finite set $\mathcal{Q}_{\epsilon_c}$ with an error

in terms of the total variation not exceeding ϵ_c. Fix an arbitrary ordering of the elements in $\mathcal{Q}_{\epsilon_c}$

$$\mathcal{Q}_{\epsilon_c} = \{Q_{X,1}, Q_{X,2}, \ldots Q_{X,|\mathcal{Q}_{\epsilon_c}|}\},$$

and let

$$\tilde{\mathcal{Q}}_i \triangleq \{\tilde{P}_X \in \tilde{\mathcal{P}}_{\epsilon_x} : \|Q_{X,i} - \tilde{P}_X\|_{\mathrm{TV}} < \epsilon_c\}$$

for $i \in \{1, \ldots, |\mathcal{Q}_{\epsilon_c}|\}$. To avoid the possibility that $\tilde{P}_X \in \tilde{\mathcal{Q}}_i$ and $\tilde{P}_X \in \tilde{\mathcal{Q}}_j$ for $i \neq j$, we define the following disjoint sets

$$\mathcal{Q}_1 \triangleq \tilde{\mathcal{Q}}_1,$$
$$\mathcal{Q}_i \triangleq \tilde{\mathcal{Q}}_i \backslash \bigcup_{j=1}^{i-1} \tilde{\mathcal{Q}}_j$$

for $i \in \{2, \ldots, |\mathcal{Q}_{\epsilon_c}|\}$. Observe that $\cup_i \mathcal{Q}_i = \tilde{\mathcal{P}}_{\epsilon_x}$. Thus, for each $\boldsymbol{x}^n \notin \mathcal{T}_{\epsilon_x}^{(n)}(\mathcal{P})$ its type $T_{\boldsymbol{x}^n}$ satisfies $T_{\boldsymbol{x}^n} \in \mathcal{Q}_i$ for exactly one $i \in \{1, \ldots, |\mathcal{Q}_{\epsilon_c}|\}$. Using this covering into disjoints sets, we write

$$\sum_{\boldsymbol{x}^n \notin \mathcal{T}_{\epsilon_x}^{(n)}(\mathcal{P})} P_{\boldsymbol{X}^n}(\boldsymbol{x}^n) = \sum_{i=1}^{|\mathcal{Q}_{\epsilon_c}|} \sum_{\boldsymbol{x}^n : T_{\boldsymbol{x}^n} \in \mathcal{Q}_i} P_{\boldsymbol{X}^n}(\boldsymbol{x}^n).$$

Now, for arbitrary $\epsilon > 0$, write

$$\sum_{\boldsymbol{z}^n \notin \mathcal{T}_{\epsilon}^{(n)}(G_Z)} P_{\boldsymbol{Z}^n}(\boldsymbol{z}^n) = \sum_{\boldsymbol{x}^n} P_{\boldsymbol{X}^n}(\boldsymbol{x}^n) \sum_{\boldsymbol{z}^n \notin \mathcal{T}_{\epsilon}^{(n)}(G_Z)} P_{\boldsymbol{Z}^n|\boldsymbol{X}^n}(\boldsymbol{z}^n|\boldsymbol{x}^n)$$

$$\geq \sum_{\boldsymbol{x}^n \notin \mathcal{T}_{\epsilon_x}^{(n)}(\mathcal{P})} P_{\boldsymbol{X}^n}(\boldsymbol{x}^n) \sum_{\boldsymbol{z}^n \notin \mathcal{T}_{\epsilon}^{(n)}(G_Z)} P_{\boldsymbol{Z}^n|\boldsymbol{X}^n}(\boldsymbol{z}^n|\boldsymbol{x}^n)$$

$$= \sum_{\boldsymbol{x}^n : T_{\boldsymbol{x}^n} \in \mathcal{Q}_1} P_{\boldsymbol{X}^n}(\boldsymbol{x}^n) \sum_{\boldsymbol{z}^n \notin \mathcal{T}_{\epsilon}^{(n)}(G_Z)} P_{\boldsymbol{Z}^n|\boldsymbol{X}^n}(\boldsymbol{z}^n|\boldsymbol{x}^n)$$

$$+ \sum_{\boldsymbol{x}^n : T_{\boldsymbol{x}^n} \in \mathcal{Q}_2} P_{\boldsymbol{X}^n}(\boldsymbol{x}^n) \sum_{\boldsymbol{z}^n \notin \mathcal{T}_{\epsilon}^{(n)}(G_Z)} P_{\boldsymbol{Z}^n|\boldsymbol{X}^n}(\boldsymbol{z}^n|\boldsymbol{x}^n)$$

$$+ \ldots \qquad (6)$$

Consider the ith term in (6). First, note that each of the sequences $\boldsymbol{x}^n$ in the sum belongs to the typical set $\mathcal{T}_{\epsilon_c}^{(n)}(Q_{X,i})$. Now, define $Q_{Z,i} \triangleq \sum_x P_{Z|X} Q_{X,i}$ and consider the set $\mathcal{T}_{\epsilon}^{(n)}(Q_{Z,i})$ of sequences $\boldsymbol{z}^n$ that are typical according to $Q_{Z,i}$.

