Interference Channel with Generalized Feedback (a.k.a. with source cooperation): Part I: Achievable Region

Shuang (Echo) Yang and Daniela Tuninetti

Abstract—**An Interference Channel with Generalized Feedback (IFC-GF) models a wireless network where the sources can sense the channel activity. The signal overheard from the channel provides information about the activity of the other sources and thus furnishes the basis for cooperation. This two-part paper studies achievable strategies (Part I) and outer bounds (Part II) for the general discrete memoryless IFC-GF with two source-destination pairs.**

In Part I, the generalized feedback is used to gain knowledge about the message sent by the other source and then exploited in two ways: (a) to *relay* the messages that can be decoded at both destinations, thus realizing the gains of beam-forming of a distributed multi-antenna system, and (b) to *hide* the messages that can not be decoded at the non-intended destination, thus leveraging the interference "pre-cancellation" property of dirty-paper coding. We show that our achievable region generalizes several known achievable regions for the IFC-GF and that it reduces to known achievable regions for the channels subsumed by the IFC-GF model. For the Gaussian channel, it is shown that source cooperation enlarges the achievable rate region of the corresponding IFC without generalized feedback/cooperation.

Index Terms—**Achievable region, Binning, Generalized feedback, Gaussian channel, Interference channel, Source cooperation, Superposition coding.**

I. Introduction

THE practical bottleneck of today's communication networks is interference. The solution of commercial available networks is to avoid interference through orthogonalization of transmissions, that is, division of the resources, such as time, spectrum, space and waveforms, among the competing users. This approach is appealing in practice because it results in simple network architectures. It might also appear a good solution in theory since it is well known that allocating all the available channel resources to the user who experiences the instantaneous highest channel gain, is sum-rate optimal in fading MAC (Multiple Access Channels) [49] and fading BC (Broadcast Channels) [38]. Perfect orthogonalization of the users is however not possible in practice. The practical solution to deal with residual interference is to treat it as noise, as if its structure could not be exploited. This negative view of interference was further reinforced by scaling law results of the early 00's that showed that the throughput of wireless networks with K users only scales as $\sqrt{K}$, thus yielding a vanishing per-user rate as the network grows [25]. It was also conjectured that even with user cooperation a better scaling could be not achieved [30].

E. Yang and D. Tuninetti are with the Electrical and Computer Engineering Department of the University of Illinois at Chicago, Chicago, IL 60607, USA. Their work was partially funded by NSF under award number 0643954. The contents of this article are solely the responsibility of the authors and do not necessarily represent the official views of the NSF.

Recently, it has become apparent that interference avoidance and treating interference as noise are highly suboptimal in certain networks; instead interference should be *managed*. It has been known since the mid 70's that there exist channels where interference does not reduce capacity [7], [13], [51], [52]. These channels have "very strong interference", that is, the power imbalance between the useful signal and the interfering signal at a destination is large enough so that a receiver can first decode the interfering signal by treating its own signal as noise, and then decode its signal as in an interference-free channel. More recently, a technique called *interference alignment* has been shown to achieve a per-user rate of $\frac{1}{2}\log(1+\mathrm{SNR})$ [3]. Under the interference alignment paradigm, users do not avoid interference (even though they still treat it as noise), but instead make sure that the interference they collectively generate at a given destination is neatly confined in half of the signal space. This leaves the remaining half of the signal space interference-free.

Another promising interference management technique is *user cooperation*. Cooperation can be among sources, or among destinations, or among any node in the network. With cooperation, a node in the network helps the other nodes by relaying what it has received and detected; despite the fact that part of the resources of a node are devoted to the transmission of other sources' signal, cooperation has been shown to improve the achievable region of cellular network without increasing the transmit power [53].

In this paper we explore the rate advantages of *source cooperation* as interference management tool. We consider networks of full-duplex nodes, where several source-destination pairs share a common channel, and where all nodes can listen to the channel activity. The key observation is that interference due to simultaneous communications provides the basis for cooperation among otherwise uncoordinated nodes. In particular, a source can "overhear" what other sources are sending and act as a relay for them. At a very high level, a message/packet that has been spread throughout the network thanks to source cooperation can be routed to the intended destination though the best available channel/path thereby avoiding bottleneck links in the network.

In the rest of this section, we will revise relevant past work and summarize our main contributions.

A. Related Work

The signal a node can "overhear" from the channel is a form of feedback. To distinguish this feedback information from the classical Shannon output feedback [54], Willems referred to it as *Generalized Feedback* (GF) [64]. GF encompasses a wide range of possible situations, examples are: (1) *noisy output feedback* (with the non-feedback case and the output feedback case at the two extremes); (2) *conferencing encoders* (where there are separate noise-free and interference-free links between the sources, each link is of finite capacity); and (3) the case of GF with independent noises, sometimes referred to as *user cooperation*. Here, we consider the general case of GF. Hence our results can be specialized to all the above mentioned special cases. Our approach shows that all the above cases, which have been considered separately in the past, can be dealt with in great generality. One of our contributions is to show a unifying way of deriving some of the results available in the literature for different types of GF.

a) MAC-GF (Multiple Access Channels with Generalized Feedback): In [64], Willems derived an achievable region for a two-source MAC-GF. The two main ingredients are *regular block-Markov superposition coding* and *backward decoding*. In Willems' scheme, each source relays part of the information of the other source by using Decode-and-Forward [14], [15]. Willems' coding scheme for Gaussian channels was popularized by Sendonaris et al. [53] under the name of *user cooperation* in the context of cellular networks. In [53] it was shown that cooperation between users achieves collectively higher data rates or, alternatively, allows to reach the same data rates with less transmit power. Since the publication of [53], the interest in cooperative strategies has not ceased to increase (we do not attempt here to review all the extensions of [53] for sake of space).

b) Early work on IFC-GF: Although the MAC-GF has been instrumental in understanding the potential of user cooperation in networks, it is not a suitable model for (decentralized) ad-hoc networks, where the absence of a central receiver exacerbate the problem of interference. Host-Madsen [29] first extended the Gaussian MAC-GF model of [53] to the case of Gaussian IFC-GF. In [29], inner and outer bounds for the sum-rate were developed for source cooperation and for destination cooperation. It was shown that the multiplexing gain of a SISO Gaussian IFC-GF with two source-destination pairs is one for both forms of cooperation, instead of two as one would expect from a distributed 2×2 MIMO system. The work in [29] only dealt with Gaussian channels; it proposed several achievable strategies, each relatively simple and tailored to a specific set of channel gains. In [29], it was first noted that in the high SNR regime, nulling the interference (i.e., an ancestor of interference alignment) is asymptotically optimal.

The work in [29] was extended to a general DM (Discrete Memoryless) IFC-GF in [5], [31], [58]. In these works, the proposed achievable regions extend the idea of rate-splitting [8] of the original Han-Kobayashi scheme for the IFC without GF [27] and combine it with regular block Markov coding & backward decoding [64].

c) IFC-GF: Cooperation on sending the common information: In [58], we proposed to further split the common message in two parts: one part (referred to as *non-cooperative common information*) is as in the Han-Kobayashi scheme, while the other part (referred to as *cooperative common information*) is decoded at the other source too. The sources use a block-Markov encoding scheme where in a given slot they relay the cooperative common information learned in the previous slot. This cooperation strategy aims to realize the beam-forming gain of a distributed MISO system.

A scheme similar to ours in [58] was independently proposed by Jiang et al. in [31] for the IFC with output feedback (i.e., not for a general GF setting). The difference lies in the way the cooperative common information is dealt with, both at the sources and at the destination. It is not clear which scheme achieves the largest achievable region (we will elaborate more on this point in Section IV-B).

Recently, Prabhakaran and Viswanath [46] developed an outer bound for the sum-rate of a symmetric Gaussian IFC-GF with independent noises, by extending the semi-deterministic model of [57]; they showed that their upper bound is achievable to within 19 bits; when the cooperation link gains are smaller than the direct link gains, their achievable strategy is a simple form of our region in [58].

Cooperation on sending the common information is the essence of the first achievable region presented in this paper.

d) IFC-GF: Cooperation on sending both common and private information: In [5], Cao and Chen proposed to further split the private message (rather than the common message) into two parts. One part (which we shall refer to as *non-cooperative private information*) is as in the Han-Kobayashi scheme, while the other part (which we shall refer to as *cooperative private information*) is decoded at the other source too. The sources use a block-Markov encoding scheme where in a given slot they "hide" the cooperative private information learned in the previous slot from their intended receiver by using Gelfand-Pinsker coding [11], [23]. An approach similar to [5] (commonly referred to as "dirty paper coding" for Gaussian channels [11]) was already used in [29] to characterize the high SNR sum-rate capacity of the Gaussian IFC-GF.

In [6], the ideas of [5] and of [58] were merged into a scheme where both the common and the private information are split into two parts. In [69] we proposed a more structured coding strategy than that in [6], which generalizes and simplifies the description of the achievable region of [6]. In addition, we also added a binning step similar to Marton's achievable scheme for a general two-user broadcast channel [44]. Broadcast-type binning is possible in IFC-GFs because each encoder knows part of the message sent by the other encoder, i.e., each transmitter is partially cognitive in the sense of [41].

The second achievable region presented in this paper is a further enhancement of the region in [69].

e) Special cases of GF: The GF setting covers a wide range of situations, such as *degraded/noisy output feedback* and *conferencing encoders*.

In a Gaussian setting, degraded output feedback refers to the case where the GF signal received at a source is a noisier version of the signal received at the intended destination. If the variance of the extra noise on the GF signal is zero, the GF is simply referred to as *output feedback*. Kramer in [34], [35] developed inner and outer bounds for the Gaussian channel with output feedback. Kramer and Gastpar [22], and more recently Tandon and Ulukus [56], derived an outer bound for the degraded output feedback case based on the dependance balance idea of Hekstra and Willems [28], which is shown to be tighter than the cut-set bound in some parameter regimes. Suh and Tse [55] developed an outer bound for the Gaussian channel with output feedback by following the approach of [18], and showed its achievability to within 1.7 bits in the symmetric case. In [55], it was also shown that the achievable region of [31] is optimal for the case of deterministic channels with output feedback.

In an IFC with conferencing encoders, the sources can communicate through noise-free rate-limited channels. Wang and Tse [63] characterized the capacity region of the symmetric Gaussian conferencing encoder channel to within 6.5 bits. Vahid and Avestimehr [61] characterized the sum-rate capacity of the high-SNR linear deterministic approximation of the Gaussian channel in the symmetric case and showed that the sum-rate capacity can increase by at most the rate of the conferencing channels.

We will discuss how our second achievable region reduces to, and extends, known schemes for the case of degraded output feedback and for the case of conferencing encoders.

f) Other channel models: The IFC-GF reduces to some well known channel models under certain conditions, such as the Broadcast Channel (BC) [44], the MAC-GF [64], the Relay Channel (RC) [14], and the Cognitive IFC (C-IFC) [47]. *We will describe how our second achievable region encompasses known results for the channels subsumed by the IFC-GF.*

B. Summary of Contributions

In Part I of this paper, we present two achievable regions for a general IFC-GF. For the first region, sources cooperate on sending the common information only; in the second region, the sources cooperate on sending both the common and the private information. Although the first region is a special case of the second, the purpose of presenting cooperation on sending the common information first is to highlight some of the key elements of our coding scheme before proceeding to describe our more general scheme, which could be otherwise difficult to grasp because of the many auxiliary random variables involved. Our contributions are as follows:

1) We propose a structured way of performing superposition coding and binning that greatly reduces the number of error events to be considered in the error analysis. This result in a description of the achievable region with fewer rate-bounds than existing ones. The systematic method we developed for error analysis is of interest in its own.
2) We propose a binning scheme similar in spirit to multiple description source coding [62], where the different codebooks are jointly binned rather than sequentially binned.
3) We perform Fourier-Motzkin elimination to obtain a region with only five types of rate bounds, as in the case of IFC without feedback. This was not immediately obvious since, with GF, each message is split into four parts, and not in two as for the case without feedback.
4) We also show that when cooperation is on sending the common information only, the achievable rate region cannot be enlarged if the sources are required to decode more information than they will eventually "relay" to the destinations.
5) We show how our region does reduce to the best known achievable region based on Decode-and-Forward for the channels subsumed by the IFC-GF model.[1]
6) For the Gaussian channel, by means of a numerical example, we show that cooperation greatly increases the achievable rates with respect to the case without GF.

C. Paper Organization

The rest of the paper is organized as follows. Section II introduces the channel model and the notation. Section III revises known results for IFC without feedback. Section IV describes the achievable region where cooperation is on sending the common message only, and Section V describes the region where cooperation is on sending both the common and the private message. Section VI gives a numerical evaluation of proposed regions for the Gaussian channel. Section VII concludes Part I of this paper. All the proofs are in the Appendix.

II. Network Model and Definitions

A. Channel Model

Fig. 1 shows an IFC-GF with two source-destination pairs. It consists of a channel with two input alphabets $(\mathcal{X}_1, \mathcal{X}_2)$, four output alphabets $(\mathcal{Y}_1, \mathcal{Y}_2, \mathcal{Y}_3, \mathcal{Y}_4)$, and a transition probability $P_{Y_1\,Y_2\,Y_3\,Y_4|X_1\,X_2}$. We assume that the channel is memoryless and that all the alphabets have finite cardinality. The extension to continuous alphabets follows from standard arguments [16]. Source u, $u \in \{1, 2\}$, has a message W_u for destination u. The messages W_1 and W_2 are independent and uniformly distributed over the set $\{1, \cdots, \mathrm{e}^{n\,R_1}\}$ and $\{1, \cdots, \mathrm{e}^{n\,R_2}\}$, respectively, where n denotes the codeword length and R_u, $u \in \{1, 2\}$, the transmission rate for user u expressed in nats per channel use. At time t, $t \in \{1, \cdots, n\}$, source u, $u \in \{1, 2\}$, maps its message W_u and its past channel observations $Y_u^{t-1} \triangleq (Y_{u,1}, Y_{u,2}, \cdots, Y_{u,t-1})$ into a channel input symbol

$$X_{u,t} = f_{u,t}^{(n)}(W_u, Y_u^{t-1}),$$
$$f_{u,t}^{(n)} : \{1, \cdots, \mathrm{e}^{n\,R_u}\} \times \mathcal{Y}_u^{t-1} \to \mathcal{X}_u.$$

At time n, destination u, $u \in \{1, 2\}$, outputs an estimate of its intended message W_u based on all its channel observations Y_{u+2}^n, i.e.,

$$\widehat{W}_u = g_u^{(n)}(Y_{u+2}^n),$$
$$g_u^{(n)} : \mathcal{Y}_{u+2}^n \to \{1, \cdots, \mathrm{e}^{n\,R_u}\}.$$

[1]Our region however does not give the largest possible achievable rate for the relay channel because it does not include Compress-and-Forward [14]. It has been shown in [2] that the gap between Decode-and-Forward and the cut-set outer bound for the Gaussian relay channel with independent noises is less than one bit.

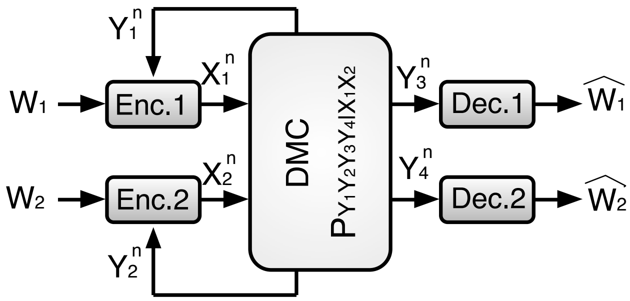

Fig. 1. The memoryless Interference Channel with Generalized Feedback (IFC-GF) with two source-destination pairs considered in this paper.

The capacity region is defined as the convex closure of all non-negative rate pairs (R_1, R_2) such that

$$\max_{u\in\{1,2\}} \Pr[\widehat{W}_u \neq W_u] \to 0 \quad \text{as} \quad n \to \infty.$$

B. Notation

We use standard notation [21], except for the mutual information between two random variables X and Y that we indicate with $I(X \wedge Y)$ rather than $I(X;Y)$ (to avoid possible confusion between colons and semicolons).

R_{xyz} indicates the information rate "from source x to destination y with the help of z", with $x \in \{1,2\}$, $y \in \{0,1,2\}$ and $z \in \{c,n\}$. In particular:

- $y = 0$: the message is decoded at both destinations (common message),
- $y = 1$: the message is decoded only at destination 1 (private message for user 1), and
- $y = 2$: the message is decoded only at destination 2 (private message for user 2);
- $z = c$: the message is sent cooperatively by both sources, and
- $z = n$: the message is sent non-cooperatively.

$T_\epsilon^{(n)}(P|\mathcal{S})$ indicates the set of length-n sequences that are strongly ϵ-typical with respect to the distribution P, conditioned on the sequences in $\mathcal{S}$ [17].

We use the following convention in the rest of the paper: a rate bound expressed as "$R_1 \leq$ (n)" means that R_1 is upper bounded by the expression on the RHS (right hand side) of equation number (n).

III. IFC WITHOUT FEEDBACK

In this section we briefly revise the best known inner bound for an IFC *without feedback*, i.e., $\mathcal{Y}_1 = \mathcal{Y}_2 = \emptyset$. This will make the exposition of our coding strategy easier to follow.

The capacity region of a general IFC without feedback is still unknown. The largest achievable region is due to Han and Kobayashi [27], whose "compact" expression appeared in [10]. In an IFC without generalized feedback, communications is as follows. Each transmitter splits its message in two parts: a *common message* and a *private message*. The two parts are superimposed and sent through the channel. Each receiver decodes its intended common and private message, as well as the common message of the other user, by treating the other user's private message as noise. The goal of decoding part of the non-intended massage is to reduce the interference level at a destination. The Han-Kobayashi scheme is optimal in strong interference [7], [12], [13], [36], [51], and it is shown to be sum-rate optimal in mixed interference [45], [60], in very weak interference [1], [45], [67], for the Z-IFC [50] (where only one receiver experiences interference), and for certain semi-deterministic channels [19], [57]. Moreover, a simple rate-splitting choice in the Han-Kobayashi scheme is optimal to within 1 bit for the Gaussian IFC [18].

Encoding and decoding in the Han and Kobayashi strategy are as follows:

Class of Input Distributions: Consider a distribution from the class:

$$P_{QU_1T_1X_1U_2T_2X_2Y_3Y_4} = P_Q P_{U_1T_1X_1|Q} P_{U_2T_2X_2|Q} P_{Y_3Y_4|X_1X_2}. \tag{1}$$

The channel transition probability $P_{Y_3Y_4|X_1X_2}$ is fixed, while the other factors in (1) can be varied.

Rate Splitting: The message $W_u \in \{1, ..., \mathrm{e}^{nR_u}\}$, $u \in \{1,2\}$, is split into two parts (W_{u0n}, W_{uun}): $W_{u0n} \in \{1, ..., \mathrm{e}^{nR_{u0n}}\}$ is the common information decoded at both receivers while $W_{uun} \in \{1, ..., \mathrm{e}^{nR_{uun}}\}$ is the private information decoded only at the intended receiver, with $R_u = R_{u0n} + R_{uun}$. Without feedback, all messages are sent non-cooperatively. In (1), the random variable Q serves as a time sharing random variable, U_u carries the common message of user u, and T_u carries the private message of user u, $u \in \{1,2\}$.

Coodbook Generation: Consider a distribution in (1). Pick uniformly at random a length-n sequence Q^n from the typical set $T_\epsilon^{(n)}(P_Q)$. For the codeword $Q^n = q^n$, pick uniformly at random $\mathrm{e}^{nR_{10n}}$ length-n sequences $U_1^n(k)$, $k \in \{1, \cdots, \mathrm{e}^{nR_{10n}}\}$, from the typical set $T_\epsilon^{(n)}(P_{U_1|Q}|q^n)$. For each pair $(Q^n, U_1^n(k)) = (q^n, u_1^n)$, pick uniformly at random $\mathrm{e}^{nR_{11n}}$ length-n sequences $T_1^n(m,k)$, $m \in \{1, \cdots, \mathrm{e}^{nR_{11n}}\}$, from the typical set $T_\epsilon^{(n)}(P_{T_1|QU_1}|q^n, u_1^n)$. For each triplet $(Q^n, U_1^n(k), T_1^n(m,k)) = (q^n, u_1^n, t_1^n)$, choose uniformly at random a sequence $X_1^n(m,k)$ from the typical set $T_\epsilon^{(n)}(P_{X_1|QU_1T_1}|q^n, u_1^n, t_1^n)$.

The generation of the codebooks at source 2 proceeds similarly.

Encoding: In order to send the message $W_u = (W_{u0n}, W_{uun})$, source u, $u \in \{1,2\}$, transmits $X_u^n(W_{u0n}, W_{uun})$.

Decoding: Destination 1 (and similarly for destination 2, but with the role of the users swapped) decodes the triplet $(W_{10n}, W_{20n}, W_{11n})$ from Y_3^n by searching for a unique pair (i_1, j_1), $j_1 \in \{1, ..., \mathrm{e}^{nR_{10n}}\}$ and $i_1 \in \{1, ..., \mathrm{e}^{nR_{11n}}\}$, and for some index j_2, $j_2 \in \{1, ..., \mathrm{e}^{nR_{20n}}\}$, such that

$$(U_1^n(j_1), T_1^n(i_1, j_1), U_2^n(j_2), Y_3^n) \in T_\epsilon^{(n)}(P^{(\mathrm{dec1})}_{U_1T_1U_2Y_3|Q}|Q^n),$$

where

$$P^{(\text{dec1})}_{U_1T_1U_2Y_3|Q} = \frac{\sum_{X_1,T_2,X_2} P_Q P_{U_1T_1X_1|Q} P_{U_2T_2X_2|Q} P_{Y_3|X_1X_2}}{P_Q}$$
$$= P_{U_1T_1|Q} P_{U_2|Q} \Big(\sum_{X_1,X_2} P_{X_1|QU_1T_1} P_{X_2|QU_2} P_{Y_3|X_1X_2} \Big).$$

If no such a pair (i_1, j_1) is found, or more than one pair is found, the receiver sets $(i_1, j_1) = (1, 1)$; in this case we say that an error at destination 1 has occurred.

