Infeasibility Proof and Information State in Network Information Theory

Amin Gohari and Venkat Anantharam

Abstract— **In this paper, we revisit the structure of infeasibility results in network information theory, based on a notion of information state. We also discuss ideas for generalizing a known outer bound for lossless transmission of independent sources over a network to one of lossy transmission of dependent sources over the same network. To concretely demonstrate this, we apply our ideas and prove new results for lossy transmission of dependent sources by generalizing: 1) the cut-set bound; 2) the best known outer bound on the capacity region of a general broadcast channel; and 3) the outer bound part of the result of Maric, Yates, and Kramer on strong interference channels with a common message.**

Index Terms— **Information state, converse proof, lossy joint source-channel coding.**

I. Introduction

PROVING a given outer bound in network information theory generally follows by starting from an arbitrary code of length n and identifying suitable auxiliary random variables, establishing that the rates supported by the code lie in the given outer bound region (Gallager type technique). The auxiliary random variables may involve the past and/or future of other random variables, and it is not always easy to assign meanings to these expressions in an operational sense. Therefore infeasibility results generally lack the transparency, meaningfulness, and structure of the achievability proofs in information theory.

In this paper we provide connections between what we call the "information state" and the structure of the converse proofs. During simulation of a code for a general network information theory problem, the information of the parties begins from the $i-th$ terminal having i.i.d. repetitions of its source, 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 with high probability or within a desired average distortion. To define the information state of the parties at a given stage during the communication process, we construct 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. The input associated to this virtual channel is the initial knowledge of the parties. The information state is then defined as the virtual channel together with the input associated to it. We show that the infeasibility proofs can be understood as basically quantifying the gradual evolution of the information state, *bounding 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 (a "one step of communication property") to 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. In order to quantify the information of the k parties at a given stage of the process, one can evaluate the outer bound expression we started with for the virtual channel and the input distribution associated to it.

Furthermore we show that it is relatively easy to extend outer bounds for the rate region in problems of lossless transmission of independent messages over a network to that of lossy transmission of dependent messages over the same network. More specifically, we consider the expression of the cut-set bound, and that of two outer bounds for broadcast and interference channels, all for lossless transmission of independent sources. We observe that the fact that these expressions are valid outer bounds for problems of transmission of independent messages implies that these functionals satisfy the "one step of communication property". This has implications beyond that of lossless transmission of independent sources: with a little more work one can show a similar result for lossy transmission of correlated sources.

In Section II we revisit the proofs for the outer bound to capacity for the point to point discrete memoryless channel and the cut-set bound; no new result is proven here, but it foreshadows the form of the subsequent proofs. In Section III, we introduce the basic notation and definitions used in this paper. The results are provided in Section IV. We end the paper in Section V by some remarks on information state, additivity, and the structure of auxiliary random variables. The proof of

Manuscript received July 17, 2010; revised November 25, 2013; accepted June 21, 2014. Date of publication August 13, 2014; date of current version September 11, 2014. This work was supported in part by the National Science Foundation (NSF) under Grant CCF-0500234, Grant CCF-0635372, Grant CNS-0627161, and Grant CNS-0910702, in part by the NSF Science and Technology Center, under Grant CCF-0939370, in part by the Science of Information, and in part by the Army Research Office, Multidisciplinary University Research Initiative, under Grant W911NF-08-1-0233 through the Project entitled Tools for the Analysis and Design of Complex Multiscale Networks. This paper was presented at the 2009 IEEE International Symposium on Information Theory.

A. Gohari was with the Department of Electrical Engineering and Computer Sciences, University of California at Berkeley, Berkeley, CA 94720 USA. He is now with the Department of Electrical Engineering, Sharif University of Technology, Tehran 14588-89694, Iran (e-mail: aminzadeh@sharif.edu).

V. Anantharam is with the Department of Electrical Engineering and Computer Sciences, University of California at Berkeley, Berkeley, CA 94720 USA (e-mail: ananth@eecs.berkeley.edu).

Communicated by S. Diggavi, Associate Editor for Shannon Theory.

Digital Object Identifier 10.1109/TIT.2014.2347301

a lemma needed in our discussion of the interference channel is relegated to the Appendix.

II. Revisiting the Infeasibility Proofs for the Point to Point Channel and the Cut-Set Bound

In this section we revisit the infeasibility proofs for the point to point discrete memoryless channel (with and without output feedback) and the cut-set bound. This section does not contain new results and aims to convey the basic intuitions. Since these are known theorems, we are being slightly loose and overlook the ϵ terms that come from the Fano inequality.

Consider the problem of transmission of a message M over a point to point discrete memoryless channel $q(y|x)$, without output feedback. We would like to revisit the proof that any achievable code attains a rate less than or equal to $C = \max_{p(x)} I(X;Y)$. Consider a code of length n represented by random variables X^n and Y^n, where $p(m, x^n, y^n) = p(m)p(x^n|m)\prod_{i=1}^{n} q(y_i|x_i)$. The information available to the transmitter at the stage i of the communication is M, and the information available to the receiver is Y^i. Here we use Y^i to denote $(Y_1,\dots,Y_i)$. To assign an information state to this stage, we may construct a virtual channel $p(y^i|m)$ together with the input pmf $p(m)$ for the input. A single use of this virtual channel directly gives the receiver Y^i; thus the reason for calling it an information state. Now, we use the same mutual information functional to assign the real number $I(M;Y^i)$ to the information state (channel $p(y^i|m)$ with input pmf $p(m)$). It is easy to verify that (1) at $i = 0$, $I(M;Y^i) = 0$, (2) at each stage i, the amount of increase in $I(M;Y^i)$, i.e. $I(M;Y^i) - I(M;Y^{i-1})$ is at most C, and (3) at $i = n$, $I(M;Y^n)$ becomes greater than or equal to $I(M;\hat{M}) \approx H(M)$, where $\hat{M}$ denotes the reconstruction of M at the receiver at the end of the n-th stage. These three properties imply that $H(M)$ is at most nC. Observe that the term $I(M;\hat{M})$ can be also understood as the result of applying the mutual information functional to the information state $p(\hat{m}|m)$ with input pmf $p(m)$ (the information state after the decoder has computed $\hat{M}$ and discarded Y^n).

Next, let us consider a point-to-point channel with output feedback. Here the information available to the transmitter also grows: at the stage i of the communication the transmitter knows MY^i, and the receiver knows Y^i. Thus, we need to revise our definition of information state. To fully capture the information state, we denote the initial information available to the two terminals by the pair (W_1, W_2) where the first coordinate $W_1 = M$ is for the transmitter and the second coordinate $W_2 = constant$ is for the receiver. The information of the two parties at stage i is equal to the pair (MY^i, Y^i). Thus the virtual channel associated to this would be $p(my^i, y^i|w_1, w_2)$, together with the input distribution $p(w_1, w_2)$. This is a channel with two inputs and two outputs, whereas $q(y|x)$ was a channel with one input and one output. However we can also interpret the original channel $q(y|x)$ as a channel $p(y_1, y_2|x_1, x_2) = q(y_1|x_1)1[y_2 = y_1]$ where x_1 is the input by the transmitter, and x_2 is the input by the receiver (which we can disable by setting the alphabet set of $\mathcal{X}_2$ to be of size one). Y_1 and Y_2 are the outputs at the transmitter and receiver respectively. The mutual information functional over this channel is still the mutual information between the transmitter's input and the receiver's output. Therefore, applying this to the virtual channel $p(my^i, y^i|w_1, w_2)$ and the input distribution $p(w_1, w_2)$, we again get back $I(M;Y^i)$ as a quantifier of the information state. It is easy to verify that the three conditions mentioned above on $I(M;Y^i)$ are still satisfied and the same upper bound proof goes through.

Lastly, let us revisit the proof of the cut-set bound for a network $q(y_1y_2\cdots y_k|x_1x_2\cdots x_k)$. Let us denote the input to the $j-th$ terminal at time stage i of a code by X_{ji}. According to this outer bound, if it is possible to communicate at rate $R^{(j_1,j_2)}$ from node j_1 to j_2 for $j_1, j_2 \in \{1,2,\dots,k\}$ with asymptotically vanishing block probability of error, then there should exist a pmf $p(x_1x_2\cdots x_k)$ such that for every $T \subset \{1,2,\dots,k\}$ we have

$$\sum_{j_1\in T, j_2\in T^c} R^{(j_1,j_2)} \le I\big(X_T; Y_{T^c}|X_{T^c}\big),$$

where $X_T = (X_i, i \in T)$. Assume that the parties initially have access to independent messages $W_1, W_2, \dots, W_k$. The information of the terminals begins from the $j-th$ terminal having W_j and gradually evolves over time with the use of the network. At the $i-th$ stage, the $j-th$ terminal has $W_jY_j^i$, where $Y_j^i = (Y_{j1}, Y_{j2}, \dots, Y_{ji})$. We represent the information state of the whole system at the $i-th$ stage by the virtual channel $p(w_1y_1^i, w_2y_2^i, \dots, w_ky_k^i|w_1,\dots,w_k)$ and the input distribution $p(w_1,\dots,w_k)$. For a given $T \subset \{1,2,\dots,k\}$, we can use this virtual channel and the input associated to it with the help of the cut-set bound functional to quantify this information state by $I\big(W_T; W_{T^c}Y_{T^c}^i|W_{T^c}\big)$ where $Y_{T^c}^i = (Y_j^i, j \in T^c) = (Y_{j1}, Y_{j2}, \dots, Y_{ji}, j \in T^c)$. We now verify the three properties (1) at $i = 0$, $I\big(W_T; W_{T^c}Y_{T^c}^i|W_{T^c}\big) = 0$; (2) at each stage i, the amount of increase in $I\big(W_T; W_{T^c}Y_{T^c}^i|W_{T^c}\big)$ is at most $I\big(X_{Ti}; Y_{T^ci}|X_{T^ci}\big)$ where $X_{Ti} = (X_{ji}, j \in T)$ and $X_{T^ci} = (X_{ji}, j \in T^c)$, i.e.

$$\begin{aligned}
&I\big(W_T; W_{T^c}Y_{T^c}^i|W_{T^c}\big) - I\big(W_T; W_{T^c}Y_{T^c}^{i-1}|W_{T^c}\big)\\
&= I\big(W_T; Y_{T^ci}|W_{T^c}Y_{T^c}^{i-1}\big)\\
&= I\big(W_T; Y_{T^ci}|W_{T^c}X_{T^ci}Y_{T^c}^{i-1}\big)\\
&\le I\big(W_TX_{Ti}; Y_{T^ci}|W_{T^c}X_{T^ci}Y_{T^c}^{i-1}\big)\\
&= I\big(X_{Ti}; Y_{T^ci}|X_{T^ci}\big);
\end{aligned}$$

and (3) at $i = n$, $I\big(W_T; W_{T^c}Y_{T^c}^n|W_{T^c}\big) = I\big(W_T; W_{T^c}Y_{T^c}^n\big) \ge I\big(W_T; W_{T^c}\hat{M}_{T^c}^n\big) \ge I\big(W_T; \hat{M}_{T^c}\big)$ becomes greater than or equal to $n\sum_{j_1\in T, j_2\in T^c} R^{(j_1,j_2)}$, for the messages from nodes $j_1 \in T$ are functions of W_T and are included in $\hat{M}_{T^c}$. Thus, by adding up how much the quantifier $I\big(W_T; W_{T^c}Y_{T^c}^i|W_{T^c}\big)$ increases in each stage and where it starts and ends up we can write:

$$n\sum_{j_1\in T, j_2\in T^c} R^{(j_1,j_2)} \le \sum_{i=1}^{n} I\big(X_{Ti}; Y_{T^ci}|X_{T^ci}\big).$$

Observe that the above is simply the same single-letterization step of the traditional cut-set bound proof that has been rewritten in a different form.