From Lemma 8 we know that, given ϵ_x, there exists a fixed $d^\star > 0$ such that $\|G_Z - Q_{Z,i}\|_{\mathrm{TV}} \geq d^\star$ for all $Q_{Z,i}$ ($i \in \{1, \ldots, |\mathcal{Q}_{\epsilon_c}|\}$). Thus, for any ϵ such that $0 < \epsilon < \frac{d^\star}{2}$, applying Lemma 9 we see that $\mathcal{T}_{\epsilon}^{(n)}(G_Z) \cap \mathcal{T}_{\epsilon}^{(n)}(Q_{Z,i}) = \emptyset$. Using this, we write

$$\sum_{\boldsymbol{x}^n : T_{\boldsymbol{x}^n} \in \mathcal{Q}_i} P_{\boldsymbol{X}^n}(\boldsymbol{x}^n) \sum_{\boldsymbol{z}^n \notin \mathcal{T}_{\epsilon}^{(n)}(G_Z)} P_{\boldsymbol{Z}^n|\boldsymbol{X}^n}(\boldsymbol{z}^n|\boldsymbol{x}^n)$$

$$\geq \sum_{\boldsymbol{x}^n : T_{\boldsymbol{x}^n} \in \mathcal{Q}_i} P_{\boldsymbol{X}^n}(\boldsymbol{x}^n) \sum_{\boldsymbol{z}^n \in \mathcal{T}_{\epsilon}^{(n)}(Q_{Z,i})} P_{\boldsymbol{Z}^n|\boldsymbol{X}^n}(\boldsymbol{z}^n|\boldsymbol{x}^n).$$

Moreover, by the conditional typicality lemma [8, Lemma 2.12], we know that

$$\sum_{\boldsymbol{z}^n \in \mathcal{T}_{\epsilon}^{(n)}(Q_{Z,i})} P_{\boldsymbol{Z}^n|\boldsymbol{X}^n}(\boldsymbol{z}^n|\boldsymbol{x}^n) \geq 1 - \delta_{\epsilon_c, \epsilon}(n)$$

for every $\boldsymbol{x}^n$ such that $T_{\boldsymbol{x}^n} \in \mathcal{Q}_i$ and where $\delta_{\epsilon_c, \epsilon}(n) \triangleq \frac{1}{4n}\left(\frac{|\mathcal{X}||\mathcal{Z}|}{\epsilon - \epsilon_c}\right)^2$. The term $\delta_{\epsilon_c, \epsilon}(n)$ goes to 0 with n and is fixed given the cover $\mathcal{Q}_{\epsilon_c}$. Thus,

$$\sum_{\boldsymbol{x}^n : T_{\boldsymbol{x}^n} \in \mathcal{Q}_i} P_{\boldsymbol{X}^n}(\boldsymbol{x}^n) \sum_{\boldsymbol{z}^n \notin \mathcal{T}_{\epsilon}^{(n)}(G_Z)} P_{\boldsymbol{Z}^n|\boldsymbol{X}^n}(\boldsymbol{z}^n|\boldsymbol{x}^n)$$

$$\geq (1 - \delta_{\epsilon_c, \epsilon}(n)) \sum_{\boldsymbol{x}^n : T_{\boldsymbol{x}^n} \in \mathcal{Q}_i} P_{\boldsymbol{X}^n}(\boldsymbol{x}^n).$$

Using this, we rewrite (6) as

$$\sum_{\boldsymbol{z}^n \notin \mathcal{T}_{\epsilon}^{(n)}(G_Z)} P_{\boldsymbol{Z}^n}(\boldsymbol{z}^n) \geq \sum_{i=1}^{|\mathcal{Q}_{\epsilon_c}|} \sum_{\boldsymbol{x}^n : T_{\boldsymbol{x}^n} \in \mathcal{Q}_i} P_{\boldsymbol{X}^n}(\boldsymbol{x}^n)(1 - \delta_{\epsilon_c, \epsilon}(n))$$

$$\geq (1 - \delta_{\epsilon_c, \epsilon}(n)) \sum_{\boldsymbol{x}^n \notin \mathcal{T}_{\epsilon_x}^{(n)}(\mathcal{P})} P_{\boldsymbol{X}^n}(\boldsymbol{x}^n).$$

Therefore, for any $0 < \epsilon < \frac{d^\star}{2}$ we have

$$\limsup_{n \to \infty} \sum_{\boldsymbol{z}^n \notin \mathcal{T}_{\epsilon}^{(n)}(G_Z)} P_{\boldsymbol{Z}^n}(\boldsymbol{z}^n) \geq \limsup_{n \to \infty} (1 - \delta_{\epsilon_c, \epsilon}(n)) \sum_{\boldsymbol{x}^n \notin \mathcal{T}_{\epsilon_x}^{(n)}(\mathcal{P})} P_{\boldsymbol{X}^n}(\boldsymbol{x}^n)$$

$$\geq \delta$$
$$> 0.$$

This contradicts our initial hypothesis that $P_{\boldsymbol{X}^n}$ induces a type $T_{\boldsymbol{Z}^n}$ that satisfies (4). Thus, we must have $\lim_{n \to \infty} \Pr(\boldsymbol{X}^n \notin \mathcal{T}_{\epsilon}^{(n)}(\mathcal{P})) = 0$ for any $\epsilon > 0$.