Error Analysis: The error analysis can be found in [10]. The probability of error at destination 1 can be driven to zero if the rates $R_u = R_{u0n} + R_{uun}$, $u \in \{1, 2\}$, are such that

$$R_{11n} \le I(Y_3 \wedge T_1|U_1, U_2, Q) \quad (2a)$$
$$R_{20n} + R_{11n} \le I(Y_3 \wedge T_1, U_2|U_1, Q) \quad (2b)$$
$$R_{10n} + R_{11n} \le I(Y_3 \wedge T_1, U_1|U_2, Q) \quad (2c)$$
$$R_{20n} + R_{10n} + R_{11n} \le I(Y_3 \wedge T_1, U_1, U_2|Q), \quad (2d)$$

and similarly, the probability of error at destination 2 can be driven to zero if

$$R_{22n} \le I(Y_4 \wedge T_2|U_1, U_2, Q) \quad (3a)$$
$$R_{10n} + R_{22n} \le I(Y_4 \wedge T_2, U_1|U_2, Q) \quad (3b)$$
$$R_{20n} + R_{22n} \le I(Y_4 \wedge T_2, U_2|U_1, Q) \quad (3c)$$
$$R_{20n} + R_{10n} + R_{22n} \le I(Y_4 \wedge T_2, U_1, U_2|Q). \quad (3d)$$

Achievable Region: The region given by the intersection of (2) and (3) can be compactly expressed after Fourier-Motzkin elimination as:

Theorem III.1 ([10]). *For any distribution in (1) the following region is achievable:*

$$R_1 \le (2c), \quad (4a)$$
$$R_2 \le (3c), \quad (4b)$$
$$R_1 + R_2 \le \min\{(2d) + (3a),\ (2a) + (3d),\ (2b) + (3b)\}, \quad (4c)$$
$$2R_1 + R_2 \le (2d) + (2a) + (3b), \quad (4d)$$
$$R_1 + 2R_2 \le (2b) + (3a) + (3d). \quad (4e)$$

Without loss of generality, one can set $T_1 = X_1$ and $T_2 = X_2$ in (4), and, by Caratheodory's theorem, choose the auxiliary random variables (Q, U_1, U_2) from alphabets with cardinality $|\mathcal{Q}| \le 7$, $|\mathcal{U}_1| \le |\mathcal{X}_1| + 4$ and $|\mathcal{U}_2| \le |\mathcal{X}_2| + 4$. ■

IV. Superposition-only achievable region: cooperation on sending the common message only

In this section we propose a scheme where the sources cooperate by "beam-forming" part of the common messages to the destinations. In this scheme, all messages are superimposed and thus we refer to it as *superposition-only* achievable region. In Section V, we extend this scheme so as to incorporate cooperation on sending part of the private messages too by using binning/dirty paper coding. We refer to the latter scheme as *superposition & binning* achievable region. The *superposition & binning* achievable region includes the *superposition-only* achievable region as a special case. The details of the proof are given for the second scheme only. We however present the superposition-only achievable scheme first to guide the reader through the different definitions, techniques, and auxiliary random variables of the coding scheme.

We start by describing the coding scheme and then show how the proposed region reduces to known achievable regions.

A. Communication scheme

Class of Input Distributions: Consider a distribution from the class

$$P_{QV_1U_1T_1X_1V_2U_2T_2X_2Y_1Y_2Y_3Y_4} = P_Q P_{V_1U_1T_1X_1|Q} P_{V_2U_2T_2X_2|Q} P_{Y_1Y_2Y_3Y_4|X_1X_2}, \quad (5)$$

that is, after conditioning on Q, the random variables (V_1, U_1, T_1, X_1) generated at source 1 are independent of the random variables (V_2, U_2, T_2, X_2) generated at source 2. The channel transition probability $P_{Y_1Y_2Y_3Y_4|X_1X_2}$ is fixed, while the other factors in the distribution in (5) can be varied.

Rate Splitting and Transmission Strategy: The message $W_u \in \{1, ..., e^{nR_u}\}$, $u \in \{1, 2\}$, is divided into three parts $(W_{u0c}, W_{u0n}, W_{uun})$: $W_{u0c} \in \{1, ..., e^{nR_{u0c}}\}$ is the part of the common message sent cooperatively by the sources, $W_{u0n} \in \{1, ..., e^{nR_{u0n}}\}$ is the part of the common message sent by source u alone (i.e., non-cooperatively), and $W_{uun} \in \{1, ..., e^{nR_{uun}}\}$ is the private message sent non-cooperatively. The total rate is $R_u = R_{u0c} + R_{u0n} + R_{uun}$.

The proposed scheme differs from the classical Han and Kobayashi scheme in that there is a new message: the *cooperative common message*. We propose that the cooperative common message sent by a source is decoded at the other source thanks to the generalized feedback. Then, the sources send both cooperative common messages to the receivers as in a virtual (decentralized) MIMO channel, thus realizing the gain of beam-forming. This is possible by using regular block Markov superposition encoding [15] at the sources and backward decoding [64] at the destinations. In particular, transmission occurs over a frame of N slots of n channel uses each. Source 1 in slot b, $b \in \{1, \cdots, N\}$, has an estimate $W'_{20c,b-1}$ of the cooperative common message sent by source 2 in the previous slot. Similarly, source 2 in slot b has an estimate $W''_{10c,b-1}$ of the cooperative common message sent by source 1 in the previous slot. The random variable Q in (5) conveys the two cooperative common messages (W_{10c}, W_{20c}) from the previous time slot to the destinations. [2] In slot b, $b \in \{1, \cdots, N\}$, the random variables V_u, U_u, and T_u in (5), $u \in \{1, 2\}$, convey the new cooperative common message $W_{u0c,b}$, the new non-cooperative common message $W_{u0n,b}$, and the new non-cooperative private message $W_{uun,b}$, respectively. By convention $W_{u0c,0} = W_{u0c,N} = 1$, $u \in \{1, 2\}$.

Codebook Generation: Pick uniformly at random $e^{n(R_{10c}+R_{20c})}$ length-n sequences $Q^n([i, j])$, $i \in \{1, \cdots, e^{nR_{10c}}\}$ and $j \in \{1, \cdots, e^{nR_{20c}}\}$, from the

[2] In the GF case, Q is not just a simple time-sharing random variable as in the non-feedback case. Here in fact Q carries two messages.

typical set $T_\epsilon^{(n)}(P_Q)$. For each $Q^n([i,j]) = q^n$, pick uniformly at random $e^{nR_{10c}}$ length-n sequences $V_1^n(k,[i,j])$, $k \in \{1,\cdots,e^{nR_{10c}}\}$, from the typical set $T_\epsilon^{(n)}(P_{V_1|Q}|q^n)$. For each pair $(Q^n([i,j]), V_1^n(k,[i,j])) = (q^n, v_1^n)$, pick uniformly at random $e^{nR_{10n}}$ length-n sequences $U_1^n(\ell,k,[i,j])$, $\ell \in \{1,\cdots,e^{nR_{10n}}\}$, from the typical set $T_\epsilon^{(n)}(P_{U_1|QV_1}|q^n,v_1^n)$. For each triplet $(Q^n([i,j]), V_1^n(k,[i,j]), U_1^n(\ell,k,[i,j])) = (q^n,v_1^n,u_1^n)$, pick uniformly at random $e^{nR_{11n}}$ length-n sequences $T_1^n(m,\ell,k,[i,j])$, $m \in \{1,\cdots,e^{nR_{11n}}\}$, from the typical set $T_\epsilon^{(n)}(P_{T_1|QV_1U_1}|q^n,v_1^n,u_1^n)$. For each quadruplet $(Q^n([i,j]), V_1^n(k,[i,j]), U_1^n(\ell,k,[i,j]), T_1^n(m,\ell,k,[i,j])) = (q^n,v_1^n,u_1^n,t_1^n)$, pick uniformly at random one sequence $X_1^n(m,\ell,k,[i,j])$ from the typical set $T_\epsilon^{(n)}(P_{X_1|QV_1U_1T_1}|q^n,v_1^n,u_1^n,t_1^n)$.

The generation of the codebooks at source 2 is similar.

Encoding: In slot b, $b \in \{1,\cdots,N\}$, given the new message $W_{u,b} = (W_{u0c,b}, W_{u0n,b}, W_{uun,b})$ for $u \in \{1,2\}$, the transmitted codewords are

$$X_1^n(W_{11n,b}, W_{10n,b}, W_{10c,b}, [W_{10c,b-1}, W'_{20c,b-1}]),$$
$$X_2^n(W_{22n,b}, W_{20n,b}, W_{20c,b}, [W''_{10c,b-1}, W_{20c,b-1}]),$$

with the "boundary" conditions $W_{u0c,0} = W_{u0c,N} = 1$, $u = \{1,2\}$, i.e., on the first slot of the frame there is no cooperative common information from a previous slot to relay, and on the last slot of the frame there is no new cooperative common information to send because there will not be a future slot to relay it. With this scheme, user u, $u \in \{1,2\}$ transmits at an actual rate $R'_u = (1-1/N)R_{u0c} + R_{u0n} + R_{uun} < R_u$ (because no new cooperative common information is sent on the last slot). The rate R'_u can be made arbitrarily close to R_u by taking the frame length N to be sufficiently large.

Fig. 2 visualizes the proposed superposition-only coding scheme: an arrow to a codebook/random variable indicates that the codebook is superimposed to *all* the codebooks that precede it. Codebooks linked by a vertical line are conditionally independent given *everything* that precedes them; for example, codebook U_1 is superimposed to Q and V_1, and it is conditionally independent of any codebook with index 2 when conditioned on Q. In the block-Markov encoding scheme, the codebook Q carries the old (i.e., from the previous slot) cooperative messages, the codebooks V_1 and V_2 carry the new (i.e., from the current slot) cooperative messages, and the codebooks U_1, U_2, T_1 and T_2 carry the new (i.e., from the current slot) non-cooperative messages.

Cooperation: In slot b, $b \in \{1,\cdots,N-1\}$, we can assume that the users' estimate of the cooperative common messages from the previous slot is exact, that is, $W''_{10c,b-1} = W_{10c,b-1}$ and $W'_{20c,b-1} = W_{20c,b-1}$; this is true with arbitrarily high probability if the rates are chosen appropriately (as explained later). Next we describe how source 2 cooperates with source 1. Source 1 proceeds similarly.

Source 2 at the end of slot b decodes user 1's new cooperative common message $W_{10c,b}$ carried by V_1^n from its channel output $Y_{2,b}^n$, knowing everything that was generated at source 2 at the beginning of the slot. Formally, at the end of slot b,

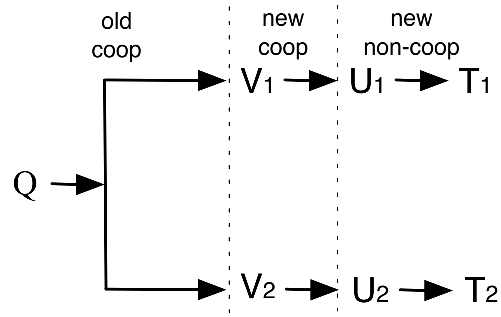

Fig. 2. A visual representation of the codebook generation for the superposition-only achievable scheme with input distribution described by (5).

$b \in \{1,\cdots,N-1\}$, source 2 has received $Y_{2,b}^n$ and looks for a unique index $i \in \{1,...,e^{nR_{10c}}\}$ such that

$$(V_1^n(i,[\cdots]), Y_{2,b}^n) \in T_\epsilon^{(n)}(P^{(\text{enc2})}_{V_1Y_2|QX_2}|\underline{X}_2^n),$$

where the dots indicate known message indices, where all that is known at source 2 is represented by

$$\underline{X}_2 = (Q, V_2, U_2, T_2, X_2),$$

and where

$$P^{(\text{enc2})}_{V_1Y_2|QX_2} = \frac{\sum_{U_1,T_1,X_1} P_Q P_{V_1U_1T_1X_1|Q} P_{V_2U_2T_2X_2|Q} P_{Y_2|X_1X_2}}{P_{QV_2U_2T_2X_2}}$$
$$= P_{V_1|Q}\Big(\sum_{X_1} P_{X_1|QV_1} P_{Y_2|X_1X_2}\Big).$$

If none or more than one such an index i is found, then source 2 sets $i = 1$; in this case we say that an error has occurred at source 2.

Error Analysis: By standard arguments, the probability of error at source 2 can be made as small as desired if

$$R_{10c} \le I(V_1 \wedge Y_2|Q, V_2, U_2, T_2, X_2). \tag{6a}$$

Decoding: The destinations wait until the last slot of the frame (i.e. slot N) has been received and then perform backward decoding. In slot $b \in \{N,\cdots,1\}$ destination 1 looks for the unique triplet $(i_1,j_1,m_1) \in \{1,...,e^{nR_{10c}}\} \times \{1,...,e^{nR_{10n}}\} \times \{1,...,e^{nR_{11n}}\}$ and for some pair $(i_2,j_2) \in \{1,...,e^{nR_{20c}}\} \times \{1,...,e^{nR_{20n}}\}$ such that

$$\Big(Q^n([i_1,i_2]), V_1^n(W'_{10c,b},[i_1,i_2]), U_1^n(j_1,W'_{10c,b},[i_1,i_2]),$$
$$T_1^n(m_1,j_1,W'_{10c,b},[i_1,i_2]),$$
$$V_2^n(W'_{20c,b},[i_1,i_2]), U_2^n(j_2,W'_{20c,b},[i_1,i_2]),$$
$$Y_{3,b}^n\Big) \in T_\epsilon^{(n)}(P^{(\text{dec1})}_{QV_1U_1T_1V_2U_2Y_3}),$$

where the pair $(W'_{10c,b}, W'_{20c,b})$ was decoded in the previous step ($W_{u0c,N} = 1$ by assumption, hence $W'_{u0c,N} = 1$ too) and where

$$P^{(\text{dec1})}_{QV_1U_1T_1V_2U_2Y_3}$$
$$= \sum_{X_1,T_2,X_2} P_Q P_{V_1U_1T_1X_1|Q} P_{V_2U_2T_2X_2|Q} P_{Y_3|X_1X_2}$$
$$= P_Q P_{V_1U_1T_1|Q} P_{V_2U_2|Q} \Big(\sum_{X_1,X_2} P_{X_1|QV_1U_1T_1} P_{X_2|QV_2U_2} P_{Y_3|X_1X_2} \Big).$$

If none or more than one such a triplet (i_1, j_1, m_1) is found, then destination 1 sets $(i_1, j_1, m_1) = (1, 1, 1)$; in this case we say that an error has occurred at destination 1.

In words, destination 1 decodes the old cooperative common messages in Q^n, the current non-cooperative common messages in (U_1^n, U_2^n), and the current non-cooperative private message in T_1^n, from its channel output $Y_{3,b}^n$. The current cooperative common messages in (V_1^n, V_2^n) are known from the decoding of slot $b+1$. The current non-cooperative private message of user 2 in T_2^n is treated as noise.

Error Analysis: By standard arguments, the probability of error at destination 1 can be made as small as desired if

$$R_{11n} \le I(Y_3 \wedge T_1|Q, V_1, V_2, U_1, U_2) \quad (6b)$$
$$R_{11n} + R_{20n} \le I(Y_3 \wedge T_1, U_2|Q, V_1, V_2, U_1) \quad (6c)$$
$$R_{11n} + R_{10n} \le I(Y_3 \wedge T_1, U_1|Q, V_1, V_2, U_2) \quad (6d)$$
$$R_{11n} + R_{10n} + R_{20n} \le I(Y_3 \wedge T_1, U_1, U_2|Q, V_1, V_2) \quad (6e)$$
$$R_{11n} + R_{10n} + R_{20n} + (R_{20c} + R_{10c}) \le I(Y_3 \wedge T_1, U_1, U_2, Q, V_1, V_2). \quad (6f)$$

Notice that

$$(6b) \le \min\{(6c), (6d)\} \le \max\{(6c), (6d)\} \le (6e) \le (6f).$$

By similar arguments, the probability of error at source 1 can be made as small as desired if

$$R_{20c} \le I(V_2 \wedge Y_1|Q, V_1, U_1, T_1, X_1), \quad (7a)$$

and the probability of error at destination 2 can be made as small as desired if

$$R_{22n} \le I(Y_4 \wedge T_2|Q, V_2, V_1, U_2, U_1) \quad (7b)$$
$$R_{22n} + R_{10n} \le I(Y_4 \wedge T_2, U_1|Q, V_2, V_1, U_2) \quad (7c)$$
$$R_{20n} + R_{22n} \le I(Y_4 \wedge T_2, U_2|Q, V_2, V_1, U_1) \quad (7d)$$
$$R_{20n} + R_{22n} + R_{10n} \le I(Y_4 \wedge T_2, U_2, U_1|Q, V_1, V_2) \quad (7e)$$
$$R_{22n} + R_{20n} + R_{10n} + (R_{20c} + R_{10c}) \le I(Y_4 \wedge T_2, U_1, U_2, Q, V_1, V_2). \quad (7f)$$

Achievable region: The intersection of the region in (6) with the region in (7) can be compactly expressed after Fourier-Motzkin elimination as follows:

Theorem IV.1. *For any distribution in (5), the following region is achievable:*

$$R_1 \le (6f) \quad (8a)$$
$$R_1 \le (6a) + (6d) \quad (8b)$$
$$R_2 \le (7f) \quad (8c)$$
$$R_2 \le (7a) + (7d) \quad (8d)$$
$$R_1 + R_2 \le (6f) + (7b) \quad (8e)$$
$$R_1 + R_2 \le (6b) + (7f) \quad (8f)$$
$$R_1 + R_2 \le (6a) + (7a) + (6e) + (7b) \quad (8g)$$
$$R_1 + R_2 \le (6a) + (7a) + (6b) + (7e) \quad (8h)$$
$$R_1 + R_2 \le (6a) + (7a) + (6c) + (7c) \quad (8i)$$
$$2R_1 + R_2 \le (6a) + (6b) + (6f) + (7c) \quad (8j)$$
$$2R_1 + R_2 \le 2\cdot(6a) + (7a) + (6b) + (6e) + (7c) \quad (8k)$$
$$R_1 + 2R_2 \le (7a) + (6c) + (7b) + (7f) \quad (8l)$$
$$R_1 + 2R_2 \le (6a) + 2\cdot(7a) + (6c) + (7b) + (7e). \quad (8m)$$

Without loss of generality, one can take $X_1 = T_1$ and $X_2 = T_2$ in (8). ■

Remark IV.2. *The proposed structured way of superimposing the codebooks in the superposition-only achievable region greatly simplifies the error analysis. Our codebook "nesting," in fact, is such that the "cloud center" codebooks (Q, V_1, V_2) are the one all terminals decode, i.e., the cooperative common codebook, to which we superimpose the non-cooperative common codebook U (to be decoded by the two destinations but not by the other source) and finally we superimpose the non-cooperative private codebook T (to be decoded at the intended receiver only). As a consequence, although a destination has to decode five messages, only 5 out of the possible $2^5 - 1 = 31$ error events matter (see (6) for destination 1 and (7) for destination 2). The region in (6) (resp. (7)) has one more rate constraint than the classical IFC without feedback region in (2) (resp. (3)).*

Remark IV.3. *The following rate constraints also appear after Fourier-Motzkin elimination:*

$$R_1 \le (6a) + (6b) + (7c), \quad (9)$$
$$R_2 \le (7a) + (7b) + (6c). \quad (10)$$

It can be shown (see Appendix A) that (9) and (10) can be removed without enlarging the achievable region. The intuitive argument is as follows. The constraint in (9) (and similarly for (10) but with the role of the users swapped) implies that the rate of user 1 is limited by the "quality" of user 2's outputs (because the mutual information in (7c) depends on Y_4). This should not be the case since user 2 is not interested in the message of user 1. Hence, when (9) is the most stringent R_1-constraint, destination 2 should not be required to decode the common information from user 1; this is equivalent to setting $U_1 = V_1 = \emptyset$. One can show that the rate points that would violate the R_1-constraint in (9) are actually included in another achievable region with $U_1 = V_1 = \emptyset$; thus, the constraints in (9) and in (10) do not limit the achievable region and can be neglected.

