A Generalized Cut-Set Bound

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 generalize the well known cut-set bound to the problem of lossy transmission of functions of arbitrarily correlated sources over a discrete memoryless multiterminal network.

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. 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 [4]. In this paper we study the limitations of joint source-channel coding strategies for lossy transmission across multiterminal networks.

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 ith party. In a general multiterminal channel with correlated sources, the m nodes are observing i.i.d. repetitions of m, possibly correlated, random variables $W^{(i)}$ for $1 \leq i \leq m$. The ith 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. $f^{(i)}(W^{(1)}, W^{(2)}, ..., W^{(m)})$ for some function $f^{(i)}(\cdot)$. 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. In some applications, each party may be interested in recovering i.i.d. repetitions of functions of the observations made at different nodes. In this case the function $f^{(i)}(W^{(1)}, W^{(2)}, ..., W^{(m)})$ takes the special form of $\left(f^{(i,1)}(W^{(1)}), f^{(i,2)}(W^{(2)}), ..., f^{(i,m)}(W^{(m)}) \right)$ for some functions $f^{(i,j)}(\cdot)$.
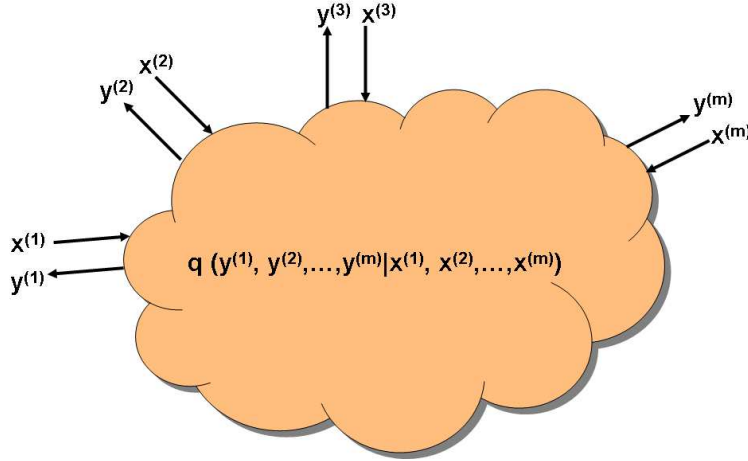
Fig. 1. The statistical description of a network.

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 the source−channel separation theorem in a network scenario breaks down [4]. In this paper, we prove a new outer bound on the admissible source region of GMNs. 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 of $(f^{(i,1)}(W^{(1)}), f^{(i,2)}(W^{(2)}), ..., f^{(i,m)}(W^{(m)}))$, 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 ith 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 ith 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 cut-set bound can be found in [2] and [5]. Furthermore some existing works show the possibility and benefit of function computation during the communication (see for instance [3][6][7][8][9]).

A main contribution of this paper is its proof technique which is based on the "potential function method" introduced in [10] and [11]. Instead of taking an arbitrary network and proving the desired outer bound while keeping the network fixed throughout, we consider a function from the set of all m-input/m-output discrete memoryless networks to subsets of $\mathbb{R}_+^c$, where $\mathbb{R}_+^c$ is the set of all c-tuples of non-negative reals. 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 generalized cut-set 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 multiterminal network. During the simulation of the code, the information of the parties begins from the ith 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 nth stage, the system can not 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. See Lemma 1 of section III and the proof of Theorem 1 of section IV for a formal formulation.

The outline of this paper is as follows. In section II, we introduce the basic notations and definitions used in this paper. Section III contains the main results of this paper followed by section IV which gives formal proofs for the results. Appendices I and II complete the proof of Theorem 1 from section III.

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 ith party is $X^{(i)}$ and the output at the ith party is $Y^{(i)}$. We assume that the ith party $(1 \le i \le m)$ has access to i.i.d. repetitions of $W^{(i)}$. The message that needs to be delivered (in a possibly lossy manner) to the ith party is taken to be $M^{(i)} = f^{(i)}(W^{(1)}, W^{(2)}, ..., W^{(m)})$ for some function $f^{(i)}(\cdot)$. We assume that for any $i \in [m]$, random variables $X^{(i)}, Y^{(i)}, W^{(i)}$ and $M^{(i)}$ take values from discrete 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 $\mathcal{M}^{(i)}$. We use $Y_{1:k}^{(i)}$ to denote $(Y_1^{(i)}, Y_2^{(i)}, ..., Y_k^{(i)})$.