TABLE I
NOTATION

Variable	Description
$\mathbb{R}$, $\mathbb{R}_+$	Real numbers, Non-negative real numbers.
$[k]$	The set $\{1, 2, 3, ..., k\}$.
k	Number of nodes of the network.
$q(y_1y_2...y_k\|x_1x_2...x_k)$ or $q(\mathbf{y}\|\mathbf{x})$	The statistical description of a general multi-terminal network.
W_j	Rv representing the source observed at the $j-th$ node.
M_j	Rv to be reconstructed in a possibly lossy way, at the $j-th$ node.
$\mathcal{X}_j, \mathcal{Y}_j, \mathcal{W}_j, \mathcal{M}_j$	Alphabets of X_j, Y_j, W_j, M_j.
$\Delta_j(\cdot,\cdot)$	Distortion function used by the $j-th$ terminal.
$\mathcal{E}_{ji}(\cdot)$ $\mathcal{D}_j(\cdot)$	The encoding function used by the $j-th$ terminal at the $i-th$ stage. The decoding function at the $j-th$ terminal.
n	Length of the code used.
Ψ	A permissible set of input distributions; Ψ is a set of joint distributions on $\mathcal{X}_1 \times \mathcal{X}_2 \times \mathcal{X}_3 \times \cdots \times \mathcal{X}_k$. Inputs to the network are required to have a joint distribution in Ψ.

III. DEFINITIONS AND NOTATION

Throughout this paper we assume that each random variable takes values in a finite alphabet (set), denoted by calligraphic scripts (e.g. $\mathcal{X}$ for random variable X, etc.). $\mathbb{R}$ denotes the set of real numbers and $\mathbb{R}_+$ denotes the set of non-negative reals. We mostly adopt the notation of the textbook of El Gamal and Kim [1]. See Table I for a summary of the description of the variables that we use in the paper. For any natural number k, let $[k] = \{1, 2, 3, \ldots, k\}$. For a set $S \subset [k]$, let S^c denote its complement, that is $[k] - S$. We use Y_j^i to denote $(Y_{j1}, Y_{j2}, \ldots, Y_{ji})$. Furthermore, $\mathbf{W}$ stands for $(W_1, W_2, \ldots, W_k)$, $\mathbf{X}$ stands for $(X_1, X_2, \ldots, X_k)$, etc. Similarly, we use $\mathbf{W}_i$ to denote $(W_{1i}, W_{2i}, \ldots, W_{ki})$ and $\mathbf{W}^i$ to denote $(W_1^i, W_2^i, \ldots, W_k^i)$, so $\mathbf{W}^n$ denotes $(W_1^n, W_2^n, \ldots, W_k^n)$, etc. For any subset $S \subset [k]$, we use X_S to denote the set of random variables $(X_j : j \in S)$. Similarly, we use X_{Si} and X_S^i to denote $(X_{ji}, j \in S)$ and $(X_j^i, j \in S)$ respectively, so X_S^n denotes $(X_j^n, j \in S)$, etc. Note that $\mathbf{W}$, $\mathbf{W}_i$, and $\mathbf{W}^i$ can be thought of as more compact notation for $W_{[k]}$, $W_{[k]i}$, and $W_{[k]}^i$ respectively. When writing joint probabilities and mutual information terms, we generally drop the commas between variables, e.g. $p(w_1w_2)$ rather than $p(w_1, w_2)$. In a few cases we keep the commas when we want to highlight certain groupings of the variables.

A (discrete memoryless) general multiterminal network (GMN) is characterized by the conditional distribution $q(y_1y_2\cdots y_k|x_1x_2\cdots x_k)$, where X_j and Y_j ($j \in [k]$) are respectively the input and the output of the channel at the $j-th$ terminal. In the previous section, we assumed that W_j was representing all the messages available at terminal j. However since we want to discuss lossy source coding throughout the rest of the paper we assume that the $j-th$ terminal ($j \in [k]$) has access to *n i.i.d. repetitions* of W_j before the beginning of the communication at block length n. The message that needs to be delivered (in a possibly lossy manner) to the $j-th$ terminal is taken to be $M_j = f_j(W_1, W_2, \ldots, W_k)$ for some function f_j. If this is asymptotically possible within a given distortion in the sense that is defined below, we call the source $(W_1, W_2, \ldots, W_k)$ admissible. Of particular interest is the case when the function $f_j(W_1, W_2, \ldots, W_k)$ takes the special form of $\big(f_{(j,1)}(W_1), f_{(j,2)}(W_2), \ldots, f_{(j,k)}(W_k)\big)$ for some functions $f_{(j_1,j_2)}$.

For any $j \in [k]$, let the distortion function Δ_j be a function $\Delta_j : \mathcal{M}_j \times \mathcal{M}_j \to [0, \infty)$. For any natural number n and vectors $(m_{j1}, m_{j2}, \ldots, m_{jn})$ and $(m'_{j1}, m'_{j2}, \ldots, m'_{jn})$ from $(\mathcal{M}_j)^n$, let $\Delta_{jn}(m_j^n, m_j'^n) = \frac{1}{n}\sum_{r=1}^{n} \Delta_j(m_{jr}, m'_{jr})$. Roughly speaking, we require the i.i.d. repetitions of random variable M_j to be reconstructed, by the $j-th$ terminal, within the average distortion D_j.

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

$$\mathcal{E}_{j1} : (\mathcal{W}_j)^n \longrightarrow \mathcal{X}_j, \quad \forall j \in [k];$$
$$\mathcal{E}_{ji} : (\mathcal{W}_j)^n \times (\mathcal{Y}_j)^{i-1} \longrightarrow \mathcal{X}_j, \quad \forall j \in [k], i \in [2:n];$$
$$\mathcal{D}_j : (\mathcal{W}_j)^n \times (\mathcal{Y}_j)^n \longrightarrow (\mathcal{M}_j)^n, \quad \forall j \in [k]. \tag{1}$$

Intuitively speaking $\mathcal{E}_{ji}$ is the encoding function of the $j-th$ terminal at the $i-th$ time instance, and $\mathcal{D}_j$ is the decoding function of the $j-th$ terminal. The joint pmf induced by an n-code is defined as follows: assume that random variables $\mathbf{W}^n$ are n i.i.d. repetitions of random variables $\mathbf{W}$ with joint distribution $p(\mathbf{w})$. Random variables X_{ji} and Y_{ji} ($i \in [n]$, $j \in [k]$) are defined according to the following constraints:

$$(\mathbf{w}^n\mathbf{x}^n\mathbf{y}^n) = \prod_{i=1}^{n} p(\mathbf{w}_i) \times \prod_{i=1}^{n} q(\mathbf{y}_i|\mathbf{x}_i) \times \prod_{i=1}^{n}\prod_{j=1}^{m} p(x_{ji}|w_j^n y_j^{i-1}); \tag{2}$$

so we may write $X_{j1} = \mathcal{E}_{j1}\big(W_j^n\big)$, and for any $i \in [n] - \{1\}$, $X_{ji} = \mathcal{E}_{ji}\big(W_j^n, Y_j^{i-1}\big)$. Random variables X_{ji} and Y_{ji} represent the input and output of the $j-th$ terminal at the $i-th$ time instance and satisfy the following Markov chains:

$$\mathbf{W}^n\mathbf{Y}^{i-1} \to W_j^n Y_j^{i-1} \to X_{ji}, \quad \mathbf{W}^n\mathbf{Y}^{i-1} \to \mathbf{X}_i \to \mathbf{Y}_i.$$

Let $M_{ji} = f_j(\mathbf{W}_i)$.

Given positive reals ϵ and D_j ($j \in [k]$), and a source marginal distribution $p(\mathbf{w})$, an n-code is said to achieve the average distortion D_j (for all $j \in [k]$) over the channel $q(\mathbf{y}|\mathbf{x})$ if for every $j \in [k]$:

$$\mathbb{E}\left[\Delta_{jn}\left(\mathcal{D}_j\big(W_j^n, Y_j^n\big), M_j^n\right)\right] \le D_j.$$

Similarly the code is said to achieve the average probability of error ϵ if for every $j \in [k]$:

$$\mathbb{P}\big(\mathcal{D}_j\big(W_j^n, Y_j^n\big) \ne M_j^n\big) \le \epsilon.$$

Definition 2: Given positive reals D_j and distortion functions Δ_j, a source marginal distribution $p(\mathbf{w})$ is called an admissible source *over the channel $q(\mathbf{y}|\mathbf{x})$ for the target reproduction functions f_j ($j \in [k]$) at the distortion levels D_j ($j \in [k]$) if for sufficiently large n, an n-code achieving the average distortion D_j exists. Similarly, a source marginal distribution is said to be* admissible *with a vanishing probability of error if every $\epsilon > 0$ and for sufficiently large n, an n-code with average probability of error of ϵ exists.*

Definition 3: Given a GMN $q(\mathbf{y}|\mathbf{x})$, and the source marginal distribution $p(\mathbf{w})$, a permissible set *of input distributions, Ψ, is a set of joint distributions on $\mathcal{X}_1 \times \mathcal{X}_2 \times \mathcal{X}_3 \times \cdots \times \mathcal{X}_k$ 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.*

The set of all probability distributions on $\mathcal{X}_1 \times \mathcal{X}_2 \times \mathcal{X}_3 \times \cdots \times \mathcal{X}_k$ is always a permissible set. However for instance in an interference channel or a multiple access channel with no output feedback, if the transmitters observe independent messages, their input to the channel remain independent. Also, constraints on the set of input distributions when the transmitters are observing i.i.d. repetitions of correlated random variables are reported in [19] (see also [36]).

Definition 1 can be extended in the following way to define the notion of an n-code for a permissible set of input distributions Ψ.

Definition 4: Given a GMN $q(\mathbf{y}|\mathbf{x})$, and the source marginal distribution $p(\mathbf{w})$, an n-code *for permissible set Ψ consists of the three mappings given in eq.* (1) *such that for any $i \in [n]$, the induced joint distribution of $\mathbf{X}_i$ (computed from eq.* (2)*) is in Ψ.*

Definition 5: Given positive reals D_j, distortion functions Δ_j and a permissible set of input distributions Ψ, a source marginal distribution $p(\mathbf{w})$ is called an admissible source *for Ψ over the channel $q(\mathbf{y}|\mathbf{x})$ for the target reproduction functions f_j ($j \in [k]$) at the distortion levels D_j ($j \in [k]$) if for sufficiently large n, an n-code for permissible set Ψ achieving the average distortion D_j exists. Similarly, a source marginal distribution is said to be* admissible *with a vanishing probability of error if every $\epsilon > 0$ and for sufficiently large n, an n-code for permissible set Ψ with average probability of error of ϵ exists.*

IV. Statement of the Results

In this section, we formally present the results. Section IV-A states a generalized cut-set bound that is applicable to all multiterminal networks. Sections IV-B and IV-C state our results for broadcast channels and strong interference channels respectively.