Remark IV.4. *In [46] it is shown that the region in Th. IV.1 is sum-rate optimal to within 19 bits for Gaussian channels with independent noises and symmetric cooperation links when the gain of the cooperation links are smaller than the gain of the interfering links.*

In the same work, it is shown that when the gain of the cooperation links are larger than the gain of the interfering links, the transmitters should decode more information from their received generalized feedback signal than they will actually use for cooperation. In our setting this amounts to: at the end of slot b, $b \in \{1, \cdots, N-1\}$*, source 1 looks for a unique index* $i \in \{1, ..., \mathrm{e}^{nR_{20c}}\}$ *and some index* $j \in \{1, ..., \mathrm{e}^{nR_{20n}}\}$ *such that the sequences*

$$\begin{aligned}&(V_2^n(i, [\cdots]), U_2^n(j, i, [\cdots]), Y_{1,b}^n)\\&\in T_\epsilon^{(n)}(P^{(\mathrm{enc1})}_{V_2U_2Y_1|Q,X_1}|\underline{X}_1^n),\end{aligned}$$

where the dots indicate known message indices, where all that is known at source 2 is represented by

$$\underline{X}_1 = (Q, V_1, U_1, T_1, X_1),$$

and where

$$P^{(\mathrm{enc1})}_{V_2U_2Y_1|QX_1} = P_{V_2U_2|Q}\Big(\sum_{X_2} P_{X_2|QV_2U_2}P_{Y_1|X_1X_2}\Big).$$

By standard arguments, decoding is successful with arbitrarily high probability if

$$R_{20n} + R_{20c} \le I(V_2, U_2 \wedge Y_1|Q, V_1, U_1, T_1, X_1). \quad (11)$$

Notice that the constraint in (11) allows $R_{20c} \le I(V_2, U_2 \wedge Y_1|Q, V_1, U_1, T_1, X_1)$ *while the constraint in (7a) only allowed for* $R_{20c} \le I(V_2 \wedge Y_1|Q, V_1, U_1, T_1, X_1)$*. However, the constraint in (11) constrains* R_{20n} *to satisfy* $R_{20n} \le I(V_2, U_2 \wedge Y_1|Q, V_1, U_1, T_1, X_1) - R_{20c}$ *while the constraint in (7a) does not constrain* R_{20n} *at all. It is not clear at priori which cooperation strategy is better, i.e., whether (7a) or (11). We will show next – as intuition suggests – that the two are equivalent, that is, with superposition-only, the sources should relay to the destinations all the common information they have acquired through the generalized feedback.*

Corollary IV.5. *After Fourier-Motzkin elimination of the intersection of the regions in (6) and (7) with (7a) replaced by (11), we get*

$$R_1 \le \min\{(6\mathrm{f}), (6\mathrm{a}) + (6\mathrm{d})\} \quad (12\mathrm{a})$$

$$R_2 \le \min\{(7\mathrm{f}), (11) + (7\mathrm{b})\} \quad (12\mathrm{b})$$

$$\begin{aligned}R_1 + R_2 \le \min\{&(6\mathrm{f}) + (7\mathrm{b}), (6\mathrm{b}) + (7\mathrm{f}),\\&(6\mathrm{a}) + (11) + (6\mathrm{b}) + (7\mathrm{c})\}\end{aligned} \quad (12\mathrm{c})$$

$$2R_1 + R_2 \le (6\mathrm{a}) + (6\mathrm{b}) + (6\mathrm{f}) + (7\mathrm{c}). \quad (12\mathrm{d})$$

It can be easily verified that the region in (12) is the same as the region in (8) computed for $P_{QV_1U_1T_1X_1V_2'U_2'T_2X_2}$ *where* $U_2' = \emptyset$ *and* $V_2' = (V_2, U_2)$*, i.e., with* $R_{20n} = 0$*; this is so because the achievable region is a function of* (V_2, U_2) *only. Hence, requiring source 1 to decode more information than actually used for cooperation does not enlarge the superposition-only achievable region in (8).* ■

It can be easily seen that the same conclusion of Corollary IV.5 holds when both sources are required to decode more information than actually used for cooperation (in this case the achievable region is a function of (V_1, U_1) *and of* (V_2, U_2) *only).*

Along the same line of reasoning, one could also require that at the end of slot b, $b \in \{1, \cdots, N-1\}$*, source 1 looks for a unique index* $i \in \{1, ..., \mathrm{e}^{nR_{20c}}\}$*, and for some pair of indices* $j \in \{1, ..., \mathrm{e}^{nR_{20n}}\}$ *and* $k \in \{1, ..., \mathrm{e}^{nR_{22n}}\}$*, such that the sequences*

$$\begin{aligned}&(V_2^n(i, [\cdots]), U_2^n(j, i, [\cdots]), T_2^n(k, j, i, [\cdots]), Y_{1,b}^n)\\&\in T_\epsilon^{(n)}(P^{(\mathrm{enc1})}_{V_2U_2T_2Y_1|QX_1}|\underline{X}_1^n),\end{aligned}$$

where

$$P^{(\mathrm{enc1})}_{V_2U_2T_2Y_1|QX_1} = P_{V_2U_2T_2|Q}\Big(\sum_{X_2} P_{X_2|QV_2U_2T_2}P_{Y_1|X_1X_2}\Big).$$

By standard arguments, decoding is successful with arbitrarily high probability if

$$\begin{aligned}R_2 &= R_{22n} + R_{20n} + R_{20c}\\&\le I(T_2, U_2, V_2 \wedge Y_1|Q, V_1, U_1, T_1, X_1).\end{aligned} \quad (13)$$

The constraint in (13) can be too restrictive for rate R_2 *when the cooperation link is weak. Indeed, consider the extreme case where* Y_1 *is independent of everything else (i.e., unilateral cooperation), then the constraint in (13) implies* $R_2 = 0$*, which can be easily beaten by ignoring the generalized feedback.*

B. Comparison with existing regions

The achievable region in (8) subsumes achievable regions for other multiuser channels. For example:

1) IFC without generalized feedback: By setting $V_1 = V_2 = \emptyset$, i.e., no cooperative common messages, our proposed encoding scheme reduces to the Han and Kobayashi region for the **IFC without feedback** in Th. III.1. In fact, (8e) is redundant because of (8g); (8f) is redundant because of (8h); (8j) is redundant because of (8k); and (8l) is redundant because of (8m).

2) IFC with output feedback: By setting $Y_1 = Y_3$ and $Y_2 = Y_4$ we obtain an **IFC with output feedback** as studied in [31], [34], [55].

The region in [31, eq.(18)-(29)] has the same codebook structure of our Th. IV.1. However, the scheme of [31] differs in the encoding and decoding of the messages in Q. In our achievable region, the sources repeat in Q the whole past cooperative common messages. In [31], the sources repeat in Q a quantized version of the past cooperative common message indices. In principle, the encoding in [31] is more general. However, in our achievable region, the destinations decode Q jointly with all other messages. In [31], Q is decoded first by treating all the rest as noise, and then all the other messages are jointly decoded. It is thus not clear a priori whether the rate-saving due to sending a quantized version of Q are wiped out by the two-step decoding (which constraints the rate of Q). A formal comparison among the two regions is however difficult, since the two could appear different for some input distribution but be actually the same when the union over all possible input distributions is taken.

The region studied by Suh and Tse [55], which is optimal for the deterministic channel with output feedback and optimal to within 1.7 bits for the symmetric Gaussian channel with output feedback, is a subset of our superposition-only achievable region obtained by setting $U_u = V_u$, $u \in \{1, 2\}$, i.e., the whole common message is sent cooperatively.

3) IFC with unilateral cooperation: By setting $(6a) = \infty$, and $(7a) = 0$ in (8), we obtained a **cognitive interference channel**, in which source 2 knows the message of source 1, while source 1 is unaware of the messages sent by source 2. The cognitive channel has also been referred to as **IFC with degraded message set** [66] and as **IFC with unidirectional cooperation** [42]. The condition $(7a) = 0$ is equivalent to $\mathcal{Y}_1 = \emptyset$, that is, source 1 cannot cooperate with source 2 because it does not have any channel observation. The condition $(6a) = \infty$ is equivalent to assume that the cooperation channel from source 1 to source 2 is never a bottle-neck in the network; this is the case for example if Y_2 is a deterministic and invertible channel [19] with $|\mathcal{Y}_2| = \mathrm{e}^{|\mathcal{Y}_3|}$, i.e., source 2 can decode from a single channel use more information than can actually be decoded at destination 1 from the whole codeword.

After setting $(6a) = \infty$, and $(7a) = 0$ in (8), we notice that the auxiliary random variables (Q, V_1, U_1, V_2) always appears together, thus without loss of generality, we can set $U_1 = V_1 = V_2 = Q$, in which case the region in (8) reduces to:

$$R_1 \le (6f) \quad (14a)$$
$$R_2 \le (7d) \quad (14b)$$
$$R_1 + R_2 \le (6f) + (7b) \quad (14c)$$
$$R_1 + R_2 \le (6b) + (7f) \quad (14d)$$
$$R_1 + 2R_2 \le (6c) + (7b) + (7f). \quad (14e)$$

The region in (14) is only a subset of the best known achievable region for a cognitive IFC by Rini *et at* [47]. The region of [47] includes binning, which was shown to be optimal for cognitive IFC with "weak interference" [66]. However, for a cognitive IFC with "very strong interference," superposition-only is optimal [42].

4) Broadcast channel: By setting $(6a) = (7a) = \infty$ (and with $U_1 = V_1 = U_2 = V_2 = Q$ without loss of generality) in (8), we obtain the achievable region with superposition-only for a **broadcast channel** (BC), namely

$$R_1 \le (6f) \quad (15a)$$
$$R_2 \le (7f) \quad (15b)$$
$$R_1 + R_2 \le (6f) + (7b) \quad (15c)$$
$$R_1 + R_2 \le (6b) + (7f). \quad (15d)$$

The region in (15) is only a subset of the largest known achievable region for a general BC by Marton [44] because the region in (8) does not use binning. The region in (15) is however optimal for the case of "BC with degraded message set" [33] and for the case of "more capable BC channels" [20] (the class of more capable BCs includes less noisy BCs and degraded BCs as special cases).

5) Multiple access channel with GF: By setting $Y_3 = Y_4 = Y$ the IFC-GF reduces to a **multiple access channel with GF** [64]. In this case we can set without loss of generality $T_1 = T_2 = \emptyset$; indeed, in a MAC all messages are to be decoded at the central receiver and thus all messages are common. With $Y_3 = Y_4 = Y$ and $T_1 = T_2 = \emptyset$ we have (6b) = (7b) = 0, (6d) = (7c), (6c) = (7d), (6e) = (7e), and (6f) = (7f), and thus the region in (8) reduces to

$$R_1 \le (6a) + (6d) \quad (16a)$$
$$R_2 \le (7a) + (6c) \quad (16b)$$
$$R_1 + R_2 \le (6f) \quad (16c)$$
$$R_1 + R_2 \le (6a) + (7a) + (6e). \quad (16d)$$

The region in (16) was first as derived in [64].

6) Multiple access channel with GF with a common/broadcast message: The region in (16) is for a MAC-GF "without common/broadcast message". The case **with common/broadcast message**, that is, with a message W_0 available at both sources and to be decoded at both destinations, can be easily incorporated by having the codebook Q also carry the common message W_0. If the rate of the common/broadcast message is R_0, the region in (16) must be modified as follows: the rate constraint (16c) becomes

$$R_0 + R_1 + R_2 \le (6f).$$

The capacity region for a MAC without feedback and with arbitrary sets of common/broadcast messages was derived in [24], [26]; the capacity is achieved by superposition-only. It is left for future work to quantify the rate improvement due to GF for MAC with common/broadcast messages.

7) IFC-GF with a common/broadcast message: The case of an **IFC-GF with common/broadcast message** can be readily obtained by using the trick described above for the MAC-GF with common/broadcast message, namely the codebook Q must carry the common/broadcast message in addition to the common cooperative messages from the previous slot. With a common/broadcast message, our Th. IV.1 must be modified as follows: the left hand side of the inequalities in (6f) and (7f) must also include the rate of the common/broadcast message.

An achievable region based on superposition-only coding for the IFC with common/broadcast message and without feedback was independently proposed in [4], [32], [40], [43]. It is left for future work to quantify the rate improvement due to GF for IFC with common/broadcast message.

8) Relay channel: Further setting $R_2 = 0$ in (16) gives

$$R_1 \le \min\{(6a) + (6d), (6f)\}, \quad (17)$$

which is the achievable rate for a **full-duplex relay channel** with partial decode-and-forward. The rate in (17) is not the largest known achievable rate for the relay channel [9], [14], which is obtained by combining decode-and-forward with compress-and-forward.

9) Multiple access channels with conferencing encoders: Channels with **conferencing encoders** are also obtained by a special choice of GF. A two-source conferencing model [64] assumes that there are two non-interfering, noise-free channels of finite capacity between the communicating source nodes, one for each direction of communication. Let C_{ij} be the capacity of the conferencing channel from node j to node i. Following the approach of [37] and with some abuse of notation, the conferencing model is captured as follows: Let the inputs be $\mathbf{X}_1 = [F_1; X_1]$ and $\mathbf{X}_2 = [F_2; X_2]$, where F_1 and F_2 are from alphabets of cardinality $\log(C_{12})$ and $\log(C_{21})$, respectively. Further set the GF signals to be $Y_1 = F_2$ and $Y_2 = F_1$, and define the channel transition probability to be

$$P_{Y_1,Y_2,Y_3,Y_4|[F_1;X_1],[F_2;X_2]} = P_{Y_3,Y_4|X_1,X_2}\, 1_{\{Y_1=F_2\}}\, 1_{\{Y_2=F_1\}},$$

where $1_{\{\mathcal{A}\}}$ is the indicator function that equals one whenever the event $\mathcal{A}$ is true and zero otherwise. In this model, the choice $V_1 = F_1$, $V_2 = F_2$, and V_1 and V_2 independent of everything else achieves (6a) $= C_{21}$ and (7a) $= C_{12}$.

For the case $Y_3 = Y_4 = Y$, the channel reduces to a **MAC with conferencing encoders**. Willems showed in [64] that the choice $V_1 = F_1$ and $V_2 = F_2$ is capacity achieving.

For the case $Y_3 \neq Y_4$, the channel is an **IFC with conferencing encoders**. This channel was studied in [61], [63]. The region [63] includes binning, thus it will be compared to our superposition & binning region in the next section.

V. Superposition & binning achievable region: cooperation on sending the common and the private messages

Because the achievable region with *superposition-only* does not reduce to the largest known achievable region when the IFC-GF reduces to a cognitive channel, a broadcast channel, or a relay channel, we now introduce binning in the superposition-only achievable scheme. In this new achievable scheme, the sources cooperate in sending part of the common messages and part of the private messages.

A. Communication scheme

Class of Input Distributions: Consider a distribution from the class

$$P_{QV_1U_1T_1S_1Z_1X_1V_2U_2T_2S_2Z_2X_2Y_1Y_2Y_3Y_4} = P_{Y_1Y_2Y_3Y_4|X_1X_2} P_{QS_1S_2}P_{V_1U_1T_1Z_1X_1|QS_1S_2}P_{V_2U_2T_2Z_2X_2|QS_1S_2}. \tag{18}$$

that is, conditioned on (Q, S_1, S_2), the random variables $(V_1, U_1, T_1, Z_1, X_1)$ generated at source 1 are independent of the random variables $(V_2, U_2, T_2, Z_2, X_2)$ generated at source 2. The channel transition probability $P_{Y_1Y_2Y_3Y_4|X_1X_2}$ is fixed, while the other factors in the distribution in (18) can be varied.

Rate Splitting and Transmission Strategy: Each message is divided into four parts: two common messages and two private messages. The sources cooperate in sending one part of the common messages and one part of the private messages. Communication again proceeds on a frame on N slots. At the end of any given slot, the sources decode the cooperative messages from their GF signal and use it in the next slot. Since the common messages are decoded at both destinations, the sources cooperate by relaying the cooperative common messages to the destinations as in a virtual MIMO channel, like for the superposition-only scheme. A private message however is decoded at the intended destination only and treated as noise at the other destination. In this case, a source treats the other source's cooperative private message as "non-causally known interference" in the next slot. Cooperation is then in the form of forwarding for the intended destination and of pre-coding/binning/dirty paper coding for the non-intended destination.

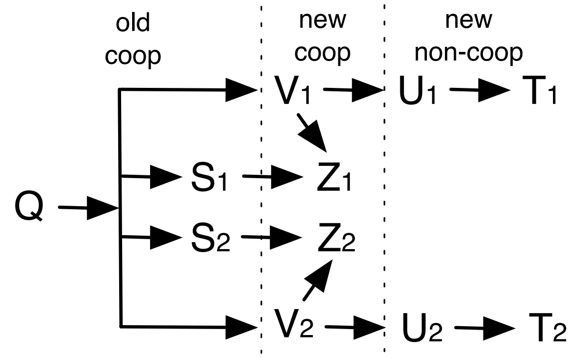

Fig. 3. A visual representation of the codebook generation for the superposition & binning achievable scheme.

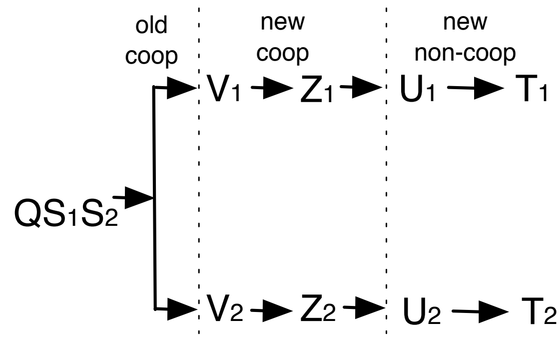

Fig. 4. A visual representation of the possible codebooks for the superposition & binning achievable scheme with input distribution described by (18).

In a block-Markov encoding scheme the random variable Q conveys the two cooperative common messages from the previous time slot; V_u, U_u, and T_u convey the new cooperative common message, the new non-cooperative common message, and the new non-cooperative private message, respectively, of the current slot; S_u and Z_u convey the previous and the new, respectively, cooperative private information of the current slot, $u = \{1, 2\}$.

Codebook Generation: From the input distribution in (18), compute the marginals $P_{V_1U_1T_1|Q}$ and $P_{V_2U_2T_2|Q}$ (i.e., drop the dependance on (S_1, S_2)), and construct codebooks Q^n, (V_1^n, U_1^n, T_1^n) and (V_2^n, U_2^n, T_2^n) as described in Section IV. The range of the indices for these codebooks will be specified later.

From the input distribution in (18), compute the marginals $P_{S_1|Q}$ (i.e., drop the dependance on S_2) and $P_{S_2|Q}$ (i.e., drop the dependance on S_1). Conditioned on each $Q^n([i,j]) = q^n$, pick uniformly at random length-n sequences $S_1^n(k_1,[i,j])$ from the typical set $T_\epsilon^{(n)}(P_{S_1|Q}|q^n)$. The range of the index k_1 will be specified later. Similarly, generate a codebook $S_2^n(k_2,[i,j])$ by picking uniformly at random length-n sequences from the typical set $T_\epsilon^{(n)}(P_{S_2|Q}|q^n)$. The range of the index k_2 will be specified later. From the input distribution in (18), compute the marginal $P_{Z_1|QS_1V_1}$. For each triplet $(Q^n([i,j]), S_1^n(k_1,[i,j]), V_1^n(\ell_1,[i,j])) = (q^n, s_1^n, v_1^n)$, generate a codebook $Z_1^n(m_1,\ell_1,k_1,[i,j])$ by picking uniformly at random length-n sequences from the typical set $T_\epsilon^{(n)}(P_{Z_1|QS_1V_1}|q^n,s_1^n,v_1^n)$. The range of the index m_1 will be specified later. Similarly, but with the role of the users swapped, construct a codebook $Z_2^n(m_2,\ell_2,k_2,[i,j])$. Finally, for each set of codewords $(Q^n,S_1^n,S_2^n,V_1^n,U_1^n,T_1^n,Z_1^n) = (q^n,s_1^n,s_2^n,v_1^n,u_1^n,t_1^n,z_1^n)$ pick uniformly at random one sequence X_1^n from the typical set $T_\epsilon^{(n)}(P_{X_1|QS_1S_2V_1U_1T_1Z_1}|q^n,s_1^n,s_2^n,v_1^n,u_1^n,t_1^n,z_1^n)$. Similarly, but with the role of the users swapped, construct a codebook X_2^n.

Fig. 3 visualizes the proposed codebook generation (the convention is the same as the one used for Fig. 2). The class of input distributions in (18) however allows for codebooks as depicted in Fig. 4. With binning/dirty paper coding we will force a codebook generated as in Fig. 3 to look like a codebook from Fig. 4.

In order to complete the codebook generation, we must specify the range of the indices. Each codebook carries a message W_{xyz} "from source x to destination y with the help of z" as explained in II-B. The index in the codebook that carries W_{xyz} is a pair of indices $[w_{xyz}, b_{xyz}]$, where $w_{xyz} \in \{1,\cdots,\mathrm{e}^{nR_{xyz}}\}$ is the "message index", and $b_{xyz} \in \{1,\cdots,\mathrm{e}^{nR'_{xyz}}\}$ is the "bin index". Notice that each bin index b_{xyz} has the same subscript as the corresponding message index w_{xyz} and has rate R'_{xyz} (i.e., notice the prime in the superscript). The only exceptions are: (a) S_1 and S_2: the rate of the bin index is indicated with a double prime superscript, that is, R''_{11c} for S_1 and R''_{22c} for S_2, so as not to confuse them with R'_{11c} for Z_1 and R'_{22c} for Z_2; and (b) Q: there is no bin index.

Encoding: We can assume correct decoding of the message indices at the sources at the end of slot $b-1$ (no error propagation), i.e.,

$$W''_{10c,b-1} = W_{10c,b-1} \text{ (carried by } V_1^n\text{; to be repeated in } Q^n),$$
$$W''_{11c,b-1} = W_{11c,b-1} \text{ (carried by } Z_1^n\text{; to be repeated in } S_1^n),$$
$$W'_{20c,b-1} = W_{20c,b-1} \text{ (carried by } V_2^n\text{; to be repeated in } Q^n),$$
$$W'_{22c,b-1} = W_{22c,b-1} \text{ (carried by } Z_2^n\text{; to be repeated in } S_2^n),$$

since the total average probability of error can be upper bounded by the sum of the decoding error probabilities at each step, under the assumption that no error propagation from the previous steps has occurred [10], [70].

The encoding process at the beginning of slot b consists of the following binning and superposition steps. The purpose of the binning steps is to allow the most general possible class of input distributions.

- *Binning codebooks S_1^n and S_2^n against each other.* At the beginning of slot b, given the past messages $(W_{10c,b-1}, W_{20c,b-1}, W_{11c,b-1}, W_{22c,b-1})$, source 1 tries to find a pair $(b_{11c,b-1}, b_{22c,b-1})$ such that

$$\Big(S_1^n([W_{11c,b-1}, b_{11c,b-1}],[\cdots]), S_2^n([W_{22c,b-1}, b_{22c,b-1}],[\cdots])\Big) \in T_\epsilon^n(P_{S_1S_2|Q}|Q^n),$$

where the dots are in place of the known messages $[W_{10c,b-1}, W_{20c,b-1}]$ from the previous slot. If more than one pair $(b_{11c,b-1}, b_{22c,b-1})$ is found, then source 1 chooses one in a pseudo-random fashion. If no such a pair $(b_{11c,b-1}, b_{22c,b-1})$ is found, source 1 sets $(b_{11c,b-1}, b_{22c,b-1}) = (1,1)$; in this case we say that an error has occurred at source 1.

This first encoding step is run in parallel at both sources, so that the two sources have the same set of past cooperative messages $(W_{10c,b-1}, W_{20c,b-1}, W_{11c,b-1}, W_{22c,b-1})$, and of bin indices $(b_{11c,b-1}, b_{22c,b-1})$. This "common knowledge" furnishes the basis for cooperation in slot b, where $S_2^n([W_{22c,b-1}, b_{22c,b-1}],[W_{10c,b-1}, b_{20c,b-1}])$ can be treated as "non-causally known interference" at source 1, and $S_1^n([W_{11c,b-1}, b_{11c,b-1}],[W_{10c,b-1}, b_{20c,b-1}])$ can be treated as "non-causally known interference" at source 2.

Notice that if more than one bin index pair $(b_{11c,b-1}, b_{22c,b-1})$ is found, the sources must agree upon which pair to choose in order to end up having the same codeword pair (S_1^n, S_2^n). In order to do so, the sources share the same pseudo-random number generator and use it to decide which bin index pair to pick. The purpose of this "randomization" is to avoid any form of "determinism" in the encoding [21].

Error analysis: The codewords (Q^n, S_1^n, S_2^n) were sampled in an i.i.d. fashion from the distribution $P_{QS_1S_2} = P_Q P_{S_1|Q} P_{S_2|Q}$. The binning process "forces" them to actually look as if they were sampled in an i.i.d. fashion from the distribution $P_Q P_{S_1S_2|Q}$. For this binning step to be successful with arbitrarily high probability we must have (see Appendix B)

$$R''_{11c} + R''_{22c} \geq I(S_1 \wedge S_2|Q). \quad (19)$$

Note: This encoding step is the same as the encoding in Marton's achievable region for a general two-user broadcast channel [44].

