Converses For Discrete Memoryless Multiterminal Networks

Amin Aminzadeh Gohari and Venkat Anantharam
Department of Electrical Engineering and Computer Science
University of California, Berkeley
{aminzade,ananth}@eecs.berkeley.edu

Abstract

In this paper we show that the "potential function method" can provide a unified framework for proving converses for problems of determining the rate region of multiterminal communication networks. In this method, for a given network structure, we *simultaneously* consider all possible networks compatible with that structure and think of the rate region as a function from such networks to subsets of the positive orthant. We then identify properties of such a function which would need to be satisfied for it to give rise to an outer bound. The desired outer bound is then proved by a verification argument. This technique also differs from the traditional ones in the single-letterizing step: instead of reducing the n-letter expression to a single-letter expression in *one shot* using time sharing and other auxiliary random variables, we effectively reduce the n-letter expression *inductively* in n steps. This approach is also useful in extending known results for problems with independent sources to ones with dependent sources. To demonstrate this, we apply the technique to recover and further generalize three known results: (1) we generalize the well known cut-set bound to the problem of lossy transmission of functions of dependent sources over a discrete memoryless multiterminal network; (2) we generalize the outer bound part of the recent result of Maric, Yates and Kramer on strong interference channels with a common message to include dependent sources; and (3) we simplify the recent outer bound of Liang, Kramer and Shamai on the capacity region of a general broadcast channel, and generalize it to include dependent sources.

I. Introduction

A general multiterminal network is a model for reliable communication of sets of messages among the nodes of a network, and has been extensively used in modeling of wireless systems. In this paper[1]

[1]Part of this work originally appeared in [6] and [7].

we describe a methodology to study the limitations of joint source-channel coding strategies for lossy transmission across multiterminal networks.

A. Channel model and Communication task

A (discrete memoryless) general multiterminal network (GMN) is characterized by the conditional distribution

$$q(y^{(1)}, y^{(2)}, ..., y^{(m)}|x^{(1)}, x^{(2)}, ..., x^{(m)}),$$

where $X^{(i)}$ and $Y^{(i)}$ $(1 \leq i \leq m)$ are respectively the input and the output of the channel at the i^{th} party. In a GMN with dependent sources, the m nodes are observing i.i.d. repetitions of m, possibly dependent, random variables $W^{(i)}$ for $1 \leq i \leq m$. The i^{th} party $(1 \leq i \leq m)$ has access to the i.i.d. repetitions of $W^{(i)}$, and wants to reconstruct, within a given distortion, the i.i.d. repetitions of a function of all the observations, i.e. $M^{(i)} = f^{(i)}(W^{(1)}, W^{(2)}, ..., W^{(m)})$ for some function $f^{(i)}$. If this is asymptotically possible within a given distortion (see section II for a formal definition), we call the source $(W^{(1)}, W^{(2)}, ..., W^{(m)})$ admissible. Of particular interest is the case when the function $f^{(i)}(W^{(1)}, W^{(2)}, ..., W^{(m)})$ takes the special form of $\big(f^{(i,1)}(W^{(1)}), f^{(i,2)}(W^{(2)}), ..., f^{(i,m)}(W^{(m)})\big)$ for some functions $f^{(i,j)}$. The communication model is rigorously defined in section II.

1) Broadcast and interference channels: In this paper, we will think of a broadcast channel as a three-input/three-output multiterminal network whose inputs are $X^{(1)}$, $X^{(2)}$, $X^{(3)}$, and whose outputs are $Y^{(1)}$, $Y^{(2)}$, $Y^{(3)}$. Furthermore the set of alphabets is assumed to belong to

$$\mathcal{A}_{broadcast} := \big\{\big(\mathcal{X}^{(1)}, \mathcal{X}^{(2)}, \mathcal{X}^{(3)}, \mathcal{Y}^{(1)}, \mathcal{Y}^{(2)}, \mathcal{Y}^{(3)}\big) : |\mathcal{X}^{(2)}| = |\mathcal{X}^{(3)}| = 1\big\},$$

and the conditional law of the network is assumed to belong to

$$\mathcal{Q}_{broadcast} := \big\{q(y^{(1)}, y^{(2)}, y^{(3)}|x^{(1)}, x^{(2)}, x^{(3)}) : H(Y^{(1)}|X^{(1)}) = 0\big\}.$$

Note that, apart from the notational change, the channel we consider is identical to what is traditionally called the broadcast channel. Similarly, we will think of an interference channel as a four-input/four-output multiterminal network whose inputs are $X^{(1)}$, $X^{(2)}$, $X^{(3)}$ and $X^{(4)}$, and whose outputs are $Y^{(1)}$, $Y^{(2)}$, $Y^{(3)}$ and $Y^{(4)}$. The set of alphabets is assumed to belong to

$$\mathcal{A}_{interference} := \big\{\big(\mathcal{X}^{(1)}, \mathcal{X}^{(2)}, \mathcal{X}^{(3)}, \mathcal{X}^{(4)}, \mathcal{Y}^{(1)}, \mathcal{Y}^{(2)}, \mathcal{Y}^{(3)}, \mathcal{Y}^{(4)}\big) : |\mathcal{X}^{(3)}| = |\mathcal{X}^{(4)}| = 1\big\},$$

and the conditional law of the network is assumed to belong to

$$\mathcal{Q}_{interference} := \big\{q(y^{(1)}, y^{(2)}, y^{(3)}, y^{(4)}|x^{(1)}, x^{(2)}, x^{(3)}, x^{(4)}) : H(Y^{(1)}|X^{(1)}) = H(Y^{(2)}|X^{(2)}) = 0\big\}.$$

Apart from notational changes, this is identical to the traditional interference channel.

B. The framework

The admissible source region of a general multiterminal network is not known when the sources are independent except in certain special cases; less is known when the sources are allowed to be arbitrarily correlated. It is known that, unlike the point-to-point scenario, in a network scenario the separation of the source and channel codings is not necessarily optimal [8], [9].

In this paper, we set up a novel framework for proving converses, and then use it to prove new outer bounds to the admissible source region of a GMN, a broadcast channel with two receivers, and an interference channel.

The proof technique is based on the "potential function method" used in [2] and [3] to bound the secrecy capacity from above. Let m and d be natural numbers and $\mathbb{R}_+^d$ the set of all d-tuples of non-negative reals. Intuitively speaking, for a given network structure, we *simultaneously* consider all possible networks compatible with that structure and think of the rate region as a function from such networks to subsets of $\mathbb{R}_+^d$. We then identify properties of such a function which would need to be satisfied in one step of the communication for it to give rise to an outer bound. The outer bound is then proved by a verification argument. Properties that such a function would need to satisfy are identified, intuitively speaking, as follows: take an arbitrary code of length say n over a GMN. During the simulation of the code, the information of the parties begins from the i^{th} party having the i.i.d. repetitions of the random variable $W^{(i)}$, gradually evolves over time with the usage of the network, and eventually after n stages of communication reaches its final state where the parties know enough to estimate their objectives within the desired average distortion. The idea is to quantify this gradual evolution of information, *bound the derivative of the information growth at each stage* from above by showing that one step of communication can buy us at most a certain amount and conclude that at the final stage, i.e. the n^{th} stage, the system cannot reach an information state better than n times the outer bound on the derivative of information growth. An implementation of this idea requires quantification of the information of the m parties at a given stage of the process. To that end, we evaluate the function we started with at a *virtual channel* whose inputs and outputs represent, roughly speaking, the initial and the gained knowledge of the parties at the given stage of the communication. We also need to make sense of the derivative of a region. This is done using Minkowski sums.

Our technique differs from the traditional ones also in the single-letterizing step: the traditional converses begin with the Fano inequality and continue with the single-letterizing step, that is, reducing the n-letter expression to a single-letter expression in *one shot* using time sharing and other auxiliary random variables. However in our technique we effectively reduce the n-letter expression *inductively* in n steps. The i^{th} step will be equivalent to bounding the derivative of the information growth at the i^{th} use of the multiterminal network (see remark 1 of subsection III-A following the proof of the main lemma). The inductive approach to the single-letterizing step is also useful in extending known results for problems with independent sources to ones with dependent sources, as will become clear after understanding the main claims of this paper.

C. New results

In this paper, we prove new outer bounds to the admissible source region of a GMN, a broadcast channel with two receivers, and a strong interference channel.

1) General multiterminal networks: The first result is a generalized cut-set bound. It claims an outer bound to the admissible source region of a GMN. Specializing by requiring zero distortion at the receivers, assuming that the functions $f^{(i)}(W^{(1)}, W^{(2)}, ..., W^{(m)})$ $(1 \leq i \leq m)$ have the form $(f^{(i,1)}(W^{(1)}), ..., f^{(i,m)}(W^{(m)}))$, and that the individual messages $f^{(i,j)}(W^{(j)})$ are mutually independent, our result reduces to the well known cut-set bound. The results can be carried over to the problem of "lossless transmission" for the following reason: requiring the i^{th} party to reconstruct the i.i.d. repetitions of $f^{(i)}(W^{(1)}, W^{(2)}, ..., W^{(m)})$ with arbitrarily small average probability of error is no stronger than requiring the i^{th} party to reconstruct the i.i.d repetitions of $f^{(i)}(W^{(1)}, W^{(2)}, ..., W^{(m)})$ with a vanishing average distortion (for details see section II). Other extensions of the cut-set bound can be found in [14] and [15]. Our formulation is also motivated by some existing works which show the possible benefits of focusing on function computation rather than just raw communication (see for instance [16], [17], [18], [19], [20]).

2) Strong interference channels: We prove a new outer bound to the admissible source region of a strong interference channel with arbitrarily correlated sources. As a special case, we recover the converse part of the capacity region given by Maric, Yates and Kramer [4] for strong interference channels with common information.

An interference channel is said to be *strong* if

$$I(X^{(1)}; Y^{(3)}|X^{(2)}) \leq I(X^{(1)}; Y^{(4)}|X^{(2)}); \tag{1}$$

$$I(X^{(2)}; Y^{(4)}|X^{(1)}) \leq I(X^{(2)}; Y^{(3)}|X^{(1)}), \tag{2}$$

for all product distributions on $\mathcal{X}^{(1)} \times \mathcal{X}^{(2)} := \{(x^{(1)}, x^{(2)}) : x^{(1)} \in \mathcal{X}^{(1)}, x^{(2)} \in \mathcal{X}^{(2)}\}$. This condition then automatically extends to all joint distributions on $\mathcal{X}^{(1)} \times \mathcal{X}^{(2)}$.

Costa and El Gamal found the capacity region of a strong interference channel when the transmitters send independent private messages to their intended receivers [21]. Recently, this result was extended by Maric, Yates and Kramer, who assumed that the transmitters additionally have a common message that needs to be sent to both the receivers [4]. In this problem a new feature arises: the inputs $X^{(1)}$ and $X^{(2)}$ are no longer guaranteed to be independent throughout the communication, since the transmitters have correlated information. Maric, Yates and Kramer found the capacity region to be the union over $p(u, x^{(1)}, x^{(2)}, y^{(3)}, y^{(4)}) = p(u)p(x^{(1)}|u)p(x^{(2)}|u)q(y^{(3)}, y^{(4)}|x^{(1)}, x^{(2)})$, of the set of all non-negative triples (R_0, R_1, R_2) satisfying

$$\begin{aligned} R_1 &\leq I(X^{(1)}; Y^{(3)}|X^{(2)}, U); \\ R_2 &\leq I(X^{(2)}; Y^{(4)}|X^{(1)}, U); \\ R_1 + R_2 &\leq \min(I(X^{(1)}X^{(2)}; Y^{(3)}|U), I(X^{(1)}X^{(2)}; Y^{(4)}|U)); \\ R_0 + R_1 + R_2 &\leq \min(I(X^{(1)}X^{(2)}; Y^{(3)}), I(X^{(1)}X^{(2)}; Y^{(4)})). \end{aligned} \tag{3}$$

In the above expressions R_0 denotes the common message rate, and R_1 and R_2 are respectively the private message rates of the first and second transmitter.

The communication task is formulated as before. In an interference channel with arbitrarily dependent sources, the transmitters are observing i.i.d. repetitions of two, possibly dependent, random variables $W^{(1)}$ and $W^{(2)}$. The transmitters would like to reliably send the i.i.d. copies of $W^{(1)}$ to the receiver $Y^{(3)}$ and the i.i.d. copies of $W^{(2)}$ to the receiver $Y^{(4)}$. To be notationally consistent, we assume that the third and fourth party are observing i.i.d. copies of $W^{(3)}$ and $W^{(4)}$, but that $|\mathcal{W}^{(3)}| = |\mathcal{W}^{(4)}| = 1$ implying that these random variables are constant. Roughly speaking, the source marginal distribution $p(w^{(1)}, w^{(2)}, w^{(3)}, w^{(4)})$ is called admissible if there exists a strategy for reliable transmission of the i.i.d copies of $W^{(1)}$ and $W^{(2)}$ to the intended receivers. A formal definition can be made by setting $M^{(1)} = f^{(1)}(W^{(1)}, W^{(2)}, W^{(3)}, W^{(4)}) = W^{(1)}$, $M^{(2)} = f^{(2)}(W^{(1)}, W^{(2)}, W^{(3)}, W^{(4)}) = W^{(2)}$, $M^{(3)} = f^{(3)}(W^{(1)}, W^{(2)}, W^{(3)}, W^{(4)}) = W^{(1)}$ and $M^{(4)} = f^{(4)}(W^{(1)}, W^{(2)}, W^{(3)}, W^{(4)}) = W^{(2)}$. Note that the definitions of $f^{(1)}$ and $f^{(2)}$ reflect the fact that neither transmitter is expected to recover the message of the other transmitter. We require the i^{th} party to reconstruct the i.i.d repetitions of $f^{(i)}(W^{(1)}, W^{(2)}, W^{(3)}, W^{(4)})$ with a vanishing average distortion. The distortion function used by the i^{th} party, $\Delta^{(i)}(m^{(i)}, m'^{(i)})$ (for $1 \leq i \leq 4$) is taken to be the indicator function $\mathbf{1}[m^{(i)} \neq m'^{(i)}]$. Since we are proving an outer bound, the main result can be carried over to the problem of "lossless transmission" since requiring the i^{th} party to reconstruct the i.i.d. repetitions of $M^{(i)} = f^{(i)}(W^{(1)}, W^{(2)}, W^{(3)}, W^{(4)})$

with arbitrarily small average probability of error is no stronger than requiring the i^{th} party to reconstruct the i.i.d repetitions of $M^{(i)}$ with a vanishing average distortion. Further, our notion of correlated source is strictly more general than that considered in [4]. For example, take three mutually independent binary random variables $L \sim Bern(\frac{1}{2})$, $N_1 \sim Bern(p_1)$ and $N_2 \sim Bern(p_2)$ and let $W^{(1)} = L \oplus N_1$ and $W^{(2)} = L \oplus N_2$ where the addition is modulo two. Since the common part of $W^{(1)}$ and $W^{(2)}$ in the sense of Gács and Körner [22] is a constant random variable, one cannot represent the dependence between the i.i.d. copies of $W^{(1)}$ and $W^{(2)}$ through a random variable that is computable by both the parties, a restriction which is required in [4]. However we only prove an outer bound whereas [4] computes the capacity region. Our outer bound reduces to the region of [4] in the case of correlated sources of the type considered there.

The admissible source region of a strong interference channel with correlated sources is not known except in certain special cases. Inner bounds to the admissible source region are reported in [23] and [25]. We are not aware of any previous work discussing any interesting outer bounds on the admissible source region of an interference channel with dependent sources. It is known that the source$-$channel separation theorem breaks down for multi-access channels [8] and multi-access channels (MACs) are special cases of strong interference channels. One can therefore not expect that finding outer bounds on an interference channel with correlated sources can be resolved by a source$-$channel separation approach. Further, since the problem of determining the admissible source region for a MAC with correlated sources is still unsolved, determining the admissible region for a strong interference channel is a difficult one.

3) Broadcast channel: In a broadcast channel with arbitrarily dependent sources, the transmitter is observing i.i.d. copies of $W^{(1)}$. To be notationally consistent, we assume that the second and third party are observing i.i.d. copies of $W^{(2)}$ and $W^{(3)}$, but that $|\mathcal{W}^{(2)}| = |\mathcal{W}^{(3)}| = 1$, implying that these random variables are constant. The transmitter aims to reliably send i.i.d. copies of $M^{(2)} = f^{(2)}(W^{(1)}, W^{(2)}, W^{(3)}) = f^{(2,1)}(W^{(1)})$ to the second party, and i.i.d. copies of $M^{(3)} = f^{(3,1)}(W^{(1)})$ to the third party. To be notationally consistent, we also set $M^{(1)} = f^{(1,1)}(W^{(1)}) = W^{(1)}$. Roughly speaking, the source marginal distribution $p(w^{(1)}, w^{(2)}, w^{(3)})$ is called admissible if there exists a strategy for reliable transmission of the i.i.d copies of $M^{(2)}$ and $M^{(3)}$ to the intended receivers. Furthermore, we require the receivers to reconstruct the i.i.d repetitions of $M^{(2)}$ and $M^{(3)}$ respectively with a vanishing average distortion. The distortion function used by the i^{th} party, $\Delta^{(i)}(m^{(i)}, m'^{(i)})$ (for $1 \leq i \leq 3$) is taken to be the indicator function $\mathbf{1}[m^{(i)} \neq m'^{(i)}]$.

The best known inner bound for the two receiver general broadcast channel with dependent sources (that is when the random variables $f^{(i,1)}(W^{(1)})$ for $i = 2, 3$ are dependent) is due to Han and Costa [9] (see also [27]). We are not aware of any previous work discussing any interesting outer bounds on the admissible source region of the general broadcast channel when the sources are allowed to be arbitrarily dependent. In this paper, we prove a new outer bound to the admissible source region of the general broadcast channel for dependent sources. Specializing by assuming independent sources, our outer bound reduces to one that is included inside the recent outer bound of Liang, Kramer and Shamai [5]. We don't know if the inclusion is strict. A preliminary version of our result originally appeared in [6]. In [26], Kramer, Liang and Shamai independently derived a similar result. The connection between the two results should become clear after reading section III-E. In that section, we analyze and simplify the earlier outer bound of [5] on the general broadcast channel.

D. *Outline*

The outline of this paper is as follows. In section II, we introduce the basic notation and definitions used in this paper. Section III contains the main results of this paper followed by the formal proofs. Appendix A and its subsections complete the proof of Theorem 1 from section III-B. Appendix B and its subsections complete the proof of Theorem 2 from section III-C. Appendix C and its subsections complete the proof of Theorem 3 from section III-D.

II. Definitions and Notation

Throughout this paper we assume that each random variable takes values in a finite set. $\mathbb{R}$ denotes the set of real numbers and $\mathbb{R}_+$ denotes the set of non-negative reals. For any natural number k, let $[k] = \{1, 2, 3, ..., k\}$. For a set $S \subset [k]$, let S^c denote its compliment, that is $[k] - S$. The context will make the ambient space of S clear.

We represent a GMN by the conditional distribution

$$q(y^{(1)}, y^{(2)}, ..., y^{(m)} | x^{(1)}, x^{(2)}, ..., x^{(m)}),$$

meaning that the input by the i^{th} party is $X^{(i)}$ and the output at the i^{th} party is $Y^{(i)}$. We assume that the i^{th} party ($1 \leq i \leq m$) has access to i.i.d. repetitions of $W^{(i)}$ before the beginning of the communication. The message that needs to be delivered (in a possibly lossy manner) to the i^{th} party is taken to be $M^{(i)} = f^{(i)}(W^{(1)}, W^{(2)}, ..., W^{(m)})$ for some function $f^{(i)}$. We assume that for any $i \in [m]$, random variables $X^{(i)}$, $Y^{(i)}$, $W^{(i)}$ and $M^{(i)}$ take values from finite sets $\mathcal{X}^{(i)}$, $\mathcal{Y}^{(i)}$, $\mathcal{W}^{(i)}$ and $\mathcal{M}^{(i)}$ respectively. For any natural number n, let $(\mathcal{X}^{(i)})^n$, $(\mathcal{Y}^{(i)})^n$, $(\mathcal{W}^{(i)})^n$ and $(\mathcal{M}^{(i)})^n$ denote the n-th product sets of $\mathcal{X}^{(i)}$, $\mathcal{Y}^{(i)}$, $\mathcal{W}^{(i)}$ and $\mathcal{M}^{(i)}$ respectively. We use $Y^{(i)}_{1:k}$ to denote $(Y^{(i)}_1, Y^{(i)}_2, ..., Y^{(i)}_k)$.

TABLE I

NOTATION

Variable	Description
$\mathbb{R}$	Real numbers.
$\mathbb{R}_+$	Non-negative real numbers.
$[k]$	The set $\{1, 2, 3, ..., k\}$.
m	Number of nodes of the network.
$q(y^{(1)}, ..., y^{(m)}\|x^{(1)}, ..., x^{(m)})$	The statistical description of a general multi-terminal network.
$W^{(i)}$	Random variable representing the source observed at the i^{th} node.
$M^{(i)}$	Random variable to be reconstructed, in a possibly lossy way, at the i^{th} node.
$\mathcal{X}^{(i)}, \mathcal{Y}^{(i)}, \mathcal{W}^{(i)}, \mathcal{M}^{(i)}$	Alphabets of $X^{(i)}$, $Y^{(i)}$, $W^{(i)}$, $M^{(i)}$.
$\Delta^{(i)}(\cdot, \cdot)$	Distortion function used by the i^{th} party.
$\zeta_k^{(i)}(\cdot)$ $\vartheta^{(i)}(\cdot)$	The encoding function used by the i^{th} party at the k^{th} stage. The decoding function at the i^{th} party.
n	Length of the code used.
$\Pi(\cdot)$ $\oplus$ $\geq$	Down-set (Definition 4). Minkowski sum of two sets (Definition 3). A vector or a set being greater than or equal another (Definition 4).
Ψ	A permissible set of input distributions; Given input sources and a GMN, Ψ is a set of joint distributions on $\mathcal{X}^{(1)} \times \mathcal{X}^{(2)} \times \mathcal{X}^{(3)} \times \cdots \times \mathcal{X}^{(m)}$. Inputs to the network are required to have a joint distribution belonging to this set.