ii) Now, we show that this implies (5). To this end, we write

$$\|\mathbb{E}\{T_{\boldsymbol{X}^n}\} - P_X^{(n)}\|_{\mathrm{TV}}$$
$$= \|\mathbb{E}\{T_{\boldsymbol{X}^n}|\boldsymbol{X}^n \in \mathcal{T}_{\epsilon}^{(n)}(\mathcal{P})\} \Pr(\boldsymbol{X}^n \in \mathcal{T}_{\epsilon}^{(n)}(\mathcal{P}))$$
$$+ \mathbb{E}\{T_{\boldsymbol{X}^n}|\boldsymbol{X}^n \notin \mathcal{T}_{\epsilon}^{(n)}(\mathcal{P})\} \Pr(\boldsymbol{X}^n \notin \mathcal{T}_{\epsilon}^{(n)}(\mathcal{P})) - P_X^{(n)}\|_{\mathrm{TV}}$$
$$\leq \|\mathbb{E}\{T_{\boldsymbol{X}^n}|\boldsymbol{X}^n \in \mathcal{T}_{\epsilon}^{(n)}(\mathcal{P})\} \Pr(\boldsymbol{X}^n \in \mathcal{T}_{\epsilon}^{(n)}(\mathcal{P})) - P_X^{(n)}\|_{\mathrm{TV}}$$
$$+ \|\mathbb{E}\{T_{\boldsymbol{X}^n}|\boldsymbol{X}^n \notin \mathcal{T}_{\epsilon}^{(n)}(\mathcal{P})\} \Pr(\boldsymbol{X}^n \notin \mathcal{T}_{\epsilon}^{(n)}(\mathcal{P}))\|_{\mathrm{TV}} \qquad (7)$$

for arbitrary $\epsilon > 0$. Note that, for any two sequences $\boldsymbol{x}^n$ and $\tilde{\boldsymbol{x}}^n$ that belong to the set $\mathcal{T}_{\epsilon}^{(n)}(\mathcal{P})$, the convex combination of their types $T_{\boldsymbol{x}^n}$ and $T_{\tilde{\boldsymbol{x}}^n}$ satisfies

$$\|\lambda T_{\boldsymbol{x}^n} + (1 - \lambda) T_{\tilde{\boldsymbol{x}}^n} - P_X\|_{\mathrm{TV}} < \epsilon$$

for some $P_X \in \mathcal{P}$ and any $\lambda \in [0, 1]$. Thus, since

$$\mathbb{E}\{T_{\boldsymbol{X}^n}|\boldsymbol{X}^n \in \mathcal{T}_{\epsilon}^{(n)}(\mathcal{P})\} \Pr(\boldsymbol{X}^n \in \mathcal{T}_{\epsilon}^{(n)}(\mathcal{P}))$$

is a convex combination of types of sequences in $\mathcal{T}_{\epsilon}^{(n)}(\mathcal{P})$, we have that

$$\|\mathbb{E}\{T_{\boldsymbol{X}^n}|\boldsymbol{X}^n \in \mathcal{T}_{\epsilon}^{(n)}(\mathcal{P})\} \Pr(\boldsymbol{X}^n \in \mathcal{T}_{\epsilon}^{(n)}(\mathcal{P})) - P_X^{(n)}\|_{\mathrm{TV}} < \epsilon$$

for some $P_X^{(n)} \in \mathcal{P}$. Regarding the second term in (7), we see that

$$\|\mathbb{E}\{T_{\boldsymbol{X}^n}|\boldsymbol{X}^n \notin \mathcal{T}_{\epsilon}^{(n)}(\mathcal{P})\} \Pr(\boldsymbol{X}^n \notin \mathcal{T}_{\epsilon}^{(n)}(\mathcal{P}))\|_{\mathrm{TV}}$$
$$= \Pr(\boldsymbol{X}^n \notin \mathcal{T}_{\epsilon}^{(n)}(\mathcal{P})) \|\mathbb{E}\{T_{\boldsymbol{X}^n}|\boldsymbol{X}^n \notin \mathcal{T}_{\epsilon}^{(n)}(\mathcal{P})\}\|_{\mathrm{TV}}$$
$$\leq \Pr(\boldsymbol{X}^n \notin \mathcal{T}_{\epsilon}^{(n)}(\mathcal{P}))$$
$$< \epsilon,$$

where the inequality is satisfied for sufficiently large n. Combining the two bounds, we see that

$$\|\mathbb{E}\{T_{\boldsymbol{X}^n}\} - P_X^{(n)}\|_{\mathrm{TV}} < 2\epsilon.$$

Finally, we complete the proof by letting $\epsilon \to 0$. ∎

We note that it is also possible to prove the preceding lemma by using the techniques in [5] (in particular, [5, Theorem 4]), adapted to our notion of convergence.