- *Joint conditional binning:* Given the new message triplet $(W_{10c,b}, W_{10n,b}, W_{11n,b})$, source 1 tries to find a set of bin indices $(b_{10c,b}, b_{10n,b}, b_{11n,b})$ such that

$$\Big(V_1^n([W_{10c,b}, b_{10c,b}],[\cdots]), U_1^n([W_{10n,b}, b_{10n,b}],[W_{10c,b}, b_{10c,b}],[\cdots]), T_1^n([W_{11n,b}, b_{11n,b}],[W_{10n,b}, b_{10n,b}],[W_{10c,b}, b_{10c,b}],[\cdots]), \in T_\epsilon^n(P_{V_1U_1T_1|QS_1S_2}|Q^n, S_1^n, S_2^n),$$

where the dots are in place of the known message pair $[W_{10c,b-1}, W_{20c,b-1}]$ from the previous slot. If more than one triplet is found, then source 1 chooses one at random. If no such a triplet $(b_{10c,b}, b_{10n,b}, b_{11n,b})$ is found, source 1 sets $(b_{10c,b}, b_{10n,b}, b_{11n,b}) = (1, 1, 1)$; in this case we say that an error has occurred at source 1.

Error analysis: The triplet (Q^n, S_1^n, S_2^n) found in the previous encoding step appears jointly typical according to $P_{QS_1S_2}$. The set of codewords $(Q^n, V_1^n, U_1^n, T_1^n)$ is jointly typical according to $P_{QV_1U_1T_1}$ by codebook generation. However, codewords (V_1^n, U_1^n, T_1^n) and (S_1^n, S_2^n) were generated independently conditioned on Q^n. The purpose of this binning step is for $(Q^n, S_1^n, S_2^n, V_1^n, U_1^n, T_1^n)$ to look jointly typical according to $P_{QS_1S_2V_1U_1T_1}$. By standard arguments (see Appendix C), this joint binning step is successful with arbitrarily high probability if

$$R'_{10c} \geq I(V_1 \wedge S_1, S_2|Q), \tag{20a}$$

$$R'_{10n} + R'_{10c} \geq I(U_1, V_1 \wedge S_1, S_2|Q), \tag{20b}$$

$$R'_{11n} + R'_{10n} + R'_{10c} \geq I(V_1, U_1, T_1 \wedge S_1, S_2|Q). \tag{20c}$$

The rate constraint in (20a) can be understood as follows. After the first (successful) binning step, the codewords $(Q^n, S_1^n, S_2^n, V_1^n)$ look as if they were sampled from the distribution $P_Q P_{S_1S_2|Q} P_{V_1|Q}$. This encoding step requires them to look as if they were sampled from the distribution $P_Q P_{S_1S_2|Q} P_{V_1|QS_1S_2}$. For this encoding step to be successful with arbitrarily high probability, the encoder needs to be able to search among an exponential (in n) number of codewords V_1^n, whose exponent must be at least $H(V_1|Q) - H(V_1|QS_1S_2) = I(V_1 \wedge S_1, S_2|Q)$. The rate constraints in (20b) and in (20c) have a similar interpretation.

Note: This encoding steps is a generalization of the "sequential binning" idea introduced in [41]. Here, instead of doing several sequential binning steps, we bin all the codewords at once similarly as for Multiple Description Coding [62].

Note: It is not possible to perform a single joint binning step of $(Q^n, S_1^n, S_2^n, V_1^n, U_1^n, T_1^n)$ at source 1, and correspondingly of $(Q^n, S_1^n, S_2^n, V_2^n, U_2^n, T_2^n)$ at source 2, because this does not guarantee that the two sources end up having the exact same triplet (Q^n, S_1^n, S_2^n) – which is required for cooperation.

- *Final binning step:* Given the new message $W_{11c,b}$, source 1 tries to find a bin index $b_{11c,b}$ such that

$$\Big(Q^n(\cdots), S_1^n(\cdots), S_2^n(\cdots), V_1^n(\cdots), U_1^n(\cdots), T_1^n(\cdots), Z_1^n([W_{11c,b}, b_{11c,b}], \cdots)\Big) \in T_\epsilon^n(P_{Z_1|QS_1S_2V_1U_1T_1}|Q^n, S_1^n, S_2^n, V_1^n, U_1^n, T_1^n),$$

where the dots are in place of the known message indices and bin indices. If more than one index $b_{11c,b}$ is found, then source 1 chooses one at random. If no such an index $b_{11c,b}$ is found, source 1 sets $b_{11c,b} = 1$; in this case we say that an error has occurred at source 1.

Error analysis: The set $(Q^n, S_1^n, S_2^n, V_1^n, U_1^n, T_1^n)$ found from the previous encoding steps, appears jointly typical according to $P_{QS_1S_2V_1U_1T_1}$. By codebook generation, $(Q^n, S_1^n, V_1^n, Z_1^n)$ are jointly typical according to $P_{QS_1V_1Z_1}$, that is, conditioned on (Q, S_1, V_1), Z_1 is independent of (S_2, U_1, T_1). The purpose of this binning step is to make the whole set jointly typical according to $P_{QS_1S_2V_1U_1T_1Z_1}$. By standard arguments (see Appendix C), this binning step is successful with arbitrary high probability if

$$R'_{11c} \geq I(Z_1 \wedge S_2, U_1, T_1|Q, S_1, V_1). \tag{20d}$$

- Finally, source 1 sends a codeword X_1^n that is jointly typical with all the sequences found in the previous binning steps.
- Encoding at source 2 proceeds similarly.

Cooperation: If the encoding steps are successful at both sources, all the transmitted and received signals are jointly typical according to the distribution in (18).

At the end of slot b, source 1 knows all the messages generated at source 1 at the beginning of the slot, including (Q^n, S_2^n) (because we can assume successfully decoding of all past cooperative messages). Source 1 (source 2 proceeds similarly) then searches for a unique pair of message indices (i, j) and for some pair of bin indices (b_i, b_j) such that

$$\Big(V_2^n([j, b_j], \cdots), Z_2^n([i, b_i], \cdots), Y_{1,b}^n\Big) \in T_\epsilon^n(P^{(\text{enc1})}_{V_2Z_2Y_1|\underline{X}_1}|\underline{X}_1^n),$$

where the dots indicate known old cooperative message indices and bin indices, where all that is known at source 1 is compactly represented by

$$\underline{X}_1 = (Q, S_1, S_2, Z_1, V_1, U_1, T_1, X_1),$$

and where

$$P^{(\text{enc1})}_{V_2Z_2Y_1|\underline{X}_1} = \sum_{U_2,T_2,X_2} \frac{P_{QS_1S_2}P_{V_1U_1T_1Z_1X_1|QS_1S_2}}{P_{QS_1S_2V_1U_1T_1Z_1X_1}} P_{V_2U_2T_2Z_2X_2|QS_1S_2}P_{Y_1|X_1X_2} = P_{V_2Z_2|QS_1S_2}\Big(\sum_{X_2} P_{X_2|QS_1S_2V_2Z_2}P_{Y_1|X_1X_2}\Big).$$

If the search fails, source 1 sets $(i, j) = (1, 1)$; in this case we say that an error has occurred at source 1.

Error analysis: From the point of view of source 1, this decoding step is equivalent to decode the codewords V_2^n and Z_2^n in a MAC-like channel with output Y_1^n and state $\underline{X}_1^n$ known at the receiver only, similar in spirit to [65]. Decoding at source 1 is depicted in Fig. 5 where the variables to be decoded are within a blue circle, the variables that are treated as noise are within a red rectangular box, and the remaining

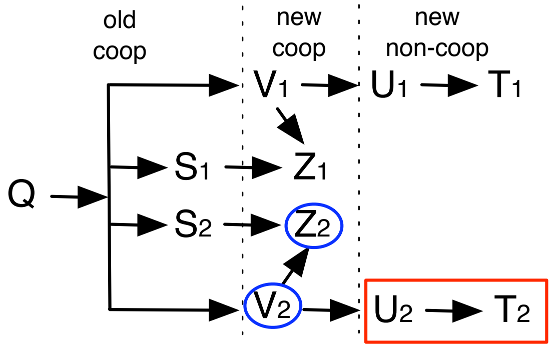

Fig. 5. Decoding at source 1: the variables to be decoded are within a blue circle, the variables that are treated as noise are within a red rectangular box, the remaining variables are known because generated by source 1.

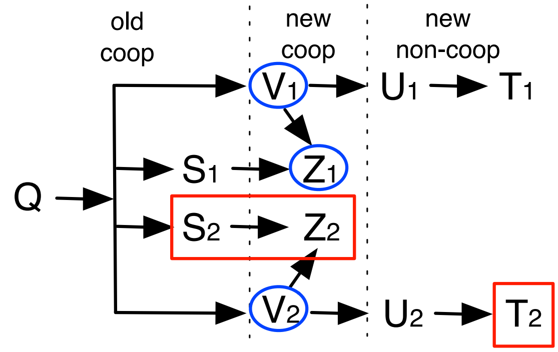

Fig. 6. Decoding at destination 1: the variables whose message index is known (but the bin index is not) are within a blue circle, the variables that are treated as noise are within a red rectangular box, the remaining variables are to be decoded.

variables are known. By standard arguments (see Appendix D), decoding is successful with arbitrarily high probability if

$$\begin{aligned} R_{Z_2} &\le I(Z_2 \wedge Y_1|\underline{X}_1, V_2) \\ &\quad + I(Z_2 \wedge S_1|Q, S_2, V_2) \qquad (21a)\\ R_{V_2} + R_{Z_2} &\le I(V_2, Z_2 \wedge Y_1|\underline{X}_1) + I(V_2 \wedge S_1, S_2|Q) \\ &\quad + I(Z_2 \wedge S_1|Q, S_2, V_2), \qquad (21b) \end{aligned}$$

with $R_{Z_2} \triangleq (R_{22c} + R'_{22c})$ and $R_{V_2} \triangleq (R_{20c} + R'_{20c})$. Intuitively, the constraints in (21) are a consequence of the following observations. Conditioned on (Q, S_1, S_2), a wrong V_2 looks sampled from $P_{V_2|QS_1S_2}$ but it was actually sampled from $P_{V_2|Q}$; this accounts for the term $I(V_2 \wedge S_1, S_2|Q)$. Similarly, conditioned on (Q, S_1, S_2), a wrong Z_2 looks sampled from $P_{Z_2|QS_1S_2V_2}$ but it was actually sampled from $P_{Z_2|Q,S_2,V_2}$; this accounts for the term $I(Z_2 \wedge S_1|Q, S_2, V_2)$.

Note: The inequalities in (21) generalizes the one in (6a), and reduce to (6a) when $S_1 = Z_1 = S_2 = Z_2 = Q$.

In (21), there is not rate constraint for R_{V_2} alone. This is so because Z_2 is superimposed to V_2; it is thus impossible that V_2 is wrong while Z_2 is correct.

Decoding: The receivers wait until the last slot of the frame has been received, and then proceed to decode by using backward decoding. We can assume that when decoding the information sent in slot b, the decoding of the information sent in the slots $b+1, \cdots, N$ was successful [10], [64]. When decoding slot b, destination 1 knows the current cooperative messages $(W_{10c,b}, W_{20c,b}, W_{11c,b})$ carried by (V_1, Z_1, V_2), and tries to decode the previous cooperative common messages $(W_{10c,b-1}, W_{20c,b-1})$ in Q, the previous cooperative private message $W_{11c,b-1}$ in S_1, and the current non-cooperative messages $W_{10n,b}$ in U_1, $W_{20n,b}$ in U_2, and $W_{11n,b}$ in T_1. Decoding at destination 1 is depicted in Fig. 6 where the variables whose message index is known (but the bin index is not) are within a blue circle, the variables that are treated as noise are within a red rectangular box, the remaining variables are to be decoded.

Formally, destination 1 looks for a unique set of message indices (q_1, s_1, u_1, t_1) and for some indices $(q_2, u_2, b_{v_1}, b_{v_2}, b_{z_1}, b_{s_1}, b_{u_1}, b_{u_2}, b_{t_1})$ such that

$$\begin{aligned} &\Big(Q^n([q_1, q_2]), \\ &S_1^n([s_1, b_{s_1}], [q_1, q_2]), \\ &V_1^n([*, b_{v_1}], [q_1, q_2]), \\ &Z_1^n([*, b_{z_1}], [s_1, b_{s_1}], [*, b_{v_1}], [q_1, q_2]), \\ &U_1^n([u_1, b_{u_1}], [*, b_{v_1}], [q_1, q_2]), \\ &T_1^n([t_1, b_{t_1}], [u_1, b_{u_1}], [*, b_{v_1}], [q_1, q_2]), \\ &V_2^n([*, b_{v_2}], [q_1, q_2]), \\ &U_2^n([u_2, b_{u_2}], [*, b_{v_2}], [q_1, q_2]), \\ &Y_{3,b}^n\Big) \in T_\epsilon^{(n)}(P^{(\mathrm{dec1})}_{QS_1V_1U_1T_1Z_1V_2U_2Y_3}), \end{aligned}$$

where a star $*$ indicates a known message index from the previous decoding step, where

$$\begin{aligned} &P^{(\mathrm{dec1})}_{QS_1V_1U_1T_1Z_1V_2U_2Y_3} \\ &= \sum_{S_2,X_1,T_2,Z_2,X_2} P_{QS_1S_2}P_{V_1U_1T_1Z_1X_1|QS_1S_2} \\ &\quad P_{V_2U_2T_2Z_2X_2|QS_1S_2}P_{Y_3|X_1X_2} \\ &= P_{QS_1}P_{V_1U_1T_1Z_1|QS_1}P_{V_2U_2|QS_1} \\ &\quad \Big(\sum_{S_2,X_1,X_2} \frac{P_{X_1S_2|QS_1V_1U_1T_1Z_1}P_{X_2S_2|QS_1V_2U_2}}{P_{S_2|QS_1}}P_{Y_3|X_1X_2}\Big), \end{aligned}$$

If none or more than one set of indices (q_1, s_1, u_1, t_1) is found, then decoder 1 sets all the indices to one; in this case we say that an error has occurred at destination 1.

Error analysis: The error analysis can be found in Appendix E. The probability of error at destination 1 can be made as small as desired if the following rate constraints are satisfied:

$$R_{V_1} + R_{V_2} + $$
$$+R_{U_1} + R_{T_1} + R_{U_2} + R_{Z_1} \leq E_0^{(1)} \quad (22a)$$
$$R_{U_1} + R_{T_1} + R_{U_2} + R_{Z_1} \leq \min\{E_1^{(1)}, E_2^{(1)}, E_4^{(1)}, E_5^{(1)}\} \quad (22b)$$
$$R_{U_1} + R_{T_1} + R_{Z_1} \leq \min\{E_3^{(1)}, E_6^{(1)}\} \quad (22c)$$
$$R_{T_1} + R_{U_2} + R_{Z_1} \leq \min\{E_7^{(1)}, E_8^{(1)}\} \quad (22d)$$
$$R_{T_1} + R_{Z_1} \leq E_9^{(1)} \quad (22e)$$
$$R_{U_2} + R_{Z_1} \leq \min\{E_{10}^{(1)}, E_{11}^{(1)}\} \quad (22f)$$
$$R_{Z_1} \leq E_{12}^{(1)} \quad (22g)$$
$$R_{U_1} + R_{T_1} + R_{U_2} \leq \min\{E_{13}^{(1)}, E_{14}^{(1)}, E_{16}^{(1)}, E_{17}^{(1)}, E_{22}^{(1)}, E_{23}^{(1)}\} \quad (22h)$$
$$R_{U_1} + R_{T_1} \leq \min\{E_{15}^{(1)}, E_{18}^{(1)}, E_{24}^{(1)}\} \quad (22i)$$
$$R_{T_1} + R_{U_2} \leq \min\{E_{19}^{(1)}, E_{20}^{(1)}, E_{25}^{(1)}, E_{26}^{(1)}\} \quad (22j)$$
$$R_{T_1} \leq \min\{E_{21}^{(1)}, E_{27}^{(1)}\}, \quad (22k)$$

where the rates $R_\star$, for $\star \in \{Q, V_1, U_1, T_1, S_1, Z_1, V_2, U_2, T_2, S_2, Z_2\}$, and the quantities $E_\ell^{(1)}$, for $\ell \in \{0, ..., 27\}$, are defined in Appendix E in Table II.

Similarly, the rate constraints at destination 2 are as in (22) but with the role of the users swapped,

Achievable region: We have the following result:

Theorem V.1. *The achievable region with superposition & binning as a function of R_1 and R_2 only, can be obtained by applying the Fourier-Motzkin elimination procedure to the intersection of (19), and (20), (21), (22), and the regions corresponding to (20), (21), (22) but with the role of the users swapped. The rates constraints are expressed as the minimum of several quantities that we do not report here for sake of space.* ■

Remark V.2. *When the binning rates are taken to satisfy the constraints in (20) and in (21) with equality for both users, the achievable region has five types of rate bounds as for the non-feedback case, i.e., for R_1, for R_2, for R_1+R_2, for $2R_1+R_2$, and for R_1+2R_2.*

Remark V.3. *As for the case of superposition-only, the error analysis for the case of binning & superposition is greatly simplified by the structured way we performed superposition and binning. In particular, only 28 error events matter out of the possible $2^8 - 1 = 255$ that would result from jointly decoding 8 messages.*

Remark V.4. *In [69], we constructed the codebooks as depicted in Fig. 7. The difference with respect to the encoding proposed in this paper and depicted in Fig. 3 is that Z_1 was superimposed to S_2 too (which carries the private message from source 2 and is sent cooperatively by source 1, but it is not decoded at destination 1), and Z_2 was superimposed to S_1 too. Thus, with the encoding of [69], Z_1 could not be*

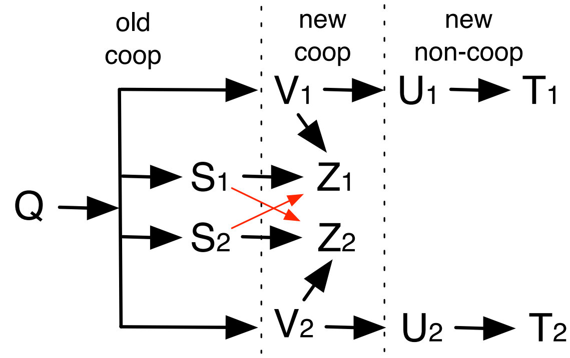

Fig. 7. A visualization of the codebook generation proposed in [69] (to be compared to the one in Fig. 3).

decoded at destination 1 for the following reason: the decoding of Z_1 implies the decoding of all messages to which Z_1 is superimposed, thus also S_2; S_2 however is a private message for user 2 that must not be decoded at destination 1; when Z_1 is not decoded, it acts as an extra noise source at the receiver. Similarly, Z_2 could not be decoded at destination 2.

With the encoding proposed in [69], the rate constraint in (20d) should be replaced by

$$R'_{11c} \geq I(Z_1 \wedge U_1, T_1 | Q, S_1, S_2, V_1), \quad (23)$$

and the rate constraints in (21) should be replaced by

$$R_{Z_2} \leq I(Z_2 \wedge Y_1 | \underline{X}_1, V_2) \quad (24a)$$
$$R_{V_2} + R_{Z_2} \leq I(V_2, Z_2 \wedge Y_1 | \underline{X}_1) + I(V_2 \wedge S_1, S_2 | Q), \quad (24b)$$

where the term $I(Z_2 \wedge S_1 | Q, S_2, V_2)$ in (21) does not appear in (24) since Z_2 is superimposed to S_1 by construction (thus it already has the desired joint distribution). In other words, the encoding of [69] is less stringent in terms of "binning rates constraints" (i.e.,$(23) \leq (20\text{d})$), but it is more stringent in terms on "decoding rate constraints" (i.e.,$(24) \leq (21)$). These two effects do compensate one another since, from (24) and (23), we have:

$$R_{22c} = R_{Z_2} - R'_{22c}$$
$$\leq I(Z_2 \wedge Y_1 | \underline{X}_1, V_2) - I(Z_2 \wedge U_2, T_2 | Q, S_1, S_2, V_2)$$
$$R_{20c} + R_{22c} = R_{V_2} - R'_{20c} + R_{Z_2} - R'_{22c}$$
$$\leq I(V_2, Z_2 \wedge Y_1 | \underline{X}_1) - I(Z_2 \wedge U_2, T_2 | Q, S_1, S_2, V_2),$$

which is the same we obtain from (21) and (20d).

The two schemes are not the same though. With the encoding scheme proposed in this paper, the message carried by Z_1 (resp. Z_2) can be decoded at destination 1 (resp. destination 2) because it is superimposed to messages that are anyway decoded by destination 1 (resp. destination 2). This enlarges the achievable region derived in [69] because Z_1 (resp. Z_2) is no longer treated as noise by destination 1 (resp. destination 2).

B. Comparison with existing regions

We showed in the previous section that the achievable region with superposition-only in (8) does not reduce to the known achievable regions for some channels subsumed by the IFC-GF model. With superposition & binning we have:

1) Broadcast channel: When the IFC-GF channel reduces to a **broadcast channel**, our region with superposition & binning reduces to Marton's inner bound for a general broadcast channel [44] by considering the auxiliary random variables (Q, S_1, S_2) only.

2) Cognitive channel: When the IFC-GF channel reduces to a **cognitive channel**, our region with superposition & binning is a still a subset of the largest known achievable region for a general cognitive channel [47]. The reason is as follows. Assume source 1 is the cognitive/secondary user and source 2 is the primary user; this scenario is captured by setting to infinity the right hand side of the different equations in (21) (i.e., no constraints on R_{V_2} and R_{Z_2}), and by setting to zero the right hand side of the different equations in (21) obtained by swapping the role of the users (i.e., $R_{V_1} = R_{Z_1} = 0$). Since the primary user (source 1) does not cooperate with the secondary user (source 2) we can set $S_1 = V_1 = Z_1 = \emptyset$. Since the secondary user (source 2) knows anti-causally the message the primary user (source 1), we can set $Z_2 = V_2 = U_2 = T_2 = \emptyset$ (i.e., no need to send new information). With these choices, our superposition & binning encoding scheme only uses (Q, U_1, T_1, S_2). If we set $Q = U_{2c}, U_1 = U_{1c}, T_1 = U_{1pb}, S_2 = U_{2pa}$, then our superposition & binning is equivalent to the scheme in [47] with $U_{2pb} = \emptyset$; however, since in general $U_{2pb} \neq \emptyset$ in [47], our regions is a subset of the region in [47]. The auxiliary random variable U_{2pb} in [47] carries part of the primary user's private message that is sent by the cognitive user only; this feature is not present in our encoding scheme.