TABLE I

NOTATIONS

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 multi-terminal network.
$W^{(i)}$	Random variable representing the source observed at the ith node.
$M^{(i)}$	Random variable to be reconstructed, in a possibly lossy way, at the ith node.
$\mathcal{X}^{(i)}, \mathcal{Y}^{(i)}, \mathcal{W}^{(i)}, \mathcal{M}^{(i)}$	Alphabet sets of $X^{(i)}, Y^{(i)}, W^{(i)}, M^{(i)}$.
$\Delta^{(i)}(\cdot, \cdot)$	Distortion function used by the ith party.
$\zeta_k^{(i)}(\cdot)$	The encoding function used by the ith party at the kth stage.
$\vartheta^{(i)}(\cdot)$	The decoding function at the ith party.
n	Length of the code used.
$\Pi(\cdot)$	Down-set (Definition 4);
$\oplus$	Minkowski sum of two sets (Definition 3).
$\geq$	A vector or a set being greater than or equal the other (Definition 4).
Ψ	A permissible set of input distributions; Given input sources and a multiterminal network, 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 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 ith party, within the average distortion of $D^{(i)}$.

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 ith party at the kth time instance, and $\vartheta^{(i)}$ is the decoding function of the ith 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 satisfy the average distortion interval $D^{(i)}$ (for all $i \in [m]$) over the channel $q(y^{(1)}, y^{(2)}, ..., y^{(m)}|x^{(1)}, x^{(2)}, ..., x^{(m)})$ if the following "average distortion" condition is satisfied:

Assume that random variables $W_{1:n}^{(i)}$ for $i \in [m]$ are n i.i.d. repetition 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:

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

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

and that $X_1^{(i)} = \zeta_1^{(i)}(W_{1:n}^{(i)})$, and for any $2 \leq k \leq n$, $X_k^{(i)} = \zeta_k^{(i)}(W_{1:n}^{(i)}, Y_{1:k-1}^{(i)})$. Random variables $X_k^{(i)}$ and $Y_k^{(i)}$ are representing the input and outputs of the ith party at the kth time instance and satisfy the following Markov chains:

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

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

We then have the following constraint for any $i \in [m]$:

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

where $M_k^{(i)} = f^{(i)}(W_k^{(1)}, W_k^{(2)}, ..., W_k^{(m)})$.

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 $q(y^{(1)}, y^{(2)}, ..., y^{(m)}|x^{(1)}, x^{(2)}, ..., x^{(m)})$ if for every positive ϵ and sufficiently large n, an (n)-code satisfying the average distortion $D^{(i)}$, exists.

The "independent messages zero distortion capacity region" of the GMN,

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

is a subset of m^2-tuples of non-negative numbers $R^{(i,j)}$ for $i, j \in [m]$ defined as follows: consider the set of all sets $\mathcal{W}^{(1)}, \mathcal{W}^{(2)}, ..., \mathcal{W}^{(m)}$, functions $f^{(i)}(W^{(1)}, W^{(2)}, ..., W^{(m)})$ $(1 \leq i \leq m)$ having the special form of

$$(f^{(i,1)}(W^{(1)}), f^{(i,2)}(W^{(2)}), ..., f^{(i,m)}(W^{(m)})),$$

the distortion functions $\Delta^{(i)}(m^{(i)}, m^{'(i)})$ (for $1 \leq i \leq m$) being equal to the indicator function $\mathbf{1}[m^{(i)} \neq m^{'(i)}]$, $D^{(i)}$ being set to be zero for all $1 \leq i \leq m$ and admissible sources $p(w^{(1)}, w^{(2)}, ..., w^{(m)})$ for which $f^{(i,j)}(W^{(j)})$'s are mutually independent of each other. The capacity region is then taken to be the set of all achievable $R^{(i,j)} = H(f^{(j,i)}(W^{(i)}))$ (for $i, j \in [m]$) given the above constraints. Intuitively speaking, $R^{(i,j)}$ is the communication rate from i^{th} party to the jth party.