For any $i \in [m]$, let the distortion function $\Delta^{(i)}$ be a function $\Delta^{(i)} : \mathcal{M}^{(i)} \times \mathcal{M}^{(i)} \to [0, \infty)$ satisfying $\Delta^{(i)}(m^{(i)}, m^{(i)}) = 0$ for all $m^{(i)} \in \mathcal{M}^{(i)}$. For any natural number n and vectors $(m_1^{(i)}, m_2^{(i)}, ..., m_n^{(i)})$ and $(m_1^{'(i)}, m_2^{'(i)}, ..., m_n^{'(i)})$ from $(\mathcal{M}^{(i)})^n$, let

$$\Delta_n^{(i)}(m_{1:n}^{(i)}, m_{1:n}^{'(i)}) = \frac{1}{n} \sum_{k=1}^{n} \Delta^{(i)}(m_k^{(i)}, m_k^{'(i)}).$$

Roughly speaking, we require the i.i.d. repetitions of random variable $M^{(i)}$ to be reconstructed, by the i^{th} party, within the average distortion $D^{(i)}$.

Definition 1: Given natural number n, an *n-code* is the following set of mappings:

$$\text{For any } i \in [m]: \quad \zeta_1^{(i)}: \quad (\mathcal{W}^{(i)})^n \longrightarrow \mathcal{X}^{(i)};$$

$$\text{For any } i \in [m], k \in [n]-\{1\}: \quad \zeta_k^{(i)}: \quad (\mathcal{W}^{(i)})^n \times (\mathcal{Y}^{(i)})^{k-1} \longrightarrow \mathcal{X}^{(i)};$$

$$\text{For any } i \in [m]: \quad \vartheta^{(i)}: \quad (\mathcal{W}^{(i)})^n \times (\mathcal{Y}^{(i)})^n \longrightarrow (\mathcal{M}^{(i)})^n.$$

Intuitively speaking $\zeta_k^{(i)}$ is the encoding function of the i^{th} party at the k^{th} time instance, and $\vartheta^{(i)}$ is the decoding function of the i^{th} party.

Given positive reals ϵ and $D^{(i)}$ ($1 \leq i \leq m$), and a source marginal distribution $p(w^{(1)}, w^{(2)}, ..., w^{(m)})$, an n-code is said to achieve the average distortion at most $D^{(i)} + \epsilon$ (for all $i \in [m]$) over the channel $q(y^{(1)}, y^{(2)}, ..., y^{(m)}|x^{(1)}, x^{(2)}, ..., x^{(m)})$ if the following condition is satisfied:

Assume that random variables $(W_{1:n}^{(i)}, i \in [m])$ are n i.i.d. repetitions of random variables $(W^{(1)}, W^{(2)}, ..., W^{(m)})$ with joint distribution $p(w^{(1)}, w^{(2)}, ..., w^{(m)})$. Random variables $X_k^{(i)}$ and $Y_k^{(i)}$ ($k \in [n]$, $i \in [m]$) are defined according to the following constraints:

$$p(w_{1:n}^{(1)}, w_{1:n}^{(2)}, ..., w_{1:n}^{(m)}, x_{1:n}^{(1)}, x_{1:n}^{(2)}, ..., x_{1:n}^{(m)}, y_{1:n}^{(1)}, y_{1:n}^{(2)}, ..., y_{1:n}^{(m)}) =$$

$$\prod_{k=1}^{n} p(w_k^{(1)}, w_k^{(2)}, ..., w_k^{(m)}) \times \prod_{k=1}^{n} q(y_k^{(1)}, y_k^{(2)}, ..., y_k^{(m)}|x_k^{(1)}, x_k^{(2)}, ..., x_k^{(m)}) \times \prod_{k=1}^{n}\prod_{i=1}^{m} p(x_k^{(i)}|w_{1:n}^{(i)}, y_{1:k-1}^{(i)});$$

so we may write $X_1^{(i)} = \zeta_1^{(i)}\big(W_{1:n}^{(i)}\big)$, and for any $2 \leq k \leq n$, $X_k^{(i)} = \zeta_k^{(i)}\big(W_{1:n}^{(i)}, Y_{1:k-1}^{(i)}\big)$. Random variables $X_k^{(i)}$ and $Y_k^{(i)}$ represent the input and output of the i^{th} party at the k^{th} time instance and satisfy the following Markov chains:

$$W_{1:n}^{(1)}...W_{1:n}^{(m)}Y_{1:k-1}^{(1)}...Y_{1:k-1}^{(m)} \rightarrow W_{1:n}^{(i)}Y_{1:k-1}^{(i)} \rightarrow X_k^{(i)},$$

$$W_{1:n}^{(1)}...W_{1:n}^{(m)}Y_{1:k-1}^{(1)}...Y_{1:k-1}^{(m)} \rightarrow X_k^{(1)}...X_k^{(m)} \rightarrow Y_k^{(1)}...Y_k^{(m)}.$$

Let $M_k^{(i)} = f^{(i)}(W_k^{(1)}, W_k^{(2)}, ..., W_k^{(m)})$. We should then have the following constraint for every $i \in [m]$:

$$\mathbb{E}\Big[\Delta_n^{(i)}\Big(\vartheta^{(i)}\big(W_{1:n}^{(i)}, Y_{1:n}^{(i)}\big), M_{1:n}^{(i)}\Big)\Big] \leq D^{(i)} + \epsilon.$$

Definition 2: Given positive reals $D^{(i)}$, a source marginal distribution $p(w^{(1)}, w^{(2)}, ..., w^{(m)})$ is called an *admissible source* over the channel $q(y^{(1)}, y^{(2)}, ..., y^{(m)}|x^{(1)}, x^{(2)}, ..., x^{(m)})$ for the target reproduction functions $f^{(i)}$ ($i \in [m]$) at the distortion levels $D^{(i)}$ ($i \in [m]$) if for every positive ϵ and sufficiently large n, an n-code achieving the average distortion at most $D^{(i)} + \epsilon$ exists.

Definition 3: For sets K and L of points in $\mathbb{R}_+^d$, let $K \oplus L$ refer to their Minkowski sum: $K \oplus L = \{v_1 + v_2 : v_1 \in K, v_2 \in L\}$. For any real number r, let $r \times K = \{r \cdot v_1 : v_1 \in K\}$. We also define $\frac{K}{r}$

as the set formed by shrinking K through scaling each point of it by a factor $\frac{1}{r}$. Note that in general $r \times K \neq (r_1 \times K) \oplus (r_2 \times K)$ when $r = r_1 + r_2$ but this is true when K is a convex set.

Definition 4: For $\overrightarrow{v_1}$ and $\overrightarrow{v_2}$ in $\mathbb{R}_+^d$, we say $\overrightarrow{v_1} \geq \overrightarrow{v_2}$ if and only if each coordinate of $\overrightarrow{v_1}$ is greater than or equal to the corresponding coordinate of $\overrightarrow{v_2}$. For sets A and B of points in $\mathbb{R}_+^d$, we say $A \leq B$ if and only if for any point $\overrightarrow{a} \in A$, there exists a point $\overrightarrow{b} \in B$ such that $\overrightarrow{a} \leq \overrightarrow{b}$. For a set $A \in \mathbb{R}_+^d$, the down-set $\Pi(A)$ is defined as: $\Pi(A) = \{\overrightarrow{v} \in \mathbb{R}_+^d : \ \overrightarrow{v} \leq \overrightarrow{w} \text{ for some } \overrightarrow{w} \in A\}$.

Definition 5: Given a network $q(y^{(1)}, y^{(2)}, ..., y^{(m)}|x^{(1)}, x^{(2)}, ..., x^{(m)})$, and the source marginal distribution $p(w^{(1)}, w^{(2)}, ..., w^{(m)})$, there may be properties that the inputs to the GMN have to satisfy throughout the communication. For instance in an interference channel or a multiple access channel with no output feedback, if the transmitters observe independent messages, the random variables representing their information stay independent of each other throughout the communication. This is because the transmitters neither interact nor receive any feedback from the outputs. Also, constraints on the set of input distributions when the transmitters are observing i.i.d. copies of correlated random variables are reported in [24]. Other constraints on the inputs to the network might come from practical requirements such as coupled magnitude constraint across inputs in each stage of the communication. Given a multiterminal network $q(y^{(1)}, y^{(2)}, ..., y^{(m)}|x^{(1)}, x^{(2)}, ..., x^{(m)})$ and assuming that $\mathcal{X}^{(i)}$ ($i \in [m]$) is the set $X^{(i)}$ is taking values from, let Ψ be a set of joint distributions on $\mathcal{X}^{(1)} \times \mathcal{X}^{(2)} \times \mathcal{X}^{(3)} \times ... \times \mathcal{X}^{(m)}$ for which the following guarantee exists: for any communication protocol, the inputs to the multiterminal network at each time stage have a joint distribution belonging to the set Ψ. Such a set will be called a *permissible set* of input distributions. Some of the results below will be stated in terms of this nebulously defined region Ψ. To get explicit results, simply replace Ψ by the set of all probability distributions on $\mathcal{X}^{(1)} \times \mathcal{X}^{(2)} \times \mathcal{X}^{(3)} \times ... \times \mathcal{X}^{(m)}$.

Definition 1 can be extended in the obvious way to define the notion of an n-code for a permissible set of input distributions Ψ. Definition 2 can also be extended in the obvious way to define the notion of an admissible source for the target reproduction functions $f^{(i)}$ ($i \in [m]$) at distortion levels $D^{(i)}$ ($i \in [m]$) for a permissible set of input distributions Ψ.

III. Statement of the results

In this section, we formally present the results. Subsection III-B states a generalized cut-set bound that is applicable to all multiterminal networks. Subsections III-C and III-D state our results for broadcast channels and strong interference channels respectively. These results are all applications of the main lemma that is stated in the subsection III-A. In the fifth subsection, we analyze and simplify the recent outer bound of Liang, Kramer and Shamai [5]. Although the analysis is referred to in subsection III-D that deals with broadcast channels with dependent sources, this subsection is stand-alone and may be of independent interest.

A. The Main Lemma

Let m and d be natural numbers. Let $\phi(p(y^{(1)},...,y^{(m)}|x^{(1)},...,x^{(m)}),\Psi)$ be a function that takes as input an arbitrary pair of

(m-input/m-output multiterminal network, a permissible set of input distributions for the network),

consistent with the structure of interest (this will be made precise below) and returns a subset of $\mathbb{R}_+^d$. We will now impose some conditions on ϕ, which we first discuss informally and then formally. Apart from these conditions nothing need be assumed about ϕ; in particular there is no need to impose any regularity conditions.

1) Intuitive discussion of the properties imposed on ϕ: Intuitively speaking, we want the function ϕ to quantify the information state during the simulation of a code: during the simulation of the code, the information of the parties begins from the i^{th} party having $W_{1:n}^{(i)}$ and gradually evolves over time with the use of the network. At the j^{th} stage, the i^{th} party has $W_{1:n}^{(i)}Y_{1:j}^{(i)}$. We represent the information state of the whole system at the j^{th} stage by the virtual channel $p(w_{1:n}^{(1)}y_{1:j}^{(1)},...,w_{1:n}^{(m)}y_{1:j}^{(m)}|w_{1:n}^{(1)},...,w_{1:n}^{(m)})$ and the single admissible input distribution $p(w_{1:n}^{(1)},...,w_{1:n}^{(m)})$. In order to quantify the information state, we map it to a subset of $\mathbb{R}_+^d$ using the function ϕ. We demand that ϕ satisfies two conditions. The first condition is intuitively saying that an additional use of the channel

$$p(y'^{(1)},y'^{(2)},...,y'^{(m)}|x'^{(1)},x'^{(2)},...,x'^{(m)})$$

restricted to input distributions from Ψ can expand ϕ by at most

$$\phi\big(p(y'^{(1)},y'^{(2)},...,y'^{(m)}|x'^{(1)},x'^{(2)},...,x'^{(m)}),\Psi\big).$$

The second condition is intuitively saying that ϕ vanishes if the parties are unable to communicate, that is, if each party receives exactly what it puts at the input of the channel.

Note that all the channels we encounter in the above process (including the virtual channels and the physical channel) fall into a certain class of m-input/m-output multiterminal networks. We can demand that the function ϕ satisfies certain conditions within this class. For instance, assume that the physical channel is a two-receiver broadcast channel. As discussed in the introduction, we think of a broadcast channel as a three-input/three-output multiterminal network whose inputs are $X^{(1)}$, $X^{(2)}$, $X^{(3)}$, and whose outputs are $Y^{(1)}$, $Y^{(2)}$, $Y^{(3)}$. Furthermore the set of alphabets is assumed to belong to

$$\mathcal{A}_{broadcast} := \big\{ \big(\mathcal{X}^{(1)}, \mathcal{X}^{(2)}, \mathcal{X}^{(3)}, \mathcal{Y}^{(1)}, \mathcal{Y}^{(2)}, \mathcal{Y}^{(3)}\big) : |\mathcal{X}^{(2)}| = |\mathcal{X}^{(3)}| = 1 \big\},$$

and the conditional law of the network is assumed to belong to

$$\mathcal{Q}_{broadcast} := \big\{ q(y^{(1)}, y^{(2)}, y^{(3)}|x^{(1)}, x^{(2)}, x^{(3)}) : H(Y^{(1)}|X^{(1)}) = 0 \big\}.$$

Note that when $|\mathcal{W}^{(2)}| = |\mathcal{W}^{(3)}| = 1$, the virtual channel $p(w^{(1)}_{1:n}y^{(1)}_{1:j}, w^{(2)}_{1:n}y^{(2)}_{1:j}, w^{(3)}_{1:n}y^{(3)}_{1:j}|w^{(1)}_{1:n}, w^{(2)}_{1:n}, w^{(3)}_{1:n})$ is also a broadcast channel. This is because $|\mathcal{W}^{(2)}_{1:n}| = |\mathcal{W}^{(3)}_{1:n}| = 1$, and $H(W^{(1)}_{1:n}Y^{(1)}_{1:j}|W^{(1)}_{1:n}) = 0$.

In general, we consider a class of m-input/m-output multiterminal networks whose set of input alphabets $(\mathcal{X}^{(1)}, ..., \mathcal{X}^{(m)}, \mathcal{Y}^{(1)}, ..., \mathcal{Y}^{(m)})$ is in a given set $\mathcal{A}$ and whose conditional law $p(y^{(1)}, ..., y^{(m)}|x^{(1)}, ..., x^{(m)})$ is in a given set $\mathcal{Q}$. Furthermore, the set $\mathcal{Q}$ is assumed to include the channel

$$p(y^{(1)}, ..., y^{(m)}|x^{(1)}, ..., x^{(m)}) = \prod_{i=1}^{m} \mathbf{1}[y^{(i)} = x^{(i)}].$$

2) Formal statement of the properties imposed on ϕ: Suppose we are given a class of m-input/m-output multiterminal networks, specified by sets $\mathcal{A}$ and $\mathcal{Q}$, as above. The formal statement of the properties imposed on ϕ is as follows. Please see Definitions 3 and 4 for the notation used.

1) Assume that the conditional law $p(y^{(1)}y'^{(1)}, y^{(2)}y'^{(2)}, ..., y^{(m)}y'^{(m)}|x^{(1)}, x^{(2)}, ..., x^{(m)})$ satisfies the following

$$p(y^{(1)}y'^{(1)}, y^{(2)}y'^{(2)}, ..., y^{(m)}y'^{(m)}|x^{(1)}, ..., x^{(m)}) =$$
$$p(y^{(1)}, y^{(2)}, ..., y^{(m)}|x^{(1)}, x^{(2)}, ..., x^{(m)}) \cdot p(y'^{(1)}, y'^{(2)}, ..., y'^{(m)}|x'^{(1)}, x'^{(2)}, ..., x'^{(m)}),$$

where $X'^{(i)}$ is a deterministic function of $X^{(i)}Y^{(i)}$ (i.e. $H(X'^{(i)}|X^{(i)}Y^{(i)}) = 0$ $(i \in [m])$). We assume that $(\mathcal{X}^{(1)}, ..., \mathcal{X}^{(m)}, \mathcal{Y}^{(1)} \times \mathcal{Y}'^{(1)}, ..., \mathcal{Y}^{(m)} \times \mathcal{Y}'^{(m)})$, $(\mathcal{X}^{(1)}, ..., \mathcal{X}^{(m)}, \mathcal{Y}^{(1)}, ..., \mathcal{Y}^{(m)})$ and $(\mathcal{X}'^{(1)}, ..., \mathcal{X}'^{(m)}, \mathcal{Y}'^{(1)}, ..., \mathcal{Y}'^{(m)})$ are in $\mathcal{A}$, and the conditional laws $p(y^{(1)}y'^{(1)}, y^{(2)}y'^{(2)}, ..., y^{(m)}y'^{(m)}|x^{(1)}, x^{(2)}, ..., x^{(m)})$, $p(y^{(1)}, y^{(2)}, ..., y^{(m)}|x^{(1)}, x^{(2)}, ..., x^{(m)})$ and $p(y'^{(1)}, y'^{(2)}, ..., y'^{(m)}|x'^{(1)}, x'^{(2)}, ..., x'^{(m)})$ are in $\mathcal{Q}$.

Take an arbitrary input distribution $q(x^{(1)}, x^{(2)}, ..., x^{(m)})$. This input distribution, together with the conditional distribution $p(y^{(1)}, y^{(2)}, ..., y^{(m)}|x^{(1)}, x^{(2)}, ..., x^{(m)})$, impose a joint distribution $q(x'^{(1)}, x'^{(2)}, ..., x'^{(m)})$ on $(X'^{(1)}, X'^{(2)}, ..., X'^{(m)})$. Then the following constraint needs to be satisfied for any arbitrary set Ψ of joint distributions on $\mathcal{X}'^{(1)} \times \mathcal{X}'^{(2)} \times \cdots \times \mathcal{X}'^{(m)}$ that contains

$q(x'^{(1)}, x'^{(2)}, ..., x'^{(m)})$:

$$\phi\Big(p(y^{(1)}y'^{(1)}, ..., y^{(m)}y'^{(m)}|x^{(1)}, ..., x^{(m)}), \{q(x^{(1)}, x^{(2)}, ..., x^{(m)})\}\Big) \subseteq$$
$$\phi\big(p(y^{(1)}, ..., y^{(m)}|x^{(1)}, ..., x^{(m)}), \{q(x^{(1)}, x^{(2)}, ..., x^{(m)})\}\big)$$
$$\oplus\ \phi\big(p(y'^{(1)}, y'^{(2)}, ..., y'^{(m)}|x'^{(1)}, ..., x'^{(m)}), \Psi\big).$$

2) Assume that

$$p(y^{(1)}, ..., y^{(m)}|x^{(1)}, ..., x^{(m)}) = \prod_{i=1}^{m} \mathbf{1}[y^{(i)} = x^{(i)}],$$

and that $(\mathcal{X}^{(1)}, ..., \mathcal{X}^{(m)}, \mathcal{Y}^{(1)}, ..., \mathcal{Y}^{(m)})$ is in $\mathcal{A}$. Then we require that for any input distribution $q(x^{(1)}, x^{(2)}, ..., x^{(m)})$, the set

$$\phi\big(p(y^{(1)}, ..., y^{(m)}|x^{(1)}, ..., x^{(m)}), \{q(x^{(1)}, x^{(2)}, ..., x^{(m)})\}\big)$$

contains only the origin in $\mathbb{R}^d$.

3) Statement of the Main Lemma:

Lemma 1: Take a physical channel $q(y^{(1)}, y^{(2)}, ..., y^{(m)}|x^{(1)}, x^{(2)}, ..., x^{(m)})$ whose set of input alphabets is in a given set $\mathcal{A}$, and whose conditional law is in a given set $\mathcal{Q}$. Let Ψ be a permissible set of input distributions for this channel. Then for any function ϕ satisfying the above two properties, target distortion levels $D^{(i)}$, and for arbitrary admissible source $W^{(i)}$ ($i \in [m]$) for these distortion levels for which for any n-code for the permissible set of input distributions Ψ, $p(w_{1:n}^{(1)}y_{1:k}^{(1)}, ..., w_{1:n}^{(m)}y_{1:k}^{(m)}|w_{1:n}^{(1)}, ..., w_{1:n}^{(m)})$ is in $\mathcal{Q}$, and $(\mathcal{W}_{1:n}^{(1)}, ..., \mathcal{W}_{1:n}^{(m)}, \mathcal{W}_{1:n}^{(1)} \times \mathcal{Y}_{1:k}^{(1)}, ..., \mathcal{W}_{1:n}^{(m)} \times \mathcal{Y}_{1:k}^{(m)})$ is in $\mathcal{A}$ for all $0 \leq k \leq n$, the following holds:

$$\phi\big(p(w_{1:n}^{(1)}y_{1:n}^{(1)}, ..., w_{1:n}^{(m)}y_{1:n}^{(m)}|w_{1:n}^{(1)}, ..., w_{1:n}^{(m)}), \{p(w_{1:n}^{(1)}, ..., w_{1:n}^{(m)})\}\big) \subseteq$$
$$n \times Convex\ Hull\big\{\phi\big(q(y^{(1)}, ..., y^{(m)}|x^{(1)}, ..., x^{(m)}), \Psi\big)\big\}.$$