3) Realy channel: When $R_2 = 0$, the IFC-GF channel reduces to a **relay channel**. Our superposition & binning region for $R_2 = 0$ however does not reduce to the largest known achievable region for relay channel [9], [14] because it does not encompass compress-and-forward relaying. The inclusion of compress-and-forward, possibly in the more general "noisy network coding" framework of [39] is an interesting open problem for future work.

4) IFC-GF with conferencing encoders: The case of **Gaussian IFC with conferencing encoders** was studied in [63]. The scheme proposed in [63] also splits each message into four parts and uses binning and superposition; however, the set of input distributions considered is less general than ours in (18) and only the Marton-like binning step (to bin S_1 and S_2 against each other) is performed. It was remarked in [63, Remark 4.3]: "We conjecture that dirty paper coding among cooperative private messages will lead to a better rate region and smaller gap to the outer bounds, while the procedure of computing the achievable region becomes complicated." Our superposition & binning achievable region exactly addresses [63, Remark 4.3]; we are currently investigating whether our scheme can reduce the 6.5 bits gap to capacity Gaussian IFC with conferencing encoders.

VI. Example: the Gaussian IFC-GF

In this section we provide a numerical evaluation of our achievable regions for the Gaussian channel.

A. Channel Model

A Gaussian channel in standard form has outputs:

$$Y_c = h_{c1} X_1 + h_{c2} X_2 + N_c,\ c \in \{1, ..., 4\},$$

where the inputs are subject to the average power constraint

$$\mathbb{E}[|X_u|^2] \le P_u,\ u \in \{1, 2\},$$

and the additive noises are independent and have distribution $N_c \sim \mathcal{N}(0, 1)$. The case of correlated noises, which includes the case of degraded/noisy output feedback as an example, will be discussed in the second part of this paper. We assume full-duplex communication and perfect knowledge of all channel gains at all terminals.

Without loss of generality we can assume that the *Direct Link* (DL) channel gains h_{31} and h_{42} are real-valued, since the destinations can compensate for the phase of the intended signal; in particular we set $h_{31} = h_{42} = 1$. Similarly, we assume that the *Cooperation Link* (CL) channel gains h_{21} and h_{12} are also real-valued. As opposed to the case without GF, the phase of the *interfering link* (IL) channel gains h_{32} and h_{41} matter because of transmitter cooperation. Since source u, $u \in \{1, 2\}$, knows its transmit signal X_u and the response of its antenna h_{uu}, it can compute $Y_u - h_{uu} X_u$, thus without loss of generality we can assume $h_{uu} = 0$.

B. Input Distribution

We consider here only jointly Gaussian inputs, that is: let $Q \sim \mathcal{N}(0, 1)$, $(\alpha_1, \alpha_2) \in \mathbb{C}^2$, and $X_m \sim \mathcal{N}(0, \sigma_m^2)$, $m \in \{10c, 10n, 11c, 11n, 20c, 20n, 22c, 22n\}$, be independent random variables that satisfy the power constraint, i.e.,

$$0 \le |\alpha_u|^2 + \sigma_{u0c}^2 + \sigma_{u0n}^2 + \sigma_{uuc}^2 + \sigma_{uun}^2 \le P_u, \qquad (25)$$

for $u \in \{1, 2\}$. For the *superposition-only* achievable scheme, we let:

$$\begin{aligned} V_u &= X_{u0c}, \\ U_u &= X_{u0n}, \\ T_u &= X_{uun}, \\ X_u &= \alpha_u\, Q + X_{u0c} + X_{u0n} + X_{uun}, \end{aligned}$$

for $u \in \{1, 2\}$. For the *superposition & binning* achievable scheme we let:

$$\begin{aligned} V_u &= X_{u0c} + \lambda_{u0c} X_{\text{interf@}u}, \\ U_u &= X_{u0n} + \lambda_{u0n} X_{\text{interf@}u}, \\ Z_u &= X_{uuc} + \lambda_{uuc} X_{\text{interf@}u}, \\ T_u &= X_{uun} + \lambda_{uun} X_{\text{interf@}u}, \\ S_u &= X_{\text{interf@}u}, \\ X_u &= \alpha_u\, Q + X_{u0c} + X_{u0n} + X_{uuc} + X_{uun}, \end{aligned}$$

for $u \in \{1, 2\}$, where $X_{\text{interf@1}} = X_{22n}$ and $X_{\text{interf@2}} = X_{11n}$, where the λ-coefficients are chosen as in [11] so as to perfectly cancel the interference produced at destination u, $u \in \{1, 2\}$, by the (not decoded) interfering message $X_{\text{interf@}u}$. We do not claim that the proposed choice of the λ-coefficients

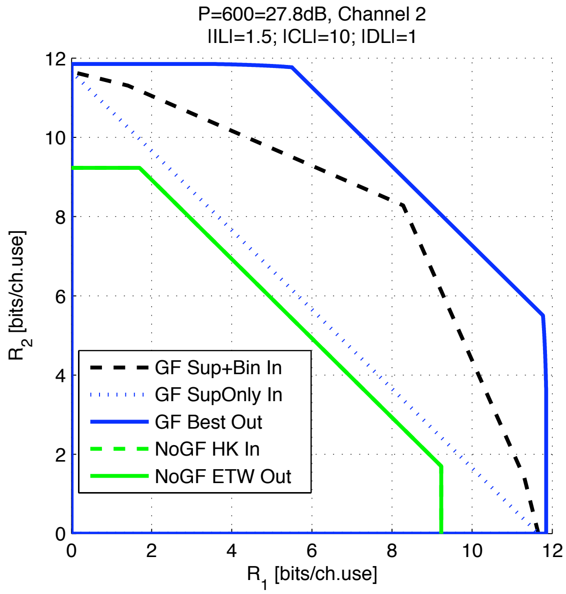

Fig. 8. Performance comparison among standard IFC and IFC-GF with strong interference and strong cooperation, i.e., IL≥DL and CL≥DL.

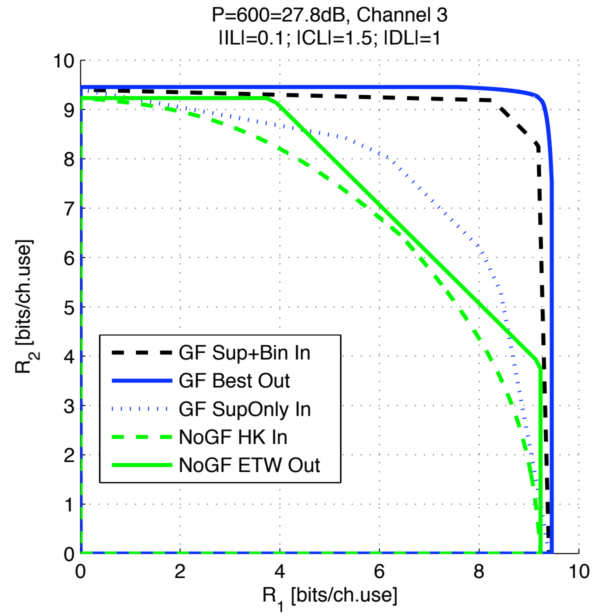

Fig. 9. Performance comparison among standard IFC and IFC-GF with weak interference and strong cooperation, i.e., IL≤DL and , i.e., CL≥DL.

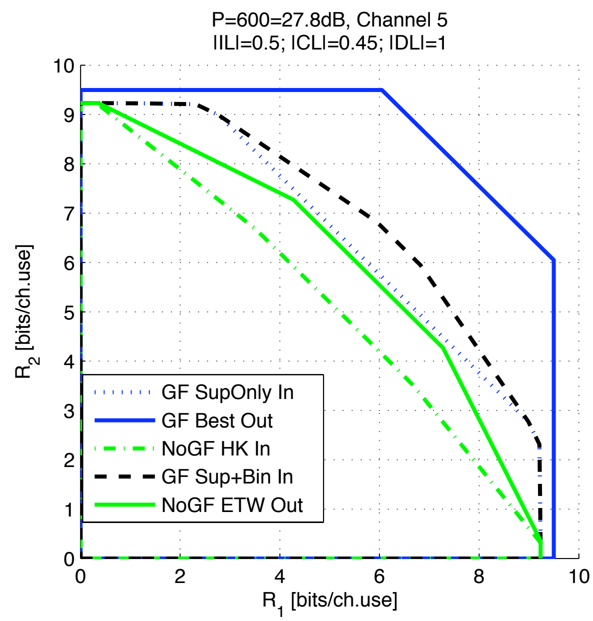

Fig. 10. Performance comparison among standard IFC and IFC-GF with weak interference and weak cooperation, i.e., IL≤DL and , i.e., CL≤DL.

is optimal; it is however convenient because the numerical optimization only involves the power allocation in (25).[3]

C. Numerical Results

In the numerical examples we consider a completely symmetric network where the direct link gains are the same (i.e., $h_{31} = h_{42} =$DL), the interfering link gains are the same (i.e., $h_{41} = h_{32} =$IL), the cooperation link gains are the same (i.e., $h_{12} = h_{21} =$CL), and the power constraints are the same (i.e., $P_1 = P_2 = P$).

Figs. 8, 9 and 10 show the achievable performance for a symmetric Gaussian IFC-GF with independent noises. Both the *superposition-only* (labeled as "GF SupOnly In") and the *superposition & binning* (labeled as "GF Sup+Bin In") achievable regions are reported. The best known outer bound (labeled as "GF Best Out"), obtained by taking the intersection of the regions we derived in [59], [68] as well as those derived in [29], [46], [56], is reported for comparison. The figures also show the Han and Kobayashi achievable region (labeled as "NoGF HK In") and the "to within one bit" outer bound of [18] (labeled as "NoGF ETW Out") for the IFC without generalized feedback in order to demonstrate the rate gains achievable through cooperation. Fig. 8 shows the achievable region for "with strong interference and strong cooperation", i.e., IL≥DL and CL≥DL, Fig. 9 shows the achievable region for "weak interference and strong cooperation," i.e., IL≤DL and , i.e., CL≥DL, and Fig. 10 shows the achievable region for "with weak interference and weak cooperation," i.e., IL≤DL and , i.e., CL≤DL. The value of the parameters used for the numerical evaluation is indicated in the figure's title.

We notice that the achievable rate regions with cooperation are larger than the capacity outer bound of the corresponding IFC without cooperation/generalized feedback. Superposition & binning greatly improves performance in "strong cooperation"; From Fig. 9 we see that the sum-rate increases from about 6.5 bits/ch.use/user ("NoGF ETW Out") to about 9 bits/ch.use/user ("GF Sup+Bin In"); we also observe that the gap between the inner bound ("GF Sup+Bin In") and the

[3] In [48] it has been shown, for the Gaussian cognitive interference channel, that choosing the λ-coefficient so that the interference produced by the primary transmitter at the secondary receiver is not perfectly pre-canceled enlarges the achievable region. We are currently investigating what rate gains can be expected from a similar choice of λ-coefficients for the general IFC-GF.

outer bound ("GF Best Out") is about 0.5 bits/ch.use/user – as opposed to the 19 bit gap predicted in [46]. Similar considerations apply to the other cases (see Figs. 8 and 10), even though the improvements are less dramatic for the "week cooperation" case. We expect larger gains by evaluating the superposition & binning achievable region with optimized "dirty paper coefficients" λ's.

VII. Conclusion

In the first part of this paper we presented a novel achievable region for a general IFC-GF. We built on the idea of splitting the information message into a common and a private part of the classical IFC and proposed a coding scheme where the sources cooperate on sending part of each message. The cooperation on sending the common message aims to realize the gains of beam-forming, as in a distributed multi-antenna system, while the cooperation on sending the private message aims to leverage the interference "pre-cancellation" property of binning/dirty-paper-type coding. Our achievable region generalizes several known achievable regions for IFC-GF and it reduces to known achievable regions for some of the channels subsumed by the IFC-GF model. Numerical results for the Gaussian channel show that source cooperation improves the achievable rate of all the involved sources.

In the second part of this paper we will present novel outer bounds for a general IFC-GF against which the achievable rate with superposition & binning will be compared.

Acknowledgment

The authors would like to thank the anonymous reviewers and the associate editor for their comments which helped improve the clarity of the paper.

Appendix A
Proof of the redundancy of two single-rate constraints

For a fixed distribution $P_{QV_1U_1T_1X_1V_2U_2T_2X_2} = P_{QV_1U_1T_1X_1}P_{V_2U_2T_2X_2|Q}$ consider another distribution $P_{Q'V_1'U_1'T_1'X_1V_2U_2T_2X_2} = P_{Q'V_1'U_1'T_1'X_1}P_{V_2U_2T_2X_2|Q'}$ with

$$V_1' = U_1' = \emptyset,\ T_1' = (T_1, U_1),\ Q' = (Q, V_1),$$

and $P_{V_2U_2T_2X_2|Q'} = P_{V_2U_2T_2X_2|Q}$, that is, under $P_{Q'V_1'U_1'T_1'X_1V_2U_2T_2X_2}$ source 1 does not send any common information. With $P_{Q'V_1'U_1'T_1'X_1V_2U_2T_2X_2}$, the achievable region in (8) reduces to:

$$R_1 \le \text{(6d)} \tag{26a}$$
$$R_2 \le \text{(7f)}' \tag{26b}$$
$$R_2 \le \text{(7a)} + \text{(7d)}' \tag{26c}$$
$$R_1 + R_2 \le \text{(6f)} + \text{(7b)}' \tag{26d}$$
$$R_1 + R_2 \le \text{(7a)} + \text{(6e)} + \text{(7b)}' \tag{26e}$$

where a prime as a superscript indicates that the mutual information in the corresponding equation must be computed for the distribution $P_{Q'V_1'U_1'T_1'X_1V_2U_2T_2X_2}$ (rather than for the distribution $P_{QV_1U_1T_1X_1V_2U_2T_2X_2}$). Notice that all the mutual information terms in (6) are larger under $P_{Q'V_1'U_1'T_1'X_1V_2U_2T_2X_2}$ than under $P_{QV_1U_1T_1X_1V_2U_2T_2X_2}$ and satisfy

$$\begin{aligned} 0 &\le (6b)' = (6d)' = (6d) \\ &\le (6c)' = (6e)' = (6e) \\ &\le (6f)' = (6f) \\ (6a)' &= 0, \end{aligned}$$

while the mutual information terms in (7) under $P_{Q'V_1'U_1'T_1'X_1V_2U_2T_2X_2}$ satisfy

$$\begin{aligned} 0 &\le (7b)' = (7c)' = I(Y_4 \wedge T_2|Q, V_1, V_2, U_2) \\ &\le (7d)' = (7e)' = I(Y_4 \wedge T_2, U_2|Q, V_1, V_2) \\ &\le (7f)' = I(Y_4 \wedge T_2, U_2, Q, V_1, V_2) \\ (7a)' &= (7a) = I(V_2 \wedge Y_1|Q, T_1, U_1, V_1, X_1). \end{aligned}$$

Consider the region in (8) (under $P_{QV_1U_1T_1X_1V_2U_2T_2X_2}$). If

$$\underbrace{\text{(6a)} + \text{(6b)} + \text{(7c)}}_{\text{in (9)}} \ge \min\{\underbrace{\text{(6f)}}_{\text{in (8a)}}, \underbrace{\text{(6a)} + \text{(6d)}}_{\text{in (8b)}}\}$$

then the rate constraint in (9) is redundant and can be omitted from the region in (8). We will now show that the rate constraint in (9) can always be omitted from the region in (8) without enlarging the achievable region. We will do so by showing that the rate points for which the rate constraint in (9) is violated, that is, when

$$\underbrace{\text{(6a)} + \text{(6b)} + \text{(7c)}}_{\text{in (9)}} < R_1 \le \min\{\underbrace{\text{(6f)}}_{\text{in (8a)}}, \underbrace{\text{(6a)} + \text{(6d)}}_{\text{in (8b)}}\} \tag{27}$$

holds together with all the rate constraint in (8), are contained in the region in (26) (under $P_{Q'V_1'U_1'T_1'X_1V_2U_2T_2X_2}$). The region in (26) is a special case of the region in (8) for which the rate constraint in (9) is redundant. This shows that the region in (8) is indeed achievable.

Assume now that (27) holds together with all the rate constraint in (8) (under $P_{QV_1U_1T_1X_1V_2U_2T_2X_2}$). We will show that (27) and (8) imply (26). We have: (26a) = (8b). Moreover

$$\begin{aligned} R_2 &\le \underbrace{\text{(6b)} + \text{(7f)}}_{\text{in (8f)}} - \Big(\text{(6a)} + \text{(6b)} + \text{(7c)}\Big) \\ &= I(Y_4; U_2, Q, V_1, V_2) - I(Y_2; V_1|\underline{X}_2) \\ &\le I(Y_4; T_2, U_2, Q, V_1, V_2) - 0 = \underbrace{\text{(7f)}'}_{\text{in (26b)}} \end{aligned}$$

and

$$\begin{aligned} R_2 &\le \underbrace{\text{(6a)} + \text{(7a)} + \text{(6b)} + \text{(7e)}}_{\text{in (8h)}} - \Big(\text{(6a)} + \text{(6b)} + \text{(7c)}\Big) \\ &= \text{(7a)} + I(Y_4; U_2|Q, V_1, V_2) \\ &\le \text{(7a)} + I(Y_4; T_2, U_2|Q, V_1, V_2) = \underbrace{\text{(7a)} + \text{(7d)}'}_{\text{in (26c)}} \end{aligned}$$

and

$$R_1 + R_2 \le \underbrace{(6a) + (6b) + (6f) + (7c)}_{\text{in (8j)}} - \Big((6a) + (6b) + (7c)\Big)$$
$$= (6f)$$
$$\le \underbrace{(6f) + (7c)'}_{\text{in (26d)}}$$

and

$$R_1 + R_2 \le \underbrace{2\cdot(6a) + (7a) + (6b) + (6e) + (7c)}_{\text{in (8k)}} +$$
$$- \Big((6a) + (6b) + (7c)\Big)$$
$$= (6a) + (7a) + (6e)$$
$$\le \underbrace{(6a) + (7a) + (6e) + (7b)'}_{\text{in (26e)}}.$$

This concludes the proof.

APPENDIX B
PROOF OF (19)

The probability that encoder 1 fails to find a good pair of indices $b_{11c,b-1} = i$, $b_{22c,b-1} = j$ is

$$\begin{aligned}\Pr\Big[\cap_{i\ge1,j\ge1} \Big(& S_1^n([W_{11c,b-1}, i], [W_{10c,b-1}, W_{20c,b-1}]),\\ & S_2^n([W_{22c,b-1}, j], [W_{10c,b-1}, W_{20c,b-1}])\Big)\\ & \notin T_\epsilon^n(P_{QS_1S_2}|Q^n)\Big]\\ & = \Big(1-p\Big)^{e^{n(R'_{11c}+R'_{22c})}}\\ & \le \exp(-e^{n(R'_{11c}+R'_{22c})}\, p)\\ & \le \exp(-e^{n(R'_{11c}+R'_{22c}-I(S_1\wedge S_2|Q)-O(\epsilon))})\end{aligned}$$

which goes to zero as $n \to \infty$ if (19) holds, where

$$\begin{aligned}p = \Pr\Big[\Big(& S_1^n([W_{11c,b-1}, i], [W_{10c,b-1}, W_{20c,b-1}]),\\ & S_2^n([W_{22c,b-1}, j], [W_{10c,b-1}, W_{20c,b-1}])\Big)\\ & \in T_\epsilon^n(P_{QS_1S_2}|Q^n)\Big]\\ & \le |T_\epsilon^n(P_{QS_1S_2}|Q^n)| P_{S_1^n|Q^n} P_{S_2^n|Q^n}\\ & \le e^{n(H(S_1S_2|Q)-H(S_1|Q)-H(S_2|Q)-O(\epsilon))}\\ & = e^{n(-I(S_1\wedge S_2|Q)-O(\epsilon))},\end{aligned}$$

where $O(\epsilon) \to 0$ as $\epsilon \to 0$.

APPENDIX C
PROOF OF (20)

Encoding fails if for all set of indices $(B_{10c,b}, B_{10n,b}, B_{11n,b})$ no triplet (V_1^n, U_1^n, T_1^n) can be found to be jointly typical with (Q^n, S_1^n, S_2^n). The probability of this event can be bounded as

$$\begin{aligned}\Pr\Big[\bigcap_{B_{10c}=1}^{e^{nR'_{10c}}} \bigcap_{B_{10n}=1}^{e^{nR'_{10n}}} \bigcap_{B_{11n}=1}^{e^{nR'_{11n}}} & \\ V_1^n([W_{10c,b}, B_{10c}], \cdots),& \\ U_1^n([W_{10n,b}, B_{10n}], [W_{10c,b}, B_{10c}], \cdots),& \\ T_1^n([W_{11n,b}, B_{11n}], [W_{10n,b}, B_{10n}], [W_{10c,b}, B_{10c}], \cdots))& \\ \notin T_\epsilon^n(P_{Q,S_1,S_2,V_1,U_1,T_1}|Q^n, S_1^n, S_2^n)\Big]& \\ = \Pr[K=0] \le \frac{\mathrm{Var}[K]}{\mathbb{E}^2[K]},&\end{aligned}$$

where

$$K = \sum_{B_{10c}=1}^{e^{nR'_{10c}}} \sum_{B_{10n}=1}^{e^{nR'_{10n}}} \sum_{B_{11n}=1}^{e^{nR'_{11n}}} K_{B_{10c},B_{10n},B_{11n}}$$

for

$$\begin{aligned}K_{B_{10c},B_{10n},B_{11n}} = 1\Big\{& \\ V_1^n([W_{10c,b}, B_{10c}], \cdots),& \\ U_1^n([W_{10n,b}, B_{10n}], [W_{10c,b}, B_{10c}], \cdots),& \\ T_1^n([W_{11n,b}, B_{11n}], [W_{10n,b}, B_{10n}], [W_{10c,b}, B_{10c}], \cdots))& \\ \in T_\epsilon^n(P_{Q,S_1,S_2,V_1,U_1,T_1}|Q^n, S_1^n, S_2^n)\Big\}&\end{aligned}$$

and $1_{\{A\}}$ is the indicator function that equals one whenever the condition expressed by A is true.