A. Cut-Set Bound for Lossy Transmission of Dependent Sources

The traditional cut-set bound says that if it is possible to communicate at rate $R^{(i,j)}$ from node i to j for $i, j \in [k]$ with asymptotically vanishing block probability of error, then we have

$$\sum_{j_1 \in T, j_2 \in T^c} R^{(j_1,j_2)} \le I\big(X_T; Y_{T^c}|X_{T^c}\big), \quad \forall T \subset [k]$$

for some $p(\mathbf{x})q(\mathbf{y}|\mathbf{x})$. Since the union in the traditional cut-set bound is taken over all $p(\mathbf{x})$, it is not tight for a MAC channel. One can slightly improve it so that it becomes tight for the MAC channel, by considering a permissible set of input distributions Ψ (e.g. independent inputs for the MAC), and express this bound as follows:

$$\sum_{j_1 \in T, j_2 \in T^c} R^{(j_1,j_2)} \le I\big(X_T; Y_{T^c}|X_{T^c}Z\big), \quad \forall T \subset [k]$$

for some $p(\mathbf{x}z)q(\mathbf{y}|\mathbf{x})$ where for any z, $p(\mathbf{x}|z) \in \Psi$. The size of the alphabet of Z can be bounded from above by $2^k - 1$ using Carathéodory's theorem. Exploiting expressions of the form $I\big(X_T; Y_{T^c}|X_{T^c}\big)$ as a quantifier for the information state, we prove the following theorem:

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

- *a joint distribution $p(\mathbf{x}z)$ where the size of the alphabet of Z is $2^k - 1$ and furthermore $p(\mathbf{x}|z)$ belongs to Ψ for any value z that the random variable Z might take;*
- *a joint distribution $p(\hat{\mathbf{m}}, \mathbf{w})$ where for each j the average distortion between $M_j = f_j(\mathbf{W})$ and $\hat{M}_j$ is less than or equal to D_j, i.e. $\mathbb{E}\big[\Delta_j(M_j, \hat{M}_j)\big] \le D_j$,*

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

$$I\big(W_T; \hat{M}_{T^c}|W_{T^c}\big) \le I\big(X_T; Y_{T^c}|X_{T^c}Z\big),$$

where $\mathbf{Y}, \mathbf{X}$ and Z are jointly distributed according to $q(\mathbf{y}|\mathbf{x}) \cdot p(\mathbf{x}z)$. Note that here the following Markov chain holds: $Z \to \mathbf{X} \to \mathbf{Y}$.

Remark 1: The fact that the expressions on both sides of the above inequality are of the same form is due to the feature of our framework where we use the same functional for quantifying the information state and for channel coding. To intuitively understand the form of the statement of the theorem, apply the outer bound expression of the channel coding problem to a virtual channel constructed using the sources as follows: given any $p(\hat{\mathbf{m}}, \mathbf{w})$ such that $\Delta_j(M_j, \hat{M}_j) \le D_j$ holds, consider the virtual channel $p(\hat{\mathbf{m}}|\mathbf{w})$ with the input distribution being $p(\mathbf{w})$. The inputs of this virtual channel, i.e. $\mathbf{W}$, and its outputs, i.e. $\hat{\mathbf{M}}$, can be understood as the raw information and acceptable information objectives at the k parties.

If we let Δ_j be the Hamming distance for each j, then a vanishing probability of error implies reconstruction with average distortion at most ϵ for any $\epsilon > 0$. Thus, by letting ϵ converge to zero, we get the following corollary, basically recovering the original cut-set bound.

Corollary 1: If it is possible to communicate at rate $R^{(j_1,j_2)}$ from node j_1 to j_2 for all $j_1, j_2 \in [k]$ with asymptotically vanishing block probability of error via n-codes for the channel

$q(\mathbf{y}|\mathbf{x})$ with permissible set of input distributions Ψ, then we must have

$$\sum_{j_1\in T, j_2\in T^c} R^{(j_1,j_2)} \leq I\big(X_T; Y_{T^c}|X_{T^c}Z\big), \qquad \forall T\subset[k]$$

for some $p(\mathbf{x},z)q(\mathbf{y}|\mathbf{x})$ where for any z, $p(\mathbf{x}|z)\in\Psi$ and the alphabet of Z is of size 2^k-1.

One can think of $R^{(j_1,j_2)}$ as $H(f_{(j_2,j_1)}(W_i))$ where $M_j = f_j(\mathbf{W})$ ($j\in[k]$) are assumed to consist of independent components $(f_{(j,1)}(W_1),\ldots,f_{(j,k)}(W_k))$. 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. The bound reduces to the traditional cut-set bound when Ψ is taken to be the set of all input distributions, and $I\big(X_T;Y_{T^c}|X_{T^c}Z\big)$ is bounded from above by $I\big(X_T;Y_{T^c}|X_{T^c}\big)$. This is valid because $I\big(X_T;Y_{T^c}|X_{T^c}Z\big) = H\big(Y_{T^c}|X_{T^c}Z\big) - H\big(Y_{T^c}|X_{[k]}Z\big) = H\big(Y_{T^c}|X_{T^c}Z\big) - H\big(Y_{T^c}|X_{[k]}\big) \leq H\big(Y_{T^c}|X_{T^c}\big) - H\big(Y_{T^c}|X_{[k]}\big) = I\big(X_T;Y_{T^c}|X_{T^c}\big)$.

Remark 2: *Other extensions of the cut-set bound can be found in [11] and [20]. Our formulation is motivated by some existing works which show the possible benefits of focusing on function computation rather than just raw communication (see for instance [13], [28], [29], [32]).*

Proof of Theorem 1: Following similar steps as in items (1) and (2) of the proof for the cut-set bound from Section II, we identify the information state at stage i as the virtual channel $p(w_1^n y_1^i, w_2^n y_2^i,\ldots,w_k^n y_k^i|w_1^n,\ldots,w_k^n)$ and the input distribution $p(w_1^n,\ldots,w_k^n)$. The only difference between this information state and that given in Section II is that instead of w_j we have w_j^n here (as discussed in the notation section). Thus given some $T\subset[k]$ the quantifier would be $I\big(W_T^n; W_{T^c}^n Y_{T^c}^i|W_{T^c}^n\big)$. At $i=0$, this expression is zero. Exactly the same proof given for item (2) in Section II shows that $I\big(W_T^n; W_{T^c}^n Y_{T^c}^i|W_{T^c}^n\big) - I\big(W_T^n; W_{T^c}^n Y_{T^c}^{i-1}|W_{T^c}^n\big)$ is at most $I\big(X_{Ti};Y_{T^ci}|X_{T^ci}\big)$. This is due to the fact that we had not used independence of W_j's there. Thus at $i=n$, $I\big(W_T^n; W_{T^c}^n Y_{T^c}^n|W_{T^c}^n\big)$ will be less than or equal to the sum of increases in the n stages of a communication protocol, i.e.

$$I\big(W_T^n; W_{T^c}^n Y_{T^c}^n|W_{T^c}^n\big) \leq \sum_{i=1}^n I\big(X_{Ti};Y_{T^ci}|X_{T^ci}\big).$$

Now, at the decoding phase terminal j computes $\hat{M}_j^n = \mathcal{D}_j(W_j^n, Y_j^n)$, and then discards $W_j^n Y_j^n$. This changes the information state to the virtual channel $p(\hat{m}_1^n,\hat{m}_2^n,\ldots,\hat{m}_k^n|w_1^n,\ldots,w_k^n)$ and the input distribution $p(w_1^n,\ldots,w_k^n)$. The quantifier for this corresponds to $I\big(W_T^n;\hat{M}_{T^c}^n|W_{T^c}^n\big)$. By data processing we have $I\big(W_T^n;\hat{M}_{T^c}^n|W_{T^c}^n\big) \leq I\big(W_T; W_{T^c}^n Y_{T^c}^n|W_{T^c}^n\big)$. Thus,

$$I\big(W_T^n;\hat{M}_{T^c}^n|W_{T^c}^n\big) \leq \sum_{i=1}^n I\big(X_{Ti};Y_{T^ci}|X_{T^ci}\big).$$

Using items 1 and 3 of Lemma 1, and Lemma 2, we have that

$$I\big(W_T^n;\hat{M}_{T^c}^n|W_{T^c}^n\big) = n\cdot I\big(W_T;\hat{M}_{T^c}|W_{T^c}\big)$$

for $\hat{M}_j$ ($j\in[k]$) jointly distributed with W_j ($j\in[k]$) where the average distortion between $M_j = f_j(\mathbf{W})$ and $\hat{M}_j$ is less than or equal to D_j. Further the induced joint pmf of $\hat{M}_j$ and $\mathbf{W}$ does not depend on the set $T\subset[k]$.

Thus, we have shown that there exists $\hat{M}_j$ jointly distributed with $W_{[k]}$ and satisfying the average distortion constraint such that for any $T\subset[k]$

$$I\big(W_T;\hat{M}_{T^c}|W_{T^c}\big) \leq \frac{1}{n}\sum_{i=1}^n I\big(X_{Ti};Y_{T^ci}|X_{T^ci}\big).$$

We can wrap up the proof by using the well-known technique of expressing the convexification on the right hand side using the auxiliary rv Z, i.e. $\frac{1}{n}\sum_{i=1}^n I\big(X_{Ti};Y_{T^ci}|X_{T^ci}\big) = I\big(X_T;Y_{T^c}|X_{T^c}Z\big)$. Observe that given any value of Z, $p(\mathbf{x}|z)$ belongs to Ψ, by construction. This property remains valid after reduction of the cardinality of Z since in cardinality reduction we fix $p(\mathbf{x}|z)$ and vary the marginal distribution of Z. □

Lemma 1: Take arbitrary random variables X^n, Y^n, Z^n such that (X^n,Y^n) is n i.i.d. repetitions of (X,Y). Let Q be uniform over $[n]$ and independent of (X^n,Y^n,Z^n). We then have

- $I(X^n;Z^n|Y^n)\geq n\cdot I(X_Q;Z_Q|Y_Q)$
- $H(X^n|Z^n)\leq n\cdot H(X_Q|Z_Q)$
- *Random variables (X_Q,Y_Q) have the same joint distribution as (X,Y).*

Proof: For the first item, we have

$$\begin{aligned}
&I(X^n;Z^n|Y^n)\\
&\quad= nH(X|Y) - H(X^n|Z^nY^n)\\
&\quad\geq \sum_{q=1}^n H(X_q|Y_q) - H(X_q|Y_qZ_q) = \sum_{q=1}^n I(X_q;Z_q|Y_q)\\
&\quad= n\cdot I(X_Q;Z_Q|QY_Q) \geq n\cdot I(X_Q;Z_Q|Y_Q).
\end{aligned}$$

For the second item, we have

$$\begin{aligned}
H(X^n|Z^n) &= \sum_{q=1}^n H(X_q|Z^nX_{1:q-1}) \leq \sum_{q=1}^n H(X_q|Z_q)\\
&= n\cdot H(X_Q|QZ_Q) \leq n\cdot H(X_Q|Z_Q).
\end{aligned}$$

The last item follows from the fact that (X^n,Y^n) is i.i.d. □

Lemma 2: For any arbitrary pair (Y^n,Z^n), the average distortion between Y_Q and Z_Q for Q uniform over $[n]$ and independent of (Y^n,Z^n), i.e. $\mathbb{E}[\Delta(Y_Q,Z_Q)]$, is equal to $\mathbb{E}[\Delta_n(Y^n,Z^n)]$.