Proof of Lemma 1: Let random variables $X_k^{(i)}$ and $Y_k^{(i)}$ ($k \in [n]$, $i \in [m]$) respectively represent the inputs and outputs of the multiterminal network. We have:

$$\phi\big(p(w_{1:n}^{(1)}y_{1:n}^{(1)}, w_{1:n}^{(2)}y_{1:n}^{(2)}, ..., w_{1:n}^{(m)}y_{1:n}^{(m)}|w_{1:n}^{(1)}, w_{1:n}^{(2)}, ..., w_{1:n}^{(m)}), \{p(w_{1:n}^{(1)}, w_{1:n}^{(2)}, ..., w_{1:n}^{(m)})\}\big) \subseteq \quad (4)$$
$$\phi\big(p(w_{1:n}^{(1)}y_{1:n-1}^{(1)}, w_{1:n}^{(2)}y_{1:n-1}^{(2)}, ..., w_{1:n}^{(m)}y_{1:n-1}^{(m)}|w_{1:n}^{(1)}, w_{1:n}^{(2)}, ..., w_{1:n}^{(m)}), \{p(w_{1:n}^{(1)}, w_{1:n}^{(2)}, ..., w_{1:n}^{(m)})\}\big) \oplus$$
$$\phi(q(y_n^{(1)}, y_n^{(2)}, ..., y_n^{(m)}|x_n^{(1)}, x_n^{(2)}, ..., x_n^{(m)}), \Psi) \subseteq$$
$$\phi\big(p(w_{1:n}^{(1)}y_{1:n-2}^{(1)}, w_{1:n}^{(2)}y_{1:n-2}^{(2)}, ..., w_{1:n}^{(m)}y_{1:n-2}^{(m)}|w_{1:n}^{(1)}, w_{1:n}^{(2)}, ..., w_{1:n}^{(m)}), \{p(w_{1:n}^{(1)}, w_{1:n}^{(2)}, ..., w_{1:n}^{(m)})\}\big) \oplus$$
$$\phi(q(y_{n-1}^{(1)}, y_{n-1}^{(2)}, ..., y_{n-1}^{(m)}|x_{n-1}^{(1)}, x_{n-1}^{(2)}, ..., x_{n-1}^{(m)}), \Psi) \oplus$$

$$\phi(q(y_n^{(1)}, y_n^{(2)}, ..., y_n^{(m)}|x_n^{(1)}, x_n^{(2)}, ..., x_n^{(m)}), \Psi) \subseteq$$

$$\cdots \subseteq$$

$$\phi\big(p(w_{1:n}^{(1)}, w_{1:n}^{(2)}, ..., w_{1:n}^{(m)}|w_{1:n}^{(1)}, w_{1:n}^{(2)}, ..., w_{1:n}^{(m)}), \{p(w_{1:n}^{(1)}, w_{1:n}^{(2)}, ..., w_{1:n}^{(m)})\}\big)\oplus$$
$$\phi(q(y_1^{(1)}, y_1^{(2)}, ..., y_1^{(m)}|x_1^{(1)}, x_1^{(2)}, ..., x_1^{(m)}), \Psi)\oplus$$
$$\phi(q(y_2^{(1)}, y_2^{(2)}, ..., y_2^{(m)}|x_2^{(1)}, x_2^{(2)}, ..., x_2^{(m)}), \Psi) \oplus \cdots$$
$$\phi(q(y_{n-1}^{(1)}, y_{n-1}^{(2)}, ..., y_{n-1}^{(m)}|x_{n-1}^{(1)}, x_{n-1}^{(2)}, ..., x_{n-1}^{(m)}), \Psi)\oplus$$
$$\phi(q(y_n^{(1)}, y_n^{(2)}, ..., y_n^{(m)}|x_n^{(1)}, x_n^{(2)}, ..., x_n^{(m)}), \Psi) \subseteq \tag{5}$$
$$\phi(q(y_1^{(1)}, y_1^{(2)}, ..., y_1^{(m)}|x_1^{(1)}, x_1^{(2)}, ..., x_1^{(m)}), \Psi)\oplus$$
$$\phi(q(y_2^{(1)}, y_2^{(2)}, ..., y_2^{(m)}|x_2^{(1)}, x_2^{(2)}, ..., x_2^{(m)}), \Psi) \oplus \cdots$$
$$\phi(q(y_{n-1}^{(1)}, y_{n-1}^{(2)}, ..., y_{n-1}^{(m)}|x_{n-1}^{(1)}, x_{n-1}^{(2)}, ..., x_{n-1}^{(m)}), \Psi)\oplus$$
$$\phi(q(y_n^{(1)}, y_n^{(2)}, ..., y_n^{(m)}|x_n^{(1)}, x_n^{(2)}, ..., x_n^{(m)}), \Psi) \subseteq \tag{6}$$
$$n \times Convex\ Hull\big\{\phi(q(y^{(1)}, y^{(2)}, ..., y^{(m)}|x^{(1)}, x^{(2)}, ..., x^{(m)}), \Psi)\big\},$$

where in equation (4) we have used property (1) because

$$p(w_{1:n}^{(1)}y_{1:n}^{(1)}, w_{1:n}^{(2)}y_{1:n}^{(2)}, ..., w_{1:n}^{(m)}y_{1:n}^{(m)}|w_{1:n}^{(1)}, w_{1:n}^{(2)}, ..., w_{1:n}^{(m)}) =$$
$$p(w_{1:n}^{(1)}y_{1:n-1}^{(1)}, w_{1:n}^{(2)}y_{1:n-1}^{(2)}, ..., w_{1:n}^{(m)}y_{1:n-1}^{(m)}|w_{1:n}^{(1)}, w_{1:n}^{(2)}, ..., w_{1:n}^{(m)}) \cdot p(y_n^{(1)}, ..., y_n^{(m)}|x_n^{(1)}, x_n^{(2)}, ..., x_n^{(m)})$$

and furthermore $H(X_n^{(i)}|W_{1:n}^{(i)}Y_{1:n-1}^{(i)}) = 0$ for all $i \in [m]$, and also

$$p(y_n^{(1)}, y_n^{(2)}, ..., y_n^{(m)}|x_n^{(1)}, x_n^{(2)}, ..., x_n^{(m)}) = q(y_n^{(1)}, y_n^{(2)}, ..., y_n^{(m)}|x_n^{(1)}, x_n^{(2)}, ..., x_n^{(m)}).$$

Since the n-code must work with the permissible set of input distributions Ψ, the joint distribution $p(x_n^{(1)}, x_n^{(2)}, ..., x_n^{(m)})$ is in Ψ. In equation (5) we have used property (2). In equation (6), we first note that the conditional distributions

$$q(y_i^{(1)}, y_i^{(2)}, ..., y_i^{(m)}|x_i^{(1)}, x_i^{(2)}, ..., x_i^{(m)})$$

for $i = 1, 2, ..., n$ are all the same. We then observe that whenever $\overrightarrow{v_i} \in \phi(q(y^{(1)}, ..., y^{(m)}|x^{(1)}, ..., x^{(m)}), \Psi)$ for $i \in [n]$, their average, $\frac{1}{n}\sum_{i=1}^{n}\overrightarrow{v_i}$, falls in the convex hull of $\phi(q(y^{(1)}, ..., y^{(m)}|x^{(1)}, ..., x^{(m)}), \Psi)$. ■

Remark 1: The second property of ϕ is used in the proof to reduce the n-letter expression

$$\phi\big(p(w_{1:n}^{(1)}y_{1:n}^{(1)}, ..., w_{1:n}^{(m)}y_{1:n}^{(m)}|w_{1:n}^{(1)}, ..., w_{1:n}^{(m)}), \{p(w_{1:n}^{(1)}, ..., w_{1:n}^{(m)})\}\big)$$

to n single-letter expressions involving $\phi(q(y^{(1)}, ..., y^{(m)}|x^{(1)}, ..., x^{(m)}), \Psi)$ in n stages.

B. A generalized cut-set bound

Theorem 1: Given any GMN $q(y^{(1)}, y^{(2)}, ..., y^{(m)}|x^{(1)}, x^{(2)}, ..., x^{(m)})$, a permissible set of input distributions Ψ, a sequence of non-negative real numbers $D^{(i)}$ $(i \in [m])$, and an arbitrary admissible source $W^{(i)}$ $(i \in [m])$ for the target reproduction functions $f^{(i)}$ $(i \in [m])$ at these distortion levels, there exists

- a joint distribution $q(x^{(1)}, x^{(2)}, ..., x^{(m)}, z)$ where the size of the alphabet of Z is $2^m - 1$ and furthermore $q(x^{(1)}, x^{(2)}, ..., x^{(m)}|z)$ belongs to Ψ for any value z that the random variable Z might take;
- a joint distribution $p(\widehat{m}^{(1)}, \widehat{m}^{(2)}, ..., \widehat{m}^{(m)}, w^{(1)}, w^{(2)}, ..., w^{(m)})$ where the average distortion between $M^{(i)} = f^{(i)}(W^{(1)}, W^{(2)}, ..., W^{(m)})$ and $\widehat{M}^{(i)}$ is less than or equal to $D^{(i)}$, i.e. $\Delta^{(i)}(M^{(i)}, \widehat{M}^{(i)}) \leq D^{(i)}$,

such that for every $T \subset [m]$ the following inequality holds:

$$I\big(W^{(i)} : i \in T \;\; ; \;\; \widehat{M}^{(j)} : j \in T^c | W^{(j)} : j \in T^c\big) \leq I\big(X^{(i)} : i \in T \;\; ; \;\; Y^{(j)} : j \in T^c | X^{(j)} : j \in T^c, Z\big),$$

where $Y^{(1)}, Y^{(2)}, ..., Y^{(m)}, X^{(1)}, X^{(2)}, ..., X^{(m)}$ and Z are jointly distributed according to

$$q(y^{(1)}, y^{(2)}, ..., y^{(m)}|x^{(1)}, x^{(2)}, ..., x^{(m)}) \cdot q(x^{(1)}, x^{(2)}, ..., x^{(m)}, z).$$

Note that here the following Markov chain holds:

$$Z \to X^{(1)}X^{(2)}...X^{(m)} \to Y^{(1)}Y^{(2)}...Y^{(m)}.$$

Discussion 1: The fact that the expressions on both sides of the above inequality are of the same form is suggestive. To any given channel $q(y^{(1)}, y^{(2)}, ..., y^{(m)}|x^{(1)}, x^{(2)}, ..., x^{(m)})$ and input distribution $q(x^{(1)}, x^{(2)}, ..., x^{(m)})$, assign the down-set of a vector in $\mathbb{R}_+^{2^m}$ whose k^{th} coordinate is defined as

$$I\big(X^{(i)} : i \in T_k \;\; ; \;\; Y^{(j)} : j \in T_k^c | X^{(j)} : j \in T_k^c\big),$$

where T_k is defined as follows: there are 2^m subsets of $[m]$; take an arbitrary ordering of these sets and take T_k to be the k^{th} subset in that ordering (though not required but for the sake of consistency with the notation used in the proof of the theorem assume that T_{2^k-1} and T_{2^k} are the empty set and the full set respectively). Next, to any channel $q(y^{(1)}, y^{(2)}, ..., y^{(m)}|x^{(1)}, x^{(2)}, ..., x^{(m)})$ and a set of permissible input distributions we assign a region by taking the convex hull of the union over all permissible input distributions, of the region associated to the channel and the varying input distribution. A channel with an associated set of permissible input distributions is said to be weaker than another channel with its associated set of permissible input distributions if the region associated to the first (channel, permissible set) pair is contained in the region associated to the second (channel, permissible set) pair.

Intuitively speaking, given a communication task one can consider a virtual channel whose inputs and outputs represent, roughly speaking, the raw and acceptable information objectives at the m parties. Furthermore, let the only permissible input distribution for this virtual channel be the one given by the statistical description of the raw information of the parties. More specifically, given any $p(\widehat{m}^{(1)}, ..., \widehat{m}^{(m)}, w^{(1)}, ..., w^{(m)})$ such that $\Delta^{(i)}(M^{(i)}, \widehat{M}^{(i)}) \leq D^{(i)}$ holds, consider the virtual channel $p(\widehat{m}^{(1)}, \widehat{m}^{(2)}, ..., \widehat{m}^{(m)}|w^{(1)}, w^{(2)}, ..., w^{(m)})$ with the only permissible input distribution being $p(w^{(1)}, w^{(2)}, ..., w^{(m)})$. The inputs of this virtual channel, i.e. $W^{(1)}, W^{(2)}, ..., W^{(m)}$, and its outputs, i.e. $\widehat{M}^{(1)}, \widehat{M}^{(2)}, ..., \widehat{M}^{(m)}$, can be understood as the raw information and acceptable information objectives at the m parties. The region associated to the virtual channel $p(\widehat{m}^{(1)}, ..., \widehat{m}^{(m)}|w^{(1)}, ..., w^{(m)})$ and the input distribution $p(w^{(1)}, w^{(2)}, ..., w^{(m)})$ would be the down-set of a vector in $\mathbb{R}_+^{2^m}$ whose k^{th} coordinate is defined as

$$I\big(W^{(i)} : i \in T \ \ ; \ \ \widehat{M}^{(j)} : j \in T^c | W^{(j)} : j \in T^c\big).$$

Theorem 1 is basically saying that there must exist a virtual channel of this kind such that this region associated to this virtual channel and the unique corresponding permissible input distribution should be included inside the region associated to the channel $q(y^{(1)}, y^{(2)}, ..., y^{(m)}|x^{(1)}, x^{(2)}, ..., x^{(m)})$ and its associated set of permissible input distributions. Here the complexity of transmission of functions of correlated messages is effectively translated into the performance region of a virtual channel at a given input distribution. There must exist a virtual channel of this kind at the given input distribution that must be, in the above mentioned sense, weaker than the physical channel in order for the channel to be fit for the communication problem.

Corollary 1: Given any GMN $q(y^{(1)}, y^{(2)}, ..., y^{(m)}|x^{(1)}, x^{(2)}, ..., x^{(m)})$ and associated set of permissible input distributions Ψ, the following region forms an outer bound on the set of vectors of rates $(R^{(i,j)}, i, j \in [m])$ at which communication is feasible in the traditional block-coding with asymptotically vanishing block probability of error formulation for independent messages:

$$\bigcup_{\substack{q(x^{(1)}, x^{(2)}, ..., x^{(m)}, z) \text{ such that for any } z \\ q(x^{(1)}, x^{(2)}, ..., x^{(m)}|z) \in \Psi \text{ and} \\ \text{the size of the alphabet of } Z \text{ is } 2^m - 1}} \Big\{ \text{non-negative } R^{(i,j)} \text{ for } i, j \in [m]\text{: for any arbitrary } T \subset [m]$$

$$\sum_{i \in T, j \in T^c} R^{(i,j)} \leq I\big(X^{(i)} : i \in T \ \ ; \ \ Y^{(j)} : j \in T^c | X^{(j)} : j \in T^c, Z\big)$$

$$\text{is satisfied.}\Big\},$$

where $Y^{(1)}, Y^{(2)}, ..., Y^{(m)}, X^{(1)}, X^{(2)}, ..., X^{(m)}$ and Z are jointly distributed according to

$$q(y^{(1)}, y^{(2)}, ..., y^{(m)}|x^{(1)}, x^{(2)}, ..., x^{(m)}) \cdot q(x^{(1)}, x^{(2)}, ..., x^{(m)}, z).$$

Remark 2: This bound is sometimes tight; for instance it is tight for a multiple access channel with independent source messages when Ψ is taken to be the set of all mutually independent input distributions.

Remark 3: This bound reduces to the traditional cut-set bound when Ψ is taken to be the set of all input distributions, and $I\big(X^{(i)} : i \in T \;\; ; \;\; Y^{(j)} : j \in T^c | X^{(i)} : i \in T^c, Z\big)$ is bounded from above by [2]

$$I\big(X^{(i)} : i \in T \;\; ; \;\; Y^{(j)} : j \in T^c | X^{(j)} : j \in T^c\big).$$

Proof of Theorem 1: The inequalities always hold for the extreme cases of the set T being either empty or $[m]$. So, it is sufficient to consider only those subsets of $[m]$ that are neither empty nor equal to $[m]$. In order to apply Lemma 1, we will define ϕ as follows: for any conditional distribution $p(y^{(1)}, y^{(2)}, ..., y^{(m)}|x^{(1)}, x^{(2)}, ..., x^{(m)})$ and an arbitrary set Ψ of distributions on $\mathcal{X}^{(1)} \times \mathcal{X}^{(2)} \times \cdots \mathcal{X}^{(m)}$, let

$$\phi(p(y^{(1)}, y^{(2)}, ..., y^{(m)}|x^{(1)}, x^{(2)}, ..., x^{(m)}), \Psi)) = \tag{7}$$

$$\bigcup_{p(x^{(1)}, x^{(2)}, ..., x^{(m)}) \in \Psi} \varphi\big(p(y^{(1)}, y^{(2)}, ..., y^{(m)}|x^{(1)}, x^{(2)}, ..., x^{(m)}) p(x^{(1)}, x^{(2)}, ..., x^{(m)})\big),$$

where $\varphi(p(y^{(1)}, y^{(2)}, ..., y^{(m)}, x^{(1)}, x^{(2)}, ..., x^{(m)}))$ is defined as the down-set [3] of a vector of size $d = 2^m - 2$ whose k^{th} coordinate equals $I\big(X^{(i)} : i \in T_k \;\; ; \;\; Y^{(j)} : j \in (T_k)^c | X^{(j)} : j \in (T_k)^c\big)$. Here T_k is defined as follows: there are $2^m - 2$ subsets of $[m]$ that are neither empty nor equal to $[m]$. Take an arbitrary ordering of these sets and take T_k to be the k^{th} subset in that ordering.

In Appendix A-A we show that the above choice of $\phi(p(y^{(1)}, y^{(2)}, ..., y^{(m)}|x^{(1)}, x^{(2)}, ..., x^{(m)}), \Psi)$ verifies the properties of Lemma 1 for $\mathcal{A}$ equal to the set allowing arbitrary finite input alphabets for each variable, and $\mathcal{Q}$ equal to the set of all conditional laws $p(y^{(1)}, ..., y^{(m)}|x^{(1)}, ..., x^{(m)})$. For any arbitrary positive ϵ and n-code for the set of permissible input distributions Ψ satisfying the average distortion condition $D^{(i)}$ (for all $i \in [m]$) (see Definition 1), the lemma implies (for the definition of multiplication of a set by a real number see Definition 3):

$$\phi\big(p(\widehat{m}^{(1)}_{1:n}, \widehat{m}^{(2)}_{1:n}, ..., \widehat{m}^{(m)}_{1:n}|w^{(1)}_{1:n}, w^{(2)}_{1:n}, ..., w^{(m)}_{1:n}), \{p(w^{(1)}_{1:n}, w^{(2)}_{1:n}, ..., w^{(m)}_{1:n})\}\big) =$$

[2]This is valid because $I\big(X^{(i)} : i \in T \;\; ; \;\; Y^{(j)} : j \in T^c | X^{(j)} : j \in T^c, Z\big) = H\big(Y^{(j)} : j \in T^c | X^{(j)} : j \in T^c, Z\big) - H\big(Y^{(j)} : j \in T^c | X^{(i)} : i \in [m], Z\big) = H\big(Y^{(j)} : j \in T^c | X^{(j)} : j \in T^c, Z\big) - H\big(Y^{(j)} : j \in T^c | X^{(i)} : i \in [m]\big) \leq H\big(Y^{(j)} : j \in T^c | X^{(j)} : j \in T^c\big) - H\big(Y^{(j)} : j \in T^c | X^{(i)} : i \in [m]\big) = I\big(X^{(i)} : i \in T \;\; ; \;\; Y^{(j)} : j \in T^c | X^{(j)} : j \in T^c\big)$.

[3]For the definition of a down-set see Definition 4.

$$\varphi\big(p(\widehat{m}_{1:n}^{(1)}, \widehat{m}_{1:n}^{(2)}, ..., \widehat{m}_{1:n}^{(m)}|w_{1:n}^{(1)}, w_{1:n}^{(2)}, ..., w_{1:n}^{(m)})p(w_{1:n}^{(1)}, w_{1:n}^{(2)}, ..., w_{1:n}^{(m)})\big) \subseteq$$
$$n \times Convex\ Hull\big\{\phi(q(y^{(1)}, y^{(2)}, ..., y^{(m)}|x^{(1)}, x^{(2)}, ..., x^{(m)}), \Psi)\big\}.$$

According to the Carathéodory theorem, every point inside the convex hull of

$$\phi(q(y^{(1)}, y^{(2)}, ..., y^{(m)}|x^{(1)}, x^{(2)}, ..., x^{(m)}), \Psi)$$

can be written as a convex combination of $d+1 = 2^m - 1$ points in the set. Corresponding to the i^{th} point in the convex combination ($i \in [2^m - 1]$) is an input distribution $q_i(x^{(1)}, x^{(2)}, ..., x^{(m)})$ such that the point lies in

$$\varphi\big(q(y^{(1)}, y^{(2)}, ..., y^{(m)}|x^{(1)}, x^{(2)}, ..., x^{(m)})q_i(x^{(1)}, x^{(2)}, ..., x^{(m)})\big).$$

Let Z be a random variable defined on the set $\{1, 2, 3, ..., 2^m - 1\}$ that takes value i with probability equal to the weight associated to the i^{th} point in the above convex combination. Z is jointly distributed with $X^{(1)}, X^{(2)}, ..., X^{(m)}, Y^{(1)}, Y^{(2)}, ..., Y^{(m)}$ as follows:

$$p(x^{(1)}, x^{(2)}, ..., x^{(m)}, z) =$$
$$p(z)q_z(x^{(1)}, x^{(2)}, ..., x^{(m)})q(y^{(1)}, y^{(2)}, ..., y^{(m)}|x^{(1)}, x^{(2)}, ..., x^{(m)}).$$

The convex hull of $\phi(q(y^{(1)}, y^{(2)}, ..., y^{(m)}|x^{(1)}, x^{(2)}, ..., x^{(m)}), \Psi)$ is therefore contained in (see Definition 3 for the definition of the summation used here):

$$\bigcup_{\substack{q(x^{(1)}, x^{(2)}, ..., x^{(m)}, z) \text{ such that for any } z\\ q(x^{(1)}, x^{(2)}, ..., x^{(m)}|z) \in \Psi \text{ and the}\\ \text{size of the alphabet set of } Z \text{ is } 2^m - 1}} \sum_z p(z) \times \varphi\big(q(y^{(1)}, ..., y^{(m)}|x^{(1)}, ..., x^{(m)})q(x^{(1)}, ..., x^{(m)}|z)\big).$$

Hence,

$$\varphi\big(p(\widehat{m}_{1:n}^{(1)}, \widehat{m}_{1:n}^{(2)}, ..., \widehat{m}_{1:n}^{(m)}|w_{1:n}^{(1)}, w_{1:n}^{(2)}, ..., w_{1:n}^{(m)})p(w_{1:n}^{(1)}, w_{1:n}^{(2)}, ..., w_{1:n}^{(m)})\big) \subseteq$$
$$n \times \bigcup_{\substack{q(x^{(1)}, x^{(2)}, ..., x^{(m)}, z) \text{ such that for any } z\\ q(x^{(1)}, x^{(2)}, ..., x^{(m)}|z) \in \Psi \text{ and the}\\ \text{size of the alphabet set of } Z \text{ is } 2^m - 1}} \sum_z p(z) \times \varphi\big(q(y^{(1)}, ..., y^{(m)}|x^{(1)}, ..., x^{(m)})q(x^{(1)}, ..., x^{(m)}|z)\big).$$

The set

$$\varphi\big(p(\widehat{m}_{1:n}^{(1)}, \widehat{m}_{1:n}^{(2)}, ..., \widehat{m}_{1:n}^{(m)}|w_{1:n}^{(1)}, w_{1:n}^{(2)}, ..., w_{1:n}^{(m)})p(w_{1:n}^{(1)}, w_{1:n}^{(2)}, ..., w_{1:n}^{(m)})\big)$$

is by definition the down-set of a vector of length $2^m - 2$, denoted here by $\overrightarrow{v}$, whose k^{th} coordinate is equal to

$$I\big(W_{1:n}^{(i)} : i \in T_k \ \ ; \ \ \widehat{M}_{1:n}^{(j)} : j \in (T_k)^c | W_{1:n}^{(j)} : j \in (T_k)^c\big).$$

The vector $\overrightarrow{v}$ is greater than or equal to $\overrightarrow{\overrightarrow{v}}$ whose k^{th} element equals:[4]

$$n \cdot I\big(\widetilde{W^{(i)}} : i \in T_k \;\; ; \;\; \widetilde{\widehat{M^{(j)}}} : j \in (T_k)^c | \widetilde{W^{(j)}} : j \in (T_k)^c\big),$$

for some $\widetilde{W^{(i)}}$ and $\widetilde{\widehat{M^{(i)}}}$ ($i \in [m]$) such that the joint distribution of $\widetilde{W^{(i)}}$ ($i \in [m]$) is the same as that of $W^{(i)}$ ($i \in [m]$), and the average distortion between $\widetilde{M^{(i)}} = f^{(i)}(\widetilde{W^{(1)}}, \widetilde{W^{(2)}}, ..., \widetilde{W^{(m)}})$ and $\widetilde{\widehat{M^{(i)}}}$ is less than or equal to $D^{(i)} + \epsilon$.[5] In Appendix A-B, we perturb random variables $\widetilde{\widehat{M^{(i)}}}$ (for $i \in [m]$) and define random variables $\widetilde{\widehat{M'^{(i)}}}$ (for $i \in [m]$) such that for every $i \in [m]$, the average distortion between $\widetilde{\widehat{M'^{(i)}}}$ and $\widetilde{M^{(i)}}$ is less than or equal to $D^{(i)}$ (rather than $D^{(i)} + \epsilon$ as in the case of $\widetilde{\widehat{M^{(i)}}}$) and furthermore for every k

$$I\big(\widetilde{W^{(i)}} : i \in T_k \;\; ; \;\; \widetilde{\widehat{M'^{(j)}}} : j \in (T_k)^c | \widetilde{W^{(j)}} : j \in (T_k)^c\big) - O(\tau(\epsilon)) \leq$$
$$I\big(\widetilde{W^{(i)}} : i \in T_k \;\; ; \;\; \widetilde{\widehat{M^{(j)}}} : j \in (T_k)^c | \widetilde{W^{(j)}} : j \in (T_k)^c\big),$$

where τ is a real-valued function that satisfies the property that $\tau(\epsilon) \rightarrow 0$ as $\epsilon \rightarrow 0$.