The mean of the random variable K is easily lower bounded as

$$\begin{aligned}\mathbb{E}[K] = \sum_{B_{10c}=1}^{e^{nR'_{10c}}} \sum_{B_{10n}=1}^{e^{nR'_{10n}}} \sum_{B_{11n}=1}^{e^{nR'_{11n}}} \Pr\Big[& \\ V_1^n([W_{10c,b}, B_{10c}], \cdots),& \\ U_1^n([W_{10n,b}, B_{10n}], [W_{10c,b}, B_{10c}], \cdots),& \\ T_1^n([W_{11n,b}, B_{11n}], [W_{10n,b}, B_{10n}], [W_{10c,b}, B_{10c}], \cdots))& \\ \in T_\epsilon^n(P_{Q,S_1,S_2,V_1,U_1,T_1}|Q^n, S_1^n, S_2^n)\Big]&\\ = \sum_{B_{10c}=1}^{e^{nR'_{10c}}} \sum_{B_{10n}=1}^{e^{nR'_{10n}}} \sum_{B_{11n}=1}^{e^{nR'_{11n}}} |T_\epsilon^n(P_{Q,S_1,S_2,V_1,U_1,T_1}|Q^n, S_1^n, S_2^n)|\cdot& \\ \cdot P_{V_1^n,U_1^n,T_1^n|Q^n}& \\ \ge e^{n[R'_{10c}+R'_{10n}+R'_{11n}-I(S_1,S_2\wedge V_1,U_1,T_1|Q)-O(\epsilon)]},&\end{aligned}$$

and similarly upper bounded as

$$\mathbb{E}[K] \le e^{n[R'_{10c}+R'_{10n}+R'_{11n}-I(S_1,S_2\wedge V_1,U_1,T_1|Q)+O(\epsilon)]}.$$

The variance of K can be computed as

$$\begin{aligned}\mathrm{Var}[K] = \sum_{B_{10c}=1}^{e^{nR'_{10c}}} \sum_{B_{10n}=1}^{e^{nR'_{10n}}} \sum_{B_{11n}=1}^{e^{nR'_{11n}}} \sum_{B'_{10c}=1}^{e^{nR'_{10c}}} \sum_{B'_{10n}=1}^{e^{nR'_{10n}}} \sum_{B'_{11n}=1}^{e^{nR'_{11n}}}& \\ \Big(\Pr[K_{B_{10c},B_{10n},B_{11n}} = 1, K_{B'_{10c},B'_{10n},B'_{11n}} = 1]& \\ - \Pr[K_{B_{10c},B_{10n},B_{11n}} = 1] \Pr[K_{B'_{10c},B'_{10n},B'_{11n}} = 1]\Big).&\end{aligned}$$

When $B_{10c} \neq B'_{10c}$, the random variables $K_{B_{10c},B_{10n},B_{11n}}$ and $K_{B'_{10c},B'_{10n},B'_{11n}}$ are independent by construction, hence they do not contribute to the summation. When $B_{10c} = B'_{10c}$, we upper-bound $\mathrm{Var}[K]$ by neglecting the non-negative term $\Pr[K_{B_{10c},B_{10n},B_{11n}} = 1] \Pr[K_{B'_{10c},B'_{10n},B'_{11n}} = 1]$. Hence we have

$$\begin{aligned}\mathrm{Var}[K] \leq & \sum_{B_{10c}=B'_{10c}} \sum_{B_{10n}=B'_{10n}} \sum_{B_{11n}=B'_{11n}} \Pr[K_{B_{10c},B_{10n},B_{11n}} = 1] \\ & + \sum_{B_{10c}=B'_{10c}} \sum_{B_{10n}=B'_{10n}} \sum_{(B'_{11n}\neq B_{11n})} \Pr[K_{B_{10c},B_{10n},B_{11n}} = 1]\mathrm{e}^{-A'} \\ & + \sum_{B_{10c}=B'_{10c}} \sum_{(B'_{10n}\neq B_{10n})} \sum_{(B_{11n},\ B'_{11n})} \Pr[K_{B_{10c},B_{10n},B_{11n}} = 1]\mathrm{e}^{-B'} \\ & \leq \mathrm{e}^{n[R'_{10c}+R'_{10n}+R'_{11n}-I(S_1S_2\wedge V_1U_1T_1|Q)-O(\epsilon)]} \\ & \quad (1+\mathrm{e}^{n[R'_{11n}-A]}+\mathrm{e}^{n[R'_{10n}+R'_{11n}-B]}),\end{aligned}$$

We now evaluate A and B. We have, neglecting the terms that go to zero as $\epsilon \to 0$,

$$\begin{aligned}\mathrm{e}^{-B'} &= \Pr[K_{B_{10c},B'_{10n},B'_{11n}} = 1 | K_{B_{10c},B_{10n},B_{11n}} = 1] \\ &= \Pr[(U_1^n, T_1^n) \in T_\epsilon^n(P_{Q,S_1,S_2,V_1,U_1,T_1}|Q^n, S_1^n, S_2^n, V_1^n)] \\ &\leq \mathrm{e}^{nH(U_1,T_1|Q,S_1,S_2,V_1)-nH(U_1,T_1|Q,V_1)} \\ &\leq \mathrm{e}^{-nI(U_1,T_1\wedge S_1,S_2|Q,V_1)} = \mathrm{e}^{-nB},\end{aligned}$$

and

$$\begin{aligned}\mathrm{e}^{-A'} &= \Pr[K_{B_{10c},B_{10n},B'_{11n}} = 1 | K_{B_{10c},B_{10n},B_{11n}} = 1] \\ &= \Pr[T_1^n \in T_\epsilon^n(P_{Q,S_1,S_2,V_1,U_1,T_1}|Q^n, S_1^n, S_2^n, V_1^n, U_1^n)] \\ &\leq \mathrm{e}^{nH(T_1|Q,S_1,S_2,V_1,U_1)-nH(T_1|Q,V_1,U_1)} \\ &\leq \mathrm{e}^{-nI(T_1\wedge S_1,S_2|Q,V_1,U_1)} = \mathrm{e}^{-nA}.\end{aligned}$$

After having evaluated A and B, we have that

$$\frac{\mathrm{Var}[K]}{\mathbb{E}^2[K]} \leq \frac{1+\mathrm{e}^{n[R'_{11n}-A]}+\mathrm{e}^{n[R'_{10n}+R'_{11n}-B]}}{\mathrm{e}^{n[R'_{10c}+R'_{10n}+R'_{11n}-I(S_1S_2\wedge V_1U_1T_1|Q)]}}$$

which goes to zero as $n \to \infty$ if (20) holds.

APPENDIX D
PROOF OF (21)

Let

$$\begin{aligned}E_{ijb_ib_j} = \Big\{ & (V_2^n([j,b_j],\cdots), Z_2^n([i,b_i]\cdots), Y_{1,b}^n) \\ & \in T_\epsilon^n(P_{V_2Z_2|QS_1S_2}P_{Y_1|QS_1S_2V_2Z_2\underline{X}_1}|\underline{X}_1^n)\Big\},\end{aligned}$$

where all that is known at transmitter 1 is represented by

$$\underline{X}_1^n = (Q^n, S_1^n, S_2^n, Z_1^n, V_1^n, U_1^n, T_1^n, X_1^n).$$

Assume that $(i, j, b_i, b_j) = (1, 1, 1, 1)$ was sent. The probability that the estimate (j, i) of $(W_{20c,b}, W_{22c,b})$ is wrong is bounded by

$$\begin{aligned}P_{\mathrm{e,enc1}}^{(n)} &= \Pr[E_{1111}^c \cup_{(i,j)\neq(11),\forall(b_i,b_j)} E_{ijk}] \\ &\leq \Pr[E_{1111}^c] + \sum_{i>1,j>1} \Pr[\cup_{\forall(b_i,b_j)} E_{ijb_ib_j}] \\ &+ \sum_{i>1} \Pr[\cup_{\forall b_i} E_{i1b_i1}] + \sum_{i>1,b_j>1} \Pr[\cup_{\forall b_i} E_{i1b_ib_j}] \\ &+ \sum_{j>1} \Pr[\cup_{\forall b_j} E_{1j1b_j}] + \sum_{j>1,b_i>1} \Pr[\cup_{\forall b_j} E_{1jb_ib_j}],\end{aligned}$$

where all probabilities are conditioned on $(i, j, b_i, b_j) = (1, 1, 1, 1)$ being sent.

The probability of E_{1111}^c is vanishing as $n \to \infty$ because the transmitted codewords are jointly typical with the received signal with high probability. For the other terms we proceed as follows.

Case $i > 1, j > 1$) In this case V_2^n is wrong; when V_2^n is wrong, it does not matter whether Z_2^n is correct or wrong because Z_2^n is anyway superimposed to a wrong V_2^n and thus the distribution to use in the computation of the probability of error is the same in either case; the most stringent error bound is when both V_2^n and Z_2^n are wrong; we thus have

$$\begin{aligned}&\sum_{i>1,j>1} \Pr[\cup_{\forall(b_i,b_j)} E_{ijb_ib_j}] \\ &\leq \sum_{i>1,j>1,b_i\geq1,b_j\geq1} |T_\epsilon^n(P_{V_2Z_2|QS_1S_2}P_{Y_1|QS_1S_2V_2Z_2\underline{X}_1}|\underline{X}_1^n)| \cdot \\ &\quad \cdot P_{V_2|Q}P_{Z_2|QS_2V_2}P_{Y_1|QS_1S_2\underline{X}_1} \\ &\leq \mathrm{e}^{n[R_{22c}+R_{20c}+R'_{22c}+R'_{20c}]} \\ &\mathrm{e}^{n[H(V_2|QS_1S_2)+H(Z_2|QS_1S_2V_2)+H(Y_1|QS_1S_2V_2Z_2\underline{X}_1)]} \\ &\mathrm{e}^{-n[H(V_2|Q)+H(Z_2|QS_2V_2)+H(Y_1|QS_1S_2\underline{X}_1)]} \\ &\leq \mathrm{e}^{n[R_{22c}+R_{20c}+R'_{22c}+R'_{20c}]} \\ &\cdot \mathrm{e}^{n[-I(V_2\wedge S_1S_2|Q)-I(Z_2\wedge S_1|QS_2V_2)-I(Y_1\wedge V_2Z_2|QS_1S_2\underline{X}_1)]},\end{aligned}$$

since, by codebook generation, V_2 and S_2 are independent conditioned on Q, and since Z_2 is superimposed to (S_2, V_2). This probability can be driven to zero if (21b) holds.

Case $i > 1, j = 1$) In this case V_2^n is correct and Z_2^n wrong; we thus have

$$\begin{aligned}&\sum_{i>1} \Pr[\cup_{\forall b_i} E_{i1b_i1}] \\ &\leq \sum_{i>1,b_i\geq1} |T_\epsilon^n(P_{V_2Z_2|QS_1S_2}P_{Y_1|QS_1S_2V_2Z_2\underline{X}_1}|\underline{X}_1^n)| \cdot \\ &\quad \cdot P_{V_2|QS_1S_2}P_{Z_2|QS_2V_2}P_{Y_1|QS_1S_2V_2\underline{X}_1} \\ &\leq \mathrm{e}^{n[R_{22c}+R'_{22c}]} \\ &\mathrm{e}^{n[H(V_2|QS_1S_2)+H(Z_2|QS_1S_2V_2)+H(Y_1|QS_1S_2V_2\underline{X}_1)]} \\ &\mathrm{e}^{-n[H(V_2|QS_1S_2)+H(Z_2|QS_2V_2)+H(Y_1|QS_1S_2V_2\underline{X}_1)]} \\ &\leq \mathrm{e}^{n[R_{22c}+R'_{22c}-I(Z_2\wedge S_1|QS_2V_2)-I(Y_1\wedge Z_2|QS_1S_2V_2\underline{X}_1)]},\end{aligned}$$

since now V_2^n has the marginal distribution imposed by the binning step during the encoding process. This probability can be driven to zero if (21a) holds.

APPENDIX E
PROOF OF (22)

In slot b, $b = N, N-1, ..., 1$, destination 1 tries to find a unique set of message indices $(q_1, q_2, s_1, u_1, u_2, t_1)$ and some bin indices $(b_{v_1}, b_{v_2}, b_{z_1}, b_{s_1}, b_{u_1}, b_{u_2}, b_{t_1})$ such that

$$\Big(Q^n([q_1,q_2]),\ S_1^n([s_1,b_{s_1}],[q_1,q_2]),\ V_1^n([1,b_{v_1}],[q_1,q_2]),\ Z_1^n([1,b_{z_1}],[s_1,b_{s_1}],[1,b_{v_1}],[q_1,q_2]),\ U_1^n([u_1,b_{u_1}],[1,b_{v_1}],[q_1,q_2]),\ T_1^n([t_1,b_{t_1}],[u_1,b_{u_1}],[1,b_{v_1}],[q_1,q_2]),\ V_2^n([1,b_{v_2}],[q_1,q_2]),\ U_2^n([u_2,b_{u_2}],[1,b_{v_2}],[q_1,q_2]),\ Y_{3,b}^n\Big) \in T_\epsilon^{(n)}(P^{(\mathrm{dec1})}_{QS_1V_1U_1T_1Z_1V_2U_2Y_3}),$$

where

$$P^{(\mathrm{dec1})}_{QS_1V_1U_1T_1Z_1V_2U_2Y_3} = P_{QS_1}P_{V_1U_1T_1Z_1|QS_1}P_{V_2U_2|QS_1} \Big(\sum_{S_2,X_1,X_2} P_{X_1S_2|QS_1V_1U_1T_1Z_1}\frac{P_{X_2S_2|QS_1V_2U_2}}{P_{S_2|QS_1}}P_{Y_3|X_1X_2}\Big).$$

Notice that, given (Q,S_1) the input variables for source 1 are not independent of the input variables for source 2.

The possible error events are listed in Table I. In Table I the symbols "1", "0" and "$\star$" have the following meaning. A "1" indicates that either the message index or the bin index are in error. A "0" indicates that both the message index and the bin index are correct. A "$\star$" indicates that it does not matter whether the message index is in error; this is so because of superposition coding; in this case in fact, the codeword selected by the decoder – even though with the correct message index – is superimposed to a wrong codeword and it is thus independent of the received signal. In case of a "$\star$", the factorization of the joint probability needed for the evaluation of the probability of error is as for the case where the message is wrong; this implies that the error event that gives the most stringent rate bound is that for which the message is wrong (i.e., as far as error bounds are concerned, a "$\star$" is equivalent to a "1"). The second to last column in Table I counts how many error events are included in the corresponding row (i.e., each "$\star$" corresponds to two possible cases).

There are several groups of error events in Table I: For event $\mathcal{E}_0^{(1)}$: Q is wrong, and hence all the decoded codewords are independent of the received signal (because of superposition coding). For events from $\mathcal{E}_1^{(1)}$ to $\mathcal{E}_{12}^{(1)}$: S_1 is wrong, and thus also Z_1 is wrong (because superimposed to a wrong S_1). For events from $\mathcal{E}_{13}^{(1)}$ to $\mathcal{E}_{21}^{(1)}$: S_1 is correct but Z_1 is wrong. For events from $\mathcal{E}_{22}^{(1)}$ to $\mathcal{E}_{27}^{(1)}$: both S_1 and Z_1 are correct. Notice that, because of the way codebooks are superimposed, out of the possible $2^8-1=255$ error events, only 28 events matter.

A way to understand the error events listed in Table I is a follows. From destination 1's perspective, conditioned on Q, decoding is as for a 3-(virtual)user multiple access channel, where each user sends a superposition of codebooks. Here, (virtual)user 1 sends (V_1,U_1,T_1), (virtual)user 2 sends (S_1,Z_1), and (virtual)user 3 sends (V_2,U_2). Thus, destination 1 must consider all possible combinations of events that consist of jointly decoding a set of messages from the first column of Fig. 11, and a set of messages from the second column of Fig. 11, and a set of messages from the third column of Fig. 11 (even though not all combinations are actual errors for destination 1). In considering such "joint-decoding events", the messages that do not appear in the "set of jointly decoded messages" must be considered as correctly decoded and stripped from the received signal, as in a standard multiple access channel.

Ø				
T_1		Ø		Ø
T_1 U_1	and	Z_1	and	U_2
T_1 U_1 V_1		Z_1 S_1		U_2 V_2

Fig. 11. From destination 1's perspective, conditioned on Q, decoding is as for a 3-(virtual)user multiple access channel, where each user sends a superposition of codebooks. Here, (virtual)user 1 sends (V_1,U_1,T_1), (virtual)user 2 sends (S_1,Z_1), and (virtual)user 3 sends (V_2,U_2). Destination 1 must consider all possible combinations of events that consist of jointly decoding a set of messages from the first column and a set of messages from the second column and a set of messages from the third column (even though not all combinations are actual errors for destination 1).

The last column in Table I is used as follows. Let $\mathbf{X}$ be the set of transmitted codewords (we do not write here the superscript n that indicates the block-length in order to have a lighter notation), and $\mathbf{X}'$ be the set of decoded codewords. Let $\mathcal{C}$ be the subset of the correctly decoded message indices such that $\mathbf{X}(\mathcal{C})=\mathbf{X}'(\mathcal{C})$ (recall that with superposition coding, the decoder might select a codeword $\mathbf{X}'$ that is different from the transmitted codeword $\mathbf{X}$ but with same message index; this happens when an error is committed on one of the "base layer" codewords). The last column of Table I lists the elements of $\mathcal{C}$, i.e., $\mathcal{C}$ contains the codewords that have a "0" in the corresponding row. The sets $\mathcal{C}$ are important for the factorization of the joint density needed for the evaluation of the probability of error. The error analysis proceeds as follows. The joint distribution of the decoded codewords and the received signal is

$$\sum_{\mathbf{X}(\mathcal{C}^c)} P_{\mathbf{X}(\mathcal{C})\mathbf{X}(\mathcal{C}^c)\mathbf{X}'(\mathcal{C}^c)Y} = \sum_{\mathbf{X}(\mathcal{C}^c)} P_{\mathbf{X}(\mathcal{C})}P_{\mathbf{X}(\mathcal{C}^c)|\mathbf{X}(\mathcal{C})}P^{(g)}_{\mathbf{X}'(\mathcal{C}^c)|\mathbf{X}(\mathcal{C})}P_{Y|\mathbf{X}(\mathcal{C})\mathbf{X}(\mathcal{C}^c)} = P^{(g)}_{\mathbf{X}'(\mathcal{C}^c)|\mathbf{X}(\mathcal{C})}P_{\mathbf{X}(\mathcal{C})Y},$$

where $\mathcal{C}^c$ is the complement of $\mathcal{C}$ with respect to the set of messages that a destination decodes, where P is a distribution from the set of possible input distributions in (18), and $P^{(g)}$ is computed from P as described in the codebook generation paragraph in Section V, that is,

$$P^{(g)}_{QV_1U_1T_1S_1Z_1V_2U_2} = P_Q P_{S_1|Q}P_{V_1U_1T_1|Q}P_{Z_1|QS_1V_1}P_{V_2U_2|Q},$$

where all the factors of $P^{(g)}$ are obtained from the corresponding marginalization of P. In the following, we shall drop the prime superscript to distinguished between the wrongly decoded codewords and the transmitted codewords. We will add a superscript "(g)" to the symbol for entropy to indicate that the entropy must be evaluated by using the distribution $P^{(g)}$; the symbol for entropy without any superscript indicates that the entropy must be evaluated by using the distribution P.

Destination 1 searches for codewords that are joint typicality with the received signal according to $P_{\mathbf{X}(\mathcal{C})\mathbf{X}(\mathcal{C}^c)Y}$; however, assuming that the messages in $\mathcal{C}$ are correctly decoded and those in $\mathcal{C}^c$ are wrongly decoded (this is the case that gives the most stringent error bound), the actual joint distribution is $P^{(g)}_{\mathbf{X}(\mathcal{C}^c)|\mathbf{X}(\mathcal{C})}P_{\mathbf{X}(\mathcal{C})Y}$. The probability of the error for the messages in $\mathcal{C}^c$ (neglecting the terms that will eventually be taken to go to zero) is:

$$\begin{aligned}
&\Pr[\text{error } \mathcal{C}^c] = \sum_{\mathbf{x}\in T_\epsilon^{(n)}(P_{\mathbf{X}(\mathcal{C})\mathbf{X}(\mathcal{C}^c)Y}|\mathbf{X}(\mathcal{C}))} P^{(g)}_{\mathbf{X}(\mathcal{C}^c)|\mathbf{X}(\mathcal{C})}P_{Y|\mathbf{X}(\mathcal{C})}\\
&\leq \exp\Big(n\Big[R(\mathcal{C}^c) + H(\mathbf{X}(\mathcal{C}^c)|\mathbf{X}(\mathcal{C})) + H(Y|\mathbf{X}(\mathcal{C})\mathbf{X}(\mathcal{C}^c))\\
&\quad - H^{(g)}(\mathbf{X}(\mathcal{C}^c)|\mathbf{X}(\mathcal{C})) - H(Y|\mathbf{X}(\mathcal{C}))\Big]\Big)\\
&= \exp\Big(n\Big[R(\mathcal{C}^c) - I(Y \wedge \mathbf{X}(\mathcal{C}^c)|\mathbf{X}(\mathcal{C}))\\
&\quad + H(\mathbf{X}(\mathcal{C}^c),\mathbf{X}(\mathcal{C})) - H^{(g)}(\mathbf{X}(\mathcal{C}^c),\mathbf{X}(\mathcal{C}))\\
&\quad - H(\mathbf{X}(\mathcal{C})) + H^{(g)}(\mathbf{X}(\mathcal{C}))\Big]\Big)
\end{aligned}$$

where $R(\mathcal{C}^c)$ is the sum of the rates corresponding to the wrongly decoded messages that are indexed by $\mathcal{C}^c$. In the following, for any two distributions P and Q, the notation $\mathbb{E}[\log(P/Q)]$ stands for the Kullback-Leibler divergence $D(P||Q)$. Let

$$\begin{aligned}
\Delta^{(1)} &\triangleq H^{(g)}(\mathbf{X}(\mathcal{C}^c),\mathbf{X}(\mathcal{C})) - H(\mathbf{X}(\mathcal{C}^c),\mathbf{X}(\mathcal{C}))\\
&= \mathbb{E}\left[\log \frac{P_{\mathbf{X}(\mathcal{C}^c\cup\mathcal{C})}}{P^{(g)}_{\mathbf{X}(\mathcal{C}^c\cup\mathcal{C})}}\right]\\
&= \mathbb{E}\left[\log \frac{P_{QV_1U_1T_1S_1Z_1V_2U_2}}{P_QP_{V_1U_1T_1|Q}P_{S_1|Q}P_{Z_1|QS_1V_1}P_{V_2U_2|Q}}\right]\\
&= \mathbb{E}\left[\log \frac{P_{S_1|QV_1U_1T_1}P_{Z_1|QV_1U_1T_1S_1}P_{V_2U_2|QV_1U_1T_1S_1Z_1}}{P_{S_1|Q}P_{Z_1|QS_1V_1}P_{V_2U_2|Q}}\right]\\
&= I(S_1 \wedge V_1U_1T_1|Q) + I(Z_1 \wedge U_1T_1|QS_1V_1)\\
&\quad + I(V_2U_2 \wedge V_1U_1T_1S_1Z_1|Q). \qquad (28)
\end{aligned}$$

Finally, $\Pr[\text{error } \mathcal{C}^c] \to 0$ as $n \to \infty$ if

$$R(\mathcal{C}^c) \leq I(Y \wedge \mathbf{X}(\mathcal{C}^c)|\mathbf{X}(\mathcal{C})) + \Delta^{(1)} - \underbrace{\mathbb{E}\left[\log \frac{P_{\mathbf{X}(\mathcal{C})}}{P^{(g)}_{\mathbf{X}(\mathcal{C})}}\right]}_{\triangleq \Delta^{(1)}_{\mathcal{C}}},$$

with $\Delta^{(1)}$ defined in (28).