We are now ready to prove Theorem 6.

Proof of Theorem 6: The achievability result follows easily from Shannon's coding theorem. For the converse result, consider a sequence of $(n, 2^{nR})$-codes that achieve the rate-interference type pair (R, G_Z). The sequence, together with the uniform distribution on the messages, induces the joint distribution

$$\frac{1}{|\mathcal{M}|} P_{\boldsymbol{X}^n|M} P_{\boldsymbol{Y}^n|\boldsymbol{X}^n} P_{\boldsymbol{Z}^n|\boldsymbol{X}^n} P_{\hat{M}|\boldsymbol{Y}^n}, \quad (8)$$

with $P_{\boldsymbol{Y}^n|\boldsymbol{X}^n} = \prod P_{Y|X}$ and $P_{\boldsymbol{Z}^n|\boldsymbol{X}^n} = \prod P_{Z|X}$. Observe that in (8), we have restricted our attention to distributions $P_{Y,Z|X} = P_{Y|X} P_{Z|X}$. As discussed before, this entails no loss of generality.

First, by the standard arguments based on Fano's inequality (e.g., see [1, eq. (3.3)]), a vanishing error probability (i.e., (1)) implies that

$$\begin{aligned}
nR &\le \sum_{q=1}^n I(X_q; Y_q) + n\epsilon_n \\
&= n \sum_{q=1}^n \frac{1}{n} I(X_q; Y_q | Q = q) + n\epsilon_n \\
&= n I(X_Q; Y_Q | Q) + n\epsilon_n \\
&\le n I(Q X_Q; Y_Q) + n\epsilon_n \\
&= n I(X_Q; Y_Q) + n\epsilon_n \quad (9)
\end{aligned}$$

where Q is a random variable uniformly distributed on $\{1, \dots, n\}$ and independent of $(\boldsymbol{X}^n, \boldsymbol{Y}^n, \boldsymbol{Z}^n)$, and $\epsilon_n \ge 0$ with $\epsilon_n \to 0$ as $n \to \infty$. The last equality in (9) is justified by the fact that the DMC establishes the Markov chain $Q - X_Q - Y_Q$. Dividing by n, we obtain

$$R \le I(X_Q; Y_Q) + \epsilon_n.$$

This mutual information is evaluated for P_{X_Q, Y_Q}, which can be written as

$$\begin{aligned}
P_{X_Q, Y_Q}(x, y) &= P_{X_Q}(x) P_{Y|X}(y|x) \\
&= \mathbb{E}\{T_{\boldsymbol{X}^n}(x)\} P_{Y|X}(y|x).
\end{aligned}$$

The first equality comes from the Markov chain $Q - X_Q - Y_Q$. The second equality is Property 2 in [6, Section VII.B.2].

Now, condition (2) on the type of the interference for a sequence of $(n, 2^{nR})$-codes that achieves the pair (R, G_Z), combined with Lemma 10, implies that the expectation of the type of the input to the channel $\mathbb{E}\{T_{\boldsymbol{X}^n}\}$ must converge to a sequence $P_X^{(n)}$ with $P_X^{(n)} \in \mathcal{P}$ for all n. That is,

$$\mathbb{E}\{T_{\boldsymbol{X}^n}(x)\} P_{Y|X}(y|x) \to P_X^{(n)}(x) P_{Y|X}(y|x)$$
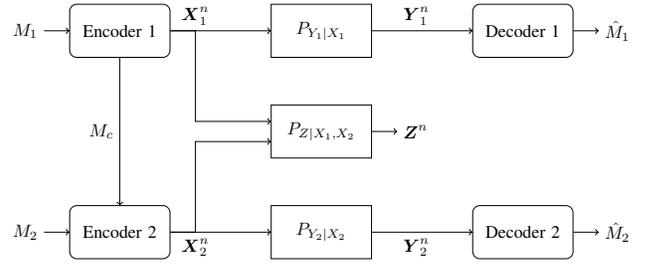

Fig. 2. Scenario for coordination of communications with interference constraints.

or, equivalently,

$$P_{X_Q, Y_Q}(x, y) \to P_X^{(n)}(x) P_{Y|X}(y|x).$$

Since the mutual information is a continuous function of the input distribution, this convergence implies that any sequence of $(n, 2^{nR})$-codes must satisfy

$$\begin{aligned}
R &\le \limsup_{n \to \infty} I(X; Y)|_{P_X^{(n)}} \\
&\le \max_{P_X \in \mathcal{P}} I(X; Y).
\end{aligned}$$

In conclusion, achievability of the pair (R, G_Z) implies that $(R, G_Z) \in \mathcal{C}$. ∎

III. MULTIPLE USERS

Consider the scenario depicted in Figure 2. Two transmitters want to communicate with their respective receivers through a channel governed by a conditional product pmf

$$P_{Y_1, Y_2, Z|X_1, X_2} = P_{Y_1|X_1} P_{Y_2|X_2} P_{Z|X_1, X_2}. \quad (10)$$

The marginals $P_{Y_1|X_1}$ and $P_{Y_2|X_2}$ model orthogonal communication channels between pairs of encoders and decoders, whereas $P_{Z|X_1, X_2}$ models the joint disturbance that the two transmissions create to the observer. That is, although the user pairs do not hamper each other's transmission, they create interference at a third external node, the observer. To control this interference, the two transmitters have access to a unidirectional rate-limited noiseless link from the first to the second encoder. They can use this resource to coordinate their transmissions and shape the type of the interference $T_{\boldsymbol{z}^n}(z)$.