Definition 3: For any natural number c and any two sets of points K and L in $\mathbb{R}_+^c$, 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 any two points $\vec{v_1}$ and $\vec{v_2}$ in $\mathbb{R}_+^c$, we say $\vec{v_1} \geq \vec{v_2}$ if and only if each coordinate of $\vec{v_1}$ is greater than or equal to the corresponding coordinate of $\vec{v_2}$. For any two sets of points A and B in $\mathbb{R}_+^c$, we say $A \leq B$ if and only if for any point $\vec{a} \in A$, there exists a point $\vec{b} \in B$ such that $\vec{a} \leq \vec{b}$. For a set $A \in \mathbb{R}_+^c$, the down-set $\Pi(A)$ is defined as: $\Pi(A) = \{\vec{v} \in \mathbb{R}_+^c : \vec{v} \leq \vec{w} \text{ for some } \vec{w} \in A\}$.

Definition 5: Given a specific network architecture $q(y^{(1)}, y^{(2)}, ..., y^{(m)}|x^{(1)}, x^{(2)}, ..., x^{(m)})$, and the source marginal distribution $p(w^{(1)}, w^{(2)}, ..., w^{(m)})$, it may be possible to find properties that the inputs to the multiterminal network throughout the communication satisfy. 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. Other constraints on the inputs to the network might come from practical requirements such as a maximum instantaneous power used up by one or a group of nodes 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 value 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 $\mathcal{X}^{(1)} \times \mathcal{X}^{(2)} \times \mathcal{X}^{(3)} \times ... \times \mathcal{X}^{(m)}$.

III. STATEMENT OF THE RESULTS

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

- joint distribution $q(x^{(1)}, x^{(2)}, ..., x^{(m)}, z)$ where size of the alphabet set 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;

- 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 any arbitrary $T \subset [m]$ 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:

$$Z - X^{(1)}, X^{(2)}, ..., X^{(m)} - Y^{(1)}, Y^{(2)}, ..., Y^{(m)}.$$

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 kth 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 is said to be weaker than another channel if the region associated to the first channel is contained in the region associated to the second channel.

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 to be 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 $p(\widehat{m}^{(1)}, \widehat{m}^{(2)}, ..., \widehat{m}^{(m)} | w^{(1)}, w^{(2)}, ..., w^{(m)})$ and the input distribution $p(w^{(1)}, w^{(2)}, ..., w^{(m)})$. 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 this region associated to this virtual channel and the corresponding input distribution should be included inside the region associated to the channel $q(y^{(1)}, y^{(2)}, ..., y^{(m)} | x^{(1)}, x^{(2)}, ..., x^{(m)})$. 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. This virtual channel at the given input distribution must be, in the above mentioned sense, weaker than any physical channel fit for the communication problem.

Corollary 1: Given any GMN $q(y^{(1)}, y^{(2)}, ..., y^{(m)} | x^{(1)}, x^{(2)}, ..., x^{(m)})$, the following region forms an outer bound on the independent messages zero distortion capacity region (see Definition 2) of the network:

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

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

$$\left. \text{is satisfied.} \right\},$$

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 1: 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 2: This bound reduces to the traditional cut-set bound when Ψ is taken to be the set of all input distributions, and $I\big(X^{(i)} : i \in T \; ; \; Y^{(i)} : i \in T^c | X^{(i)} : i \in T^c, Z\big)$ is bounded from above by [1]

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

A. The Main Lemma

During the simulation of the code, the information of the parties begins from the ith party having $W_{1:n}^{(i)}$ and gradually evolves over time with the usage of the network. At the jth stage, the ith party has $W_{1:n}^{(i)} Y_{1:j}^{(i)}$. We represent the information state of the whole system at the jth 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 input distribution $p(w_{1:n}^{(1)}, ..., w_{1:n}^{(m)})$. In order to quantify the information state, we map the information state to a subset of $\mathbb{R}_+^c$ (c is a natural number) using a function $\phi(.)$. A formal definition of ϕ and the properties we require it to satisfy are as follows:

Let $\phi(p(y^{(1)}, ..., y^{(m)} | x^{(1)}, ..., x^{(m)}), \Psi)$ be a function that takes as input an arbitrary m-input/m-output GMN and a subset of probability distributions on the inputs of this network and returns a subset of $\mathbb{R}_+^c$ where c is a natural number. $\phi(.)$ is thus a function from the set of all conditional probability distributions defined on finite sets and a corresponding set of input distributions, to subsets of $\mathbb{R}_+^c$.