We now evaluate $\Delta^{(1)}_{\mathcal{C}}$ for all possible error events in Table I. For $\mathcal{E}^{(1)}_0$: Q is wrong and hence – because of superposition encoding – the most stringent error event is when the messages carried by Q and all the messages superimposed to Q are wrong. In this case $\mathcal{C} = \emptyset$ and thus $\Delta^{(1)}_{\emptyset} = 0$. It can be also easily verified that

$$\begin{aligned}
&\underbrace{\Delta^{(1)}_{\emptyset}}_{\text{for } \mathcal{E}^{(1)}_0} = \underbrace{\Delta^{(1)}_{\{Q\}}}_{\text{for } \mathcal{E}^{(1)}_1} = \underbrace{\Delta^{(1)}_{\{Q,V_2\}}}_{\text{for } \mathcal{E}^{(1)}_2} = \underbrace{\Delta^{(1)}_{\{Q,V_2,U_2\}}}_{\text{for } \mathcal{E}^{(1)}_3} = \underbrace{\Delta^{(1)}_{\{Q,V_1\}}}_{\text{for } \mathcal{E}^{(1)}_4}\\
&= \underbrace{\Delta^{(1)}_{\{Q,V_1,U_1\}}}_{\text{for } \mathcal{E}^{(1)}_7} = \underbrace{\Delta^{(1)}_{\{Q,V_1,U_1,T_1\}}}_{\text{for } \mathcal{E}^{(1)}_{10}} = \underbrace{\Delta^{(1)}_{\{Q,S_1\}}}_{\text{for } \mathcal{E}^{(1)}_{13}} = 0,
\end{aligned}$$

because in these cases $P_{\mathbf{X}(\mathcal{C})} = P^{(g)}_{\mathbf{X}(\mathcal{C})}$. Then we have:

$$\begin{aligned}
\mathcal{E}^{(1)}_5 : \Delta^{(1)}_{Q,V_1,V_2} &= \mathbb{E}\left[\log \frac{P_{QV_1V_2}}{P_QP_{V_1|Q}P_{V_2|Q}}\right]\\
&= I(V_1 \wedge V_2|Q),
\end{aligned}$$

$$\begin{aligned}
\mathcal{E}^{(1)}_6 : \Delta^{(1)}_{Q,V_1,V_2,U_2} &= \mathbb{E}\left[\log \frac{P_{QV_1V_2U_2}}{P_QP_{V_1|Q}P_{V_2U_2|Q}}\right]\\
&= I(V_1 \wedge V_2,U_2|Q),
\end{aligned}$$

$$\begin{aligned}
\mathcal{E}^{(1)}_8 : \Delta^{(1)}_{Q,V_1,U_1,V_2} &= \mathbb{E}\left[\log \frac{P_{QV_1U_1V_2}}{P_QP_{V_1U_1|Q}P_{V_2|Q}}\right]\\
&= I(V_1,U_1 \wedge V_2|Q),
\end{aligned}$$

$$\begin{aligned}
\mathcal{E}^{(1)}_9 : \Delta^{(1)}_{Q,V_1,U_1,V_2,U_2} &= \mathbb{E}\left[\log \frac{P_{QV_1U_1V_2U_2}}{P_QP_{V_1U_1|Q}P_{V_2U_2|Q}}\right]\\
&= I(V_1,U_1 \wedge V_2,U_2|Q),
\end{aligned}$$

$$\begin{aligned}
\mathcal{E}^{(1)}_{11} : \Delta^{(1)}_{Q,V_1,U_1,T_1,V_2} &= \mathbb{E}\left[\log \frac{P_{QV_1U_1T_1V_2}}{P_QP_{V_1U_1T_1|Q}P_{V_2|Q}}\right]\\
&= I(V_1,U_1,T_1 \wedge V_2|Q),
\end{aligned}$$

$$\begin{aligned}
\mathcal{E}^{(1)}_{12} : \Delta^{(1)}_{Q,V_1,U_1,T_1,V_2,U_2} &= \mathbb{E}\left[\log \frac{P_{QV_1U_1T_1V_2U_2}}{P_QP_{V_1U_1T_1|Q}P_{V_2U_2|Q}}\right]\\
&= I(V_1,U_1,T_1 \wedge V_2,U_2|Q),
\end{aligned}$$

$$\begin{aligned}
\mathcal{E}^{(1)}_{14} : \Delta^{(1)}_{Q,S_1,V_2} &= \mathbb{E}\left[\log \frac{P_{QS_1V_2}}{P_QP_{S_1|Q}P_{V_2|Q}}\right]\\
&= I(S_1 \wedge V_2|Q),
\end{aligned}$$

$$\begin{aligned}
\mathcal{E}^{(1)}_{15} : \Delta^{(1)}_{Q,S_1,V_2,U_2} &= \mathbb{E}\left[\log \frac{P_{QS_1V_2U_2}}{P_QP_{S_1|Q}P_{V_2U_2|Q}}\right]\\
&= I(S_1 \wedge V_2,U_2|Q),
\end{aligned}$$

$$\begin{aligned}
\mathcal{E}^{(1)}_{16} : \Delta^{(1)}_{Q,S_1,V_1} &= \mathbb{E}\left[\log \frac{P_{QS_1V_1}}{P_QP_{S_1|Q}P_{V_1|Q}}\right]\\
&= I(S_1 \wedge V_1|Q),
\end{aligned}$$

$$\begin{aligned}
\mathcal{E}^{(1)}_{17} : \Delta^{(1)}_{Q,S_1,V_1,V_2} &= \mathbb{E}\left[\log \frac{P_{QS_1V_1V_2}}{P_QP_{S_1|Q}P_{V_1|Q}P_{V_2|Q}}\right]\\
&= \mathbb{E}\left[\log \frac{P_{V_1|QS_1}P_{V_2|QS_1V_1}}{P_{V_1|Q}P_{V_2|Q}}\right]\\
&= I(S_1 \wedge V_1|Q) + I(S_1,V_1 \wedge V_2|Q),
\end{aligned}$$

$$\mathcal{E}_{18}^{(1)} : \Delta_{Q,S_1,V_1,V_2,U_2}^{(1)} = \mathbb{E}\left[\log \frac{P_{QS_1V_1V_2U_2}}{P_QP_{S_1|Q}P_{V_1|Q}P_{V_2U_2|Q}}\right] = I(S_1 \wedge V_1|Q) + I(S_1, V_1 \wedge V_2, U_2|Q),$$

$$\mathcal{E}_{19}^{(1)} : \Delta_{Q,S_1,V_1,U_1}^{(1)} = \mathbb{E}\left[\log \frac{P_{QS_1V_1U_1}}{P_QP_{S_1|Q}P_{V_1U_1|Q}}\right] = I(S_1 \wedge V_1, U_1|Q),$$

$$\mathcal{E}_{20}^{(1)} : \Delta_{Q,S_1,V_1,U_1,V_2}^{(1)} = \mathbb{E}\left[\log \frac{P_{QS_1V_1U_1V_2}}{P_QP_{S_1|Q}P_{V_1U_1|Q}P_{V_2|Q}}\right] = I(S_1 \wedge V_1, U_1|Q) + I(S_1, V_1, U_1 \wedge V_2|Q),$$

$$\mathcal{E}_{21}^{(1)} : \Delta_{Q,S_1,V_1,U_1,V_2,U_2}^{(1)} = \mathbb{E}\left[\log \frac{P_{QS_1V_1U_1V_2U_2}}{P_QP_{S_1|Q}P_{V_1U_1|Q}P_{V_2U_2|Q}}\right] = I(S_1 \wedge V_1, U_1|Q) + I(S_1, V_1, U_1 \wedge V_2, U_2|Q),$$

$$\mathcal{E}_{22}^{(1)} : \Delta_{Q,S_1,Z_1,V_1}^{(1)} = \mathbb{E}\left[\log \frac{P_{QS_1Z_1V_1}}{P_QP_{S_1|Q}P_{V_1|Q}P_{Z_1|QS_1V_1}}\right] = \mathbb{E}\left[\log \frac{P_{V_1|QS_1}}{P_{V_1|Q}}\right] = I(S_1 \wedge V_1|Q),$$

$$\begin{aligned}\mathcal{E}_{23}^{(1)} : \Delta_{Q,S_1,Z_1,V_2,V_1}^{(1)} &= \mathbb{E}\left[\log \frac{P_{QS_1Z_1V_1V_2}}{P_QP_{S_1|Q}P_{V_1|Q}P_{Z_1|QS_1V_1}P_{V_2|Q}}\right] \\ &= \mathbb{E}\left[\log \frac{P_{V_1|QS_1}P_{V_2|QS_1Z_1V_1}}{P_{V_1|Q}P_{V_2|Q}}\right] \\ &= I(S_1 \wedge V_1|Q) + I(V_2 \wedge S_1, Z_1, V_1|Q),\end{aligned}$$

$$\begin{aligned}\mathcal{E}_{24}^{(1)} : \Delta_{Q,S_1,Z_1,V_2,U_2,V_1}^{(1)} &= \mathbb{E}\left[\log \frac{P_{QS_1Z_1V_1V_2U_2}}{P_QP_{S_1|Q}P_{V_1|Q}P_{Z_1|QS_1V_1}P_{V_2U_2|Q}}\right] \\ &= I(S_1 \wedge V_1|Q) + I(V_2, U_2 \wedge S_1, Z_1, V_1|Q),\end{aligned}$$

$$\begin{aligned}\mathcal{E}_{25}^{(1)} : \Delta_{Q,S_1,Z_1,V_1,U_1}^{(1)} &= \mathbb{E}\left[\log \frac{P_{QS_1Z_1V_1U_1}}{P_QP_{S_1|Q}P_{V_1U_1|Q}P_{Z_1|QS_1V_1}}\right] \\ &= \mathbb{E}\left[\log \frac{P_{V_1|QS_1}P_{U_1|QS_1V_1Z_1}}{P_{V_1|Q}P_{U_1|QV_1}}\right] \\ &= I(S_1 \wedge V_1|Q) + I(U_1 \wedge S_1, Z_1|Q, V_1),\end{aligned}$$

$$\begin{aligned}\mathcal{E}_{26}^{(1)} : \Delta_{Q,S_1,Z_1,V_2,V_1,U_1}^{(1)} &= \mathbb{E}\left[\log \frac{P_{QS_1Z_1V_1U_1V_2}}{P_QP_{S_1|Q}P_{V_1U_1|Q}P_{Z_1|QS_1V_1}P_{V_2|Q}}\right] \\ &= I(S_1 \wedge V_1|Q) + I(U_1 \wedge S_1, Z_1|Q, V_1) \\ &\quad + I(V_2 \wedge S_1, V_1, Z_1, U_1|Q),\end{aligned}$$

$$\begin{aligned}\mathcal{E}_{27}^{(1)} : \Delta_{Q,S_1,Z_1,V_2,U_2,V_1,U_1}^{(1)} &= \mathbb{E}\left[\log \frac{P_{QS_1Z_1V_1U_1V_2U_2}}{P_QP_{S_1|Q}P_{V_1U_1|Q}P_{Z_1|QS_1V_1}P_{V_2U_2|Q}}\right] \\ &= I(S_1 \wedge V_1|Q) + I(U_1 \wedge S_1, Z_1|Q, V_1) \\ &\quad + I(V_2, U_2 \wedge S_1, V_1, Z_1, U_1|Q).\end{aligned}$$

Let

$$\begin{aligned}
& & R_Q &= R_{10c} + R_{20c} \\
R'_{V_1} &= R'_{10c}, & R_{V_1} &= R_{10c} + R'_{10c} \\
R'_{V_2} &= R'_{20c}, & R_{V_2} &= R_{20c} + R'_{20c} \\
& & R_{U_1} &= R_{10n} + R'_{10n} \\
& & R_{T_1} &= R_{11n} + R'_{11n} \\
R'_{S_1} &= R''_{11c}, & R_{S_1} &= R_{11c} + R''_{11c} \\
R'_{Z_1} &= R'_{11c}, & R_{Z_1} &= R_{11c} + R'_{11c} \\
& & R_{U_2} &= R_{20n} + R'_{20n} \\
& & R_{T_2} &= R_{22n} + R'_{22n} \\
R'_{S_2} &= R''_{22c}, & R_{S_2} &= R_{22c} + R''_{22c} \\
R'_{Z_2} &= R'_{22c}, & R_{Z_2} &= R_{22c} + R'_{22c},
\end{aligned}$$

and hence

$$\begin{aligned}R_{V_1} + R_{V_2} &= R_Q + (R'_{V_1} + R'_{V_2}) \\ R_{Z_1} &= R_{S_1} - R'_{S_1} + R'_{Z_1} \\ R_{Z_2} &= R_{S_2} - R'_{S_2} + R'_{Z_2}.\end{aligned}$$

With these definition we have that the rate constraints arising from decoding at destination 1 are as in (22) and where the quantities $E_\ell^{(1)}$, for $\ell \in \{0, ..., 27\}$, are listed in Table II.

Remark E.1. *Subsets of the above achievable region with less rate constraints can be obtained as follows:*

- *If $R'_{V_1} = R'_{V_2} = 0$, that is, V_1 and V_2 are not binned against the known interference, then V_1 and V_2 are correct whenever Q is correct. In this case, 16 of the 31 error events listed in Table I are impossible (all those for which the bin index in either V_1 or V_2 is wrong).*
- *If $R'_{Z_1} = 0$ (similar observation can be made if $R'_{Z_2} = 0$), that is, Z_1 is not binned against the known interference, then Z_1 is corect whenever Q and V_1 are correct. In this case, the 9 error events from $\mathcal{E}_{13}^{(1)}$ to $\mathcal{E}_{21}^{(1)}$ listed in Table I are impossible and the achievable region becomes*

$$R_{V_1} + R_{V_2} + R_{U_1} + R_{T_1} + R_{U_2} \le E_0^{(1)} \tag{29a}$$

$$R_{U_1} + R_{T_1} + R_{U_2} \le \min\{E_1^{(1)}, E_2^{(1)}, E_4^{(1)}, E_5^{(1)}, E_{22}^{(1)}, E_{23}^{(1)}\} \tag{29b}$$

$$R_{U_1} + R_{T_1} \le \min\{E_3^{(1)}, E_6^{(1)}, E_{24}^{(1)}\} \tag{29c}$$

$$R_{T_1} + R_{U_2} \le \min\{E_7^{(1)}, E_8^{(1)} E_{25}^{(1)}, E_{26}^{(1)}\} \tag{29d}$$

$$R_{T_1} \le \min\{E_9^{(1)}, E_{27}^{(1)}\}, \tag{29e}$$

with only five rate constraints, as for the case of superposition only. Notice that the rate bound on R_{U_2} can be dropped since an error on U_2 alone is not an error from the point of view of source 1.

- *Instead of joint decoding of all the messages, one can perform a two-step decoding as follows.*
  First step: decode Q and S_1 jointly, and then strip them from the received signal. This the first decoding step is successful if

$$\begin{aligned}R_{S_1} &\le I(Y_3 \wedge S_1|Q) \\ R_Q + R_{S_1} &\le I(Y_3 \wedge S_1, Q).\end{aligned}$$

Second step: jointly decode all the other messages . For this second step, one only needs to consider the error events from $\mathcal{E}_{13}^{(1)}$ *to* $\mathcal{E}_{27}^{(1)}$*. This would reduce the number of constraints in the achievable region.*

TABLE I
ERROR EVENTS AT DESTINATION 1.

	1 $[q_1, q_2]$	2 $[1, b_{v_1}]$	3 $[u_1, b_{u_1}]$	4 $[t_1, b_{t_1}]$	5 $[s_1, b_{s_1}]$	6 $[1, b_{z_1}]$	7 $[1, b_{v_2}]$	8 $[u_2, b_{u_2}]$	N	$\mathcal{C}$
$\mathcal{E}_0^{(1)}$	1	*	*	*	*	*	*	*	2^7	$\emptyset$
$\mathcal{E}_1^{(1)}$	0	1	*	*	1	*	1	*	2^4	Q
$\mathcal{E}_2^{(1)}$	0	1	*	*	1	*	0	1	2^3	Q, V_2
$\mathcal{E}_3^{(1)}$	0	1	*	*	1	*	0	0	2^3	Q, V_2, U_2
$\mathcal{E}_4^{(1)}$	0	0	1	*	1	*	1	*	2^3	Q, V_1
$\mathcal{E}_5^{(1)}$	0	0	1	*	1	*	0	1	2^2	Q, V_2, V_1
$\mathcal{E}_6^{(1)}$	0	0	1	*	1	*	0	0	2^2	Q, V_2, U_2, V_1
$\mathcal{E}_7^{(1)}$	0	0	0	1	1	*	1	*	2^2	Q, V_1, U_1
$\mathcal{E}_8^{(1)}$	0	0	0	1	1	*	0	1	2^1	Q, V_2, V_1, U_1
$\mathcal{E}_9^{(1)}$	0	0	0	1	1	*	0	0	2^1	Q, V_2, U_2, V_1, U_1
$\mathcal{E}_{10}^{(1)}$	0	0	0	0	1	*	1	*	2^2	Q, V_1, U_1, T_1
$\mathcal{E}_{11}^{(1)}$	0	0	0	0	1	*	0	1	2^1	Q, V_2, V_1, U_1, T_1
$\mathcal{E}_{12}^{(1)}$	0	0	0	0	1	*	0	0	2^1	$Q, V_2, U_2, V_1, U_1, T_1$
$\mathcal{E}_{13}^{(1)}$	0	1	*	*	0	*	1	*	2^4	Q, S_1
$\mathcal{E}_{14}^{(1)}$	0	1	*	*	0	*	0	1	2^3	Q, S_1, V_2
$\mathcal{E}_{15}^{(1)}$	0	1	*	*	0	*	0	0	2^3	Q, S_1, V_2, U_2
$\mathcal{E}_{16}^{(1)}$	0	0	1	*	0	1	1	*	2^2	Q, S_1, V_1
$\mathcal{E}_{17}^{(1)}$	0	0	1	*	0	1	0	1	2^1	Q, S_1, V_2, V_1
$\mathcal{E}_{18}^{(1)}$	0	0	1	*	0	1	0	0	2^1	Q, S_1, V_2, U_2, V_1
$\mathcal{E}_{19}^{(1)}$	0	0	0	1	0	1	1	*	2^1	Q, S_1, V_1, U_1
$\mathcal{E}_{20}^{(1)}$	0	0	0	1	0	1	0	1	2^0	Q, S_1, V_2, V_1, U_1
$\mathcal{E}_{21}^{(1)}$	0	0	0	1	0	1	0	0	2^0	$Q, S_1, V_2, U_2, V_1, U_1$
$\mathcal{E}_{22}^{(1)}$	0	0	1	*	0	0	1	*	2^2	Q, S_1, Z_1, V_1
$\mathcal{E}_{23}^{(1)}$	0	0	1	*	0	0	0	1	2^1	Q, S_1, Z_1, V_2, V_1
$\mathcal{E}_{24}^{(1)}$	0	0	1	*	0	0	0	0	2^1	$Q, S_1, Z_1, V_2, U_2, V_1$
$\mathcal{E}_{25}^{(1)}$	0	0	0	1	0	0	1	*	2^1	Q, S_1, Z_1, V_1, U_1
$\mathcal{E}_{26}^{(1)}$	0	0	0	1	0	0	0	1	2^0	$Q, S_1, Z_1, V_2, V_1, U_1$
$\mathcal{E}_{27}^{(1)}$	0	0	0	1	0	0	0	0	2^0	$Q, S_1, Z_1, V_2, U_2, V_1, U_1$
OK	0	0	0	0	0	*	*	*	2^3	all correct

TABLE II
RATE BOUNDS AT DESTINATION 1.