Hence the vector $\overrightarrow{\overrightarrow{v}}$ is coordinate by coordinate greater than or equal to a vector $\overrightarrow{\overrightarrow{v'}}$ whose k^{th} element is defined as

$$\max\Bigg(n \cdot I\big(\widetilde{W^{(i)}} : i \in T_k \;\; ; \;\; \widetilde{\widehat{M'^{(j)}}} : j \in (T_k)^c | \widetilde{W^{(j)}} : j \in (T_k)^c\big) - n \cdot O(\tau(\epsilon)), 0\Bigg).$$

The vector $\overrightarrow{\overrightarrow{v'}}$ must lie in

$$\varphi\big(p((\widehat{m}^{(1)}_{1:n}, \widehat{m}^{(2)}_{1:n}, ..., \widehat{m}^{(m)}_{1:n} | w^{(1)}_{1:n}, w^{(2)}_{1:n}, ..., w^{(m)}_{1:n}) p(w^{(1)}_{1:n}, w^{(2)}_{1:n}, ..., w^{(m)}_{1:n})\big),$$

since it is coordinate by coordinate less than or equal to $\overrightarrow{\overrightarrow{v}}$. It must therefore also lie in

$$n \times \bigcup_{\substack{q(x^{(1)},...,x^{(m)},z) \text{ such that for any } z \\ q(x^{(1)},...,x^{(m)}|z) \in \Psi \text{ and the} \\ \text{size of the alphabet set of } Z \text{ is } 2^m - 1}} \sum_z p(z) \times \varphi\big(q(y^{(1)}, ..., y^{(m)} | x^{(1)}, ..., x^{(m)}) q(x^{(1)}, ..., x^{(m)} | z)\big).$$

Note that since φ is the down-set of a non-negative vector, the above Minkowski sum inside the union would itself be the down-set of a vector.[6] The left hand side can be therefore written as the union over

[4]This is because for any arbitrary random variables X^n, Y^n, Z^n such that (X^n, Y^n) is n i.i.d. repetitions of (X, Y), we have: $I(X^n; Z^n|Y^n) = nH(X|Y) - H(X^n|Z^n Y^n) \geq \sum_{g=1}^n H(X_g|Y_g) - H(X_g|Y_g Z_g) = \sum_{g=1}^n I(X_g; Z_g|Y_g) = n \cdot I(X_G; Z_G|GY_G) \geq n \cdot I(X_G; Z_G|Y_G)$ where G is uniform over $\{1, 2, ..., n\}$ and independent of (X^n, Y^n, Z^n). Random variables (X_G, Y_G) have the same joint distribution as (X, Y).

[5]This is because for any arbitrary pair (Y^n, Z^n), the average distortion between Y_G and Z_G for G uniform over $\{1, 2, ..., n\}$ and independent of (Y^n, Z^n), is equal to $\mathbb{E}[\Delta(Y_G, Z_G)] = \mathbb{E}[\mathbb{E}[\Delta(Y_G, Z_G)|G]] = \sum_{g=1}^n \frac{1}{n}\mathbb{E}[\Delta(Y_g, Z_g)] = \mathbb{E}[\Delta_n(Y^n, Z^n)]$.

[6]This is because for every two non-negative vectors $\overrightarrow{v_1}$ and $\overrightarrow{v_2}$, we have $\lambda \times \Pi(\overrightarrow{v_1}) \oplus (1-\lambda) \times \Pi(\overrightarrow{v_2}) = \Pi(\lambda\overrightarrow{v_1} + (1-\lambda)\overrightarrow{v_2})$ for any $\lambda \in [0, 1]$.

all $q(x^{(1)}, x^{(2)}, ..., x^{(m)}, z)$ such that $q(x^{(1)}, x^{(2)}, ..., x^{(m)}|z) \in \Psi$ for every z, of the down-set of a vector whose k^{th} coordinate equals $I\big(X^{(i)} : i \in T_k \;\; ; \;\; Y^{(j)} : j \in (T_k)^c | X^{(j)} : j \in (T_k)^c, Z\big)$. Since $\overrightarrow{v'}$ falls inside this union, there must exist a particular $q(x^{(1)}, x^{(2)}, ..., x^{(m)}, z)$ whose corresponding vector is coordinate by coordinate greater than or equal to $\overrightarrow{v'}$. The proof ends by recalling the definition of $\overrightarrow{v'}$ and letting ϵ converge to zero. ■

C. Strong interference channels

An interference channel is a four-input/four-output multiterminal network whose set of alphabets is assumed to belong to

$$\mathcal{A}_{interference} := \big\{\big(\mathcal{X}^{(1)}, \mathcal{X}^{(2)}, \mathcal{X}^{(3)}, \mathcal{X}^{(4)}, \mathcal{Y}^{(1)}, \mathcal{Y}^{(2)}, \mathcal{Y}^{(3)}, \mathcal{Y}^{(4)}\big) : |\mathcal{X}^{(3)}| = |\mathcal{X}^{(4)}| = 1\big\},$$

and whose conditional law is assumed to belong to

$$\mathcal{Q}_{interference} := \big\{q(y^{(1)}, y^{(2)}, y^{(3)}, y^{(4)}|x^{(1)}, x^{(2)}, x^{(3)}, x^{(4)}) : H(Y^{(1)}|X^{(1)}) = H(Y^{(2)}|X^{(2)}) = 0\big\}.$$

As discussed in subsection I-C2, we consider strong interference channels and assume that the third and fourth party are observing i.i.d. copies of $W^{(3)}$ and $W^{(4)}$, but that $|\mathcal{W}^{(3)}| = |\mathcal{W}^{(4)}| = 1$ implying that these random variables are constant. Furthermore, we consider the special case of

$$M^{(1)} = f^{(1)}(W^{(1)}, W^{(2)}, W^{(3)}, W^{(4)}) = W^{(1)},$$
$$M^{(2)} = f^{(2)}(W^{(1)}, W^{(2)}, W^{(3)}, W^{(4)}) = W^{(2)}$$
$$M^{(3)} = f^{(3)}(W^{(1)}, W^{(2)}, W^{(3)}, W^{(4)}) = W^{(1)}$$
$$M^{(4)} = f^{(4)}(W^{(1)}, W^{(2)}, W^{(3)}, W^{(4)}) = W^{(2)},$$

and $D^{(i)} = 0$ for $i = 1, 2, 3, 4$. The distortion function used by the i^{th} party, $\Delta^{(i)}(m^{(i)}, m'^{(i)})$ (for $1 \leq i \leq 4$) is taken to be the indicator function $\mathbf{1}[m^{(i)} \neq m'^{(i)}]$.

We state our main result in terms of the nebulously defined permissible set of input distributions Ψ. To get explicit results, simply replace Ψ by the set of all probability distributions on $\mathcal{X}^{(1)} \times \mathcal{X}^{(2)} \times \mathcal{X}^{(3)} \times \mathcal{X}^{(4)}$. Our main result is the following:

Theorem 2: Take an arbitrary interference channel $q(y^{(1)}, y^{(2)}, y^{(3)}, y^{(4)}|x^{(1)}, x^{(2)}, x^{(3)}, x^{(4)})$, a permissible set of input distributions Ψ, and an admissible source marginal distribution $p(w^{(1)}, w^{(2)}, w^{(3)}, w^{(4)})$ at zero distortion levels satisfying $|\mathcal{W}^{(3)}| = |\mathcal{W}^{(4)}| = 1$. Then for any random variable L where $W^{(1)} \rightarrow L \rightarrow W^{(2)}$ holds, there must exist

$$p(u, x^{(1)}, x^{(2)}, x^{(3)}, x^{(4)}, y^{(1)}, y^{(2)}, y^{(3)}, y^{(4)})$$

such that

$$p(x^{(1)}, x^{(2)}, x^{(3)}, x^{(4)}) \text{ is in the convex hull of } \Psi,$$

$$p(u, x^{(1)}, x^{(2)}, y^{(1)}, y^{(2)}, y^{(3)}, y^{(4)}) = p(u)p(x^{(1)}|u)p(x^{(2)}|u)q(y^{(1)}, y^{(2)}, y^{(3)}, y^{(4)}|x^{(1)}, x^{(2)}),$$

$\big($ note that $q(y^{(1)}, y^{(2)}, y^{(3)}, y^{(4)}|x^{(1)}, x^{(2)})$ can be used for $q(y^{(1)}, y^{(2)}, y^{(3)}, y^{(4)}|x^{(1)}, x^{(2)}, x^{(3)}, x^{(4)})$ because of the standing assumption that $|\mathcal{X}^{(3)}| = |\mathcal{X}^{(4)}| = 1\big)$ such the following inequalities hold:

$$\begin{aligned}
H(W^{(1)}|L) &\leq I(X^{(1)}; Y^{(3)}|X^{(2)}, U);\\
H(W^{(2)}|L) &\leq I(X^{(2)}; Y^{(4)}|X^{(1)}, U);\\
H(W^{(1)}|L) + H(W^{(2)}|L) &\leq \min(I(X^{(1)}X^{(2)}; Y^{(3)}|U), I(X^{(1)}X^{(2)}; Y^{(4)}|U));\\
H(W^{(1)}W^{(2)}) &\leq \min(I(X^{(1)}X^{(2)}; Y^{(3)}), I(X^{(1)}X^{(2)}; Y^{(4)}));\\
H(W^{(1)}|W^{(2)}) &\leq I(X^{(1)}; Y^{(3)}|X^{(2)});\\
H(W^{(2)}|W^{(1)}) &\leq I(X^{(2)}; Y^{(4)}|X^{(1)}).
\end{aligned} \tag{8}$$

Remark 4: One can use the strengthened Carathéodory theorem of Fenchel to bound the cardinality of U from above by $|\mathcal{X}^{(1)}||\mathcal{X}^{(2)}| + 3$.

Proof of Theorem 2: In order to apply Lemma 1, we will define $\phi(q(y^{(1)}, y^{(2)}, y^{(3)}, y^{(4)}|x^{(1)}, x^{(2)}, x^{(3)}, x^{(4)}), \Psi)$, a function that takes as input an arbitrary 4-input/4-output multiterminal network from $\mathcal{Q}_{interference}$, where the alphabets are from $\mathcal{A}_{interference}$ and a subset of probability distributions on $\mathcal{X}^{(1)} \times \mathcal{X}^{(2)} \times \mathcal{X}^{(3)} \times \mathcal{X}^{(4)}$ and returns a subset of $\mathbb{R}^8_+$. We let:

$$\phi(q(y^{(1)}, y^{(2)}, y^{(3)}, y^{(4)}|x^{(1)}, x^{(2)}, x^{(3)}, x^{(4)}), \Psi) =$$

$$\bigcup_{p(x^{(1)}, x^{(2)}, x^{(3)}, x^{(4)}) \in \Psi} \quad \bigcup_{\substack{U \text{ such that} \\ X^{(1)} \rightarrow U \rightarrow X^{(2)}, \\ U \rightarrow X^{(1)}X^{(2)} \rightarrow Y^{(3)}Y^{(4)}}}$$

$$\Pi\Big(\big\{\big(I(X^{(1)}; Y^{(3)}|X^{(2)}, U),\ I(X^{(2)}; Y^{(4)}|X^{(1)}, U),\ I(X^{(1)}X^{(2)}; Y^{(3)}|U),$$
$$I(X^{(1)}X^{(2)}; Y^{(4)}|U),\ I(X^{(1)}X^{(2)}; Y^{(3)}),\ I(X^{(1)}X^{(2)}; Y^{(4)}),$$
$$I(X^{(1)}; Y^{(3)}|X^{(2)}),\ I(X^{(2)}; Y^{(4)}|X^{(1)})\big)\big\}\Big).$$

In Appendix B-A, we show that the above choice of $\phi(q(y^{(1)}, y^{(2)}, y^{(3)}, y^{(4)}|x^{(1)}, x^{(2)}, x^{(3)}, x^{(4)}), \Psi)$ verifies the properties of Lemma 1 for $\mathcal{A}_{interference}$ and $\mathcal{Q}_{interference}$. Take an arbitrary interference channel

$$q(y^{(1)}, y^{(2)}, y^{(3)}, y^{(4)}|x^{(1)}, x^{(2)}, x^{(3)}, x^{(4)}),$$

an arbitrary admissible source marginal distribution

$$p(w^{(1)}, w^{(2)}, w^{(3)}, w^{(4)})$$

satisfying $|\mathcal{W}^{(3)}| = |\mathcal{W}^{(4)}| = 1$, and a permissible set of input distributions, Ψ. Let $\overline{\Psi}$ denote the convex hull of Ψ; $\overline{\Psi}$ itself is a permissible set of input distributions. The conditions of Lemma 1 are satisfied for $\mathcal{A}_{interference}$ and $\mathcal{Q}_{interference}$ since $|\mathcal{W}^{(3)}_{1:n}| = |\mathcal{W}^{(4)}_{1:n}| = 1$ and $H(W^{(1)}_{1:n}Y^{(1)}_{1:k}|W^{(1)}_{1:n}) = H(W^{(2)}_{1:n}Y^{(2)}_{1:k}|W^{(2)}_{1:n}) = 0$ for any $1 \leq k \leq n$. For any arbitrary positive ϵ and n-code for which at each stage the inputs belong to the permissible set of input distributions Ψ, the lemma implies

$$\phi\big(p(w^{(1)}_{1:n}y^{(1)}_{1:n}, w^{(2)}_{1:n}y^{(2)}_{1:n}, w^{(3)}_{1:n}y^{(3)}_{1:n}, w^{(4)}_{1:n}y^{(4)}_{1:n}|w^{(1)}_{1:n}, w^{(2)}_{1:n}, w^{(3)}_{1:n}, w^{(4)}_{1:n}), \{p(w^{(1)}_{1:n}, w^{(2)}_{1:n}, w^{(3)}_{1:n}, w^{(4)}_{1:n})\}\big) \subseteq$$
$$n \times Convex\ Hull\big\{\phi\big(q(y^{(1)}, y^{(2)}, y^{(3)}, y^{(4)}|x^{(1)}, x^{(2)}, x^{(3)}, x^{(4)}), \overline{\Psi}\big)\big\}. \quad (9)$$

In Appendix B-B we show that

$$n \times \Big(H(W^{(1)}|L) - O(h(\epsilon)), H(W^{(2)}|L) - O(h(\epsilon)), H(W^{(1)}|L) + H(W^{(2)}|L) - O(h(\epsilon)),$$
$$H(W^{(1)}|L) + H(W^{(2)}|L) - O(h(\epsilon)), H(W^{(1)}W^{(2)}) - O(h(\epsilon)),$$
$$H(W^{(1)}W^{(2)}) - O(h(\epsilon)), H(W^{(1)}|W^{(2)}) - O(h(\epsilon)), H(W^{(2)}|W^{(1)}) - O(h(\epsilon))\Big) \in$$
$$\phi\big(p(w^{(1)}_{1:n}y^{(1)}_{1:n}, w^{(2)}_{1:n}y^{(2)}_{1:n}, w^{(3)}_{1:n}y^{(3)}_{1:n}, w^{(4)}_{1:n}y^{(4)}_{1:n}|w^{(1)}_{1:n}, w^{(2)}_{1:n}, w^{(3)}_{1:n}, w^{(4)}_{1:n}), \{p(w^{(1)}_{1:n}, w^{(2)}_{1:n}, w^{(3)}_{1:n}, w^{(4)}_{1:n})\}\big). \quad (10)$$

In the above expression $O(h(\epsilon))$ we mean a constant (that depends only on the network $q(y^{(1)}, ..., y^{(4)}|x^{(1)}, ..., x^{(4)})$) times $h(\epsilon)$. Here $h(\cdot)$ is the binary entropy function. Equations (9) and (10) imply that

$$\Big(H(W^{(1)}|L) - O(h(\epsilon)), H(W^{(2)}|L) - O(h(\epsilon)), H(W^{(1)}|L) + H(W^{(2)}|L) - O(h(\epsilon)),$$
$$H(W^{(1)}|L) + H(W^{(2)}|L) - O(h(\epsilon)), H(W^{(1)}W^{(2)}) - O(h(\epsilon)),$$
$$H(W^{(1)}W^{(2)}) - O(h(\epsilon)), H(W^{(1)}|W^{(2)}) - O(h(\epsilon)), H(W^{(2)}|W^{(1)}) - O(h(\epsilon))\Big) \in$$
$$Convex\ Hull\big\{\phi\big(q(y^{(1)}, y^{(2)}, y^{(3)}, y^{(4)}|x^{(1)}, x^{(2)}, x^{(3)}, x^{(4)}), \overline{\Psi}\big)\big\}.$$

In Appendix B-C we show that for any interference channel $q(y^{(1)},...,y^{(4)}|x^{(1)},...,x^{(4)})$, $\phi\big(q(y^{(1)},y^{(2)},y^{(3)},y^{(4)}|x^{(1)},x^{(2)},x^{(3)},x^{(4)}),\Psi\big)$ is convex when the set Ψ is convex. Hence

$$\Big(H(W^{(1)}|L)-O(h(\epsilon)),H(W^{(2)}|L)-O(h(\epsilon)),H(W^{(1)}|L)+H(W^{(2)}|L)-O(h(\epsilon)),$$

$$H(W^{(1)}|L)+H(W^{(2)}|L)-O(h(\epsilon)),H(W^{(1)}W^{(2)})-O(h(\epsilon)),$$

$$H(W^{(1)}W^{(2)})-O(h(\epsilon)),H(W^{(1)}|W^{(2)})-O(h(\epsilon)),H(W^{(2)}|W^{(1)})-O(h(\epsilon))\Big)\in$$

$$\phi\big(q(y^{(1)},y^{(2)},y^{(3)},y^{(4)}|x^{(1)},x^{(2)},x^{(3)},x^{(4)}),\overline{\Psi}\big).$$

Since $\phi\big(q(y^{(1)},y^{(2)},y^{(3)},y^{(4)}|x^{(1)},x^{(2)},x^{(3)},x^{(4)}),\overline{\Psi}\big)$ is closed (since the cardinality of U can be bounded, as mentioned in remark 4), letting ϵ converge to zero, we get

$$\Big(H(W^{(1)}|L),H(W^{(2)}|L),H(W^{(1)}|L)+H(W^{(2)}|L),H(W^{(1)}|L)+H(W^{(2)}|L),$$

$$H(W^{(1)}W^{(2)}),H(W^{(1)}W^{(2)}),H(W^{(1)}|W^{(2)}),H(W^{(2)}|W^{(1)})\Big)\in$$

$$\phi\big(q(y^{(1)},y^{(2)},y^{(3)},y^{(4)}|x^{(1)},x^{(2)},x^{(3)},x^{(4)}),\overline{\Psi}\big).$$

The above equation completes the proof of Theorem 2. ■

D. Broadcast channel

A broadcast channel is a three-input/three-output multiterminal network whose set of alphabets is assumed to belong to

$$\mathcal{A}_{broadcast}:=\big\{\big(\mathcal{X}^{(1)},\mathcal{X}^{(2)},\mathcal{X}^{(3)},\mathcal{Y}^{(1)},\mathcal{Y}^{(2)},\mathcal{Y}^{(3)}\big):|\mathcal{X}^{(2)}|=|\mathcal{X}^{(3)}|=1\big\},$$

and whose conditional law is assumed to belong to

$$\mathcal{Q}_{broadcast}:=\big\{q(y^{(1)},y^{(2)},y^{(3)}|x^{(1)},x^{(2)},x^{(3)}):H(Y^{(1)}|X^{(1)})=0\big\}.$$

Theorem 3: Take an arbitrary broadcast channel $q(y^{(1)},y^{(2)},y^{(3)}|x^{(1)},x^{(2)},x^{(3)})$, a permissible set of input distributions Ψ, and an associated admissible source marginal distribution $p(w^{(1)},w^{(2)},w^{(3)})$ for zero distortion levels satisfying $|\mathcal{W}^{(2)}|=|\mathcal{W}^{(3)}|=1$. Take an arbitrary $\epsilon>0$, and a random variable L satisfying $H(L|M^{(2)})=H(L|M^{(3)})=0$. Then there must exist $p(u,v,w_0,w_1,w_2,x^{(1)},x^{(2)},x^{(3)})q(y^{(1)},y^{(2)},y^{(3)}|x^{(1)})$[7]

[7]Note that $q(y^{(1)},y^{(2)},y^{(3)}|x^{(1)})$ can be used for $q(y^{(1)},y^{(2)},y^{(3)}|x^{(1)},x^{(2)},x^{(3)})$ because of the standing assumption that $|\mathcal{X}^{(2)}|=|\mathcal{X}^{(3)}|=1$.

such that

$$p(x^{(1)}, x^{(2)}, x^{(3)}) \text{ is in the convex hull of } \Psi,$$
$$H(W_0|W_1U) = H(W_0|W_2V) = H(X^{(1)}|W_0W_1W_2UV) = 0,$$

and such the following inequalities hold:

$$\begin{aligned}
H(L) - \epsilon &\leq \min\{I(W_0; Y^{(2)}|U), I(W_0; Y^{(3)}|V)\};\\
H(M^{(2)}) - \epsilon &\leq I(W_1; Y^{(2)}|U);\\
H(M^{(3)}) - \epsilon &\leq I(W_2; Y^{(3)}|V);\\
H(M^{(2)}) - \epsilon &\leq I(W_1; Y^{(2)}|W_0UV) + I(W_0U; Y^{(3)}|V);\\
H(M^{(3)}) - \epsilon &\leq I(W_2; Y^{(3)}|W_0UV) + I(W_0V; Y^{(2)}|U);\\
H(M^{(2)}M^{(3)}) - \epsilon &\leq I(W_1; Y^{(2)}|W_0W_2UV) + I(W_2U; Y^{(3)}|V);\\
H(M^{(2)}M^{(3)}) - \epsilon &\leq I(W_2; Y^{(3)}|W_0W_1UV) + I(W_1V; Y^{(2)}|U);\\
H(M^{(2)}M^{(3)}) - \epsilon &\leq I(W_0V; Y^{(2)}|U) + I(W_1; Y^{(2)}|W_0W_2UV) + I(W_2; Y^{(3)}|W_0UV);\\
H(M^{(2)}M^{(3)}) - \epsilon &\leq I(W_0U; Y^{(3)}|V) + I(W_2; Y^{(3)}|W_0W_1UV) + I(W_1; Y^{(2)}|W_0UV).
\end{aligned}$$

Remark 5: Note that random variable L is representing the common part between $M^{(2)}$ and $M^{(3)}$ in the sense of Gács and Körner [22]. Random variable L represents the common message that needs to be transmitted to both receivers.