Proof: $\mathbb{E}[\Delta(Y_Q,Z_Q)] = \mathbb{E}[\mathbb{E}[\Delta(Y_Q,Z_Q)|Q]] = \sum_{q=1}^n \frac{1}{n}\mathbb{E}[\Delta(Y_q,Z_q)] = \mathbb{E}[\Delta_n(Y^n,Z^n)]$. □

B. Lossy Transmission of Dependent Sources Over a Broadcast Channel

Theorem 2: Take an arbitrary broadcast channel $q(y_2y_3|x_1)$, a permissible set of input distributions Ψ, and an associated admissible source marginal distribution $p(w_1)$ with distortion levels D_2 and D_3 for given distortion *functions Δ_2 and Δ_3. Take a random variable L arbitrarily jointly distributed with (M_2, M_3).*[1] *Then there must exist a joint distribution $p(\hat{m}_2\hat{m}_3|m_2m_3l)$ such that*

$$\mathbb{E}\big[\Delta_j(M_j, \hat{M}_j)\big] \le D_j, \qquad j = 2, 3,$$

and a joint pmf $p(uvwx_1)q(y_2y_3|x_1)$ such that $p(x_1)$ is in the convex hull of Ψ and the following inequalities hold:

$$\begin{aligned}
I(L; \hat{M}_2) &\le I(W; Y_2);\\
I(L; \hat{M}_3) &\le I(W; Y_3);\\
I(LM_2; \hat{M}_2) &\le I(UW; Y_2);\\
I(M_2; \hat{M}_2|L) + I(L; \hat{M}_3) &\le I(U; Y_2|W) + I(W; Y_3);\\
I(M_3L; \hat{M}_3) &\le I(VW; Y_3);\\
I(M_3; \hat{M}_3|L) + I(L; \hat{M}_2) &\le I(V; Y_3|W) + I(W; Y_2);\\
I(L; \hat{M}_2) + I(M_3; \hat{M}_3|L) &\le I(W; Y_2) + I(V; Y_3|W)\\
+I(M_2; \hat{M}_2|M_3L) &\quad +I(U; Y_2|VW);\\
I(M_2; \hat{M}_2|M_3L) &\le I(U; Y_2|VW)\\
+I(M_3L; \hat{M}_3) &\quad +I(VW; Y_3);\\
I(L; \hat{M}_3) + I(M_2; \hat{M}_2|L) &\le I(W; Y_3) + I(U; Y_2|W)\\
+I(M_3; \hat{M}_3|M_2L) &\quad +I(V; Y_3|UW);\\
I(M_3; \hat{M}_3|M_2L) &\le I(V; Y_3|UW)\\
+I(M_2L; \hat{M}_2) &\quad +I(UW; Y_2).
\end{aligned}$$

Remark 3: The above theorem is an extension of the outer bound of [26] for lossy transmission of dependent sources. Nair uses the same expressions involving U, V, W on the right hand side of the above inequalities (see also the Nair and El Gamal outer bound in [27]). The same argument as in [26] shows that we can restrict to $p(uvwx_1) = p(w)p(u|w)p(v|w)p(x_1|uvw)$, X_1 being a function of (U, V, W), and can use the strengthened Carathéodory theorem of Fenchel to bound the cardinality of U, V, W from above by $|\mathcal{U}| \le |\mathcal{X}_1| + 1$, $|\mathcal{V}| \le |\mathcal{X}_1| + 1$, $|\mathcal{W}| \le |\mathcal{X}_1| + 5$.

Remark 4: The right hand side and left hand side of the above equations are of the same form. To see this consider a virtual channel corresponding to the communication task as $p(m_2m_3|w_1)$ together with input distribution $p(w_1)$. We are applying the same functional for the virtual channel (source coding) and channel coding.

Corollary 2: Take an arbitrary broadcast channel $q(y_2y_3|x_1)$, a permissible set of input distributions Ψ, and an associated admissible source marginal distribution $p(w_1)$ with vanishing probability of error at both receivers. Take an arbitrary random variable L satisfying $H(L|M_2) = H(L|M_3) = 0$.[2] *Then there must exist $p(w)p(u|w)p(v|w)p(x_1|uvw)q(y_2y_3|x_1)$ such that X_1 is a function of (U, V, W) and $p(x_1)$ is in the convex hull of Ψ and the following inequalities hold:*

$$\begin{aligned}
H(L) &\le \min\{I(W; Y_2), I(W; Y_3)\};\\
H(M_2) &\le I(U; Y_2|W) + \min\{I(W; Y_2), I(W; Y_3)\};\\
H(M_3) &\le I(V; Y_3|W) + \min\{I(W; Y_2), I(W; Y_3)\};\\
H(M_2M_3) &\le \min\{I(W; Y_2), I(W; Y_3)\}\\
&\quad +I(U; Y_2|VW) + I(V; Y_3|W);\\
H(M_2M_3) &\le \min\{I(W; Y_2), I(W; Y_3)\}\\
&\quad +I(V; Y_3|UW) + I(U; Y_2|W).
\end{aligned}$$

Remark 5: Prior to a preliminary conference version of this article [16], there had been no previous work discussing any interesting outer bounds on the admissible source region of the general broadcast channel with arbitrarily dependent sources. In a later work [21], Kramer, Liang and Shamai independently derived a similar result. All of these works consider lossless transmission. In a subsequent work, the best known outer bound on the general broadcast channel was simplified by Nair in [26]. The best known inner bound for the two receiver general broadcast channel with dependent sources is due to Han and Costa [31]. See [35] for a correction to the proof and also [25] for an alternative proof.

Remark 6: During the review process for this paper, the authors have learnt about a related result that has been independently derived by Khezeli and Chen in [37, Th. 3]. The two bounds differ in two ways: the bound given above has only one auxiliary random variable L whereas the bound given in [37] has three random variables. Furthermore, $p(\hat{m}_2, \hat{m}_3|m_2, m_3l)$ can depend on L but the bound given in [37] does not have such dependence. The bound of [37] can be however shown from the same proof given below (See Remark 7).

Proof of Theorem 2: Since there is only one source W_1^n observed by the transmitter and the information of the second and third terminals increases over time, we may identify the information state at stage i as the virtual channel $p(y_2^i, y_3^i|w_1^n)$ and the input distribution $p(w_1^n)$.

Based on the statement of the theorem, given $q(y_2y_3|x_1)$ and $q(x_1)$, the information state can be quantified by any of the following expressions showing up on the right hand side of the equations for any arbitrary $p(uvw|x_1)$:

$$\Big(I(W; Y_2),\ I(W; Y_3),\ I(UW; Y_2),$$
$$I(U; Y_2|W) + I(W; Y_3), I(V; Y_3|W) + I(W; Y_2),$$
$$I(VW; Y_3), I(U; Y_2|VW) + I(V; Y_3|W) + I(W; Y_2),$$
$$I(U; Y_2|VW) + I(V; Y_3|W) + I(W; Y_3),$$
$$I(V; Y_3|UW) + I(U; Y_2|W) + I(W; Y_2),$$
$$I(V; Y_3|UW) + I(U; Y_2|W) + I(W; Y_3)\Big).$$

This gives the following vector of quantifiers for the information state at stage i given any $p(\tilde{u}_i\tilde{v}_i\tilde{w}_i|w_1^n)$:

$$\Big(I(\tilde{W}_i; Y_2^i),\ I(\tilde{W}_i; Y_3^i),\ I(\tilde{U}_i\tilde{W}_i; Y_2^i),$$

[1] M_2 and M_3 are functions of W_1 that are to be reconstructed at the receivers.

[2] Random variable L is representing the common part between M_2 and M_3 in the sense of Gács and Körner [10] and the common message that needs to be transmitted to both receivers.

$$I(\tilde{U}_i; Y_2^i|\tilde{W}_i) + I(\tilde{W}_i; Y_3^i),$$
$$I(\tilde{V}_i; Y_3^i|\tilde{W}_i) + I(\tilde{W}_i; Y_2^i), I(\tilde{V}_i\tilde{W}_i; Y_3^i),$$
$$I(\tilde{W}_i; Y_2^i) + I(\tilde{V}_i; Y_3^i|\tilde{W}_i) + I(\tilde{U}_i; Y_2^i|\tilde{V}_i\tilde{W}_i),$$
$$I(\tilde{U}_i; Y_2^i|\tilde{V}_i\tilde{W}_i) + I(\tilde{V}_i\tilde{W}_i; Y_3^i),$$
$$I(\tilde{W}_i; Y_3^i) + I(\tilde{U}_i; Y_2^i|\tilde{W}_i) + I(\tilde{V}_i; Y_3^i|\tilde{U}_i\tilde{W}_i),$$
$$I(\tilde{V}_i; Y_3^i|\tilde{U}_i\tilde{W}_i) + I(\tilde{U}_i\tilde{W}_i; Y_2^i)\Big). \quad (3)$$

Let us denote the above vector by $b_i(\tilde{U}_i, \tilde{V}_i, \tilde{W}_i)$ where, with a slight abuse of notation, we have used the variables $\tilde{U}_i, \tilde{V}_i, \tilde{W}_i$ instead of $p(\tilde{u}_i\tilde{v}_i\tilde{w}_i|w_1^n)$ (just like when we use $H(X)$ to denote $H(p(x))$).

Clearly all of these quantifiers are zero when $i = 0$, i.e. $b_0(\tilde{U}_0, \tilde{V}_0, \tilde{W}_0)$ is always zero. Furthermore, to compare the quantifiers at stage i and $i-1$, one can verify that

$$\begin{aligned}
b_i(\tilde{U}_i, \tilde{V}_i, \tilde{W}_i) &\le b_{i-1}(\tilde{U}_{i-1}, \tilde{V}_{i-1}, \tilde{W}_{i-1}) \\
&+ \big(I(W_i; Y_{2i}), I(W_i; Y_{3i}), I(U_iW_i; Y_{2i}), I(U_i; Y_{2i}|W_i) \\
&+ I(W_i; Y_{3i}), I(V_i; Y_{3i}|W_i) + I(W_i; Y_{2i}), \\
&\quad I(V_iW_i; Y_{3i}), I(U_i; Y_{2i}|V_iW_i) + I(V_i; Y_{3i}|W_i) \\
&+ I(W_i; Y_{2i}), I(U_i; Y_{2i}|V_iW_i) + I(V_i; Y_{3i}|W_i) \\
&+ I(W_i; Y_{3i}) I(V_i; Y_{3i}|U_iW_i) + I(U_i; Y_{2i}|W_i) \\
&+ I(W_i; Y_{2i}), I(V_i; Y_{3i}|U_iW_i) + I(U_i; Y_{2i}|W_i) \\
&+ I(W_i; Y_{3i})\big),
\end{aligned}$$

where $\le$ is coordinatewise and $\tilde{W}_{i-1} = \tilde{W}_i Y_{3i}$, $W_i = \tilde{W}_i Y_2^{i-1}$, $\tilde{U}_i = \tilde{U}_{i-1} = U_i, \tilde{V}_i = \tilde{V}_{i-1} = V_i$. Further the Markov chain $\tilde{W}_i\tilde{U}_i\tilde{V}_i \to W_1^n \to Y_2^iY_3^i$ implies $U_iV_iW_i \to X_{1i} \to Y_{2i}Y_{3i}$ and $\tilde{W}_{i-1}\tilde{U}_{i-1}\tilde{V}_{i-1} \to W_1^n \to Y_2^{i-1}Y_3^{i-1}$. In other words, having $\tilde{W}_i, \tilde{U}_i, \tilde{V}_i$, we can find a valid $\tilde{W}_{i-1}, \tilde{U}_{i-1}, \tilde{V}_{i-1}$ and W_i, U_i, V_i in terms of it.