Assume that the function $\phi(.)$ satisfies the following three properties. The intuitive description of the properties is provided after their formal statement. Please see Definitions 3 and 4 for the notations used.

1) Assume that the conditional distribution $p(y^{(1)} y'^{(1)}, y^{(2)} y'^{(2)}, ..., y^{(m)} y'^{(m)} | x^{(1)}, x^{(2)}, ..., x^{(m)})$ 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 $Y^{(i)}$ (i.e. $H(X'^{(i)} | Y^{(i)}) = 0$ ($i \in [m]$)). Random variable $X'^{(i)}$ (for $i \in [m]$) is assumed to take value from set $\mathcal{X}'^{(i)}$. 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

---

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

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

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

$$, \{q(x_1, ..., x_m)\}\bigg) \subseteq$$

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

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

2) Assume that

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

Then we require that for any input distribution $q(x_1, x_2, ..., x_m)$, the set

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

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

3) Assume that

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

$$p(y^{(1)}, ..., y^{(m)}|x^{(1)}, ..., x^{(m)}) \prod_{i=1}^{m} p(z_i|y_i).$$

Then we require that for any input distribution $q(x_1, x_2, ..., x_m)$,

$$\phi\big(p(z^{(1)}, ..., z^{(m)}|x^{(1)}, ..., x^{(m)}), \{q(x_1, ..., x_m)\}\big) \subseteq$$

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

The first condition is intuitively saying that additional use of the channel

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

can expand $\phi(.)$ 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 $\phi(.)$ vanishes if the parties are unable to communicate, that is each party receives exactly what it puts at the input of the channel. The third condition is basically saying that making a channel weaker at each party can not cause $\phi(.)$ expand.

Lemma 1: For any function $\phi(.)$ satisfying the above three properties, and for any multiterminal network

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

distortions $D^{(i)}$ and arbitrary admissible source $W^{(i)}$ ($i \in [m]$), positive ϵ and (n)-code satisfying the distortion constraints and a permissible set Ψ of input distributions, we have (for the definition of multiplication of a set by a real number see Definition 3):

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

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

where $W_{1:n}^{(i)}$ ($i \in [m]$) are the messages observed at the nodes; $\widehat{M}_{1:n}^{(i)}$ ($i \in [m]$) are the reconstructions by the parties at the end of the communication satisfying

$$\mathbb{E}\left[\Delta_n^{(i)}\left((\widehat{m}_{1:n}^{(i)}, m_{1:n}^{(i)})\right)\right] \leq D^{(i)} + \epsilon,$$

for any $i \in [m]$.

IV. PROOFS

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

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

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

$$\phi\big(p(w_{1:n}^{(1)}y_{1:n-1}^{(1)}, w_{1:n}^{(2)}y_{1:n-1}^{(2)}, ..., w_{1:n}^{(m)}y_{1:n-1}^{(m)}|w_{1:n}^{(1)}, w_{1:n}^{(2)}, ..., w_{1:n}^{(m)}), \{p(w_{1:n}^{(1)}, w_{1:n}^{(2)}, ..., w_{1:n}^{(m)})\}\big) \oplus$$

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

$$\phi\big(p(w_{1:n}^{(1)}y_{1:n-2}^{(1)}, w_{1:n}^{(2)}y_{1:n-2}^{(2)}, ..., w_{1:n}^{(m)}y_{1:n-2}^{(m)}|w_{1:n}^{(1)}, w_{1:n}^{(2)}, ..., w_{1:n}^{(m)}), \{p(w_{1:n}^{(1)}, w_{1:n}^{(2)}, ..., w_{1:n}^{(m)})\}\big) \oplus$$

$$\phi\big(q(y_{n-1}^{(1)}, y_{n-1}^{(2)}, ..., y_{n-1}^{(m)}|x_{n-1}^{(1)}, x_{n-1}^{(2)}, ..., x_{n-1}^{(m)}), \Psi\big) \oplus$$