Remark 6: Although we are claiming the statement for every $\epsilon > 0$, we cannot set $\epsilon = 0$ by taking the limit as $\epsilon \to 0$ since the set of all possible joint distributions $p(u, v, w_0, w_1, w_2, x^{(1)}, x^{(2)}, x^{(3)})$ is not closed, because it is not possible to restrict ourselves to joint distributions on random variables with finite cardinality bounds, since no cardinality bounds on the auxiliary random variables of the outer bound exist at the moment.

Corollary 2: The closure[8] of the following region forms an outer bound on the capacity region of the

[8]Taking the closure is necessary since cardinality bounds are not known to exist for this region, so it is not clear if the region is closed.

general broadcast channel:

$$\begin{aligned} R_0, R_1, R_2 &\geq 0; \\ R_0 &\leq \min\{I(W_0;Y^{(2)}|U), I(W_0;Y^{(3)}|V)\}; \\ R_0+R_1 &\leq I(W_1;Y^{(2)}|U); \\ R_0+R_2 &\leq I(W_2;Y^{(3)}|V); \\ R_0+R_1 &\leq I(W_1;Y^{(2)}|W_0UV) + I(W_0U;Y^{(3)}|V); \\ R_0+R_2 &\leq I(W_2;Y^{(3)}|W_0UV) + I(W_0V;Y^{(2)}|U); \\ R_0+R_1+R_2 &\leq I(W_1;Y^{(2)}|W_0W_2UV) + I(W_2U;Y^{(3)}|V); \\ R_0+R_1+R_2 &\leq I(W_2;Y^{(3)}|W_0W_1UV) + I(W_1V;Y^{(2)}|U); \\ R_0+R_1+R_2 &\leq I(W_0V;Y^{(2)}|U) + I(W_1;Y^{(2)}|W_0W_2UV) + I(W_2;Y^{(3)}|W_0UV); \\ R_0+R_1+R_2 &\leq I(W_0U;Y^{(3)}|V) + I(W_2;Y^{(3)}|W_0W_1UV) + I(W_1;Y^{(2)}|W_0UV), \end{aligned}$$

for some $p(u, v, w_0, w_1, w_2, x^{(1)}, x^{(2)}, x^{(3)})q(y^{(1)}, y^{(2)}, y^{(3)}|x^{(1)})$ satisfying the aforementioned constraints. Here R_0 denotes the common rate, R_1 the private rate to receiver one, and R_2 the private rate to receiver two.

Remark 7: The above outer bound is included inside that of Liang, Kramer and Shamai (see section III-E where the simplified form of Liang, Kramer and Shamai's outer bound is discussed). We don't know if this bound is strictly better than that of [5].

Proof of Theorem 3: In order to apply Lemma 1, we will define $\phi(q(y^{(1)}, y^{(2)}, y^{(3)}|x^{(1)}, x^{(2)}, x^{(3)}), \Psi)$, a function that takes as input an arbitrary 3-input/3-output broadcast channel, i.e. one belonging to $\mathcal{Q}_{broadcast}$ where the alphabets are from $\mathcal{A}_{broadcast}$, and a subset of probability distributions on $\mathcal{X}^{(1)} \times \mathcal{X}^{(2)} \times \mathcal{X}^{(3)}$, and returns a subset of $\mathbb{R}_+^{10}$. We let:

$$\phi(q(y^{(1)}, y^{(2)}, y^{(3)}|x^{(1)}, x^{(2)}, x^{(3)}), \Psi) =$$

$$\bigcup_{p(x^{(1)},x^{(2)},x^{(3)})\in\Psi} \quad \bigcup_{\substack{U,V,W_0,W_1,W_2 \text{ such that} \\ UVW_0W_1W_2 \to X^{(1)} \to Y^{(2)}Y^{(3)}, \\ H(W_0|W_1U) = H(W_0|W_2V) = H(X^{(1)}|W_0W_1W_2UV) = 0}}$$

$$\Pi\bigg(\big\{\big(I(W_0;Y^{(2)}|U),\ I(W_0;Y^{(3)}|V),\ I(W_1;Y^{(2)}|U),\ I(W_2;Y^{(3)}|V),$$

$$I(W_1;Y^{(2)}|W_0UV)+I(W_0U;Y^{(3)}|V),\ I(W_2;Y^{(3)}|W_0UV)+I(W_0V;Y^{(2)}|U),$$

$$I(W_1;Y^{(2)}|W_0W_2UV)+I(W_2U;Y^{(3)}|V),\ I(W_2;Y^{(3)}|W_0W_1UV)+I(W_1V;Y^{(2)}|U)$$

$$I(W_0V;Y^{(2)}|U)+I(W_1;Y^{(2)}|W_0W_2UV)+I(W_2;Y^{(3)}|W_0UV),$$

$$I(W_0U;Y^{(3)}|V)+I(W_2;Y^{(3)}|W_0W_1UV)+I(W_1;Y^{(2)}|W_0UV)\big)\big\}\bigg).$$

In Appendix C-A, we show that the above choice of $\phi(q(y^{(1)},y^{(2)},y^{(3)}|x^{(1)},x^{(2)},x^{(3)}),\Psi)$ verifies the properties of Lemma 1 for $\mathcal{A}_{broadcast}$ and $\mathcal{Q}_{broadcast}$. Take an arbitrary broadcast channel

$$q(y^{(1)},y^{(2)},y^{(3)}|x^{(1)},x^{(2)},x^{(3)}),$$

a permissible set of input distributions Ψ, and an arbitrary admissible source marginal distribution $p(w^{(1)},w^{(2)},w^{(3)})$ satisfying $|\mathcal{W}^{(2)}|=|\mathcal{W}^{(3)}|=1$. Let $\overline{\Psi}$ denote the convex hull of Ψ; $\overline{\Psi}$ itself is a permissible set of input distributions. The conditions of Lemma 1 are satisfied for $\mathcal{A}_{broadcast}$ and $\mathcal{Q}_{broadcast}$ since $|\mathcal{W}^{(2)}_{1:n}|=|\mathcal{W}^{(3)}_{1:n}|=1$ and $H(W^{(1)}_{1:n}Y^{(1)}_{1:k}|W^{(1)}_{1:n})=0$ for any $1\leq k\leq n$ (this verifies that at each stage of the communication the virtual channel alphabets are in $\mathcal{A}_{broadcast}$ and the virtual channel is in $\mathcal{Q}_{broadcast}$). For any arbitrary positive ϵ and n-code for which at each stage the inputs belong to the permissible set of input distributions Ψ, the lemma implies

$$\phi\big(p(w^{(1)}_{1:n}y^{(1)}_{1:n},w^{(2)}_{1:n}y^{(2)}_{1:n},w^{(3)}_{1:n}y^{(3)}_{1:n}|w^{(1)}_{1:n},w^{(2)}_{1:n},w^{(3)}_{1:n}),\{p(w^{(1)}_{1:n},w^{(2)}_{1:n},w^{(3)}_{1:n})\}\big)\subseteq$$
$$n\times Convex\ Hull\big\{\phi\big(q(y^{(1)},y^{(2)},y^{(3)}|x^{(1)},x^{(2)},x^{(3)}),\overline{\Psi}\big)\big\}. \tag{11}$$

In Appendix C-B we show that

$$n\times\bigg(H(L)-O(h(\epsilon)),H(L)-O(h(\epsilon)),H(M^{(2)})-O(h(\epsilon)),H(M^{(3)})-O(h(\epsilon)),$$

$$H(M^{(2)})-O(h(\epsilon)),H(M^{(3)})-O(h(\epsilon)),H(M^{(2)}M^{(3)})-O(h(\epsilon)),$$

$$H(M^{(2)}M^{(3)})-O(h(\epsilon)),H(M^{(2)}M^{(3)})-O(h(\epsilon)),H(M^{(2)}M^{(3)})-O(h(\epsilon))\bigg)\in$$

$$\phi\big(p(w^{(1)}_{1:n}y^{(1)}_{1:n},w^{(2)}_{1:n}y^{(2)}_{1:n},w^{(3)}_{1:n}y^{(3)}_{1:n}|w^{(1)}_{1:n},w^{(2)}_{1:n},w^{(3)}_{1:n}),\{p(w^{(1)}_{1:n},w^{(2)}_{1:n},w^{(3)}_{1:n})\}\big). \tag{12}$$

where by $O(h(\epsilon))$ we mean a constant (that depends only on the network $q(y^{(1)},y^{(2)},y^{(3)}|x^{(1)},x^{(2)},x^{(3)})$) times $h(\epsilon)$. Here $h(\cdot)$ is the binary entropy function. Equations (11) and (12) imply that

$$\bigg(H(L)-O(h(\epsilon)),H(L)-O(h(\epsilon)),H(M^{(2)})-O(h(\epsilon)),H(M^{(3)})-O(h(\epsilon)),$$

$$H(M^{(2)}) - O(h(\epsilon)), H(M^{(3)}) - O(h(\epsilon)), H(M^{(2)}M^{(3)}) - O(h(\epsilon)),$$

$$H(M^{(2)}M^{(3)}) - O(h(\epsilon)), H(M^{(2)}M^{(3)}) - O(h(\epsilon)), H(M^{(2)}M^{(3)}) - O(h(\epsilon))\Big) \in$$

$$Convex\ Hull\big\{\phi\big(q(y^{(1)}, y^{(2)}, y^{(3)}|x^{(1)}, x^{(2)}, x^{(3)}), \overline{\Psi}\big)\big\}.$$

In Appendix C-C we show that for any broadcast channel $q(y^{(1)}, y^{(2)}, y^{(3)}|x^{(1)}, x^{(2)}, x^{(3)})$, $\phi\big(q(y^{(1)}, y^{(2)}, y^{(3)}|x^{(1)}, x^{(2)}, x^{(3)}), \Psi\big)$ is convex when the set Ψ is convex. Hence

$$\Big(H(L) - O(h(\epsilon)), H(L) - O(h(\epsilon)), H(M^{(2)}) - O(h(\epsilon)), H(M^{(3)}) - O(h(\epsilon)),$$

$$H(M^{(2)}) - O(h(\epsilon)), H(M^{(3)}) - O(h(\epsilon)), H(M^{(2)}M^{(3)}) - O(h(\epsilon)),$$

$$H(M^{(2)}M^{(3)}) - O(h(\epsilon)), H(M^{(2)}M^{(3)}) - O(h(\epsilon)), H(M^{(2)}M^{(3)}) - O(h(\epsilon))\Big) \in$$

$$\phi\big(q(y^{(1)}, y^{(2)}, y^{(3)}|x^{(1)}, x^{(2)}, x^{(3)}), \overline{\Psi}\big).$$

Since $h(\epsilon)$ converges zero when ϵ converges zero, the above equation completes the proof of Theorem 3. ∎

E. Analysis of Liang, Kramer and Shamai's outer bound

Given the broadcast channel $q(y^{(2)}, y^{(3)}|x^{(1)})$, Liang, Kramer and Shamai define their outer bound on its capacity region as follows [5]: let $\varrho(q(y^{(2)}, y^{(3)}|x^{(1)}))$ be the union over all joint distributions $p(w_0, w_1, w_2, u, v, x^{(1)}, y^{(2)}, y^{(3)}) = p(w_0, w_1, w_2, u, v, x^{(1)})q(y^{(2)}, y^{(3)}|x^{(1)})$ for which W_0, W_1 and W_2 are both mutually independent and uniform, and $X^{(1)}$ is a deterministic function of (W_0, W_1, W_2, U, V), of the region:

$$\begin{aligned}
R_0, R_1, R_2 &\geq 0; \\
R_0 &\leq \min\{I(W_0; Y^{(2)}|U), I(W_0; Y^{(3)}|V)\}; \\
R_1 &\leq I(W_1; Y^{(2)}|U); \\
R_2 &\leq I(W_2; Y^{(3)}|V); \\
R_0 + R_1 &\leq \min(I(W_0W_1; Y^{(2)}|U), I(W_1; Y^{(2)}|W_0UV) + I(W_0U; Y^{(3)}|V)); \\
R_0 + R_2 &\leq \min(I(W_0W_2; Y^{(3)}|V), I(W_2; Y^{(3)}|W_0UV) + I(W_0V; Y^{(2)}|U)); \\
R_0 + R_1 + R_2 &\leq \min(I(W_1; Y^{(2)}|W_0W_2UV) + I(W_0W_2U; Y^{(3)}|V), \\
&\quad I(W_2; Y^{(3)}|W_0W_1UV) + I(W_0W_1V; Y^{(2)}|U), \\
&\quad I(W_0UV; Y^{(2)}) + I(W_1; Y^{(2)}|W_0W_2UV) + I(W_2; Y^{(3)}|W_0UV), \\
&\quad I(W_0UV; Y^{(3)}) + I(W_2; Y^{(3)}|W_0W_1UV) + I(W_1; Y^{(2)}|W_0UV)).
\end{aligned}$$

Theorem 4: The region $\varrho(q(y^{(2)}, y^{(3)}|x^{(1)}))$ equals $\varrho_1(q(y^{(2)}, y^{(3)}|x^{(1)}))$ defined as follows: let $\varrho_1(q(y^{(2)}, y^{(3)}|x^{(1)}))$ be the region defined as above except that the extra constraint $H(W_0|W_1U) = H(W_0|W_2V) = 0$ is imposed on W_0, W_1, W_2, U and V (so we still require W_0, W_1, W_2 to be mutually independent and uniform on their respective ranges, and $X^{(1)}$ to be a deterministic function of (W_0, W_1, W_2, U, V)) and that the set of inequalities is simplified by replacing them with the following apparently stronger set of inequalities:

$$\begin{aligned}
R_0, R_1, R_2 &\geq 0; \\
R_0 &\leq \min\{I(W_0; Y^{(2)}|U), I(W_0; Y^{(3)}|V)\}; \\
R_0 + R_1 &\leq I(W_1; Y^{(2)}|U); \\
R_0 + R_2 &\leq I(W_2; Y^{(3)}|V); \\
R_0 + R_1 &\leq I(W_1; Y^{(2)}|W_0UV) + I(W_0U; Y^{(3)}|V); \\
R_0 + R_2 &\leq I(W_2; Y^{(3)}|W_0UV) + I(W_0V; Y^{(2)}|U);
\end{aligned}$$

$$
\begin{aligned}
R_0+R_1+R_2 \quad \leq \quad & \min(I(W_1;Y^{(2)}|W_0W_2UV)+I(W_2U;Y^{(3)}|V),\\
& I(W_2;Y^{(3)}|W_0W_1UV)+I(W_1V;Y^{(2)}|U),\\
& I(W_0UV;Y^{(2)})+I(W_1;Y^{(2)}|W_0W_2UV)+I(W_2;Y^{(3)}|W_0UV),\\
& I(W_0UV;Y^{(3)})+I(W_2;Y^{(3)}|W_0W_1UV)+I(W_1;Y^{(2)}|W_0UV)).
\end{aligned}
$$

Remark 8: In order to get the original set of inequalities, replace the inequality $R_0+R_1 \leq I(W_1;Y^{(2)}|U)$ with two weaker inequalities $R_1 \leq I(W_1;Y^{(2)}|U)$ and $R_0+R_1 \leq I(W_0W_1;Y^{(2)}|U)$. Similarly, replace $R_0+R_2 \leq I(W_2;Y^{(3)}|V)$ with $R_2 \leq I(W_2;Y^{(3)}|V)$ and $R_0+R_2 \leq I(W_0W_2;Y^{(3)}|V)$. Furthermore, weaken the first and second inequality on $R_0+R_1+R_2$ by adding respectively terms $I(W_0;Y^{(3)}|W_2UV)$ and $I(W_0;Y^{(2)}|W_1UV)$ to the left hand side of these inequalities.

Proof of Theorem 4: Clearly $\varrho_1(q(y^{(2)},y^{(3)}|x^{(1)})) \subseteq \varrho(q(y^{(2)},y^{(3)}|x^{(1)}))$ because we have replaced the set of inequalities with a stronger one, and have further restricted the set of permissible (W_0,W_1,W_2,U,V). Below we will show that $\varrho(q(y^{(2)},y^{(3)}|x^{(1)})) \subseteq \varrho_1(q(y^{(2)},y^{(3)}|x^{(1)}))$. Take an arbitrary $(R_0,R_1,R_2) \in \varrho(q(y^{(2)},y^{(3)}|x^{(1)}))$. Corresponding to (R_0,R_1,R_2), there exists $p(x^{(1)})$ and $p(w_0,w_1,w_2,u,v,x^{(1)},y^{(2)},y^{(3)}) = p(w_0,w_1,w_2,u,v,x^{(1)})q(y^{(2)},y^{(3)}|x^{(1)})$ for which W_0, W_1 and W_2 are both mutually independent and uniform, and $X^{(1)}$ is a deterministic function of (W_0,W_1,W_2,U,V) such that the inequalities in $\varrho(q(y^{(2)},y^{(3)}|x^{(1)}))$ are satisfied. We will define an appropriate $(\widetilde{U},\widetilde{V},\widetilde{W_0},\widetilde{W_1},\widetilde{W_2},\widetilde{X}^{(1)},\widetilde{Y}^{(2)},\widetilde{Y}^{(3)})$ that would imply that $(R_0,R_1,R_2) \in \varrho_1(q(y^{(2)},y^{(3)}|x^{(1)}))$ (i.e. the corresponding inequalities for (R_0,R_1,R_2) in $\varrho_1(q(y^{(2)},y^{(3)}|x^{(1)}))$ would be satisfied with this choice).

Let random variables A_0, A_1, A_2 and A_3 be uniform on the alphabet of W_0 (without loss of generality assumed to be $\{1,2,...,M_0'\}$ for some M_0'). Let A_2 and A_3 and $(W_0,W_1,W_2,U,V,X^{(1)},Y^{(2)},Y^{(3)})$ be mutually independent. Random variables A_0 and A_1 are then defined as follows:

$$
A_i = 1+(W_0+A_{i+2} \mod M_0') \quad i=0,1.
$$

Let $(\widetilde{U},\widetilde{V},\widetilde{W_0},\widetilde{W_1},\widetilde{W_2},\widetilde{X}^{(1)},\widetilde{Y}^{(2)},\widetilde{Y}^{(3)})$ be equal to $(UA_0,VA_1,W_0,W_1A_2,W_2A_3,X^{(1)},Y^{(2)},Y^{(3)})$.

It can be verified that $p(\widetilde{x}) = p(x)$. Furthermore $(\widetilde{U},\widetilde{V},\widetilde{W_0},\widetilde{W_1},\widetilde{W_2},\widetilde{X}^{(1)},\widetilde{Y}^{(2)},\widetilde{Y}^{(3)})$ satisfies the required properties (in particular $H(\widetilde{W_0}|\widetilde{W_1}\widetilde{U}) = H(\widetilde{W_0}|\widetilde{W_2}\widetilde{V}) = 0$). One can use $(\widetilde{U},\widetilde{V},\widetilde{W_0},\widetilde{W_1},\widetilde{W_2},\widetilde{X}^{(1)},\widetilde{Y}^{(2)},\widetilde{Y}^{(3)})$ in the definition of ϱ_1 to show that $(R_0,R_1,R_2) \in \varrho_1(q(y^{(2)},y^{(3)}|x^{(1)}))$.

More specifically, one can verify the following inequalities:

$$\begin{cases} R_0 \geq 0, R_1 \geq 0, R_2 \geq 0; \\ R_0 \leq \min\{I(\widetilde{W_0};\widetilde{Y}^{(2)}|\widetilde{U}), I(\widetilde{W_0};\widetilde{Y}^{(3)}|\widetilde{V})\}; \\ R_0 + R_1 \leq I(\widetilde{W_1};\widetilde{Y}^{(2)}|\widetilde{U}); \\ ... \\ R_0 + R_1 + R_2 \leq \\ \quad I(\widetilde{W_0}\widetilde{U}\widetilde{V};\widetilde{Y}^{(3)}) + I(\widetilde{W_2};\widetilde{Y}^{(3)}|\widetilde{W_0}\widetilde{W_1}\widetilde{U}\widetilde{V}) + I(\widetilde{W_1};\widetilde{Y}^{(2)}|\widetilde{W_0}\widetilde{U}\widetilde{V}). \end{cases}$$

■

Theorem 5: The region $\varrho_1(q(y^{(2)}, y^{(3)}|x^{(1)}))$ with or without the constraint on W_0, W_1 and W_2 being mutually independent and uniform is the same.

Remark 9: Theorem 5 originally appeared in [6]. The technique used to relax the mutually independence constraint on W_0, W_1 and W_2 is similar to the one used in an earlier work by Nair and Zizhou in [10]. Nair and El Gamal in [13] had proposed two outer bounds, defined in equation (3.1) and in Theorem 3.1 of [13], and had suspected that one is strictly tighter than the other. In [10], Nair and Zizhou showed that this is not the case.