$$
\begin{aligned}
&\mathcal{E}_0^{(1)} : E_0^{(1)} = I(Y_3 \wedge Q, V_1, U_1, T_1, S_1, Z_1, V_2, U_2) - (R'_{S_1}) + \Delta^{(1)}\\
&\mathcal{E}_1^{(1)} : E_1^{(1)} = I(Y_3 \wedge V_1, U_1, T_1, S_1, Z_1, V_2, U_2|Q) - (R'_{V_1} + R'_{S_1} + R'_{V_2}) + \Delta^{(1)}\\
&\mathcal{E}_2^{(1)} : E_2^{(1)} = I(Y_3 \wedge V_1, U_1, T_1, S_1, Z_1, U_2|Q, V_2) - (R'_{V_1} + R'_{S_1}) + \Delta^{(1)}\\
&\mathcal{E}_3^{(1)} : E_3^{(1)} = I(Y_3 \wedge V_1, U_1, T_1, S_1, Z_1|Q, V_2, U_2) - (R'_{V_1} + R'_{S_1}) + \Delta^{(1)}\\
&\mathcal{E}_4^{(1)} : E_4^{(1)} = I(Y_3 \wedge U_1, T_1, S_1, Z_1, V_2, U_2|Q, V_1) - (R'_{S_1} + R'_{V_2}) + \Delta^{(1)}\\
&\mathcal{E}_5^{(1)} : E_5^{(1)} = I(Y_3 \wedge U_1, T_1, S_1, Z_1, U_2|Q, V_1, V_2) - (R'_{S_1}) + \Delta^{(1)} - I(V_1 \wedge V_2|Q)\\
&\mathcal{E}_6^{(1)} : E_6^{(1)} = I(Y_3 \wedge U_1, T_1, S_1, Z_1|Q, V_1, V_2, U_2) - (R'_{S_1}) + \Delta^{(1)} - I(V_1 \wedge V_2, U_2|Q)\\
&\mathcal{E}_7^{(1)} : E_7^{(1)} = I(Y_3 \wedge T_1, S_1, Z_1, V_2, U_2|Q, V_1, U_1) - (R'_{S_1} + R'_{V_2}) + \Delta^{(1)}\\
&\mathcal{E}_8^{(1)} : E_8^{(1)} = I(Y_3 \wedge T_1, S_1, Z_1, U_2|Q, V_1, U_1, V_2) - (R'_{S_1}) + \Delta^{(1)} - I(V_1, U_1 \wedge V_2|Q)\\
&\mathcal{E}_9^{(1)} : E_9^{(1)} = I(Y_3 \wedge T_1, S_1, Z_1|Q, V_1, U_1, V_2, U_2) - (R'_{S_1}) + \Delta^{(1)} - I(V_1, U_1 \wedge V_2, U_2|Q)\\
&\mathcal{E}_{10}^{(1)} : E_{10}^{(1)} = I(Y_3 \wedge S_1, Z_1, V_2, U_2|Q, V_1, U_1, T_1) - (R'_{S_1} + R'_{V_2}) + \Delta^{(1)}\\
&\mathcal{E}_{11}^{(1)} : E_{11}^{(1)} = I(Y_3 \wedge S_1, Z_1, U_2|Q, V_1, U_1, T_1, V_2) - (R'_{S_1}) + \Delta^{(1)} - I(V_1, U_1, T_1 \wedge V_2|Q)\\
&\mathcal{E}_{12}^{(1)} : E_{12}^{(1)} = I(Y_3 \wedge S_1, Z_1|Q, V_1, U_1, T_1, V_2, U_2) - (R'_{S_1}) + \Delta^{(1)} - I(V_1, U_1, T_1 \wedge V_2, U_2|Q)\\
&\mathcal{E}_{13}^{(1)} : E_{13}^{(1)} = I(Y_3 \wedge V_1, U_1, T_1, Z_1, V_2, U_2|Q, S_1) - (R'_{V_1} + R'_{Z_1} + R'_{V_2}) + \Delta^{(1)},\\
&\mathcal{E}_{14}^{(1)} : E_{14}^{(1)} = I(Y_3 \wedge V_1, U_1, T_1, Z_1, U_2|Q, S_1, V_2) - (R'_{V_1} + R'_{Z_1}) + \Delta^{(1)} - I(S_1 \wedge V_2|Q),\\
&\mathcal{E}_{15}^{(1)} : E_{15}^{(1)} = I(Y_3 \wedge V_1, U_1, T_1, Z_1|Q, S_1, V_2, U_2) - (R'_{V_1} + R'_{Z_1}) + \Delta^{(1)} - I(S_1 \wedge V_2, U_2|Q)\\
&\mathcal{E}_{16}^{(1)} : E_{16}^{(1)} = I(Y_3 \wedge U_1, T_1, Z_1, V_2, U_2|Q, S_1, V_1) - (R'_{Z_1} + R'_{V_2}) + \Delta^{(1)} - I(S_1 \wedge V_1|Q),\\
&\mathcal{E}_{17}^{(1)} : E_{17}^{(1)} = I(Y_3 \wedge U_1, T_1, Z_1, U_2|Q, S_1, V_1, V_2) - (R'_{Z_1}) + \Delta^{(1)} - I(S_1 \wedge V_1|Q) - I(S_1, V_1 \wedge V_2|Q),\\
&\mathcal{E}_{18}^{(1)} : E_{18}^{(1)} = I(Y_3 \wedge U_1, T_1, Z_1|Q, S_1, V_1, V_2, U_2) - (R'_{Z_1}) + \Delta^{(1)} - I(S_1 \wedge V_1|Q) - I(S_1, V_1 \wedge V_2, U_2|Q)\\
&\mathcal{E}_{19}^{(1)} : E_{19}^{(1)} = I(Y_3 \wedge T_1, Z_1, V_2, U_2|Q, S_1, V_1, U_1) - (R'_{Z_1} + R'_{V_2}) + \Delta^{(1)} - I(S_1 \wedge V_1, U_1|Q),\\
&\mathcal{E}_{20}^{(1)} : E_{20}^{(1)} = I(Y_3 \wedge T_1, Z_1, U_2|Q, S_1, V_1, U_1, V_2) - (R'_{Z_1}) + \Delta^{(1)} - I(S_1 \wedge V_1, U_1|Q) - I(S_1, V_1, U_1 \wedge V_2|Q),\\
&\mathcal{E}_{21}^{(1)} : E_{21}^{(1)} = I(Y_3 \wedge T_1, Z_1|Q, S_1, V_1, U_1, V_2, U_2) - (R'_{Z_1}) + \Delta^{(1)} - I(S_1 \wedge V_1, U_1|Q) - I(S_1, V_1, U_1 \wedge V_2, U_2|Q)\\
&\mathcal{E}_{22}^{(1)} : E_{22}^{(1)} = I(Y_3 \wedge U_1, T_1, V_2, U_2|Q, S_1, Z_1, V_1) - (R'_{V_2}) + \Delta^{(1)} - I(S_1 \wedge V_1|Q),\\
&\mathcal{E}_{23}^{(1)} : E_{23}^{(1)} = I(Y_3 \wedge U_1, T_1, U_2|Q, S_1, Z_1, V_1, V_2) + \Delta^{(1)} - I(S_1 \wedge V_1|Q) - I(S_1, Z_1, V_1 \wedge V_2|Q),\\
&\mathcal{E}_{24}^{(1)} : E_{24}^{(1)} = I(Y_3 \wedge U_1, T_1|Q, S_1, Z_1, V_1, V_2, U_2) + \Delta^{(1)} - I(S_1 \wedge V_1|Q) - I(S_1, Z_1, V_1 \wedge V_2, U_2|Q)\\
&\mathcal{E}_{25}^{(1)} : E_{25}^{(1)} = I(Y_3 \wedge T_1, V_2, U_2|Q, S_1, Z_1, V_1, U_1) - (R'_{V_2}) + \Delta^{(1)} - I(S_1 \wedge V_1|Q) - I(S_1, Z_1 \wedge U_1|Q),\\
&\mathcal{E}_{26}^{(1)} : E_{26}^{(1)} = I(Y_3 \wedge T_1, U_2|Q, S_1, Z_1, V_1, U_1, V_2) + \Delta^{(1)} - I(S_1 \wedge V_1|Q) - I(S_1, Z_1 \wedge U_1|Q) - I(S_1, Z_1, V_1, U_1 \wedge V_2|Q),\\
&\mathcal{E}_{27}^{(1)} : E_{27}^{(1)} = I(Y_3 \wedge T_1|Q, S_1, Z_1, V_1, U_1, V_2, U_2) + \Delta^{(1)} - I(S_1 \wedge V_1|Q) - I(S_1, Z_1 \wedge U_1|Q) - I(S_1, Z_1, V_1, U_1 \wedge V_2, U_2|Q)
\end{aligned}
$$

REFERENCES

[1] V. Sreekanth Annapureddy and Venugopal V. Veeravalli. Sum capacity of the gaussian interference channel in the low interference regime. In *Proceedings of ITA Workshop, San Diego, CA, arXiv:0801.0452*, Jan 2008.

[2] S. Avestimehr, S. Diggavi, and D. Tse. Deterministic approach to wireless relay networks. In *The Allerton Conference on Communication, Control, and computing.*, 2007.

[3] V. Cadambe and S. Jafar. Degrees of freedom of wireless networks with relays, feedback, cooperation and full duplex operation. *IEEE Trans. Inform. Theory*, 55(5):2334–2344, May 2009.

[4] Y. Cao, B. Chen, and J. Zhang. A new achievable rate region for interference channels with common information. *Proc. IEEE Wireless Commun. Netw. Conf. (WCNC 07), Hong Kong*, March 2007.

[5] Yi Cao and Biao Chen. An achievable rate region for interference channel with conferencing. In *Proceedings of 2007 IEEE International Symposium on Information Theory*, pages 1251–1255, June 2007.

[6] Yi Cao and Biao Chen. An achievable rate region for interference channel with generalized feedback. In *Submitted to IEEE Trans. Inform. Theory*, 2008.

[7] A. B. Carleial. A case where interference does not reduce capacity. In *IEEE Trans. Inform. Theory*, volume 21(5), pages 569–570, Sept 1975.

[8] A. B. Carleial. Interference channels. In *IEEE Trans. Inform. Theory*, volume 24(1), pages 60–70, Jan 1978.

[9] H.F. Chong, M. Motani, and H.K. Garg. Generalized backward decoding strategies for the relay channel. *IEEE Trans. Inform. Theory*, 53(1):394 – 401, Jan 2007.

[10] H.F. Chong, M. Motani, H.K. Garg, and H. El Gamal. On the han-kobayashi region for the interference channel. *IEEE Trans. Inform. Theory*, 54(7):3188–3195, 2008.

[11] M. H. Costa. Writing on dirty paper. In *IEEE Trans. Inform. Theory*, volume 29(3), pages 439 – 441, May 1983.

[12] M. H. M. Costa. On the gaussian interference channel. In *IEEE Trans. Inform. Theory*, volume 31(5), pages 607–615, Sept 1985.

[13] M. H. M. Costa and A. A. El Gamal. The capacity region of the discrete memoryless interference channel with strong interference. In *IEEE Trans. Inform. Theory*, volume 33(5), pages 710–711, Sept 1987.

[14] T. M. Cover and A. El Gamal. Capacity theorems for the relay channel. In *IEEE Trans. Inform. Theory*, volume 25(5), page 572584, Sept. 1979.

[15] T. M. Cover and C. S. K. Leung. An achievable rate region for the multiple-access channel with feedback. *IEEE Trans. Inform. Theory*, 27(3):292298, May 1981.

[16] T.M. Cover and J. Thomas. *Elements of information theory*. Wiley, New York, 1991.

[17] I. Csiszar and J. Korner. *Coding Theorems for Discrete Memoryless Systems*. Akademiai Kiado, Budapest, 1981.

[18] R. H. Etkin, D. N. C. Tse, and Hua Wang. Gaussian interference channel capacity to within one bit. *IEEE Trans. Inform. Theory*, 54(12):5534 – 5562, 2008.

[19] A. A. El Gamal and M. H. M. Costa. The capacity region of a class of deterministic interference channels. In *IEEE Trans. Inform. Theory*, volume 28(2), pages 343–346, March 1982.

[20] A. El Gamal. The capacity of a class of broadcast channels. In *IEEE Trans. Inform. Theory*, volume IT-25, pages 166–169, Mar 1979.

[21] Abbas El Gamal and Young-Han Kim. Lecture notes on network information theory. *online at arXiv:1001.3404*, 2010.

[22] M. Gastpar and G. Kramer. On noisy feedback for interference channels. In *Asilomar Conference on Signals, Systems, and Computers*, pages 216–220, Oct 2006.

[23] S. I. Gelfand and M. S. Pinsker. Coding for channel with random parameters. *Problem of Control and Information Theory*, 9(1):19–31, 1980.

[24] D. Gunduz and O. Simeone. On the capacity region of a multiple access channel with common messages. *Proceedings of 2010 IEEE International Symposium on Information Theory*, June 2010.

[25] P. Gupta and P.R. Kumar. Towards an information theory of large networks: an achievable rate region. *IEEE Trans. Inform. Theory*, 49(8):1877–1894, Aug 2001.

[26] T. S. Han. The capacity region of general multiple-access channel with certain correlated sources. *Inform. and Cont.*, 40(1), Jan 1979.

[27] T. S. Han and K. Kobayashi. A new achievable rate region for the interference channel. In *IEEE Trans. Inform. Theory*, volume 27(1), pages 49 –60, Jan 1981.

[28] A. P. Hekstra and F. M. J. Willems. Dependence balance bounds for single-output two-way channels. *IEEE Trans. Inform. Theory*, 35(1):44–53, 1989.

[29] A. Host-Madsen. Capacity bounds for cooperative diversity. In *IEEE Trans. Inform. Theory*, volume 52(4), pages 1522 – 1544, April 2006.

[30] A. Host-Madsen and A. Nosratinia. The multiplexing gain of wireless networks. In *IEEE International Symposium on Information Theory*, pages 2065 – 2069, July 2005.

[31] J. Jiang, Y. Xin, and H. K. Garg. Discrete memoryless interference channels with feedback. In *Proc. 41st Annual Conference on Information Sciences and Systems (CISS 2007), , Baltimore, MD, Mar. 1416, 2007*, 2008.

[32] J. Jiang, Y. Xin, and H. K. Garg. Interference channels with common information. *IEEE Trans. Inform. Theory*, 54(1):171187, Jan 2008.

[33] J. Korner and K. Marton. General broadcast channels with degraded message sets. In *IEEE Trans. Inform. Theory*, volume IT-23, pages 60–64, Jan 1977.

[34] G. Kramer. Feedback strategies for white gaussian interference networks. *IEEE Trans. Inform. Theory*, 48:1423–1438, June 2002.

[35] G. Kramer. Correction to ”feedback strategies for white gaussian interference networks”, and a capacity theorem for gaussian interference channels with feedback. *IEEE Trans. Inform. Theory*, 50(6):1373 – 1374, June 2004.

[36] G. Kramer. Outer bounds on the capacity of gaussian interference channels. In *IEEE Trans. Inform. Theory*, volume 50(3), pages 581–586, March 2004.

[37] G. Kramer. Topics in multi-user information theory. *Foundations and Trends in Communications and Information Theory*, 4(4-5):265–444, 2007.

[38] L. Li and A.J. Goldsmith. Capacity and optimal resource allocation for fading broadcast channels .i. ergodic capacity. *IEEE Trans. Inform. Theory*, 47(3):1083 – 1102, March 2001.

[39] Sung Hoon Lim, Young-Han Kim, Abbas El Gamal, and Sae-Young Chung. Noisy network coding. *submitted to IEEE Transactions on Information Theory*, 2010. online at arXiv:1002.3188.

[40] Wei Liu and Biao Chen. Interference channels with arbitrarily correlated sources. *Communication, Control, and Computing, 2009. Allerton 2009. 47th Annual Allerton Conference on*, Sept 2009.

[41] I. Maric, A. Goldsmith, G. Kramer, and S. Shamai (Shitz). On the capacity of interference channels with one cooperating transmitter. In *European Transactions on Telecommunications*, 2007.

[42] I. Maric, R. D. Yates, and G. Kramer. The strong interference channel with unidirectional cooperation. In *Proceedings of the UCSD Workshop on Information Theory and its Applications, San Diego, CA*, Feb 2006.

[43] I. Maric, R. D. Yates, and G. Kramer. Capacity of interference channels with partial transmitter cooperation. *IEEE Trans. Inform. Theory*, 53(10):3536 – 3548, 2007.

[44] K. Marton. A coding theorem for the discrete memoryless broadcast channel. *IEEE Trans. Inform. Theory*, 25(3):306– 311, May 1979.

[45] A. S. Motahari and A. K. Khandani. Capacity bounds for the gaussian interference channel. *IEEE Trans. Inform. Theory*, 55(2):620–643, 2009.

[46] Vinod Prabhakaran and Pramod Viswanath. Interference channels with source cooperation. *submitted to IEEE Trans. Info. Theory in May 2009, Arxiv preprint arXiv:0905.3109v1*, 2009.

[47] S. Rini, D. Tuninetti, and N. Devroye. New inner and outer bounds for the discrete memoryless cognitive interference channel and some capacity results. *Submitted to IT in March 2010, arxiv preprint arXiv:1003.4328*, 2010.

[48] S. Rini, D. Tuninetti, and N. Devroye. New results on the capacity of the gaussian cognitive interference channel. *Proceedings of the 2010 Allerton Conference, Monticello, IL USA, Sept 2010; Arxiv preprint arXiv:1007.1243*, 2010.

[49] R.Knopp and P.A.Humblet. Information capacity and power control in single-cell multiuser communications. In *IEEE International Conference on Communications, 1995 (ICC ’95), ’Gateway to Globalization’*, volume 1, pages 331–335, Seattle, July 1999.

[50] I. Sason. On achievable rate regions for the gaussian interference channel. In *IEEE Trans. Inform. Theory*, volume 50(6), pages 1345–1356, June 2004.

[51] H. Sato. On the capacity region of a discrete two-user channel for strong interference. In *IEEE Trans. Inform. Theory*, volume 24(3), pages 377–379, May 1978.

[52] H. Sato. The capacity of the gaussian interference channel under strong interference (corresp.). *IEEE Trans. Inform. Theory*, 27(6):786 – 788, 1981.

[53] A. Sendonaris, E. Erkip, and B. Aazhang. User cooperation diversity. Part I. System description. In *IEEE Trans. Commun.*, volume 51(1), pages 1927 – 1938, Nov 2003.

[54] C. E. Shannon. The zero-error capacity of a noisy channel. In *IRE Trans. Inform. Theory*, volume IT-2, pages 8–19, Ulm, Germany, June-July 1956.

[55] C. Suh and D. N. C. Tse. Feedback capacity of the gaussian interference channel to within 1.7075 bits: the symmetric case. *IEEE International Symposium on Information Theory, June 2009. Submitted to Transaction on Info.Theory. preprint arXiv:0901.3580*, June 2009.

[56] Ravi Tandon and Sennur Ulukus. Dependence balance based outer bounds for gaussian networks with cooperation and feedback. *Submitted to IEEE Trans. Inform. Theory, preprint arXiv:0812.1857*, Dec 2008.

[57] E. Telatar and D. N. C. Tse. Bounds on the capacity region of a class of interference channels. *IEEE International Symposium on Information Theory, June 2007. Submitted to Transaction on Info.Theory.*, pages 2871–2874, June 2007.

[58] D. Tuninetti. On interference channels with generalized feedback. In *Proceedings of 2006 IEEE International Symposium on Information Theory*, June 2007.

[59] D. Tuninetti. An outer bound region for interference channels with generalized feedback. *Proceedings of the Information Theory and Applications Workshop, 2010, San Diego, CA USA*, Jan 2010.

[60] D. Tuninetti and Y. Weng. On gaussian mixed interference channels. In *IEEE International Symposium Information Theory*, Toronto, Canada, July 2008.

[61] A. Vahid and S. Avestimehr. The two-user deterministic interference channel with rate-limited feedback. In *Proceedings of 2009 IEEE International Symposium on Information Theory, preprint on arXiv:1004.5132*, 2010.

[62] R. Venkataramani, G. Kramer, and V.K. Goyal. Multiple description coding with many channels. *IEEE Trans. Inform. Theory*, 49(9), Sept 2003.

[63] I-Hsiang Wang and D. Tse. Interference mitigation through limited receiver cooperation. In *submitted to IEEE Trans. Inform. Theory*, 2009.

[64] F. M. J. Willems. *Information theoretical Results for Multiple Access Channels*. Ph.d. dissertation, K.U. Leuven, 1982.

[65] Wei Wu, SriramVishwanath, and Ari Arapostathis. On the capacityof multiple access channels with state information and feedback. *Online at www.arxiv.org/pdf/cs/0606014*, 2006.

[66] Wei Wu, Sriram Vishwanath, and Ari Arapostathis. On the capacity of interference channel with degraded message sets. In *Submitted to IEEE Trans. Inform. Theory*, 2003.

[67] X.Shang, G.Kramer, and B.Chen. A new outer bound and noisy-interference sum-rate capcity for gaussian interference channels. In *Proceedings of 2008 IEEE International Symposium on Information Theory*, volume 2008, June 2008.

[68] S. Yang and D. Tuninetti. A new sum-rate outer bound for gaussian interference channels with generalized feedback. In *Proceedings of 2009 IEEE International Symposium on Information Theory (ISIT 2009), Seoul, South Korea*, June 2009.

[69] Shuang (Echo) Yang and D. Tuninetti. A new achievable region for interference channels with generalized feedback. In *Proceedings of CISS 2008*, March 2008.

[70] C. M. Zeng, F. Kuhlmann, and A. Buzo. Achievability proof of some multiuser channel coding theorems using backward decoding. In *IEEE Trans. Inform. Theory*, volume 35(6), pages 1160–1165, Nov 1989.

PLACE PHOTO HERE

Shuang (Echo) Yang is currently a Ph.D. student with the Department of Electrical and Computer Engineering at the University of Illinois at Chicago, Chicago, IL (USA). She received her M.S. in Electrical Engineering from Beijing Institute of Technology, China. Prior to joining UIC, Mrs. Yang was a Member of Technical Staff with Bell Labs - Lucent Technologies, Mobility R&D, China, in the capacity of software engineer working on CDMA and UMTS projects. Since 2007 she is with Alcatel-Lucent, Naperville, IL USA, where she has been working on UMTS software development and data design.

Her research interests are in user cooperation for ad-hoc/interference wireless networks.

PLACE PHOTO HERE

Daniela Tuninetti Dr Tuninetti received her M.S. in Telecommunication Engineering from Politecnico di Torino, Turin (Italy) in 1998, and her Ph.D. in Electrical Engineering from ENST/Telecom ParisTech, Paris (France) (work done at the Eurecom Institute in Sophia Antipolis, France) in 2002. From 2002 to 2004 she was a postdoctoral research associate at the School of Communication and Computer Science at the EPFL/Swiss Federal Institute of Technology in Lausanne, Lausanne (Switzerland). Since January 2005, she is an assistant professor with the Department of Electrical and Computer Engineering at the University of Illinois at Chicago, Chicago, IL (USA).

Dr Tuninetti was the editor-in-chief of the IEEE Information Theory Society Newsletter from 2006 to 2008, and an associate editor for IEEE Communication Letters from 2006 to 2009. She currently an editor for IEEE Transactions on Wireless Communications. She served on the Technical Program Committee of several IEEE workshops and conferences, and was the Communication Theory symposium co-chair at the IEEE ICC 2010 conference.

Dr Tuninetti received the best student paper award at the European Wireless Conference in 2002, and was the recipient of an NSF CAREER award in 2007.

Her research interests are in the ultimate information theoretic performance of wireless interference networks, with special emphasis on cognition and user cooperation.