Observe that our model makes no assumption on how the two transmitters interfere with the observer, beyond the structure in (10) (i.e., memoryless interference at symbol level). By choosing appropriately $P_{Z|X_1, X_2}$, we can model a scenarios ranging from symbol-level synchronization to carrier level synchronization, among others.

We now introduce the necessary definitions and state our main results for this scenario.

Definition 11 (Code). An $(n, 2^{nR_1}, 2^{nR_2}, 2^{nR_c})$-code for the scenario in Figure 2 consists of:

- three sets of messages:

$$\begin{aligned}
\mathcal{M}_j &\triangleq \{1, \dots, \lceil 2^{nR_j} \rceil\} \text{ for } j \in \{1, 2\}, \\
\mathcal{M}_c &\triangleq \{1, \dots, \lfloor 2^{nR_c} \rfloor\},
\end{aligned}$$

- two encoding functions

$$\boldsymbol{x}_1^n : \mathcal{M}_1 \to \mathcal{X}_1^n,$$
$$\boldsymbol{x}_2^n : \mathcal{M}_2 \times \mathcal{M}_c \to \mathcal{X}_2^n,$$

- a coordination function $c : \mathcal{M}_1 \to \mathcal{M}_c$,
- and two decoding functions $\hat{m}_j : \mathcal{Y}_j^n \to \mathcal{M}_j \cup \{e\}$ for $j \in \{1, 2\}$.

$\diamondsuit$

We assume that the message pair (M_1, M_2) is uniformly distributed over the set $\mathcal{M}_1 \times \mathcal{M}_2$. The notion of achievability and the definition of the communication-interference capacity region $\mathcal{C}$ are straightforward extensions of those introduced in the single user case. As for that case, the communication-interference capacity region $\mathcal{C}$ is convex. However, observe that the factorization in (10) entails a loss of generality.

Consider the following set:

$$\mathcal{R} \triangleq \left\{ \begin{array}{l} (R_1, R_2, R_c, Q_Z) \text{ s.t. } \exists\, P_U P_{X_1|U} P_{X_2|U} \text{ s.t.} \\ \quad R_1 < I(X_1; Y_1), \\ \quad R_2 < [I(X_2; Y_2) - I(U; X_2)]^+, \\ \quad R_c > I(U; X_1), \\ \quad \sum_{u, x_1, x_2} P_U P_{X_1|U} P_{X_2|U} P_{Z|X_1, X_2} = Q_Z \end{array} \right\}$$

where $[x]^+ \triangleq \max(x, 0)$. Let $\mathrm{conv}(\mathcal{R})$ denote the convex hull of $\mathcal{R}$. Our main result for the channel model in Figure 2 is the following partial characterization.

Theorem 12. *The communication-interference capacity region $\mathcal{C}$ satisfies*

$$\mathrm{conv}(\mathcal{R}) \subseteq \mathcal{C}.$$

$\square$

Before proving the theorem, we make the following two observations about $\mathcal{R}$: i) The random variable U plays the role of the coordination message sent from Encoder 1 to Encoder 2. By setting $U = \emptyset$, we obtain $R_c = 0$ and recover the case where the users are not coordinated (i.e., X_1 and X_2 are independent). For most distributions $P_{Z|X_1, X_2}$, our strategy strictly improves upon uncoordinated communication. ii) The coordination message U couples the rates R_1 and R_2 in two ways. First, the choices of input distributions have to be compatible in the sense that they yield the desired G_Z. In addition, the rate for Encoder 2 has a penalty term that reflects that the transmitted signals are correlated. That is, X_2 carries information about X_1. This is similar to the situation in Gel'fand Pinsker coding, where the transmission is aligned with the channel state and thus carries information about it [9]. These considerations are illustrated by the following example.

Example 13. Consider the scenario in which each of the two encoders can make use of the set of 16 symbols depicted in Figure 3 as inputs to the channel. Assume that the observer tolerates only low and mild levels of interference. This means that the two encoders are not allowed to use the black-circle symbols simultaneously. For simplicity, assume that the channels $P_{Y_1|X_1}$ and $P_{Y_2|X_2}$ are noiseless.
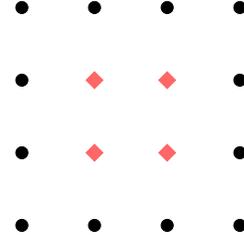

Fig. 3. Constellation with 16 symbols in Example 13. The constraint on the interference at the observer precludes transmission of black-circle symbols by both encoders at the same time.