Proof of Theorem 5: Take an arbitrary joint distribution $p(w_0, w_1, w_2, u, v, x^{(1)}, y^{(2)}, y^{(3)}) = p(w_0, w_1, w_2, u, v, x^{(1)})q(y^{(2)}, y^{(3)}|x^{(1)})$ for which $X^{(1)}$ is a deterministic function of (W_0, W_1, W_2, U, V) and $H(W_0|W_1U) = H(W_0|W_2V) = 0$. We will define $(\widetilde{U}, \widetilde{V}, \widetilde{W_0}, \widetilde{W_1}, \widetilde{W_2}, \widetilde{X}^{(1)}, \widetilde{Y}^{(2)}, \widetilde{Y}^{(3)})$ that yield the same region of triples of (R_0, R_1, R_2) as in the definition of $\varrho_1(q(y^{(2)}, y^{(3)}|x^{(1)}))$, but furthermore satisfy the additional constraint that $\widetilde{W_0}, \widetilde{W_1}, \widetilde{W_2}$ are both mutually independent and uniform.

Let random variables A_0, A_1, A_2 be uniform on the alphabets of W_0, W_1 and W_2 respectively (without loss of generality assumed to be $\{1, 2, ..., M_i'\}$ for some M_i' ($i = 0, 1, 2$)). Furthermore assume that A_0, A_1, A_2 and $(W_0, W_1, W_2, U, V, X^{(1)}, Y^{(2)}, Y^{(3)})$ are mutually independent. Random variables A_3, A_4 and A_5 are then defined as follows:

$$A_{i+3} = 1 + (W_i + A_i \mod M_i') \quad i = 0, 1, 2.$$

It can be verified that A_3, A_4, A_5 and $(W_0, W_1, W_2, U, V, X^{(1)}, Y^{(2)}, Y^{(3)})$ are mutually independent. Let $(\widetilde{U}, \widetilde{V}, \widetilde{W_0}, \widetilde{W_1}, \widetilde{W_2}, \widetilde{X}^{(1)}, \widetilde{Y}^{(2)}, \widetilde{Y}^{(3)})$ be equal to $(UA_3A_4A_5, VA_3A_4A_5, A_0, A_1, A_2, X^{(1)}, Y^{(2)}, Y^{(3)})$.

It can be verified that $p(\widetilde{x}) = p(x)$. Furthermore $(\widetilde{U}, \widetilde{V}, \widetilde{W_0}, \widetilde{W_1}, \widetilde{W_2}, \widetilde{X}^{(1)}, \widetilde{Y}^{(2)}, \widetilde{Y}^{(3)})$ satisfies all the required properties; in particular $\widetilde{W_0}, \widetilde{W_1}, \widetilde{W_2}$ are both mutually independent and uniform.

■

Acknowledgement

The research of the authors was supported by the ARO MURI grant W911NF-08-1-0233, ”Tools for the Analysis and Design of Complex Multi-Scale Networks”, by the NSF grants CCF-0500234, CCF-0635372, CNS-0627161, and CNS-0910702, by Marvell Semiconductor Inc., and by the U. C. Discovery program.

References

[1] T. M. Cover and J. A. Thomas, *Elements of Information Theory*, John Wiley and Sons, 1991.

[2] Amin A. Gohari and V. Anantharam, “Information-Theoretic Key Agreement of Multiple Terminals – Part I: Source Model,” accepted for publication in *IEEE Trans. Information Theory*. Available at http://www.eecs.berkeley.edu/∼aminzade/SourceModel.pdf

[3] Amin A. Gohari and V. Anantharam, “Information-Theoretic Key Agreement of Multiple Terminals – Part II: Channel Model,” accepted for publication in *IEEE Trans. Information Theory*. Available at http://www.eecs.berkeley.edu/∼aminzade/ChannelModel.pdf

[4] I. Maric, R. D. Yates and G. Kramer, “The Capacity Region of the Strong Interference Channel with Common Information”, *Asilomar Conference On Signals, Systems and Computers*, Pacific Grove, CA, pp. 1737-1741 (2005).

[5] Y. Liang, G. Kramer, and S. Shamai (Shitz), “Capacity outer bounds for broadcast channels,” *IEEE Inf. Theory Workshop*, Porto, Portugal, May 5-9, pp. 2-4 (2008).

[6] A. A. Gohari and V. Anantharam,, “An Outer Bound to the Admissible Source Region of Broadcast Channels with Arbitrarily Correlated Sources and Channel Variations,” *Proceedings of the 46th Annual Allerton Conference on Communications, Control and Computing*, Urbana, Illinois, September 23 -26, pp. 301-308 (2008).

[7] A. A. Gohari and V. Anantharam, “A Generalized Cut-Set Bound,” *Proceedings of the International Symposium on Information Theory (ISIT)*, South Korea, pp. 99-103 (2009).

[8] T. M. Cover, A. El Gamal, and M. Salehi, “Multiple access channels with arbitrarily correlated sources,” *IEEE Trans. Information Theory*, IT-26, pp. 648-657 (1980).

[9] Te Han, M. Costa, “Broadcast Channels with Arbitrarily Correlated Sources,” *IEEE Trans. Information Theory* 33(5), pp. 641- 650 (1987)

[10] C. Nair and V.W. Zizhou, “On the inner and outer bounds for 2-receiver discrete memoryless broadcast channels,” *Proceedings of the ITA workshop*, San Diego, pp. 226-229 (2008).

[11] Y. Cao, B. Chen, J. Zhang, “A New Achievable Rate Region for Interference Channels with Common Information,” *Wireless Communications and Networking Conference*, IEEE 11-15: pp. 2069 - 2073 (2007).

[12] C. Nair, “An outer bound for 2-receiver discrete memoryless broadcast channels,” Available at http://chandra.ie.cuhk.edu.hk/pub/papers/outerbound.pdf

[13] C. Nair and A. A. El Gamal, “An outer bound to the capacity region of the broadcast channel,” *IEEE Trans. Information Theory*, 53(1), pp. 350-355 (2007).

[14] M. Gastpar, “Cut-set Arguments For Source-Channel Networks,” *Proceedings of the International Symposium on Information Theory (ISIT)*, 34, (2004).

[15] G. Kramer and S. A. Savari, "Cut sets and information flow in networks of two-way channels," *Proceedings of the International Symposium on Information Theory (ISIT)*, 33, (2004).

[16] B. Nazer and M. Gastpar, "Computation over multiple-access channels," *IEEE Trans. Information Theory*, 53(10) pp. 3498-3516 (2007).

[17] A. Giridhar and P. R. Kumar, "Computing and communicating functions over sensor networks," *IEEE Journal on Selected Areas in Communications*, 23(4), pp. 755-764 (2005).

[18] A. Orlitsky and J. R. Roche, "Coding for computing," *IEEE Trans. Information Theory*, 47 (3), pp. 903-917 (2001).

[19] H. Yamamoto, "Wyner-Ziv theory for a general function of the correlated sources," *IEEE Trans. Information Theory*, 28 (5), pp. 803-807 (1982).

[20] R. Appuswamy, M. Franceschetti, N. Karamchandani, and K. Zeger, "Network Coding for Computing", *Proceedings of the 46th Annual Allerton Conference on Communications, Control and Computing*, Urbana, Illinois, September 23 -26, pp. 1-6 (2008).

[21] M. H. M. Costa and A. El Gamal, "The capacity region of the discrete memoryless interference channel with strong interference", *IEEE Trans. Information Theory*, vol. 33 (5), pp. 710-711 (1987).

[22] P. Gács and J, Körner, "Common information is much less than mutual information," *Probl. Contr. Inf. Theory*, 150 (2), pp. 149-162 (1973).

[23] Deniz Gündüz, Elza Erkip, Andrea Goldsmith, H. Vincent Poor, "Source and Channel Coding for Correlated Sources Over Multiuser Channels," *IEEE Trans. Information Theory*, 55 (9), pp. 3927-3944 (2009).

[24] W. Kang and S. Ulukus, "A new data processing inequality and its applications in distributed source and channel coding," submitted to *IEEE Trans. Information Theory*.

[25] S. S. Pradhan, S. Choi and K. Ramchandran, "A graph-based framework for transmission of correlated sources over multiple access channels," *IEEE Trans. Information Theory*, 53 (12), pp. 4583-4604 (2007).

[26] G. Kramer, Y. Liang and S. Shamai (Shitz), "Outer bounds on the admissible source region for broadcast channels with dependent sources," *Inform. Theory and Applications Workshop*, pp. 162-179 (2009).

[27] P. Minero and Y.-H. Kim, "Correlated Sources over Broadcast Channels," *Proceedings of the International Symposium on Information Theory (ISIT)*, South Korea, pp. 2780 - 2784 (2009).

APPENDIX A

COMPLETING THE PROOF OF THEOREM 1 ON THE CUT-SET BOUND

A. Verifying the properties of Lemma 1

1) Checking the first property of Lemma 1: Given the definition of ϕ in equation (7), one needs to verify that:

$$\varphi\big(p(y^{(1)}y'^{(1)},...,y^{(m)}y'^{(m)}|x^{(1)},x^{(2)},...,x^{(m)})p(x^{(1)},x^{(2)},...,x^{(m)})\big) \subseteq$$

$$\varphi\big(p(y^{(1)},y^{(2)},...,y^{(m)}|x^{(1)},x^{(2)},...,x^{(m)})p(x^{(1)},x^{(2)},...,x^{(m)})\big)\oplus$$

$$\bigcup_{p(x'^{(1)},x'^{(2)},...,x'^{(m)})\in\Psi} \varphi\big(p(y'^{(1)},y'^{(2)},...,y'^{(m)}|x'^{(1)},x'^{(2)},...,x'^{(m)})p(x'^{(1)},x'^{(2)},...,x'^{(m)})\big).$$

Take an arbitrary point $\overrightarrow{v}$ inside

$$\varphi\big(p(y^{(1)}y^{'(1)},...,y^{(m)}y^{'(m)}|x^{(1)},x^{(2)},...,x^{(m)})p(x^{(1)},x^{(2)},...,x^{(m)})\big).$$

We would like to prove that there exists

$$\overrightarrow{v_1}\in\varphi\big(p(y^{(1)},y^{(2)},...,y^{(m)}|x^{(1)},x^{(2)},...,x^{(m)})p(x^{(1)},x^{(2)},...,x^{(m)})\big),$$

and

$$\overrightarrow{v_2}\in\varphi\big(p(y^{'(1)},y^{'(2)},...,y^{'(m)}|x^{'(1)},x^{'(2)},...,x^{'(m)})p(x^{'(1)},x^{'(2)},...,x^{'(m)})\big),$$

for some $p(x^{'(1)},x^{'(2)},...,x^{'(m)})\in\Psi$ such that $\overrightarrow{v_1}+\overrightarrow{v_2}\geq\overrightarrow{v}$.

Since $\overrightarrow{v}$ is inside

$$\varphi\big(p(y^{(1)}y^{'(1)},...,y^{(m)}y^{'(m)}|x^{(1)},x^{(2)},...,x^{(m)})p(x^{(1)},x^{(2)},...,x^{(m)})\big),$$

the k^{th} coordinate of $\overrightarrow{v}$ is less than or equal to $I\big(X^{(i)}:i\in T_k\ ;\ Y^{(j)}Y^{'(j)}:j\in(T_k)^c|X^{(j)}:j\in(T_k)^c\big)$ where T_k is defined as in the proof of Theorem 1.

We have:

$$I\big(X^{(i)}:i\in T_k\ ;\ Y^{(j)}Y^{'(j)}:j\in(T_k)^c|X^{(j)}:j\in(T_k)^c\big)=$$

$$I\big(X^{(i)}:i\in T_k\ ;\ Y^{(j)}:j\in(T_k)^c|X^{(j)}:j\in(T_k)^c\big)+$$

$$I\big(X^{(i)}:i\in T_k\ ;\ Y^{'(j)}:j\in(T_k)^c|X^{(j)}:j\in(T_k)^c,Y^{(j)}:j\in(T_k)^c\big).$$

The second term can be written as:

$$I\big(X^{(i)}:i\in T_k\ ;\ Y^{'(j)}:j\in(T_k)^c|X^{(j)}:j\in(T_k)^c,Y^{(j)}:j\in(T_k)^c\big)\leq \tag{13}$$

$$I\big(X^{(i)}X^{'(i)}:i\in T_k\ ;\ Y^{'(j)}:j\in(T_k)^c|X^{(j)}:j\in(T_k)^c,Y^{(j)}X^{'(j)}:j\in(T_k)^c\big)= \tag{14}$$

$$I\big(X^{'(i)}:i\in T_k\ ;\ Y^{'(j)}:j\in(T_k)^c|X^{(j)}X^{'(j)}Y^{(j)}:j\in(T_k)^c\big)+0=$$

$$I\big(X^{'(i)}:i\in T_k,X^{(j)}Y^{(j)}:j\in(T_k)^c\ ;\ Y^{'(j)}:j\in(T_k)^c|X^{'(j)}:j\in(T_k)^c\big)-$$

$$I\big(X^{(j)}Y^{(j)}:j\in(T_k)^c\ ;\ Y^{'(j)}:j\in(T_k)^c|X^{'(j)}:j\in(T_k)^c\big)= \tag{15}$$

$$I\big(X^{'(i)}:i\in T_k\ ;\ Y^{'(j)}:j\in(T_k)^c|X^{'(j)}:j\in(T_k)^c\big)-$$

$$I\big(X^{(j)}Y^{(j)}:j\in(T_k)^c\ ;\ Y^{'(j)}:j\in(T_k)^c|X^{'(j)}:j\in(T_k)^c\big)\leq$$

$$I\big(X^{'(i)}:i\in T_k\ ;\ Y^{'(j)}:j\in(T_k)^c|X^{'(j)}:j\in(T_k)^c\big)$$

where in inequality (13) we have used the fact that $H(X'^{(j)}|X^{(j)}Y^{(j)})$=0 to add $X'^{(j)}: \ j \in (T_k)^c$ in the conditioning part of the mutual information term. We have also added $X'^{(i)} : i \in T_k$, but this cannot cause the expression decrease. In equation (14) and equation (15) we have used the following Markov chain

$$\big(Y'^{(i)} : i \in [m]\big) - (X'^{(i)} : i \in [m]) - (Y^{(i)}X^{(i)} : i \in [m]).$$

The the k^{th} coordinate of $\overrightarrow{v}$ is thus less than or equal to

$$I\big(X^{(i)} : i \in T_k \ \ ; \ \ Y^{(j)} : j \in (T_k)^c | X^{(j)} : j \in (T_k)^c\big)+$$
$$I\big(X'^{(i)} : i \in T_k \ \ ; \ \ Y'^{(j)} : j \in (T_k)^c | X'^{(j)} : j \in (T_k)^c\big).$$

Let the k^{th} coordinate of $\overrightarrow{v_1}$ be

$$I\big(X^{(i)} : i \in T_k \ \ ; \ \ Y^{(j)} : j \in (T_k)^c | X^{(j)} : j \in (T_k)^c\big),$$

and the k^{th} coordinate of $\overrightarrow{v_2}$ be

$$I\big(X'^{(i)} : i \in T_k \ \ ; \ \ Y'^{(j)} : j \in (T_k)^c | X'^{(j)} : j \in (T_k)^c\big).$$

■

2) Checking the second property of Lemma 1: Our choice of ϕ implies

$$\phi\big(p(y^{(1)}, ..., y^{(m)}|x^{(1)}, ..., x^{(m)}), \{q(x_1, ..., x_m)\}\big) = \varphi(p(y^{(1)}, ..., y^{(m)}|x^{(1)}, ..., x^{(m)})p(x^{(1)}, ..., x^{(m)})).$$

Take an arbitrary point $\overrightarrow{v}$ inside the above set. The k^{th} coordinate of $\overrightarrow{v}$ is less than or equal to $I\big(X^{(i)} : i \in T_k \ \ ; \ \ Y^{(j)} : j \in (T_k)^c | X^{(j)} : j \in (T_k)^c\big)$ where T_k is defined as in the proof of Theorem 1. Since $Y^{(j)} = X^{(j)}$ for $j \in [m]$, the k^{th} coordinate of $\overrightarrow{v}$ would be less than or equal to zero. But $\overrightarrow{v}$ also lies in $\mathbb{R}_+^c$, hence it has to be equal to the all zero vector. ■

B. Exchanging distortion for mutual information

We will define random variables $\widetilde{\widehat{M}'^{(i)}}$ (for $i \in [m]$) such that for any $i \in [m]$

$$\mathbb{E}\big[\Delta_i(\widetilde{\widehat{M}'^{(i)}}, \widetilde{\widehat{M}^{(i)}})\big] \leq D^{(i)},$$

and furthermore

$$I\big(\widetilde{W}^{(i)} : i \in T_k \ \ ; \ \ \widetilde{\widehat{M}'^{(j)}} : j \in (T_k)^c | \widetilde{W}^{(j)} : j \in (T_k)^c\big) - O(\tau(\epsilon)) \leq$$
$$I\big(\widetilde{W}^{(i)} : i \in T_k \ \ ; \ \ \widetilde{\widehat{M}^{(j)}} : j \in (T_k)^c | \widetilde{W}^{(j)} : j \in (T_k)^c\big),$$

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

Intuitively speaking, the algorithm for creating $\widetilde{\widehat{M'}^{(i)}}$ is to begin with $\widetilde{\widehat{M}^{(i)}}$ ($i \in [m]$), and then perturb this set of m random variables in m stages as follows: at the r^{th} stage, we perturb the r^{th} random variable so that the average distortion constraint is satisfied while making sure that changes in the mutual information terms are under control.

More precisely, let $(G_0^{(1)}, G_0^{(2)}, ..., G_0^{(m)})$ be equal to $(\widetilde{\widehat{M}^{(1)}}, \widetilde{\widehat{M}^{(2)}}, ..., \widetilde{\widehat{M}^{(m)}})$. We define random variables $(G_r^{(1)}, G_r^{(2)}, ..., G_r^{(m)})$ for $r \in [m]$ using $(G_{r-1}^{(1)}, G_{r-1}^{(2)}, ..., G_{r-1}^{(m)})$ in a sequential manner as follows: let $G_r^{(i)} := G_{r-1}^{(i)}$ for all $i \in [m]$, $i \neq r$. Random variable $G_r^{(r)}$ is defined below by perturbing $G_{r-1}^{(i)}$ in such a way that the average distortion between $G_r^{(r)}$ and $\widetilde{M}^{(r)}$ is less than or equal to $D^{(r)}$ while making sure that for any $k \in [2^m - 2]$,

$$I\big(\widetilde{W}^{(i)} : i \in T_k \ ; \ G_r^{(j)} : j \in (T_k)^c | \widetilde{W}^{(j)} : j \in (T_k)^c\big) - I\big(\widetilde{W}^{(i)} : i \in T_k \ ; \ G_{r-1}^{(j)} : j \in (T_k)^c | \widetilde{W}^{(j)} : j \in (T_k)^c\big)$$

is of order $O(\tau_r(\epsilon))$ where τ_r is a real-valued function that satisfies the property that $\tau_r(\epsilon) \to 0$ as $\epsilon \to 0$. Once this is done, we can take $\widetilde{\widehat{M'}^{(i)}} = G_m^{(i)}$ for all $i \in [m]$ and let $\tau(\epsilon) = \sum_{r=1}^{m} \tau_r(\epsilon)$.

For any $k \in [2^m - 2]$, as long as r does not belong to $(T_k)^c$, the expression

$$I\big(\widetilde{W}^{(i)} : i \in T_k \ ; \ G_r^{(j)} : j \in (T_k)^c | \widetilde{W}^{(j)} : j \in (T_k)^c\big) -$$
$$I\big(\widetilde{W}^{(i)} : i \in T_k \ ; \ G_{r-1}^{(j)} : j \in (T_k)^c | \widetilde{W}^{(j)} : j \in (T_k)^c\big),$$

would be zero no matter how $G_r^{(r)}$ is defined. We should therefore consider only the cases where r belongs to $(T_k)^c$. In order to define $G_r^{(r)}$, we consider two cases:

1) Case $D^{(r)} \neq 0$: Take a binary random variable Q_r independent of all other random variables defined in previous stages. Assume that $P(Q_r = 0) = \frac{\epsilon}{D^{(r)}+\epsilon}$ and $P(Q_r = 1) = \frac{D^{(r)}}{D^{(r)}+\epsilon}$. Let $G_r^{(r)}$ be equal to $G_{r-1}^{(r)}$ if $Q_r = 1$, and be equal to $\widetilde{M}^{(r)}$ if $Q_r = 0$. It can be verified that the average distortion between $G_r^{(r)}$ and $\widetilde{M}^{(r)}$ is less than or equal to $D^{(r)}$.[9]

   Take an arbitrary $k \in [2^m - 2]$ such that $r \in T_k$. Since for any five random variables A, B, B', C, D where D is independent of (A, B, C) we have $I(A; B'|C) - I(A; B|C) \leq I(A; B'|BCD)$, [10] we can write:

$$I\big(\widetilde{W}^{(i)} : i \in T_k \ ; \ G_r^{(j)} : j \in (T_k)^c | \widetilde{W}^{(j)} : j \in (T_k)^c\big) -$$

[9] This is because $\mathbb{E}\big[\Delta_r(G_r^{(r)}, \widetilde{M}^{(r)})\big] = \mathbb{E}\big[\mathbb{E}\big[\Delta_r(G_r^{(r)}, \widetilde{M}^{(r)})|Q_r\big]\big] = P(Q_r = 1)\mathbb{E}\big[\Delta_r(G_{r-1}^{(r)}, \widetilde{M}^{(r)})\big] \leq \frac{D^{(r)}}{D^{(r)}+\epsilon} \cdot (D^{(r)} + \epsilon) = D^{(r)}$.