Thus, for any arbitrary $p(\tilde{u}_n\tilde{v}_n\tilde{w}_n|w_1^n)$ we have

$$\begin{aligned}
&b_n(\tilde{U}_n, \tilde{V}_n, \tilde{W}_n) \\
&\le \sum_{i=1}^n \Big(I(W_i; Y_{2i}), I(W_i; Y_{3i}), I(U_iW_i; Y_{2i}), I(U_i; Y_{2i}|W_i) \\
&\quad + I(W_i; Y_{3i}), I(V_i; Y_{3i}|W_i) + I(W_i; Y_{2i}), I(V_iW_i; Y_{3i}), \\
&\quad I(U_i; Y_{2i}|V_iW_i) + I(V_i; Y_{3i}|W_i) + I(W_i; Y_{2i}), \\
&\quad I(U_i; Y_{2i}|V_iW_i) + I(V_i; Y_{3i}|W_i) + I(W_i; Y_{3i}), \\
&\quad I(V_i; Y_{3i}|U_iW_i) + I(U_i; Y_{2i}|W_i) + I(W_i; Y_{2i}), \\
&\quad I(V_i; Y_{3i}|U_iW_i) + I(U_i; Y_{2i}|W_i) + I(W_i; Y_{3i})\Big).
\end{aligned}$$

where U_i, V_i, W_i are uniquely specified by $\tilde{U}_n, \tilde{V}_n, \tilde{W}_n$. Letting Q be uniform over $[n]$ and independent of all previously defined variables we can write

$$\begin{aligned}
&b_n(\tilde{U}_n, \tilde{V}_n, \tilde{W}_n) \\
&\le n \times \Big(I(W_Q; Y_{2Q}|Q), I(W_Q; Y_{3Q}|Q), \\
&\quad I(U_QW_Q; Y_{2Q}|Q), \\
&\quad I(U_Q; Y_{2Q}|W_QQ) + I(W_Q; Y_{3Q}|Q), \\
&\quad I(V_Q; Y_{3Q}|W_QQ) + I(W_Q; Y_{2Q}|Q), \\
&\quad I(V_QW_Q; Y_{3Q}|Q), \\
&\quad I(U_Q; Y_{2Q}|V_QW_QQ) + I(V_Q; Y_{3Q}|W_QQ) \\
&\qquad + I(W_Q; Y_{2Q}|Q), \\
&\quad I(U_Q; Y_{2Q}|V_QW_QQ) + I(V_Q; Y_{3Q}|W_QQ) \\
&\qquad + I(W_Q; Y_{3Q}|Q), \\
&\quad I(V_Q; Y_{3Q}|U_QW_QQ) + I(U_Q; Y_{2Q}|W_QQ) \\
&\qquad + I(W_Q; Y_{2Q}|Q), \\
&\quad I(V_Q; Y_{3Q}|U_QW_QQ) + I(U_Q; Y_{2Q}|W_QQ) \\
&\qquad + I(W_Q; Y_{3Q}|Q)\Big).
\end{aligned}$$

Letting $U = U_Q, V = V_Q, W = (W_Q, Q), X_1 = X_{1Q}, Y_2 = Y_{2Q}, Y_3 = Y_{3Q}$ one gets that

$$\begin{aligned}
&b_n(\tilde{U}_n, \tilde{V}_n, \tilde{W}_n) \\
&\le n \times \Big(I(W; Y_2), I(W; Y_3), \\
&\quad I(UW; Y_2|W), I(U; Y_2|W) + I(W; Y_3), \\
&\quad I(V; Y_3|W) + I(W; Y_2), I(VW; Y_3), \\
&\quad I(U; Y_2|VW) + I(V; Y_3|W) + I(W; Y_2), \\
&\quad I(U; Y_2|VW) + I(V; Y_3|W) + I(W; Y_3), \\
&\quad I(V; Y_3|UW) + I(U; Y_2|W) + I(W; Y_2), \\
&\quad I(V; Y_3|UW) + I(U; Y_2|W) + I(W; Y_3)\Big), \quad (4)
\end{aligned}$$

for some $p(uvwx_1)$ that depends on $\tilde{U}_n, \tilde{V}_n, \tilde{W}_n$, but satisfies the property that $p(x_1) = \frac{1}{n}\sum_{q=1}^n p(x_{1q})$ belongs to the convex hull of Ψ.

So far, we have basically shown that one can follow the proof of [26] while interpreting its steps with the concept of information state.

Next, at the decoding phase terminal j ($j = 2, 3$) computes $\hat{M}_j^n = \mathcal{D}_j(Y_j^n)$, and then discards Y_j^n. This changes the information state to the virtual channel $p(\hat{m}_2^n, \hat{m}_3^n|w_1^n)$ and the input distribution $p(w_1^n)$. The quantifier for this corresponds to the following for any given $p(\tilde{u}_n\tilde{v}_n\tilde{w}_n|w_1^n)$:

$$\begin{aligned}
&\Big(I(\tilde{W}_n; \hat{M}_2^n), I(\tilde{W}_n; \hat{M}_3^n), I(\tilde{W}_n\tilde{U}_n; \hat{M}_2^n), \\
&\quad I(\tilde{U}_n; \hat{M}_2^n|\tilde{W}_n) + I(\tilde{W}_n; \hat{M}_3^n), I(\tilde{V}_n; \hat{M}_3^n|\tilde{W}_n) \\
&\quad + I(\tilde{W}_n; \hat{M}_2^n), I(\tilde{V}_n\tilde{W}_n; \hat{M}_3^n), I(\tilde{W}_n; \hat{M}_2^n) + I(\tilde{V}_n; \hat{M}_3^n|\tilde{W}_n) \\
&\quad + I(\tilde{U}_n; \hat{M}_2^n|\tilde{V}_n\tilde{W}_n), I(\tilde{U}_n; \hat{M}_2^n|\tilde{V}_n\tilde{W}_n) + I(\tilde{V}_n\tilde{W}_n; \hat{M}_3^n), \\
&\quad I(\tilde{W}_n; \hat{M}_3^n) + I(\tilde{U}_n; \hat{M}_2^n|\tilde{W}_n) + I(\tilde{V}_n; \hat{M}_3^n|\tilde{U}_n\tilde{W}_n), \\
&\quad I(\tilde{V}_n; \hat{M}_3^n|\tilde{U}_n\tilde{W}_n) + I(\tilde{U}_n\tilde{W}_n; \hat{M}_2^n)\Big). \quad (5)
\end{aligned}$$

By the data processing inequality, the vector given in eq. (5) is coordinatewise less than or equal to $b_n(\tilde{U}_n, \tilde{V}_n, \tilde{W}_n)$, i.e. the vector given in eq. (3) for $i = n$. Let us choose

$$\tilde{U}_n = M_2^n, \tilde{V}_n = M_3^n, \tilde{W}_n = L^n \tag{6}$$

where L^n is created such that (M_2^n, M_3^n, L^n) is n i.i.d. repetitions from $p(m_2m_3l)$ and that $L^n \to M_2^n M_3^n \to X_1^n \to Y_2^n Y_3^n$ form a Markov chain. Then using equation (4) we have that

$$\begin{aligned}
&\Big(I(L^n; \hat{M}_2^n), I(L^n; \hat{M}_3^n), I(L^n M_2^n; \hat{M}_2^n),\\
&\quad I(M_2^n; \hat{M}_2^n|L^n) + I(L^n; \hat{M}_3^n),\\
&\quad I(M_3^n; \hat{M}_3^n|L^n) + I(L^n; \hat{M}_2^n),\\
&\quad I(M_3^n L^n; \hat{M}_3^n),\\
&I(L^n; \hat{M}_2^n) + I(M_3^n; \hat{M}_3^n|L^n) + I(M_2^n; \hat{M}_2^n|M_3^n L^n),\\
&\quad I(M_2^n; \hat{M}_2^n|M_3^n L^n) + I(M_3^n L^n; \hat{M}_3^n),\\
&I(L^n; \hat{M}_3^n) + I(M_2^n; \hat{M}_2^n|L^n) + I(M_3^n; \hat{M}_3^n|M_2^n L^n),\\
&\quad I(M_3^n; \hat{M}_3^n|M_2^n L^n) + I(M_2^n L^n; \hat{M}_2^n)\Big)\\
&\leq n \times \Big(I(W; Y_2), I(W; Y_3), I(U; Y_2|W) + I(W; Y_2),\\
&\quad I(U; Y_2|W) + I(W; Y_3), I(V; Y_3|W)\\
&\quad + I(W; Y_2), I(V; Y_3|W) + I(W; Y_3),\\
&\quad I(U; Y_2|VW) + I(V; Y_3|W) + I(W; Y_2),\\
&\quad I(U; Y_2|VW) + I(V; Y_3|W) + I(W; Y_3),\\
&\quad I(V; Y_3|UW) + I(U; Y_2|W) + I(W; Y_2),\\
&\quad I(V; Y_3|UW) + I(U; Y_2|W) + I(W; Y_3)\Big)
\end{aligned} \tag{7}$$

for some $p(uvwx_1)$, where $p(x_1)$ belongs to the convex hull of Ψ. As in the previous section, items 1 and 3 of Lemma 1, and Lemma 2, imply that the vector on the left hand side of equation (7) is coordinatewise greater than or equal to

$$\begin{aligned}
&\Big(I(L; \hat{M}_2), I(L; \hat{M}_3), I(LM_2; \hat{M}_2),\\
&\quad I(M_2; \hat{M}_2|L) + I(L; \hat{M}_3),\\
&\quad I(M_3; \hat{M}_3|L) + I(L; \hat{M}_2), I(L; \hat{M}_2) + I(M_3; \hat{M}_3|L)\\
&\ + I(M_2; \hat{M}_2|M_3L), I(M_2; \hat{M}_2|M_3L) + I(M_3L; \hat{M}_3),\\
&\quad I(L; \hat{M}_3) + I(M_2; \hat{M}_2|L) + I(M_3; \hat{M}_3|M_2L),\\
&\quad I(M_3; \hat{M}_3|M_2L) + I(M_2L; \hat{M}_2)\Big)
\end{aligned}$$

for some $\hat{M}_2, \hat{M}_3$ jointly distributed with M_2, M_3, L where the average distortion between M_j and $\hat{M}_j$ is less than or equal to D_j, $i = 2, 3$. This completes the proof of Theorem 2. □