Without coordination, one of the two users is restricted to use only the subset of red-diamond symbols. Assume that the restriction is placed on the second user. This yields the rate pair $(R_1, R_2) = (4, 2)$. In contrast, if Encoder 1 uses the coordination link to declare whether it will use a black-circle or a red-diamond symbol, Encoder 2 can opportunistically choose its constellation to boost its communication rate. For example, if Encoder 1 makes use of all 16 symbols with equal frequency, then Encoder 2 is forced to use the red-diamond symbols (i.e., transmit 2 [bpcu]) 75% of the times. However, in the remaining 25%, it can use any of the black-circle symbols (i.e., $\log_2 12$ [bpcu]). This yields

$$R_2 = \frac{3}{4} 2 + \frac{1}{4} \log_2 12 \approx 2.4 \text{ [bpcu]}.$$

Thus, we have $(R_1, R_2) = (4, 2.4)$. Observe that the constraint placed by the observer does not preclude Encoder 2 from using *any* of the symbols in Figure 3 when Encoder 1 sends a red-diamond symbol. However, Decoder 2 needs to know whether the transmitted symbol corresponds to 2 or 4 bits. By restricting its input to belong to the set of black-circle symbols, Encoder 2 is conveying information about the message of Encoder 1, namely that the current input consists of one of the red-diamond symbols.

A coordination rate equal to $R_c = 0.81$ [bpa] is sufficient to implement this protocol if Encoder 1 uses a lossless source coding algorithm to declare its intentions for a batch of channel uses. $\diamondsuit$

Proof of Theorem 12: Fix arbitrary $\epsilon > 0$ and let $\delta(\epsilon) > 0$ be some positive function such that $\delta(\epsilon) \to 0$ as $\epsilon \to 0$. Choose a tuple $(R_1, R_2, R_c, Q_Z) \in \mathcal{R}$ and let $\tilde{R}_2 > R_2$. Let $P_U P_{X_1|U} P_{X_2|U}$ be the corresponding distribution.

Codebook generation

- For every $m_c \in \mathcal{M}_c$, generate a sequence $\boldsymbol{u}^n(m_c)$ according to $\prod_{i=1}^n P_U(u_i)$.
- For every $m_1 \in \mathcal{M}_1$, generate a codeword $\boldsymbol{x}_1^n(m_1)$ according to $\prod_{i=1}^n P_{X_1}(x_{1i})$.
- For every $m_2 \in \mathcal{M}_2$ and every $l \in \{1, .., \lceil 2^{n(\tilde{R}_2 - R_2)} \rceil\}$, generate a codeword $\boldsymbol{x}_2^n(l, m_2)$ according to $\prod_{i=1}^n P_{X_2}(x_{2i})$.

Encoding

1) To transmit the message m_1, Encoder 1 puts the codeword $\boldsymbol{x}_1^n(m_1)$ into the channel.

2) To generate the coordination message given $\boldsymbol{x}_1^n(m_1)$, Encoder 1 searches for an index m_c such that $(\boldsymbol{u}^n(m_c), \boldsymbol{x}_1^n(m_1)) \in \mathcal{T}_\epsilon^{(n)}(P_{U,X_1})$. If more than one such m_c exists, it chooses one at random among the candidates. If none exists, then it chooses $m_c = 1$. Finally, it conveys the index m_c to Encoder 2.

3) To transmit the message m_2, Encoder 2 searches for an index l such that $(\boldsymbol{u}^n(m_c), \boldsymbol{x}_2^n(l, m_2)) \in \mathcal{T}_\epsilon^{(n)}(P_{U,X_2})$. If more than one such l exists, it chooses one at random among the candidates. If none exists, then it chooses $l = 1$. Finally, it puts the codeword $\boldsymbol{x}_2^n(l, m_c)$ into the channel.

Decoding

- Given the observation $\boldsymbol{y}_1^n$, Decoder 1 searches for a unique index $\hat{m}_1$ such that $(\boldsymbol{x}_1^n(\hat{m}_1), \boldsymbol{y}_1^n) \in \mathcal{T}_\epsilon^{(n)}(P_{X_1,Y_1})$. If no such $\hat{m}_1$ is found or if it is not unique, the decoder declares an error.
- Given the observation $\boldsymbol{y}_2^n$, Decoder 2 searches for a unique index $\hat{m}_2$ such that $(\boldsymbol{x}_2^n(\hat{l}, \hat{m}_2), \boldsymbol{y}_2^n) \in \mathcal{T}_\epsilon^{(n)}(P_{X_2,Y_2})$ for some $\hat{l} \in \{1, \ldots, \lceil 2^{n(\tilde{R}_2 - R_2)} \rceil\}$. If no such $\hat{m}_2$ is found or if it is not unique, the decoder declares an error.