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

$$\cdots \subseteq$$

$$\phi\big(p(w_{1:n}^{(1)}, w_{1:n}^{(2)}, ..., w_{1:n}^{(m)}|w_{1:n}^{(1)}, w_{1:n}^{(2)}, ..., w_{1:n}^{(m)}), \{p(w_{1:n}^{(1)}, w_{1:n}^{(2)}, ..., w_{1:n}^{(m)})\}\big) \oplus$$

$$\phi\big(q(y_1^{(1)}, y_1^{(2)}, ..., y_1^{(m)}|x_1^{(1)}, x_1^{(2)}, ..., x_1^{(m)}), \Psi\big) \oplus$$

$$\phi(q(y_2^{(1)}, y_2^{(2)}, ..., y_2^{(m)} | x_2^{(1)}, x_2^{(2)}, ..., x_2^{(m)}), \Psi) \oplus \cdots$$

$$\phi(q(y_{n-1}^{(1)}, y_{n-1}^{(2)}, ..., y_{n-1}^{(m)} | x_{n-1}^{(1)}, x_{n-1}^{(2)}, ..., x_{n-1}^{(m)}), \Psi) \oplus$$

$$\phi(q(y_n^{(1)}, y_n^{(2)}, ..., y_n^{(m)} | x_n^{(1)}, x_n^{(2)}, ..., x_n^{(m)}), \Psi) \subseteq \tag{3}$$

$$\phi(q(y_1^{(1)}, y_1^{(2)}, ..., y_1^{(m)} | x_1^{(1)}, x_1^{(2)}, ..., x_1^{(m)}), \Psi) \oplus$$

$$\phi(q(y_2^{(1)}, y_2^{(2)}, ..., y_2^{(m)} | x_2^{(1)}, x_2^{(2)}, ..., x_2^{(m)}), \Psi) \oplus \cdots$$

$$\phi(q(y_{n-1}^{(1)}, y_{n-1}^{(2)}, ..., y_{n-1}^{(m)} | x_{n-1}^{(1)}, x_{n-1}^{(2)}, ..., x_{n-1}^{(m)}), \Psi) \oplus$$

$$\phi(q(y_n^{(1)}, y_n^{(2)}, ..., y_n^{(m)} | x_n^{(1)}, x_n^{(2)}, ..., x_n^{(m)}), \Psi) \subseteq \tag{4}$$

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

where in equation 1 we have used property (3); in equation 2 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 that

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

The definition of permissible sets implies that the joint distribution $p(x_n^{(1)}, x_n^{(2)}, ..., x_n^{(m)})$ is in Ψ; in equation 3 we have used property (2). In equation 4, 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 $\phi(q(y^{(1)}, ..., y^{(m)} | x^{(1)}, ..., x^{(m)}), \Psi)$. ∎

Proof: [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 $[m]$. Take an arbitrary $\epsilon > 0$ and an (n)-code satisfying the average distortion condition $D^{(i)}$ (for all $i \in [m]$) over the channel $q(y^{(1)}, y^{(2)}, ..., y^{(m)} | x^{(1)}, x^{(2)}, ..., x^{(m)})$. Let random variables $X_k^{(i)}$ and $Y_k^{(i)}$ ($k \in [n]$, $i \in [m]$) respectively represent the inputs to the multiterminal network and the outputs at the nodes of the network. Also assume that $W_{1:n}^{(i)}$ ($i \in [m]$) are the messages observed at the nodes. Let $\widehat{M}_{1:n}^{(i)}$ ($i \in [m]$) be the reconstructions by the parties at the end of the communication satisfying

$$\mathbb{E}\left[\Delta_n^{(i)}\left((\widehat{m}_{1:n}^{(i)}, m_{1:n}^{(i)})\right)\right] \leq D^{(i)} + \epsilon,$$

for any $i \in [m]$. Lastly, let Ψ be a permissible set of input distributions.

We define a function $\phi(.)$ 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

$$\phi(p(y^{(1)}, y^{(2)}, ..., y^{(m)}|x^{(1)}, x^{(2)}, ..., x^{(m)}), \Psi)) = \tag{5}$$

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

where $\varphi(p(y^{(1)}, y^{(2)}, ..., y^{(m)}, x^{(1)}, x^{(2)}, ..., x^{(m)}))$ is defined as the down-set [2] of a vector of size $c = 2^m - 2$ whose k^{th} coordinate equals $I\big(X^{(i)} : i \in T_k \; ; \; Y^{(j)} : j \in (T_k)^c|X^{(j)} : j \in (T_k)^c\big)$ where 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 kth subset in that ordering.