Remark 7: A proof for [37, Th. 3] can be shown along the same lines as above if we use a more general choice of $\tilde{U}$ and $\tilde{V}$ in equation (6) as i.i.d. repetitions of two random variables, arbitrarily distributed with M_2, M_3 and L. The dependence of $p(\hat{m}_2, \hat{m}_3|m_2, m_3l)$ on L is also unnecessary to impose from the way $\hat{M}_2$ and $\hat{M}_3$ are constructed in the proof, as $\hat{M}_{2Q}$ and $\hat{M}_{3Q}$ for some random time index Q. The Markov chain $L^n \to M_2^n M_3^n \to X_1^n \to Y_2^n Y_3^n \to \hat{M}_2^n \hat{M}_3^n$ implies that $L^n \to M_2^n M_3^n \to \hat{M}_2^n \hat{M}_3^n$. Since (M_2^n, M_3^n, L^n) is i.i.d., we get that $L_q \to M_{2q} M_{3q} \to \hat{M}_2^n \hat{M}_3^n$ forms a Markov chain for any $q \in [n]$. Hence $L_Q \to M_{2Q} M_{3Q} \to \hat{M}_{2Q} \hat{M}_{3Q}$ also forms a Markov chain. Further the choice of $p(\hat{m}_{2Q}\hat{m}_{3Q}|m_{2Q}m_{3Q})$ depends only on the code, and not on $p(l|m_2m_3)$.

C. Lossy Transmission of Dependent Sources Over a Strong Interference Channel

Definition 6: An interference channel (IC) $q(y_3y_4|x_1x_2)$ is said to be strong *if*

$$I(X_1; Y_3|X_2) \leq I(X_1; Y_4|X_2); \tag{8}$$
$$I(X_2; Y_4|X_1) \leq I(X_2; Y_3|X_1), \tag{9}$$

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 $\mathcal{X}_1 \times \mathcal{X}_2$.

A permissible set of input distributions is taken to be a subset of joint distributions on X_1 and X_2 (the receivers do not have inputs to the channel, or rather their input alphabets are of size one). As in the traditional IC, the goal of receivers Y_3 and Y_4 is to recover $M_3 = W_1$ and $M_4 = W_2$ within distortions D_3 and D_4 respectively.

Theorem 3: Take an arbitrary strong interference channel $q(y_3y_4|x_1x_2)$, a permissible set of input distributions Ψ, and an admissible source marginal distribution $p(w_1w_2)$ with distortions D_3 and D_4 for $M_3 = W_1$ and $M_4 = W_2$. Then there must exist a joint distribution $p(\hat{m}_3\hat{m}_4|m_3m_4)$ such that

$$\mathbb{E}\big[\Delta_j(M_j, \hat{M}_j)\big] \leq D_j, \qquad i = 3, 4,$$

holds and for any random variable L where $p(lm_3m_4\hat{m}_3\hat{m}_4) = p(l)p(m_3|l)p(m_4|l)p(\hat{m}_3\hat{m}_4|m_3m_4)$, there must exist a joint pmf $p(ux_1x_2y_3y_4)$ such that

$p(x_1x_2)$ is in the convex hull of Ψ,

$$p(ux_1x_2y_3y_4) = p(u)p(x_1|u)p(x_2|u)q(y_3y_4|x_1x_2),$$

such the following inequalities hold:

$$\begin{aligned}
I(M_3; \hat{M}_3|M_4L) &\leq I(X_1; Y_3|X_2U);\\
I(M_4; \hat{M}_4|M_3L) &\leq I(X_2; Y_4|X_1U);\\
I(M_3; \hat{M}_3|M_4) &\leq I(X_1; Y_3|X_2);\\
I(M_4; \hat{M}_4|M_3) &\leq I(X_2; Y_4|X_1);\\
I(M_3M_4; \hat{M}_3\hat{M}_4) &\leq \min(I(X_1X_2; Y_3)\\
-H(M_3|\hat{M}_3) - H(M_4|\hat{M}_4) &\qquad , I(X_1X_2; Y_4));\\
I(M_3; \hat{M}_3|L) + I(M_4; \hat{M}_4|L) &\leq \min(I(X_1X_2; Y_3|U)\\
-H(M_3M_4|\hat{M}_3\hat{M}_4L) &\qquad , I(X_1X_2; Y_4|U)).
\end{aligned}$$

Remark 8: One can use the strengthened Carathéodory theorem of Fenchel to bound the cardinality of U from above by $|\mathcal{X}_1||\mathcal{X}_2| + 3$.

Corollary 3: Take an arbitrary strong interference channel $q(y_3y_4|x_1x_2)$, a permissible set of input distributions Ψ, and an admissible source marginal distribution $p(w_1w_2)$ with a vanishing probability of error. Then for any random *variable L where $M_3 \to L \to M_4$ holds, there must exist $p(ux_1x_2y_3y_4)$ such that*

$$p(x_1x_2) \text{ is in the convex hull of } \Psi,$$
$$p(ux_1x_2y_3y_4) = p(u)p(x_1|u)p(x_2|u)q(y_3y_4|x_1x_2),$$

such the following inequalities hold:

$$\begin{aligned} H(M_3|L) &\leq I(X_1;Y_3|X_2U);\\ H(M_4|L) &\leq I(X_2;Y_4|X_1U);\\ H(M_3|M_4) &\leq I(X_1;Y_3|X_2);\\ H(M_4|M_3) &\leq I(X_2;Y_4|X_1);\\ H(M_3M_4) &\leq \min(I(X_1X_2;Y_3), I(X_1X_2;Y_4));\\ H(M_3|L)+H(M_4|L) &\leq \min(I(X_1X_2;Y_3|U),\\ &\qquad I(X_1X_2;Y_4|U)). \end{aligned}$$

Remark 9: Intuitively speaking random variable L satisfying $W_1 \to L \to W_2$ for sources W_1 and W_2 inevitably shows up in the result because we are using the outer bound expression in the original result of [24] to define a quantifier for the information state, and because the outer bound expression contains an auxiliary random variable U satisfying $X_1 \to U \to X_2$ (thus roughly speaking, the same functional is applied for source coding and channel coding). The virtual channel corresponding to the communication task is $p(m_3m_4|w_1w_2)$ together with input distribution $p(w_1w_2)$. One has to apply the same *functional to the physical and virtual channels.*

Remark 10: The above theorem subsumes the converse part of the result of Maric, Yates and Kramer [24] as a special case, where the capacity region of a strong interference channels with common information is derived as the union over $p(ux_1x_2y_3y_4) = p(u)p(x_1|u)p(x_2|u)q(y_3y_4|x_1x_2)$, of the set of all non-negative triples (R_0, R_1, R_2) satisfying

$$\begin{aligned} R_1 &\leq I(X_1;Y_3|X_2U);\\ R_2 &\leq I(X_2;Y_4|X_1U);\\ R_1+R_2 &\leq \min(I(X_1X_2;Y_3|U), I(X_1X_2;Y_4|U));\\ R_0+R_1+R_2 &\leq \min(I(X_1X_2;Y_3), I(X_1X_2;Y_4)). \end{aligned} \tag{10}$$

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. We are not aware of any previous work discussing any interesting outer bounds on the admissible source region of an interference channel with general dependent sources. However inner bounds for this problem are reported in [18] and [30]. Since multi-access channels (MACs) are special cases of strong interference channels and 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 also a difficult problem. It is known that the source−channel separation theorem breaks down for multi-access channels [8].

Proof of Theorem 3: Since there is only one pair source (W_1^n, W_2^n), whose components are respectively observed by the two transmitters, and the information of the third and fourth terminals increases over time, we may identify the information state at stage i as the virtual channel $p(y_3^i, y_4^i|w_1^n, w_2^n)$ and the input distribution $p(w_1^n, w_2^n)$.

Based on the statement of the theorem, given $q(y_3y_4|x_1x_2)$ and $q(x_1x_2)$, the information state can be quantified by any of the following expressions showing up on the right hand side of the equations for any arbitrary $p(u|x_1x_2)$ where $p(ux_1x_2)$ factorizes as $p(u)p(x_1|u)p(x_2|u)$:

$$\Big(I(X_1;Y_3|X_2U), I(X_2;Y_4|X_1U), I(X_1;Y_3|X_2), I(X_2;Y_4|X_1), I(X_1X_2;Y_3|U), I(X_1X_2;Y_4|U), I(X_1X_2;Y_3), I(X_1X_2;Y_4)\Big).$$

This gives the following vector of quantifiers for the information state at stage i given any $p(\tilde{u}_i|w_1^nw_2^n)$ under which W_1^n and W_2^n are conditionally independent given $\tilde{U}_i$:

$$\Big(I(W_1^n;Y_3^n|W_2^n\tilde{U}_i), I(W_2^n;Y_4^n|W_1^n\tilde{U}_i), I(W_1^n;Y_3^n|W_2^n), I(W_2^n;Y_4^n|W_1^n), I(W_1^nW_2^n;Y_3^n|\tilde{U}_i), I(W_1^nW_2^n;Y_4^n|\tilde{U}_i), I(W_1^nW_2^n;Y_3^n), I(W_1^nW_2^n;Y_4^n)\Big). \tag{11}$$

Let us denote the above vector by $b_i(\tilde{U}_i)$ where with a slight abuse of notation, we have used the variable $\tilde{U}_i$ instead of $p(\tilde{u}_i|w_1^nw_2^n)$.

Clearly all of these quantifiers are zero when $i = 0$, i.e. $b_0(\tilde{U}_0)$ is always zero. Furthermore, to compare the quantifiers at stage i and $i-1$, we have that

$$\begin{aligned} b_i(\tilde{U}_i) &\leq b_{i-1}(\tilde{U}_{i-1})\\ &+\Big(I(X_{1i};Y_{3i}|X_{2i}U_i), I(X_{2i};Y_{4i}|X_{1i}U_i), I(X_{1i};Y_{3i}|X_{2i}),\\ &\quad I(X_{2i};Y_{4i}|X_{1i}), I(X_{1i}X_{2i};Y_{3i}|U_i),\\ &\quad I(X_{1i}X_{2i};Y_{4i}|U_i), I(X_{1i}X_{2i};Y_{3i}), I(X_{1i}X_{2i};Y_{4i})\Big), \end{aligned}$$

where $\leq$ is coordinatewise and $\tilde{U}_i = \tilde{U}_{i-1} = U_i$. Further the Markov chain $W_1^n - \tilde{U}_i - W_2^n$ implies $W_1^n - \tilde{U}_{i-1} - W_2^n$ and $X_{1i} - \tilde{U}_i - X_{2i}$. In other words, having $\tilde{U}_i$, we can find a valid $\tilde{U}_{i-1}$ and U_i in terms of it.