[10] This is because $I(A; B|C) \geq I(A; B'|C) - I(A; B'|BC) \geq I(A; B'|C) - I(A; B'D|BC) \geq I(A; B'|C) - I(A; D|BC) - I(A; B'|BCD) = I(A; B'|C) - 0 - I(A; B'|BCD) = I(A; B'|C) - I(A; B'|BCD)$.

$$I\big(\widetilde{W}^{(i)}: i \in T_k \;\; ; \;\; G_{r-1}^{(j)}: j \in (T_k)^c | \widetilde{W}^{(j)}: j \in (T_k)^c\big) \leq$$

$$I\big(\widetilde{W}^{(i)}: i \in T_k \;\; ; \;\; G_{r}^{(j)}: j \in (T_k)^c | G_{r-1}^{(j)} \widetilde{W}^{(j)}: j \in (T_k)^c, Q_r\big).$$

We would like to prove that the last term is of order $\tau_r(\epsilon) := O(\frac{\epsilon}{D^{(r)}+\epsilon})$. Clearly then $\tau_r(\epsilon) \to 0$ as $\epsilon \to 0$ since $D^{(r)}$ is assumed to be non-zero. The last term above is of order $\frac{\epsilon}{D^{(r)}+\epsilon}$ because:

$$I\big(\widetilde{W}^{(i)}: i \in T_k \;\; ; \;\; G_{r}^{(j)}: j \in (T_k)^c | G_{r-1}^{(j)} \widetilde{W}^{(j)}: j \in (T_k)^c, Q_r\big) =$$

$$0 \cdot P(Q_r = 1) +$$

$$I\big(\widetilde{W}^{(i)}: i \in T_k \;\; ; \;\; G_{r}^{(j)}: j \in (T_k)^c | G_{r-1}^{(j)} \widetilde{W}^{(j)}: j \in (T_k)^c, Q_r = 0\big) \cdot P(Q_r = 0) \leq$$

$$H(\widetilde{W}^{(i)}: i \in [m]) \cdot P(Q_r = 0) = O(\frac{\epsilon}{D^{(i)}+\epsilon}).$$

2) Case $D^{(r)} = 0$: Let the binary random variable Q_r be the indicator function $\mathbf{1}[\Delta_r(G_{r-1}^{(r)}, \widetilde{M}^{(r)}) = 0]$. Let $G_r^{(r)}$ be equal to $G_{r-1}^{(r)}$ if $Q_r = 1$, and be equal to $\widetilde{M}^{(r)}$ if $Q_r = 0$. The average distortion between $G_r^{(r)}$ and $\widetilde{M}^{(r)}$ is clearly zero. Since the average distortion between $G_{r-1}^{(r)}$ and $\widetilde{M}^{(r)}$ is less than or equal to ϵ, we get that $P(Q_r = 0) \leq \frac{\epsilon}{\delta_{min}}$ where δ_{min} is defined as follows: ($\widetilde{\mathcal{M}}^{(r)}$ here refers to the set $\widetilde{M}^{(r)}$ is taking values from)

$$\delta_{min} = \min_{\substack{i,j \in \widetilde{\mathcal{M}}^{(r)} \text{ such that} \\ \Delta_r(i,j) \neq 0}} \Delta_r(i,j).$$

Take an arbitrary $k \in [2^m - 2]$ such that $r \in T_k$.

$$I\big(\widetilde{W}^{(i)}: i \in T_k \;\; ; \;\; G_{r}^{(j)}: j \in (T_k)^c | \widetilde{W}^{(j)}: j \in (T_k)^c\big) -$$

$$I\big(\widetilde{W}^{(i)}: i \in T_k \;\; ; \;\; G_{r-1}^{(j)}: j \in (T_k)^c | \widetilde{W}^{(j)}: j \in (T_k)^c\big) =$$

$$H\big(\widetilde{W}^{(i)}: i \in T_k | G_r^{(j)} \widetilde{W}^{(j)}: j \in (T_k)^c\big) -$$

$$H\big(\widetilde{W}^{(i)}: i \in T_k | G_{r-1}^{(j)} \widetilde{W}^{(j)}: j \in (T_k)^c\big) \leq$$

$$H(Q_r) + H\big(\widetilde{W}^{(i)}: i \in T_k | G_r^{(j)} \widetilde{W}^{(j)}: j \in (T_k)^c, Q_r\big) -$$

$$H\big(\widetilde{W}^{(i)}: i \in T_k | G_{r-1}^{(j)} \widetilde{W}^{(j)}: j \in (T_k)^c, Q_r\big) \leq$$

$$H(Q_r) + P(Q_r = 0) \cdot H\big(\widetilde{W}^{(i)}: i \in T_k | G_r^{(j)} \widetilde{W}^{(j)}: j \in (T_k)^c, Q_r = 0\big) \leq$$

$$H(Q_r) + P(Q_r = 0) \cdot H(\widetilde{W}^{(i)}: i \in [m]).$$

Let $\tau_r(\epsilon) := H(Q_r) + P(Q_r = 0) \cdot H(\widetilde{W}^{(i)}: i \in [m])$. Since $P(Q_r = 0)$ is bounded from above by $\frac{\epsilon}{\delta_{min}}$ that converges to zero as $\epsilon \to 0$, $\tau_r(\epsilon)$ too would converge to zero as $\epsilon \to 0$.

∎

APPENDIX B

COMPLETING THE PROOF OF THEOREM 2 ON THE INTERFERENCE CHANNEL

A. Verifying the properties of Lemma 1

In this appendix we show that our choice of ϕ verifies the two properties of the main lemma for $m = 4$ when restricted to the class of interference channels. For the first property of the lemma, take a channel

$$p(y^{(1)}y'^{(1)}, ..., y^{(4)}y'^{(4)}|x^{(1)}, ..., x^{(4)}) =$$
$$p(y^{(1)}, ..., y^{(4)}|x^{(1)}, ..., x^{(4)}) \cdot p(y'^{(1)}, ..., y'^{(4)}|x'^{(1)}, ..., x'^{(4)})$$

where $H(X'^{(i)}|X^{(i)}Y^{(i)}) = 0$ for $1 \leq i \leq 4$. Take an arbitrary point from

$$\phi\big(p(y^{(1)}y'^{(1)}, ..., y^{(4)}y'^{(4)}|x^{(1)}, ..., x^{(4)}), \{q(x^{(1)}, ..., x^{(4)})\}\big).$$

The set of input distributions contains only one distribution $q(x^{(1)}, ..., x^{(4)})$. Therefore we define random variables $X^{(1)}, X^{(2)}, ..., X^{(4)}, Y^{(1)}, Y'^{(1)}, ..., Y^{(4)}, Y'^{(4)}$ jointly distributed according to

$$p(y^{(1)}y'^{(1)}, ..., y^{(4)}y'^{(4)}|x^{(1)}, ..., x^{(4)}) \cdot q(x^{(1)}, ..., x^{(4)}) =$$
$$p(y^{(1)}, ..., y^{(4)}|x^{(1)}, ..., x^{(4)}) \cdot p(y'^{(1)}, ..., y'^{(4)}|x'^{(1)}, ..., x'^{(4)}) \cdot q(x^{(1)}, ..., x^{(4)}).$$

Corresponding to the arbitrary point in

$$\phi\big(p(y^{(1)}y'^{(1)}, ..., y^{(4)}y'^{(4)}|x^{(1)}, ..., x^{(4)}), \{q(x^{(1)}, ..., x^{(4)})\}\big)$$

is a random variable U satisfying $X^{(1)} \to U \to X^{(2)}$ and $U \to X^{(1)}X^{(2)} \to Y^{(3)}Y^{(4)}Y'^{(3)}Y'^{(4)}$ such that the point is coordinate by coordinate less than or equal to the point

$$\overrightarrow{v} = \big(I(X^{(1)}; Y^{(3)}Y'^{(3)}|X^{(2)}, U), I(X^{(2)}; Y^{(4)}Y'^{(4)}|X^{(1)}, U), I(X^{(1)}X^{(2)}; Y^{(3)}Y'^{(3)}|U),$$
$$I(X^{(1)}X^{(2)}; Y^{(4)}Y'^{(4)}|U), I(X^{(1)}X^{(2)}; Y^{(3)}Y'^{(3)}), I(X^{(1)}X^{(2)}; Y^{(4)}Y'^{(4)}),$$
$$I(X^{(1)}; Y^{(3)}Y'^{(3)}|X^{(2)}), I(X^{(2)}; Y^{(4)}Y'^{(4)}|X^{(1)})\big).$$

Since $p(y^{(1)}, ..., y^{(4)}|x^{(1)}, ..., x^{(4)})$ and $p(y'^{(1)}, ..., y'^{(4)}|x'^{(1)}, ..., x'^{(4)})$ belong to the class of interference channels, we have $H(Y^{(1)}|X^{(1)}) = H(Y^{(2)}|X^{(2)}) = 0$. $H(X'^{(1)}|X^{(1)}Y^{(1)}) = H(X'^{(2)}|X^{(2)}Y^{(2)}) = 0$ then implies that $H(X'^{(1)}|X^{(1)}) = 0$ and $H(X'^{(2)}|X^{(2)}) = 0$. Since $|\mathcal{X}^{(3)}| = |\mathcal{X}^{(4)}| = |\mathcal{X}'^{(3)}| = |\mathcal{X}'^{(4)}| = 1$

$$p(y^{(3)}y'^{(3)}, y^{(4)}y'^{(4)}|x^{(1)}, x^{(2)}, x^{(3)}, x^{(4)}) =$$
$$p(y^{(3)}, y^{(4)}|x^{(1)}, x^{(2)}) \cdot p(y'^{(3)}, y'^{(4)}|x'^{(1)}, x'^{(2)}).$$

Since $X^{(1)} \to U \to X^{(2)}$ and $U \to X^{(1)}X^{(2)} \to Y^{(3)}Y^{(4)}$, the point

$$\overrightarrow{v_1} = \big(I(X^{(1)};Y^{(3)}|X^{(2)},U),\ I(X^{(2)};Y^{(4)}|X^{(1)},U),\ I(X^{(1)}X^{(2)};Y^{(3)}|U),$$
$$I(X^{(1)}X^{(2)};Y^{(4)}|U),\ I(X^{(1)}X^{(2)};Y^{(3)}),\ I(X^{(1)}X^{(2)};Y^{(4)}),$$
$$I(X^{(1)};Y^{(3)}|X^{(2)}),\ I(X^{(2)};Y^{(4)}|X^{(1)})\big)$$

belongs to $\phi\big(p(y^{(1)},...,y^{(4)}|x^{(1)},...,x^{(4)}),\{q(x^{(1)},x^{(2)},...,x^{(4)})\}\big)$. Next, note that $X^{(1)} \to U \to X^{(2)}$ implies $X'^{(1)} \to U \to X'^{(2)}$ since $H(X'^{(1)}|X^{(1)}) = 0$ and $H(X'^{(2)}|X^{(2)}) = 0$. Furthermore $U \to X'^{(1)}X'^{(2)} \to Y'^{(3)}Y'^{(4)}$ since

$$p(u,x^{(1)},x^{(2)},x'^{(1)},x'^{(2)},y'^{(3)},y'^{(4)}) =$$
$$p(x^{(1)},x^{(2)})p(u|x^{(1)},x^{(2)})p(x'^{(1)},x'^{(2)}|x^{(1)},x^{(2)})p(y'^{(3)},y'^{(4)}|x'^{(1)},x'^{(2)}).$$

Thus the point

$$\overrightarrow{v_2} = \big(I(X'^{(1)};Y'^{(3)}|X'^{(2)},U),\ I(X'^{(2)};Y'^{(4)}|X'^{(1)},U),\ I(X'^{(1)}X'^{(2)};Y'^{(3)}|U),$$
$$I(X'^{(1)}X'^{(2)};Y'^{(4)}|U),\ I(X'^{(1)}X'^{(2)};Y'^{(3)}),\ I(X'^{(1)}X'^{(2)};Y'^{(4)}),$$
$$I(X'^{(1)};Y'^{(3)}|X'^{(2)}),\ I(X'^{(2)};Y'^{(4)}|X'^{(1)})\big)$$

belongs to $\phi\big(p(y'^{(1)},y'^{(2)},...,y'^{(4)}|x'^{(1)},...,x'^{(4)}),\Psi\big)$ where Ψ is a given set that contains $q(x'^{(1)},...,x'^{(4)})$.

It suffices to show that $\overrightarrow{v} \leq \overrightarrow{v_1} + \overrightarrow{v_2}$. This is straightforward, once one observes that

$$p(y'^{(3)},y'^{(4)}|x^{(1)},x'^{(1)},x^{(2)},x'^{(2)},u) = p(y'^{(3)},y'^{(4)}|x'^{(1)},x^{(2)})$$

and

$$p(x^{(2)},x'^{(2)},x^{(1)},x'^{(1)},y'^{(3)},y'^{(4)}|u) = p(x^{(2)}|u)p(x'^{(2)}|x^{(2)})p(x^{(1)}|u)p(x'^{(1)}|x^{(1)})p(y'^{(3)},y'^{(4)}|x'^{(1)},x'^{(2)}).$$

The proof for the second property is straightforward since $Y^{(3)} = X^{(3)} = constant$, and $Y^{(4)} = X^{(4)} = constant$ implying that all the mutual information terms are zero. ■

B.

In this appendix we show that

$$n \times \Bigg(H(W^{(1)}|L) - O(h(\epsilon)), H(W^{(2)}|L) - O(h(\epsilon)), H(W^{(1)}|L) + H(W^{(2)}|L) - O(h(\epsilon)),$$
$$H(W^{(1)}|L) + H(W^{(2)}|L) - O(h(\epsilon)), H(W^{(1)}W^{(2)}) - O(h(\epsilon)),$$

$$H(W^{(1)}W^{(2)}) - O(h(\epsilon)), H(W^{(1)}|W^{(2)}) - O(h(\epsilon)), H(W^{(2)}|W^{(1)}) - O(h(\epsilon))\Big) \in$$

$$\phi\big(p(w_{1:n}^{(1)}y_{1:n}^{(1)}, w_{1:n}^{(2)}y_{1:n}^{(2)}, w_{1:n}^{(3)}y_{1:n}^{(3)}, w_{1:n}^{(4)}y_{1:n}^{(4)}|w_{1:n}^{(1)}, w_{1:n}^{(2)}, w_{1:n}^{(3)}, w_{1:n}^{(4)}), \{p(w_{1:n}^{(1)}, w_{1:n}^{(2)}, w_{1:n}^{(3)}, w_{1:n}^{(4)})\}\big). \quad (16)$$

Here $O(h(\epsilon))$ is equal to a constant (that depends only on the network architecture $q(y^{(1)}, ..., y^{(4)}|x^{(1)}, ..., x^{(4)})$) times $h(\epsilon)$. Here $h(\cdot)$ is the binary entropy function.

Here, because of the assumptions on the alphabets, we can think of the overall virtual channel as being $p(y_{1:n}^{(3)}, y_{1:n}^{(4)}|w_{1:n}^{(1)}, w_{1:n}^{(2)})$ and the set of admissible distributions as being $p(w_{1:n}^{(1)}, w_{1:n}^{(2)})$ where random variables $W_{1:n}^{(1)}, W_{1:n}^{(2)}, Y_{1:n}^{(3)}, Y_{1:n}^{(4)}$ are distributed according to

$$p(y_{1:n}^{(3)}, y_{1:n}^{(4)}|w_{1:n}^{(1)}, w_{1:n}^{(2)}) \cdot p(w_{1:n}^{(1)}, w_{1:n}^{(2)})$$

and $L_{1:n}$ is jointly distributed with $W_{1:n}^{(1)}, W_{1:n}^{(2)}, Y_{1:n}^{(3)}, Y_{1:n}^{(4)}$ according to

$$p(y_{1:n}^{(3)}, y_{1:n}^{(4)}, w_{1:n}^{(1)}, w_{1:n}^{(2)}) \cdot \prod_{i=1}^{n} p(l_i|w_i^{(1)}, w_i^{(2)}),$$

and such that we have $W_{1:n}^{(1)} \to L_{1:n} \to W_{1:n}^{(2)}$ and $L_{1:n} \to W_{1:n}^{(1)}W_{1:n}^{(2)} \to Y_{1:n}^{(3)}Y_{1:n}^{(4)}$. The set

$$\phi\big(p(w_{1:n}^{(1)}y_{1:n}^{(1)}, w_{1:n}^{(2)}y_{1:n}^{(2)}, w_{1:n}^{(3)}y_{1:n}^{(3)}, w_{1:n}^{(4)}y_{1:n}^{(4)}|w_{1:n}^{(1)}, w_{1:n}^{(2)}, w_{1:n}^{(3)}, w_{1:n}^{(4)}), \{p(w_{1:n}^{(1)}, w_{1:n}^{(2)}, w_{1:n}^{(3)}, w_{1:n}^{(4)})\}\big)$$

by definition contains the following point:

$$\overrightarrow{v} = \big(I(W_{1:n}^{(1)}; Y_{1:n}^{(3)}|W_{1:n}^{(2)}, L_{1:n}), I(W_{1:n}^{(2)}; Y_{1:n}^{(4)}|W_{1:n}^{(1)}, L_{1:n}), I(W_{1:n}^{(1)}W_{1:n}^{(2)}; Y_{1:n}^{(3)}|L_{1:n}),$$

$$I(W_{1:n}^{(1)}W_{1:n}^{(2)}; Y_{1:n}^{(4)}|L_{1:n}), I(W_{1:n}^{(1)}W_{1:n}^{(2)}; Y_{1:n}^{(3)}), I(W_{1:n}^{(1)}W_{1:n}^{(2)}; Y_{1:n}^{(4)}), I(W_{1:n}^{(1)}; Y_{1:n}^{(3)}|W_{1:n}^{(2)}), I(W_{1:n}^{(2)}; Y_{1:n}^{(4)}|W_{1:n}^{(1)})\big).$$

We show that $\overrightarrow{v}$ is pointwise greater than or equal to

$$n \times \Big(H(W^{(1)}|L) - O(h(\epsilon)), H(W^{(2)}|L) - O(h(\epsilon)), H(W^{(1)}|L) + H(W^{(2)}|L) - O(h(\epsilon)),$$

$$H(W^{(1)}|L) + H(W^{(2)}|L) - O(h(\epsilon)), H(W^{(1)}W^{(2)}) - O(h(\epsilon)),$$

$$H(W^{(1)}W^{(2)}) - O(h(\epsilon)), H(W^{(1)}|W^{(2)}) - O(h(\epsilon)), H(W^{(2)}|W^{(1)}) - O(h(\epsilon))\Big).$$

We only give the proof for the first, third, fifth and seventh element of $\overrightarrow{v}$. The proof for the other coordinates is similar.

- The first coordinate:

$I(W_{1:n}^{(1)}; Y_{1:n}^{(3)}|W_{1:n}^{(2)}L_{1:n}) = H(W_{1:n}^{(1)}|W_{1:n}^{(2)}L_{1:n}) - H(W_{1:n}^{(1)}|W_{1:n}^{(2)}L_{1:n}Y_{1:n}^{(3)}) = n \cdot [H(W^{(1)}|W^{(2)}L) - O(h(\epsilon))$ because of lemma 2 mentioned at the end of this appendix. Here $h(\cdot)$ denotes the binary entropy function. Note that $H(W^{(1)}|W^{(2)}L) = H(W^{(1)}|L)$.

- The third coordinate:

$I(W^{(1)}_{1:n}W^{(2)}_{1:n};Y^{(3)}_{1:n}|L_{1:n}) = H(W^{(1)}_{1:n}W^{(2)}_{1:n}|L_{1:n}) - H(W^{(1)}_{1:n}W^{(2)}_{1:n}|L_{1:n}Y^{(3)}_{1:n}) = n\cdot[H(W^{(1)}W^{(2)}|L) - O(h(\epsilon))]$ because of lemma 2 mentioned at the end of this appendix. Here $h(\cdot)$ denotes the binary entropy function. Furthermore, note that $H(W^{(1)}W^{(2)}|L) = H(W^{(1)}|L) + H(W^{(2)}|L)$.

• The fifth coordinate:

$I(W^{(1)}_{1:n}W^{(2)}_{1:n};Y^{(3)}_{1:n}) = H(W^{(1)}_{1:n}W^{(2)}_{1:n}) - H(W^{(1)}_{1:n}W^{(2)}_{1:n}|Y^{(3)}_{1:n}) = n\cdot[H(W^{(1)}W^{(2)}) - O(h(\epsilon))]$ because of lemma 2 mentioned at the end of this appendix.

• The seventh coordinate:

$I(W^{(1)}_{1:n};Y^{(3)}_{1:n}|W^{(2)}_{1:n}) = H(W^{(1)}_{1:n}|W^{(2)}_{1:n}) - H(W^{(1)}_{1:n}|W^{(2)}_{1:n}Y^{(3)}_{1:n}) = n\cdot[H(W^{(1)}|W^{(2)}) - O(h(\epsilon))$ because of lemma 2 mentioned at the end of this appendix. Here $h(\cdot)$ denotes the binary entropy function.

Lemma 2: Given any admissible source marginal distribution $p(w^{(1)},w^{(2)},w^{(3)},w^{(4)})$ for a strong interference channel $q(y^{(1)},y^{(2)},y^{(3)},y^{(4)}|x^{(1)},x^{(2)},x^{(3)},x^{(4)})$ satisfying $|\mathcal{W}^{(3)}| = |\mathcal{W}^{(4)}| = 1$, and an arbitrary positive ϵ and n-code, we have:

$$\tfrac{1}{n}H(W^{(1)}_{1:n}W^{(2)}_{1:n}|Y^{(3)}_{1:n}) = O(h(\epsilon)), \tag{17}$$

$$\tfrac{1}{n}H(W^{(1)}_{1:n}W^{(2)}_{1:n}|Y^{(4)}_{1:n}) = O(h(\epsilon)). \tag{18}$$

Proof: We prove the equation (17); the proof of equation (18) is similar. Since $q(.|.)$ is a strong interference channel, lemma of section III of [21] implies that the n-letter product channel

$$q(y^{(3)}_{1:n},y^{(4)}_{1:n}|x^{(1)}_{1:n},x^{(2)}_{1:n}) = \textstyle\prod_{i=1}^{n} q(y^{(3)}_i,y^{(4)}_i|x^{(1)}_i,x^{(2)}_i)$$

is also a strong interference channel. Therefore [11]

$$I(X^{(2)}_{1:n};Y^{(3)}_{1:n}|X^{(1)}_{1:n}) \geq I(X^{(2)}_{1:n};Y^{(4)}_{1:n}|X^{(1)}_{1:n}).$$

This implies that $H(X^{(2)}_{1:n}|Y^{(3)}_{1:n}X^{(1)}_{1:n}) \leq H(X^{(2)}_{1:n}|Y^{(4)}_{1:n}X^{(1)}_{1:n})$. We have $H(X^{(2)}_{1:n}|Y^{(4)}_{1:n}X^{(1)}_{1:n}) \leq H(W^{(2)}_{1:n}|Y^{(4)}_{1:n}X^{(1)}_{1:n}) \leq H(W^{(2)}_{1:n}|\widehat{M}^{(4)}_{1:n})$ since $H(X^{(2)}_{1:n}|W^{(2)}_{1:n}) = 0$. Here $\widehat{M}^{(4)}_{1:n}$ is the reconstruction of $W^{(2)}_{1:n}$ by the third party (the average distortion between $\widehat{M}^{(4)}_{1:n}$ and $\widehat{W}^{(2)}_{1:n}$ is less than or equal to ϵ). $H(W^{(2)}_{1:n}|\widehat{M}^{(4)}_{1:n})$ is equal to $n\cdot O(h(\epsilon))$ for $\epsilon < \frac{1}{2}$ since

$$H(W^{(2)}_{1:n}|\widehat{M}^{(4)}_{1:n}) \leq \textstyle\sum_{i=1}^{n} H(W^{(2)}_i|\widehat{M}^{(4)}_i) \leq \tag{19}$$

$$\textstyle\sum_{i=1}^{n}[h(p(W^{(2)}_i \neq \widehat{M}^{(4)}_i)) + p(W^{(2)}_i \neq \widehat{M}^{(4)}_i)\log(|\mathcal{W}^{(2)}|)] \leq \tag{20}$$

$$n\big[h(\tfrac{1}{n}\textstyle\sum_{i=1}^{n} p(W^{(2)}_i \neq \widehat{M}^{(4)}_i)) + \tfrac{1}{n}\sum_{i=1}^{n} p(W^{(2)}_i \neq \widehat{M}^{(4)}_i)\log(|\mathcal{W}^{(2)}|)\big] = n\cdot O(h(\epsilon)), \tag{21}$$

[11]Note that equations (1) and (2) hold for all arbitrary distributions on the inputs for a strong interference channel.

where in equation (19) we have used the Fano inequality, in equation (20) we have used the concavity of the binary entropy function, and lastly, in equation (21), we have used the fact that the average distortion between $\widehat{M}^{(4)}_{1:n}$ and $W^{(2)}_{1:n}$ is less than or equal to ϵ. Thus, $H(W^{(2)}_{1:n}|\widehat{M}^{(4)}_{1:n}) \leq n \cdot O(h(\epsilon))$.