Analysis of the error probability

We consider the error probability averaged over the ensemble of codebooks. Let $\mathcal{E}$ denote the error event and consider a fixed n. Due to the symmetry in the generation of the codebooks, we can assume that $M_1 = M_2 = 1$ without loss of generality. That is,

$$\Pr(\mathcal{E}) = \Pr(\mathcal{E}|(M_1, M_2) = (1, 1)).$$

To bound the error probability, consider the following events:

$$\mathcal{E}_Z \triangleq \{\|T_{\boldsymbol{Z}^n} - Q_Z\|_{\mathrm{TV}} \geq \epsilon\},$$
$$\mathcal{E}_i \triangleq \{\hat{M}_i \neq 1\}$$

for $i = \{1, 2\}$. The error probability satisfies

$$\Pr(\mathcal{E}) \leq \Pr(\mathcal{E}_Z|(M_1, M_2) = (1, 1))$$
$$+ \Pr(\mathcal{E}_1|M_1 = 1) + \Pr(\mathcal{E}_2|M_2 = 1). \quad (11)$$

We bound each of the three terms individually. For the first term in (11), consider the event

$$\mathcal{E}_{Z0} \triangleq \{\|T_{\boldsymbol{U}^n, \boldsymbol{X}_1^n(1), \boldsymbol{X}_2^n(L,1), \boldsymbol{Z}^n} - P_{Z|X_1,X_2} P_{X_1|U} P_{X_2|U} P_U\|_{\mathrm{TV}} \geq \epsilon\}$$

and note that, by the basic properties of strong typicality, for every $(\boldsymbol{u}^n, \boldsymbol{x}_1^n, \boldsymbol{x}_2^n, \boldsymbol{z}^n)$ such that

$$\|T_{\boldsymbol{u}^n, \boldsymbol{x}_1^n, \boldsymbol{x}_2^n, \boldsymbol{z}^n} - P_{Z|X_1,X_2} P_{X_1|U} P_{X_2|U} P_U\|_{\mathrm{TV}} < \epsilon,$$

we have

$$\|T_{\boldsymbol{z}^n} - Q_Z\|_{\mathrm{TV}} < \epsilon.$$

Therefore,

$$\Pr(\mathcal{E}_Z|(M_1, M_2) = (1, 1)) \leq \Pr(\mathcal{E}_{Z0}).$$

Now, let $\epsilon' = \frac{\epsilon}{4}$ and

$$\mathcal{E}_{Z1} \triangleq \{(\boldsymbol{U}^n(m_c), \boldsymbol{X}_1^n(1)) \notin \mathcal{T}_{\epsilon'}^{(n)}(P_{U,X_1}) \text{ for all } m_c \in \mathcal{M}_c\},$$
$$\mathcal{E}_{Z2} \triangleq \{(\boldsymbol{U}^n(M_c), \boldsymbol{X}_2^n(l,1)) \notin \mathcal{T}_{\epsilon'}^{(n)}(P_{U,X_2})$$
$$\text{for all } l \in \{1, \ldots, \lceil 2^{n(\tilde{R}_2 - R_2)} \rceil\}\},$$
$$\mathcal{E}_{Z3} \triangleq \{(\boldsymbol{U}^n(M_c), \boldsymbol{X}_1^n(1), \boldsymbol{X}_2^n(L,1)) \notin \mathcal{T}_\epsilon^{(n)}(P_{U,X_1,X_2})\},$$
$$\mathcal{E}_{Z4} \triangleq \{(\boldsymbol{U}^n(M_c), \boldsymbol{X}_1^n(1), \boldsymbol{X}_2^n(L,1), \boldsymbol{Z}^n) \notin \mathcal{T}_\epsilon^{(n)}(P_{U,X_1,X_2,Z})\}.$$

Here M_c and L are the random variables corresponding to the coordination index and the index chosen by Encoder 2, respectively. We have that

$$\Pr(\mathcal{E}_{Z0}) \leq \Pr(\mathcal{E}_{Z1}) + \Pr(\mathcal{E}_{Z2})$$
$$+ \Pr(\mathcal{E}_{Z3} \cap (\mathcal{E}_{Z1}^c \cap \mathcal{E}_{Z2}^c)) + \Pr(\mathcal{E}_{Z4} \cap \mathcal{E}_{Z3}^c). \quad (12)$$

By the covering lemma [1, Lemma 3.3], $\Pr(\mathcal{E}_{Z1}) \to 0$ as $n \to \infty$ if $R_c > I(U; X_1) - \delta(\epsilon')$. For the second term in (12), note that the distribution of $(\boldsymbol{U}^n(M_c), \boldsymbol{X}_2^n(l,1))$ is the same for all values of M_c and l; they are independent. Thus, again by the covering lemma, $\Pr(\mathcal{E}_{Z2}) \to 0$ as $n \to \infty$ if $\tilde{R}_2 - R_2 > I(U; X_2) - \delta(\epsilon')$.