Thus, for any $p(\tilde{u}_n|w_1^nw_2^n)$ under which W_1^n and W_2^n are conditionally independent given $\tilde{U}_i$, we have

$$\begin{aligned} b_n(\tilde{U}_n) \leq \sum_{i=1}^{n}\Big(&I(X_{1i};Y_{3i}|X_{2i}U_i), I(X_{2i};Y_{4i}|X_{1i}U_i),\\ &I(X_{1i};Y_{3i}|X_{2i}), I(X_{2i};Y_{4i}|X_{1i}),\\ &I(X_{1i}X_{2i};Y_{3i}|U_i), I(X_{1i}X_{2i};Y_{4i}|U_i),\\ &I(X_{1i}X_{2i};Y_{3i}), I(X_{1i}X_{2i};Y_{4i})\Big). \end{aligned}$$

where U_i are uniquely specified by $\tilde{U}_n$. Taking a random variable Q that is uniform over $[n]$ and independent of all previously defined variables, we can follow similar steps as in the previous section to show that

$$b_n(\tilde{U}_n) \leq n \times \Big(I(X_1;Y_3|X_2U),\, I(X_2;Y_4|X_1U),\, I(X_1;Y_3|X_2),\, I(X_2;Y_4|X_1),\, I(X_1X_2;Y_3|U),\, I(X_1X_2;Y_4|U),\, I(X_1X_2;Y_3),\, I(X_1X_2;Y_4) \Big), \tag{12}$$

for some $p(ux_1x_2)$ that depends on $\tilde{U}_n$, but satisfies the property that $p(x_1x_2) = \frac{1}{n}\sum_{q=1}^{n} p(x_{1q}x_{2q})$ belongs to the convex hull of Ψ.

Next, we proceed as in the previous section except with a change. At the decoding phase terminal j ($j = 3, 4$) computes $\hat{M}_j^n = \mathcal{D}_j(Y_j^n)$, and then discards Y_j^n. This changes the information state to the virtual channel $p(\hat{m}_3^n\hat{m}_4^n|w_1^nw_2^n)$ and the input distribution $p(w_1^nw_2^n)$. The quantifier for this corresponds to the following for any given $p(\tilde{u}_n|w_1^nw_2^n)$ under which W_1^n and W_2^n are conditionally independent given $\tilde{U}_i$:

$$\Big(I(W_1^n;\hat{M}_3^n|W_2^n\tilde{U}_n),\, I(W_2^n;\hat{M}_4^n|W_1^n\tilde{U}_n),\, I(W_1^n;\hat{M}_3^n|W_2^n),\, I(W_2^n;\hat{M}_4^n|W_1^n),\, I(W_1^nW_2^n;\hat{M}_3^n|\tilde{U}_n),\, I(W_1^nW_2^n;\hat{M}_4^n|\tilde{U}_n),\, I(W_1^nW_2^n;\hat{M}_3^n),\, I(W_1^nW_2^n;\hat{M}_4^n) \Big). \tag{13}$$

By data processing the vector given in eq. (13) is coordinatewise less than or equal to $b_n(\tilde{U}_n)$, i.e. the vector given in eq. (11) for $i = n$. However we know that in a strong interference channel, each receiver may obtain significant side information about the message intended for the other receiver. Therefore, instead of bounding all of the coordinates of $b_n(\tilde{U}_n)$ by the vector given in eq. (13), we can use Lemma 3 in the appendix to find a better lower bound for the last four coordinates. Therefore, noting that $W_1 = M_3$ and $W_2 = M_4$ we can write

$$\begin{aligned} b_n(\tilde{U}_n) \geq \Big(& I(M_3^n;\hat{M}_3^n|M_4^n\tilde{U}_n),\, I(M_4^n;\hat{M}_4^n|M_3^n\tilde{U}_n), \\ & I(M_3^n;\hat{M}_3^n|M_4^n), I(M_4^n;\hat{M}_4^n|M_3^n), \\ & I(M_3^nM_4^n;\hat{M}_3^n\hat{M}_4^n|\tilde{U}_n) - H(M_4^n|\hat{M}_4^n\tilde{U}_n) \\ & \quad - H(M_3^n|\hat{M}_3^n\tilde{U}_n), I(M_3^nM_4^n;\hat{M}_3^n\hat{M}_4^n|\tilde{U}_n) \\ & \quad - H(M_4^n|\hat{M}_4^n\tilde{U}_n) - H(M_3^n|\hat{M}_3^n\tilde{U}_n), \\ & I(M_3^nM_4^n;\hat{M}_3^n\hat{M}_4^n) - H(M_4^n|\hat{M}_4^n) - H(M_3^n|\hat{M}_3^n), \\ & I(M_3^nM_4^n;\hat{M}_3^n\hat{M}_4^n) - H(M_4^n|\hat{M}_4^n) - H(M_3^n|\hat{M}_3^n) \Big). \end{aligned} \tag{14}$$

Let us choose $\tilde{U}_n = L^n$, where L^n is created such that (M_3^n, M_4^n, L^n) is n i.i.d. repetitions from $p(m_3m_4l)$ given in the statement of the theorem and that $L^n \to M_3^nM_4^n \to X_1^nX_2^n \to Y_3^nY_4^n$ form a Markov chain. This implies that $M_3^n \to L^n \to M_4^n$ forms a Markov chain. Then using eqs. (12) and (14) we have that

$$\begin{aligned} \Big(& I(M_3^n;\hat{M}_3^n|M_4^nL^n),\, I(M_4^n;\hat{M}_4^n|M_3^nL^n), \\ & I(M_3^n;\hat{M}_3^n|M_4^n),\, I(M_4^n;\hat{M}_4^n|M_3^n), \\ & I(M_3^nM_4^n;\hat{M}_3^n\hat{M}_4^n|L^n) - H(M_4^n|\hat{M}_4^nL^n) - H(M_3^n|\hat{M}_3^nL^n), \\ & I(M_3^nM_4^n;\hat{M}_3^n\hat{M}_4^n|L^n) - H(M_4^n|\hat{M}_4^nL^n) - H(M_3^n|\hat{M}_3^nL^n), \\ & I(M_3^nM_4^n;\hat{M}_3^n\hat{M}_4^n) - H(M_4^n|\hat{M}_4^n) - H(M_3^n|\hat{M}_3^n), \\ & I(M_3^nM_4^n;\hat{M}_3^n\hat{M}_4^n) - H(M_4^n|\hat{M}_4^n) - H(M_3^n|\hat{M}_3^n) \Big) \\ \leq n \times \Big(& I(X_1;Y_3|X_2U),\, I(X_2;Y_4|X_1U), \\ & I(X_1;Y_3|X_2),\, I(X_2;Y_4|X_1),\, I(X_1X_2;Y_3|U), \\ & I(X_1X_2;Y_4|U),\, I(X_1X_2;Y_3),\, I(X_1X_2;Y_4) \Big), \end{aligned} \tag{15}$$

for some $p(ux_1x_2) = p(u)p(x_1|u)p(x_2|u)$ where $p(x_1x_2) =$ belongs to the convex hull of Ψ. As in the previous section, Lemmas 1 and 2 imply that the vector on the left hand side of equation (15) is coordinatewise greater than or equal to

$$\begin{aligned} \Big(& I(M_3;\hat{M}_3|M_4L),\, I(M_4;\hat{M}_4|M_3L), \\ & I(M_3;\hat{M}_3|M_4),\, I(M_4;\hat{M}_4|M_3), \\ & I(M_3M_4;\hat{M}_3\hat{M}_4|L) - H(M_4|\hat{M}_4L) - H(M_3|\hat{M}_3L), \\ & I(M_3M_4;\hat{M}_3\hat{M}_4|L) - H(M_4|\hat{M}_4L) - H(M_3|\hat{M}_3L), \\ & I(M_3M_4;\hat{M}_3\hat{M}_4) - H(M_4|\hat{M}_4) - H(M_3|\hat{M}_3), \\ & I(M_3M_4;\hat{M}_3\hat{M}_4) - H(M_4|\hat{M}_4) - H(M_3|\hat{M}_3) \Big), \end{aligned}$$

where $(\hat{M}_3, \hat{M}_4, M_3, M_4, L) = (\hat{M}_{3Q}, \hat{M}_{4Q}, M_{3Q}, M_{4Q}, L_Q)$ where Q is a uniform random index in $[n]$. Using the same argument as the one given in Remark 7 one can see that the joint pmf of $\hat{M}_3, \hat{M}_4$ and M_3, M_4, L satisfies $L \to M_3M_4 \to \hat{M}_3, \hat{M}_4$, and $p(\hat{m}_3\hat{m}_4|m_3m_4)$ depends only on the code and not on $p(l|m_3, m_4)$. Further the average distortion between M_j and $\hat{M}_j$ is less than or equal to D_j, $i = 3, 4$. This completes the proof of Theorem 3. □

V. Concluding Remarks: Information State, Auxiliary Random Variables and Additivity

The notion of information state was originally motivated by the *additivity* property of capacity regions. Further, the particular structure of auxiliary random variables can be understood as a consequence of additivity of an outer bound expression, when viewed as a functional of the underlying communication channel. This perspective is generally implicit in the textbooks on classical information theory, where a communication channel is kept fixed throughout. The literature on quantum information theory is more explcit about treating the communication channel as a variable (e.g. see [2]), although the connection between additivity and auxiliary random variables is generally implicit.

The idea of additivity when there is no feedback or dependent input sources is simple: a certain functional constitutes the capacity region, or an outer bound to it, if it is additive for product channels. For instance consider the general broadcast channel $q(yz|x)$ and let a convex region $\mathcal{R}\big(q(yz|x)\big)$ be a candidate outer bound. We use the tensor product $q^{\otimes n}$ to denote the product of n i.i.d. repetitions of the broadcast channel. The region $\mathcal{R}\big(q(yz|x)\big)$ would then constitute an outer bound to the capacity region if on the one hand $\bigcup_n \frac{1}{n}\mathcal{R}\big(q^{\otimes n}(y^n z^n|x^n)\big)$ is an outer bound, and on the other hand $\mathcal{R}\big(q^{\otimes n}(y^n z^n|x^n)\big)$ is equal to $n\mathcal{R}\big(q(yz|x)\big)$ for all n. The latter statement is the single-letterization step. It holds if one can inductively show that

$$\mathcal{R}\big(q^{\otimes n}(y^n z^n|x^n)\big) \subseteq \mathcal{R}\big(q^{\otimes(n-1)}(y^{n-1}z^{n-1}|x^{n-1})\big) + \mathcal{R}\big(q(y_n z_n|x_n)\big), \quad (16)$$

where $+$ stands for the vector-by-vector sum (Minkowski sum) of the regions. The latter holds if for any two channels $q(yz|x)$ and $q(\tilde{y}\tilde{z}|\tilde{x})$ one has

$$\mathcal{R}\big(q(yz|x)q(\tilde{y}\tilde{z}|\tilde{x})\big) \subseteq \mathcal{R}\big(q(yz|x)\big) + \mathcal{R}\big(q(\tilde{y}\tilde{z}|\tilde{x})\big), \quad (17)$$

because one can think of $q(\tilde{y}\tilde{z}|\tilde{x})$ being the product of $n-1$ copies of $q(yz|x)$, and $q(yz|x)q(\tilde{y}\tilde{z}|\tilde{x})$ being the product of n copies of $q(yz|x)$. This is the additivity relation. Note that eq. (17) is stated only in terms of six random variables, namely X, Y, Z, $\tilde{X}$, $\tilde{Y}$ and $\tilde{Z}$; it is only a property of joint distributions on six random variables. If one is only interested in the type of auxiliary random variables showing up in the traditional Gallager-type converses, there is an algorithmic way to deal with this problem where a computer program could be written to generate the auxiliary random variables. The idea is that each of the auxiliary random variables in the definition of $\mathcal{R}$ should take the form of a combination of some of these six random variables. In other words, in eq. (16) auxiliary variables for $\mathcal{R}\big(q^{\otimes(n-1)}(y^{n-1}z^{n-1}|x^{n-1})\big)$ and $\mathcal{R}\big(q(y_n z_n|x_n)\big)$ may be expressed in terms of those of $\mathcal{R}_n\big(q(yz|x)\big)$ and of Y^{n-1}, Z^{n-1}, X^{n-1}, Y_n, Z_n, X_n. Should we then unfold the inductive n iterations we get back the Gallager-type identification of auxiliary random variables having the form of past and/or future of other variables.