Therefore $H(X^{(2)}_{1:n}|Y^{(3)}_{1:n}X^{(1)}_{1:n}) \leq n \cdot O(h(\epsilon))$. Since the average distortion between $\widehat{M}^{(3)}_{1:n}$ and $W^{(1)}_{1:n}$ is less than or equal to ϵ, we have: $H(X^{(1)}_{1:n}|Y^{(3)}_{1:n}) \leq H(W^{(1)}_{1:n}|Y^{(3)}_{1:n}) \leq H(W^{(1)}_{1:n}|\widehat{M}^{(3)}_{1:n}) = n \cdot O(h(\epsilon))$. Thus, $H(X^{(1)}_{1:n}X^{(2)}_{1:n}|Y^{(3)}_{1:n}) \leq n \cdot O(h(\epsilon))$. Next, note that

$$H(W^{(1)}_{1:n}W^{(2)}_{1:n}|Y^{(3)}_{1:n}) \leq H(W^{(1)}_{1:n}W^{(2)}_{1:n}|X^{(1)}_{1:n}X^{(2)}_{1:n}) + H(X^{(1)}_{1:n}X^{(2)}_{1:n}|Y^{(3)}_{1:n})$$

since for any three random variables A, B and C we have $H(A|C) \leq H(A|B) + H(B|C)$. It suffices to show that $H(W^{(1)}_{1:n}W^{(2)}_{1:n}|X^{(1)}_{1:n}X^{(2)}_{1:n}) = n \cdot O(h(\epsilon))$. This is true because $H(W^{(1)}_{1:n}W^{(2)}_{1:n}|X^{(1)}_{1:n}X^{(2)}_{1:n}) = H(W^{(1)}_{1:n}W^{(2)}_{1:n}|X^{(1)}_{1:n}X^{(2)}_{1:n}Y^{(3)}_{1:n}Y^{(4)}_{1:n}) \leq H(W^{(1)}_{1:n}W^{(2)}_{1:n}|Y^{(3)}_{1:n}Y^{(4)}_{1:n}) \leq H(W^{(1)}_{1:n}|Y^{(3)}_{1:n}) + H(W^{(2)}_{1:n}|Y^{(4)}_{1:n})$. ■

C. Convexity of ϕ

In this appendix, we show that for any interference channel $q(y^{(1)}, y^{(2)}, y^{(3)}, y^{(4)}|x^{(1)}, x^{(2)}, x^{(3)}, x^{(4)})$, $\phi\big(q(y^{(1)}, y^{(2)}, y^{(3)}, y^{(4)}|x^{(1)}, x^{(2)}, x^{(3)}, x^{(4)}), \Psi\big)$ is convex when the set Ψ is convex. Take two arbitrary points $\overrightarrow{v_1}$ and $\overrightarrow{v_2}$ from this set. Corresponding to these two points are random variables $U_1, X^{(i)}_1, Y^{(i)}_1$ for $i = 1, ..., 4$ and $U_2, X^{(i)}_2, Y^{(i)}_2$ for $i = 1, ..., 4$, where $X^{(1)}_i - U_i - X^{(2)}_i$ and $U_i - X^{(1)}_i X^{(2)}_i - Y^{(3)}_i Y^{(4)}_i$ for $i = 1, 2$ and $\overrightarrow{v_i}$ is coordinate by coordinate less than or equal to

$$\big(I(X^{(1)}_i; Y^{(3)}_i|X^{(2)}_i, U_i),\ I(X^{(2)}_i; Y^{(4)}_i|X^{(1)}_i, U_i),\ I(X^{(1)}_i X^{(2)}_i; Y^{(3)}_i|U_i),$$
$$I(X^{(1)}_i X^{(2)}_i; Y^{(4)}_i|U_i),\ I(X^{(1)}_i X^{(2)}_i; Y^{(3)}_i),\ I(X^{(1)}_i X^{(2)}_i; Y^{(4)}_i),\ I(X^{(1)}_i; Y^{(3)}_i|X^{(2)}_i),\ I(X^{(2)}_i; Y^{(4)}_i|X^{(1)}_i)\big)$$

for $i = 1, 2$. Take a binary and uniform random variable Q on $\{1, 2\}$ and let $U = (U_Q, Q)$, $X^{(i)} = X^{(i)}_Q$ and $Y^{(i)} = Y^{(i)}_Q$ for $i = 1, ..., 4$. It can be then verified that $p(x^{(1)}, x^{(2)}, x^{(3)}, x^{(4)})$ is in Ψ, that the conditional law of $Y^{(1)}, Y^{(2)}, Y^{(3)}, Y^{(4)}$ given $X^{(1)}, X^{(2)}, X^{(3)}, X^{(4)}$ is described by $q(y^{(1)}, y^{(2)}, y^{(3)}, y^{(4)}|x^{(1)}, x^{(2)}, x^{(3)}, x^{(4)})$, and that the point

$$\overrightarrow{v} = \big(I(X^{(1)}; Y^{(3)}|X^{(2)}, U),\ I(X^{(2)}; Y^{(4)}|X^{(1)}, U),\ I(X^{(1)}X^{(2)}; Y^{(3)}|U),$$
$$I(X^{(1)}X^{(2)}; Y^{(4)}|U),\ I(X^{(1)}X^{(2)}; Y^{(3)}),\ I(X^{(1)}X^{(2)}; Y^{(4)}),\ I(X^{(1)}; Y^{(3)}|X^{(2)}),\ I(X^{(2)}; Y^{(4)}|X^{(1)})\big)$$

is coordinate by coordinate greater than or equal to $\frac{1}{2}(\overrightarrow{v_1} + \overrightarrow{v_2})$. The proof for the first four coordinates is straightforward. For the fifth coordinate, we use the fact that $I(X^{(1)}X^{(2)}; Y^{(3)}) = I(UX^{(1)}X^{(2)}; Y^{(3)}) \geq I(X^{(1)}X^{(2)}; Y^{(3)}|U)$. The proof for the sixth, seventh and eighth coordinates is the similar. Since $\overrightarrow{v}$ can be seen to belong to

$$\phi\big(q(y^{(1)}, y^{(2)}, y^{(3)}, y^{(4)}|x^{(1)}, x^{(2)}, x^{(3)}, x^{(4)}), \Psi\big),$$

this region must be convex. ■

APPENDIX C
COMPLETING THE PROOF OF THEOREM 3 ON THE BROADCAST CHANNEL

A. Verifying the properties of Lemma 1

In this appendix we show that our choice of ϕ verifies the two properties of the main lemma for $m = 3$ when restricted to the class of broadcast channels.

For the first property of the lemma, take a channel

$$p(y^{(1)}y'^{(1)}, y^{(2)}y'^{(2)}, y^{(3)}y'^{(3)}|x^{(1)}, x^{(2)}, x^{(3)}) =$$
$$p(y^{(1)}, y^{(2)}, y^{(3)}|x^{(1)}, x^{(2)}, x^{(3)}) \cdot p(y'^{(1)}, y'^{(2)}, y'^{(3)}|x'^{(1)}, x'^{(2)}, x'^{(3)})$$

where $H(X'^{(i)}|X^{(i)}Y^{(i)}) = 0$ for $1 \leq i \leq 3$. Take an arbitrary point from

$$\phi\big(p(y^{(1)}y'^{(1)}, y^{(2)}y'^{(2)}, y^{(3)}y'^{(3)}|x^{(1)}, x^{(2)}, x^{(3)}), \{q(x^{(1)}, x^{(2)}, x^{(3)})\}\big).$$

The set of input distributions contains only one distribution $q(x^{(1)}, x^{(2)}, x^{(3)})$. Therefore we define random variables $X^{(1)}, X^{(2)}, X^{(3)}, Y^{(1)}, Y'^{(1)}, Y^{(2)}, Y'^{(2)}, Y^{(3)}, Y'^{(3)}$ jointly distributed according to

$$p(y^{(1)}y'^{(1)}, y^{(2)}y'^{(2)}, y^{(3)}y'^{(3)}|x^{(1)}, x^{(2)}, x^{(3)}) \cdot q(x^{(1)}, x^{(2)}, x^{(3)}) =$$
$$p(y^{(1)}, y^{(2)}, y^{(3)}|x^{(1)}, x^{(2)}, x^{(3)}) \cdot p(y'^{(1)}, y'^{(2)}, y'^{(3)}|x'^{(1)}, x'^{(2)}, x'^{(3)}) \cdot q(x^{(1)}, x^{(2)}, x^{(3)}).$$

Corresponding to the arbitrary point in

$$\phi\big(p(y^{(1)}y'^{(1)}, y^{(2)}y'^{(2)}, y^{(3)}y'^{(3)}|x^{(1)}, x^{(2)}, x^{(3)}), \{q(x^{(1)}, x^{(2)}, x^{(3)})\}\big)$$

are random variables U, V, W_0, W_1, W_2 satisfying $UVW_0W_1W_2 \rightarrow X^{(1)} \rightarrow Y^{(2)}Y^{(3)}$ and $H(W_0|W_1U) = H(W_0|W_2V) = H(X^{(1)}|W_0W_1W_2UV) = 0$ such that the point is coordinate by coordinate less than or equal to the point

$$\overrightarrow{v} = \big(I(W_0; Y^{(2)}Y'^{(2)}|U), I(W_0; Y^{(3)}Y'^{(3)}|V), I(W_1; Y^{(2)}Y'^{(2)}|U), I(W_2; Y^{(3)}Y'^{(3)}|V),$$
$$I(W_1; Y^{(2)}Y'^{(2)}|W_0UV) + I(W_0U; Y^{(3)}Y'^{(3)}|V), I(W_2; Y^{(3)}Y'^{(3)}|W_0UV) + I(W_0V; Y^{(2)}Y'^{(2)}|U),$$
$$I(W_1; Y^{(2)}Y'^{(2)}|W_0W_2UV)+I(W_2U; Y^{(3)}Y'^{(3)}|V), I(W_2; Y^{(3)}Y'^{(3)}|W_0W_1UV)+I(W_1V; Y^{(2)}Y'^{(2)}|U)$$
$$I(W_0V; Y^{(2)}Y'^{(2)}|U) + I(W_1; Y^{(2)}Y'^{(2)}|W_0W_2UV) + I(W_2; Y^{(3)}Y'^{(3)}|W_0UV),$$
$$I(W_0U; Y^{(3)}Y'^{(3)}|V) + I(W_2; Y^{(3)}Y'^{(3)}|W_0W_1UV) + I(W_1; Y^{(2)}Y'^{(2)}|W_0UV)\big).$$

Since $p(y^{(1)}, y^{(2)}, y^{(3)}|x^{(1)}, x^{(2)}, x^{(3)})$ and $p(y'^{(1)}, y'^{(2)}, y'^{(3)}|x'^{(1)}, x'^{(2)}, x'^{(3)})$ belong to the class of broadcast channels, we have $H(Y^{(1)}|X^{(1)}) = H(X'^{(1)}|X^{(1)}) = 0$. Since $|\mathcal{X}^{(2)}| = |\mathcal{X}^{(3)}| = |\mathcal{X}'^{(2)}| = |\mathcal{X}'^{(3)}| = 1$, we have

$$p(y^{(2)}y'^{(2)}, y^{(3)}y'^{(3)}|x^{(1)}, x^{(2)}, x^{(3)}) =$$
$$p(y^{(2)}, y^{(3)}|x^{(1)}, x^{(2)}, x^{(3)}) \cdot p(y'^{(2)}, y'^{(3)}|x^{(1)}, x^{(2)}, x^{(3)}, y^{(2)}, y^{(3)}) =$$
$$p(y^{(2)}, y^{(3)}|x^{(1)}) \cdot p(y'^{(2)}, y'^{(3)}|x'^{(1)}).$$

Let $W'_0 = \widetilde{W_0} = W_0$, $W'_1 = \widetilde{W_1} = W_1$, $W'_2 = \widetilde{W_2} = W_2$, $V' = VY^{(3)}$, $U' = U$, $\widetilde{V} = V$, $\widetilde{U} = UY'^{(2)}$. Observe that the following properties hold:

- The Markov chain $U'V'W'_0W'_1W'_2X'^{(1)} \to X'^{(1)} \to X'^{(1)}Y'^{(2)}Y'^{(3)}$ holds. Also, $H(W'_0|W'_1U') = H(W'_0|W'_2V') = 0$ and $X'^{(1)}$ is a deterministic function of $(W'_0, W'_1, W'_2, U', V')$.
- The Markov chain $\widetilde{U}\widetilde{V}\widetilde{W_0}\widetilde{W_1}\widetilde{W_2}X^{(1)} \to X^{(1)} \to X^{(1)}Y^{(2)}Y^{(3)}$ holds. Also, $H(\widetilde{W_0}|\widetilde{W_1}\widetilde{U}) = H(\widetilde{W_0}|\widetilde{W_2}\widetilde{V}) = 0$ and $X^{(1)}$ is a deterministic function of $(\widetilde{W_0}, \widetilde{W_1}, \widetilde{W_2}, \widetilde{U}, \widetilde{V})$.

Using the above auxiliary random variables, we can define two points

$$\overrightarrow{v_1} \in \phi\big(p(y^{(1)}, y^{(2)}, y^{(3)}|x^{(1)}, x^{(2)}, x^{(3)}), \{q(x^{(1)}, x^{(2)}, x^{(3)})\}\big),$$

and

$$\overrightarrow{v_2} \in \phi\big(p(y'^{(1)}, y'^{(2)}, y'^{(3)}|x'^{(1)}, x'^{(2)}, x'^{(3)}), \Psi\big),$$

where Ψ is any given set that contains $q(x'^{(1)}, ..., x'^{(4)})$. It can be verified that $\overrightarrow{v_1} + \overrightarrow{v_2}$ is coordinate by coordinate greater than or equal to $\overrightarrow{v}$.

The proof for the second property is straightforward since $Y^{(2)} = X^{(2)} = constant$, and $Y^{(3)} = X^{(3)} = constant$, implying that all the mutual information terms are zero. ■

B.

In this appendix, we show that for any n-code and any positive ϵ,

$$n \times \Bigg(H(L) - O(h(\epsilon)), H(L) - O(h(\epsilon)), H(M^{(2)}) - O(h(\epsilon)), H(M^{(3)}) - O(h(\epsilon)),$$
$$H(M^{(2)}) - O(h(\epsilon)), H(M^{(3)}) - O(h(\epsilon)), H(M^{(2)}M^{(3)}) - O(h(\epsilon)),$$
$$H(M^{(2)}M^{(3)}) - O(h(\epsilon)), H(M^{(2)}M^{(3)}) - O(h(\epsilon)), H(M^{(2)}M^{(3)}) - O(h(\epsilon))\Bigg) \in$$
$$\phi\big(p(w_{1:n}^{(1)}y_{1:n}^{(1)}, w_{1:n}^{(2)}y_{1:n}^{(2)}, w_{1:n}^{(3)}y_{1:n}^{(3)}|w_{1:n}^{(1)}, w_{1:n}^{(2)}, w_{1:n}^{(3)}), \{p(w_{1:n}^{(1)}, w_{1:n}^{(2)}, w_{1:n}^{(3)})\}\big),$$

where by $O(h(\epsilon))$ we mean a constant (that depends only on the network $q(y^{(1)}, y^{(2)}, y^{(3)}|x^{(1)}, x^{(2)}, x^{(3)})$) times $h(\epsilon)$, where $h(\cdot)$ is the binary entropy function.

Here, because of the assumptions on the alphabets, we can think of the overall virtual channel as being $p(y_{1:n}^{(2)}, y_{1:n}^{(3)}|w_{1:n}^{(1)})$ and the set of admissible distributions as being $\{p(w_{1:n}^{(1)})\}$, where $W_{1:n}^{(1)}, Y_{1:n}^{(2)}, Y_{1:n}^{(3)}$ are distributed according to

$$p(y_{1:n}^{(2)}, y_{1:n}^{(3)}|w_{1:n}^{(1)}) \cdot p(w_{1:n}^{(1)}).$$

Let $W_0 = L_{1:n}$, $W_1 = M_{1:n}^{(2)}$ and $W_2 = M_{1:n}^{(3)}$, U a constant, and V be a random variable independent of $M_{1:n}^{(2)}M_{1:n}^{(3)}$ such that $H(W_{1:n}^{(1)}|VM_{1:n}^{(2)}M_{1:n}^{(3)}) = 0$, and $VM_{1:n}^{(2)}M_{1:n}^{(3)} \to W_{1:n}^{(1)} \to Y_{1:n}^{(2)}Y_{1:n}^{(3)}$. These auxiliary random variables imply that the set

$$\phi\big(p(w_{1:n}^{(1)}y_{1:n}^{(1)}, w_{1:n}^{(2)}y_{1:n}^{(2)}, w_{1:n}^{(3)}y_{1:n}^{(3)}|w_{1:n}^{(1)}, w_{1:n}^{(2)}, w_{1:n}^{(3)}), \{p(w_{1:n}^{(1)}, w_{1:n}^{(2)}, w_{1:n}^{(3)})\}\big)$$

contains the following point

$$\overrightarrow{v} = \big(I(L_{1:n}; Y_{1:n}^{(2)}),\ I(L_{1:n}; Y_{1:n}^{(3)}|V),\ I(M_{1:n}^{(2)}; Y_{1:n}^{(2)}),\ I(M_{1:n}^{(3)}; Y_{1:n}^{(3)}|V),$$
$$I(M_{1:n}^{(2)}; Y_{1:n}^{(2)}|L_{1:n}V) + I(L_{1:n}; Y_{1:n}^{(3)}|V),\ I(M_{1:n}^{(3)}; Y_{1:n}^{(3)}|L_{1:n}V) + I(L_{1:n}V; Y_{1:n}^{(2)}),$$
$$I(M_{1:n}^{(2)}; Y_{1:n}^{(2)}|L_{1:n}M_{1:n}^{(3)}V) + I(M_{1:n}^{(3)}; Y_{1:n}^{(3)}|V),\ I(M_{1:n}^{(3)}; Y_{1:n}^{(3)}|L_{1:n}M_{1:n}^{(2)}V) + I(M_{1:n}^{(2)}V; Y_{1:n}^{(2)})$$
$$I(L_{1:n}V; Y_{1:n}^{(2)}) + I(M_{1:n}^{(2)}; Y_{1:n}^{(2)}|L_{1:n}M_{1:n}^{(3)}V) + I(M_{1:n}^{(3)}; Y_{1:n}^{(3)}|L_{1:n}V),$$
$$I(L_{1:n}; Y_{1:n}^{(3)}|V) + I(M_{1:n}^{(3)}; Y_{1:n}^{(3)}|L_{1:n}M_{1:n}^{(2)}V) + I(M_{1:n}^{(2)}; Y_{1:n}^{(2)}|L_{1:n}V)\big).$$

One can get the desired result using the Fano inequality. ■

C. *Convexity of* ϕ

In this appendix, we show that for any broadcast channel $q(y^{(1)}, y^{(2)}, y^{(3)}|x^{(1)}, x^{(2)}, x^{(3)})$, $\phi\big(q(y^{(1)}, y^{(2)}, y^{(3)}|x^{(1)}, x^{(2)}, x^{(3)}), \Psi\big)$ is convex when the set Ψ is convex.

Take two points $\overrightarrow{a}$ and $\overrightarrow{b}$ in $\phi\big(q(y^{(1)}, y^{(2)}, y^{(3)}|x^{(1)}, x^{(2)}, x^{(3)}), \Psi\big)$. Corresponding to $\overrightarrow{a}$, one can find random variables $U_a, V_a, W_{0a}, W_{1a}, W_{2a}, X_a^{(1)}, Y_a^{(2)}, Y_a^{(3)}$ such that the constraints given in the definition of ϕ are satisfied. Similarly, one can find $U_b, V_b, W_{0b}, W_{1b}, W_{2b}, X_b^{(1)}, Y_b^{(2)}, Y_b^{(3)}$ corresponding to the point $\overrightarrow{b}$. Without loss of generality, one can assume that $(U_a, V_a, W_{0a}, W_{1a}, W_{2a}, X_a^{(1)}, Y_a^{(2)}, Y_a^{(3)})$ is independent of $(U_b, V_b, W_{0b}, W_{1b}, W_{2b}, X_b^{(1)}, Y_b^{(2)}, Y_b^{(3)})$.

Take a binary random variable T on $\{0, 1\}$, satisfying $p(T = 0) = t$, that is independent of all the above mentioned random variables. Let $(U, V, W_0, W_1, W_2, X^{(1)}, Y^{(2)}, Y^{(3)})$ be equal to

$$(TU_a, TV_a, W_{0a}, W_{1a}, W_{2a}, X_a^{(1)}, Y_a^{(2)}, Y_a^{(3)})$$

if $T=0$, and be equal to

$$(TU_b, TV_b, W_{0b}, W_{1b}, W_{2b}, X_b^{(1)}, Y_b^{(2)}, Y_b^{(3)})$$

if $T=1$. The convexity of Ψ implies that $p(x^{(1)})$ is in this set. Furthermore it can be verified that the conditional law $p(y^{(2)}, y^{(3)}|x^{(1)})$ is the one we started with, and random variables $(U, V, W_0, W_1, W_2, X^{(1)}, Y^{(2)}, Y^{(3)})$ satisfy all the constraints given in the definition of ϕ. Furthermore this choice of variables gives us a point in $\phi\big(q(y^{(1)}, y^{(2)}, y^{(3)}|x^{(1)}, x^{(2)}, x^{(3)}), \Psi\big)$ that coordinatewise dominates $t\overrightarrow{a} + (1-t)\overrightarrow{b}$. ■