Regarding the third term in (12), we observe the following. Given $\mathcal{E}_{Z1}^c$, we have that $(\boldsymbol{U}^n(M_c), \boldsymbol{X}_1^n(1)) \in \mathcal{T}_\epsilon^{(n)}(P_{U,X_1})$. Similarly, given $\mathcal{E}_{Z2}^c$, we have that $(\boldsymbol{U}^n(M_c), \boldsymbol{X}_2^n(L,1)) \in \mathcal{T}_\epsilon^{(n)}(P_{U,X_2})$. Thus, by the strong Markov Lemma [6, Theorem 12], $\Pr(\mathcal{E}_{Z3} \cap (\mathcal{E}_{Z1}^c \cap \mathcal{E}_{Z2}^c)) \to 0$ as $n \to \infty$. The conditions of the lemma are satisfied because $X_1 - U - X_2$ form a Markov chain and the distribution of $\boldsymbol{X}_2^n$ is permutation invariant (as defined in [6]) with respect to $\boldsymbol{u}^n$.

Finally, for the last term in (12), we have that $\boldsymbol{Z}^n$ is generated by passing a ϵ-typical pair $(\boldsymbol{X}_1^n, \boldsymbol{X}_2^n)$ through the channel $P_{Z|X_1,X_2}$. Thus, by the law of large numbers, $\Pr(\mathcal{E}_{Z4} \cap \mathcal{E}_{Z3}^c) \to 0$ as $n \to 0$.

We now turn our attention to the term $\Pr(\mathcal{E}_1|M_1 = 1)$ in (11). Consider the following events

$$\mathcal{E}_{11} \triangleq \{(\boldsymbol{X}_1^n(1), \boldsymbol{Y}_1^n) \notin \mathcal{T}_\epsilon^{(n)}(P_{X_1,Y_1})\},$$
$$\mathcal{E}_{12} \triangleq \{(\boldsymbol{X}_1^n(\hat{m}_1), \boldsymbol{Y}_1^n) \in \mathcal{T}_\epsilon^{(n)}(P_{X_1,Y_1}) \text{ for some } \hat{m}_1 \neq 1\}.$$

We have that

$$\Pr(\mathcal{E}_1|M_1 = 1) \leq \Pr(\mathcal{E}_{11}) + \Pr(\mathcal{E}_{12}),$$

where $\Pr(\mathcal{E}_{11}) \to 0$ as $n \to 0$ by the law of large numbers, and $\Pr(\mathcal{E}_{12}) \to 0$ as $n \to 0$ if $R_1 < I(X_1; Y_1) - \delta(\epsilon)$ by the packing lemma [1, Lemma 3.1].

Similarly, if $\tilde{R}_2 < I(X_2; Y_2) - \delta(\epsilon)$ then $\Pr(\mathcal{E}_2|M_2 = 1) \to 0$ as $n \to 0$. Combining all the terms and letting $\epsilon \to 0$, we obtain

$$R_c > I(U; X_1),$$
$$R_1 < I(X_1; Y_1),$$
$$R_2 < [\tilde{R}_2 - I(U; X_2)]^+ < [I(X_2; Y_2) - I(U; X_2)]^+,$$

as desired. The remaining tuples in the convex hull are achieved by time sharing. ∎

IV. Conclusion

We have proposed a generic model in terms of types (i.e., empirical distributions) for studying the effect of the interference induced by a communication process. First, we have considered the case of a single communication link and shown the existence of a tradeoff between the rate of communication and the type of the induced interference. To quantify this tradeoff, we have introduced the notion of communication-interference capacity region and we have explicitly characterized it. Then, we have studied a multiple-user scenario with unidirectional coordination of the transmitters. In this case, we have shown that the tradeoff involves the interference type and the communication rate as well as the coordination rate. We have established an inner bound to the communication-interference capacity region as a partial characterization of the tradeoff.

References

[1] A. El Gamal and Y.-H. Kim, *Network information theory*. Cambridge, UK: Cambridge University Press, 2011.

[2] B. Bandemer and A. El Gamal, "Communication with disturbance constraints," in *Proc. IEEE Int. Symp. on Information Theory (ISIT)*, Jul. 2011, pp. 2090–2094.

[3] C. Shannon, "A mathematical theory of communication," *Bell Systems Technical Journal*, vol. 27, pp. 379–423 & 623–656, 1948.

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

[5] S. Shamai and S. Verdú, "The empirical distribution of good codes," *IEEE Transactions on Information Theory*, vol. 43, no. 3, pp. 836–846, 1997.

[6] P. Cuff, H. Permuter, and T. Cover, "Coordination capacity," *IEEE Transactions on Information Theory*, vol. 56, no. 9, pp. 4181–4206, Sep. 2010.

[7] W. Rudin, *Principles of mathematical analysis*, 3rd ed. New York, USA: McGraw-Hill, 1976.

[8] I. Csiszár and J. Körner, *Information theory: Coding theorems for discrete memoryless channels*, Budapest, Hungary, 1981.

[9] S. Gel'fand and M. Pinsker, "Coding for channel with random parameters," *Prob. Contr. and Inform. Theory*, vol. 9, no. 1, pp. 19–31, 1980.