In appendices I-A, I-B and I-C, we verify that $\phi(.)$ satisfies the three properties of Lemma 1 for the choice of $c = 2^m - 2$. Lemma 1 thus implies that (for the definition of multiplication of a set by a real number see Definition 3):

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

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

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

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 $c + 1 = 2^m - 1$ points in the set. Corresponding to the ith 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 $p(x^{(1)}, x^{(2)}, ..., x^{(m)}, z) = p(z) \cdot q_z(x^{(1)}, x^{(2)}, ..., x^{(m)})$ where Z is a random variable defined on the set $\{1, 2, 3, ..., 2^m - 1\}$, taking value i with probability equal to the weight associated to the ith point

---

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

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

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

Conversely, the above set only involves convex combination of points in $\phi(q(y^{(1)}, ..., y^{(m)}|x^{(1)}, ..., x^{(m)}), \Psi)$ and hence is always contained in the convex hull of $\phi(q(y^{(1)}, ..., y^{(m)}|x^{(1)}, ..., x^{(m)}), \Psi)$. Therefore it must be equal to the convex hull region.

Hence,

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

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

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 $\overrightarrow{\widetilde{v}}$ whose k^{th} element equals:[3]

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

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 that 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 $D^{(i)} + \epsilon$.[4] In Appendix II, we perturb random variables $\widetilde{\widehat{M}^{(i)}}$ (for $i \in [m]$) and

---

[3]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. repetition of (X, Y), we have: $I(X^n; Z^n|Y^n) = nH(X|Y) - H(X^n|Z^nY^n) \geq \sum_{g=1}^n H(X_g|Y_g) - H(X_g|Y_gZ_g) = \sum_{g=1}^n I(X_g; Z_g|Y_g) = n \cdot I(X_G; Z_G|GY_G) \geq n \cdot I(X_G; Z_G|Y_G)$ 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).

[4]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 $\mathbb{E}[\Delta(Y_G, Z_G)] = \mathbb{E}[\mathbb{E}[\Delta(Y_G, Z_G)|G]] = \sum_{g=1}^n \frac{1}{n}\mathbb{E}[\Delta(Y_g, Z_g)] = \mathbb{E}[\Delta_n(Y^n, Z^n)]$.

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

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

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

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

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

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

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

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

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

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

Please note that since $\varphi(.)$ 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.[5] The left hand side can be therefore written as union over 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 $I\big(X^{(i)} : i \in T_k \quad ; \quad Y^{(j)} : j \in (T_k)^c | X^{(j)} : j \in (T_k)^c, Z\big)$. Since the $\overrightarrow{\widetilde{v'}}$ 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 $\overrightarrow{\widetilde{v'}}$. The proof ends by recalling the definition of $\overrightarrow{\widetilde{v'}}$ and letting ϵ converge zero. ∎

---

[5] 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 $\lambda \in [0, 1]$.

APPENDIX I

COMPLETING THE PROOF OF THEOREM 1

A. Checking the first property of Lemma 1

Given the definition of $\phi(.)$ in equation 5, 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

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

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 kth 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:

$$I\big(X^{(i)}:i\in T_k \; ; \; Y^{'(j)}:j\in(T_k)^c|X^{(j)}:j\in(T_k)^c,Y^{(j)}:j\in(T_k)^c\big) \leq \tag{6}$$

$$I\big(X^{(i)}X^{'(i)}:i\in T_k \; ; \; Y^{'(j)}:j\in(T_k)^c|X^{(j)}:j\in(T_k)^c,Y^{(j)}X^{'(j)}:j\in(T_k)^c\big) = \tag{7}$$

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

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

$$I\big(X^{(j)} Y^{(j)} : j \in (T_k)^c \ ; \ Y^{'(j)} : j \in (T_k)^c | X^{'(j)} : j \in (T_k)^c\big) = \tag{8}$$

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

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

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

where in inequality 6 we have used the fact that $H(X^{'(i)}|Y^{(i)})$=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 can not cause the expression decrease. In the equations 7 and 8 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 kth 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 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 kth coordinate of $\overrightarrow{v_2}$ be

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

∎

B. Checking the second property of Lemma 1

Our choice of $\phi(.)$ 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 kth 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 kth 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. ∎

C. Checking the third property of Lemma 1

Given the definition of $\phi(.)$ in equation 5, one needs to verify that:

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

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

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

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

The kth coordinate of $\overrightarrow{v}$ is less than or equal to $I\big(X^{(i)} : i \in T_k \; ; \; Z^{(j)} : j \in (T_k)^c|X^{(j)} : j \in (T_k)^c\big)$ where T_k is defined as in the proof of Theorem 1. The latter vector itself is less than or equal to a vector, denoted here by $\overrightarrow{v'}$, whose k^{th} coordinate is equal to $I\big(X^{(i)} : i \in T_k \; ; \; Y^{(j)} : j \in (T_k)^c|X^{(j)} : j \in (T_k)^c\big)$ because

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

$$p(y^{(1)}, y^{(2)}, ..., y^{(m)}|x^{(1)}, x^{(2)}, ..., x^{(m)})\prod_{i=1}^{m} p(z_i|y_i),$$

implying that for every $i \in [m]$, $I(Z^{(i)};G^{(i)}|Y^{(i)})$ is zero for $G^{(i)}$ defined as follows:

$$G^{(i)} = (Z^{(1)}, Z^{(2)}, ..., Z^{(i-1)}, Z^{(i+1)}, ..., Z^{(m)}, Y^{(1)}, Y^{(2)}, ..., Y^{(i-1)}, Y^{(i+1)}, ..., Y^{(m)}, X^{(1)}, X^{(2)}, ..., X^{(m)}).$$

Since the point $\overrightarrow{v'}$ is inside

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

we conclude that

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

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

∎

APPENDIX II

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

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

and furthermore

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

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

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

Intuitively speaking, the algorithm for creating $\widetilde{M'^{(i)}}$ is to begin with $\widetilde{M^{(i)}}$ ($i \in [m]$), and then perturbs this set of m random variables in m stages as follows: at the rth stage, we perturb the rth 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 $(\widetilde{M^{(1)}}, \widetilde{M^{(2)}}, ..., \widetilde{M^{(m)}})$. 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 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 $\tau_r(.)$ 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 $\widetilde{M'^{(i)}} = G_m^{(i)}$ for all $i \in [m]$ and let $\tau(\epsilon) = \sum_{r=1}^m \tau_r(\epsilon)$.

For any arbitrary $k \in [2^m - 2]$, 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 $D^{(r)}$.[6]

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 $I(A; B'|C) - I(A; B|C) \leq I(A; B'|BCD)$, [7] we can write:

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

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

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

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 value 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 $r \in T_k$.

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

---

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

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

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

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

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

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

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

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

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

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 $\epsilon \to 0$.

∎

REFERENCES

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

[2] M. Gastpar, "Cut-set Arguments For Source-Channel Networks," Proc IEEE Int Symp Info Theory, 34, 2004.

[3] B. Nazer and M. Gastpar, "Computation over multiple-access channels," IEEE Trans. IT, 53(10): 3498-3516, 2007.

[4] T. M. Cover, A. El Gamal, and M. Salehi, "Multiple access channels with arbitrarily correlated sources," IEEE Trans. IT, 26 (6): 648-657, (1980).

[5] G. Kramer and S. A. Savari, "Cut sets and information flow in networks of two-way channels," Proc IEEE Int Symp IT, 33, 2004.

[6] A. Giridhar and P. R. Kumar, "Computing and communicating functions over sensor networks," IEEE J. Sel. Areas Commun., 23(4): 755764, (2005).

[7] A. Orlitsky and J. R. Roche, "Coding for computing," IEEE Trans. IT, 47 (3): 903917 (2001).

[8] H. Yamamoto, "Wyner-Ziv theory for a general function of the correlated sources," IEEE Trans. IT, 28 (5): 803807 (1982).

[9] R. Appuswamy, M. Franceschetti, N. Karamchandani, and K. Zeger, "Network Coding for Computing", 46th Annual Allerton Conf. on Commun., Control and Comp, 1-6, 2008.

[10] A. A. Gohari and V. Anantharam, "Information-Theoretic Key Agreement of Multiple Terminals – Part I: Source Model," *Preprint*, Dec. 2007. Available at http://www.eecs.berkeley.edu/~aminzade/SourceModel.pdf

[11] A. A. Gohari and V. Anantharam, "Information-Theoretic Key Agreement of Multiple Terminals – Part II: Channel Model," *Preprint*, Dec. 2007. Available at http://www.eecs.berkeley.edu/~aminzade/ChannelModel.pdf