One can recover the Gallager-type results using the inductive additivity approach. On the other hand the additivity framework has been used so far to disprove the optimality of a given expression by explicitly showing that it is not additive (see [3]). However there is evidence suggesting that the inductive single-letterization technique may outperform the traditional approach. Consider the sum rate for Marton's inner bound for the general broadcast channel. It is not known whether the n-letter expression for Marton's sum rate factorizes. However, it is shown in [12] that the 2-letter expression factorizes into two single letter expressions in *a non-traditional way* in the so-called "randomized time-division strategy" region. As shown in [12, p. 12], the choice of U and V in the factorization depend on the value of the auxiliary w! If the two-letter expression were to factorize in the traditional Gallager-type way (which it doesn't), one would have expected U to consist of the past/future of certain random variables. However in this example, this past/future structure itself is changing depending on the very value of these auxiliary random variables. Therefore one should not always expect the auxiliary random variables to be composed of the past/future of certain random variables, which one can get using standard manipulations such as the chain rule on n-letter expressions. The traditional technique for assigning single-letter auxiliary random variables is thus insufficient. It is also worthwhile to note that non-Shannon-type inequalities are necessary for finding the capacity region of wired networks [9]. Wired networks are special cases of general noisy networks (see [22]). However, such inequalities are not easy to employ in single-letterization when we directly start from a code of length n. It would be simpler to employ non-Shannon-type inequalities when one is dealing with a finite number of random variables showing up in each iteration of the algorithm (the number of which is not scaling up with n).

Another point to note is that having a functional that satisfies eq. (17), one can attempt to perturb it to find another functional that satisfies this equation. This trial and error process is easier to run for eq. (17), compared to when we start from an n-code and directly identify the auxiliary random variables through a longer chain of equations. This was done in deriving Theorem 5 of the authors' earlier work [14] and [15].

The notion of additivity for cases when there is feedback or dependent sources is not as straightforward and has not been studied systematically in the literature, even though certain forms of it have appeared in the literature for certain special problems, e.g. "secret key reservoir" [14], "monotones" [33], [34], and a paper on the feedback capacity of wiretap channels [4]. See [23]. The reason that formulating additivity can be challenging is that output feedback causes the information available to the parties to evolve in a wide range of ways. To be more specific, take a channel $q(y|x)$ assisted with full feedback. Suppose we would like to show that any achievable communication rate is less than or equal to $\max_{p(x)} I(X;Y)$. Note that the n-letter $q(y_1y_2\ldots y_n|x_1x_2\ldots x_n)$ would not factorize as $\prod_{i=1}^n q(y_i|x_i)$. This is because Y_i can depend on X_i, X_{i+1}, $X_{i+2}, \ldots, X_n$. Therefore we cannot talk about additivity and factorization in the sense of eq. (17). Furthermore, the single-letter definition of $\mathcal{R}$ does *not* involve any auxiliary random variable reflecting the feedback information. Another difference is the fact that feedback is imposing a particular *time order* on indices $1, 2, 3, \ldots, n$.

We believe that our notion of information state can be helpful in formalizing additivity in general scenarios. In the above point-to-point example there is a non-symmetric time flow that can be captured by the notion of information state. We believe that our construction of a virtual channel and applying the original outer bound expression to it, is novel (see remarks 1, 9 and 4). Our idea that adapting known lossless results to lossy ones with dependent sources basically comes for free, to best of our knowledge, is also novel.

Appendix

Lemma 3: For any n-code for a strong interference channel $q(y_3y_4|x_1x_2)$ with $M_3 = W_1$ and $M_4 = W_2$ and for any

$\tilde{U}_n - W_1^n W_2^n - Y_3^n Y_4^n$ *we have:*

$$I(M_3^n M_4^n; Y_3^n) \geq I(M_3^n M_4^n; \hat{M}_3^n \hat{M}_4^n) - H(M_4^n|\hat{M}_4^n) - H(M_3^n|\hat{M}_3^n),$$
$$I(M_3^n M_4^n; Y_4^n) \geq I(M_3^n M_4^n; \hat{M}_3^n \hat{M}_4^n) - H(M_4^n|\hat{M}_4^n) - H(M_3^n|\hat{M}_3^n),$$
$$I(M_3^n M_4^n; Y_3^n|\tilde{U}_n) \geq I(M_3^n M_4^n; \hat{M}_3^n \hat{M}_4^n|\tilde{U}_n) - H(M_4^n|\hat{M}_4^n \tilde{U}_n) - H(M_3^n|\hat{M}_3^n \tilde{U}_n),$$
$$I(M_3^n M_4^n; Y_4^n|\tilde{U}_n) \geq I(M_3^n M_4^n; \hat{M}_3^n \hat{M}_4^n|\tilde{U}_n) - H(M_4^n|\hat{M}_4^n \tilde{U}_n) - H(M_3^n|\hat{M}_3^n \tilde{U}_n).$$

Proof: The above constraints are equivalent to

$$H(M_3^n M_4^n|Y_3^n) \leq H(M_3^n M_4^n|\hat{M}_3^n \hat{M}_4^n) + H(M_4^n|\hat{M}_4^n) + H(M_3^n|\hat{M}_3^n),$$
$$H(M_3^n M_4^n|Y_4^n) \leq H(M_3^n M_4^n|\hat{M}_3^n \hat{M}_4^n) + H(M_4^n|\hat{M}_4^n) + H(M_3^n|\hat{M}_3^n),$$
$$H(M_3^n M_4^n|Y_3^n \tilde{U}_n) \leq H(M_3^n M_4^n|\hat{M}_3^n \hat{M}_4^n \tilde{U}_n) + H(M_4^n|\hat{M}_4^n \tilde{U}_n) + H(M_3^n|\hat{M}_3^n \tilde{U}_n),$$
$$H(M_3^n M_4^n|Y_4^n \tilde{U}_n) \leq H(M_3^n M_4^n|\hat{M}_3^n \hat{M}_4^n \tilde{U}_n) + H(M_4^n|\hat{M}_4^n \tilde{U}_n) + H(M_3^n|\hat{M}_3^n \tilde{U}_n).$$

By symmetry we only need to prove the first and third equations; the proof for the second equation is similar. Since $q(\mathbf{y}|\mathbf{x})$ is a strong interference channel, the lemma of Section III of [6] implies that the n-letter product channel

$$q(y_3^n y_4^n|x_1^n x_2^n) = \prod_{i=1}^n q(y_{3i} y_{4i}|x_{1i} x_{2i})$$

is also a strong interference channel. Therefore $I(X_2^n; Y_3^n|X_1^n) \geq I(X_2^n; Y_4^n|X_1^n)$ since equations (8) and (9) hold for *any* arbitrary distributions on the inputs of a strong interference channel. This implies that $H(X_2^n|Y_3^n X_1^n) \leq H(X_2^n|Y_4^n X_1^n)$.

We have $H(X_2^n|Y_4^n X_1^n) \leq H(M_4^n|Y_4^n X_1^n) \leq H(M_4^n|\hat{M}_4^n)$ since $H(X_2^n|M_4^n) = H(X_2^n|W_2^n) = 0$. Here $\hat{M}_4^n$ is the reconstruction of M_4^n by the third terminal (the average distortion between $\hat{M}_4^n$ and M_4^n is less than or equal to D_4). Therefore $H(X_2^n|Y_3^n X_1^n) \leq H(M_4^n|\hat{M}_4^n)$.

We further have $H(X_1^n|Y_3^n) \leq H(M_3^n|Y_3^n) \leq H(M_3^n|\hat{M}_3^n)$ since X_1^n is a function of $W_1^n = M_3^n$. Thus,

$$H(X_1^n X_2^n|Y_3^n) \leq H(M_4^n|\hat{M}_4^n) + H(M_3^n|\hat{M}_3^n).$$

Next, note that

$$H(M_3^n M_4^n|Y_3^n) \leq H(M_3^n M_4^n|X_1^n X_2^n) + H(X_1^n X_2^n|Y_3^n),$$

since for any three random variables A, B and C we have $H(A|C) \leq H(A|B) + H(B|C)$. We have

$$H(M_3^n M_4^n|X_1^n X_2^n) = H(M_3^n M_4^n|X_1^n X_2^n \hat{M}_3^n \hat{M}_4^n) \leq H(M_3^n M_4^n|\hat{M}_3^n \hat{M}_4^n).$$

Therefore

$$H(M_3^n M_4^n|Y_3^n) \leq H(M_3^n M_4^n|\hat{M}_3^n \hat{M}_4^n) + H(M_4^n|\hat{M}_4^n) + H(M_3^n|\hat{M}_3^n).$$

The proof for the third inequality is exactly the same, except that we have to condition all the entropy terms on $\tilde{U}_n$. In particular the equation $I(X_2^n; Y_3^n|X_1^n \tilde{U}_n) \geq I(X_2^n; Y_4^n|X_1^n \tilde{U}_n)$ holds since it is valid for any value of $\tilde{U}_n = \tilde{u}_n$. All other equations hold for exactly same reasons. □

ACKNOWLEDGEMENT

The authors would like to thank the anonymous reviewers for their helpful comments.

REFERENCES

[1] A. El Gamal and Y.-H. Kim, *Network Information Theory*. Cambridge, U.K.: Cambridge Univ. Press, 2011.

[2] A. S. Holevo, "The additivity problem in quantum information theory," in *Proc. Int. Congr. Math.*, Madrid, Spain, 2006, pp. 999–1018.

[3] M. B. Hastings, "Superadditivity of communication capacity using entangled inputs," *Nature Phys.*, vol. 5, pp. 255–257, Apr. 2009.

[4] E. Ardestanizadeh, M. Franceschetti, T. Javidi, and Y.-H. Kim, "Wiretap channel with secure rate-limited feedback," *IEEE Trans. Inf. Theory*, vol. 55, no. 12, pp. 5353–5361, Dec. 2009.

[5] Y. Cao, B. Chen, and J. Zhang, "A new achievable rate region for interference channels with common information," in *Proc. IEEE Wireless Commun. Network. Conf.*, Mar. 2007, pp. 2069–2073.

[6] M. H. M. Costa and A. El Gamal, "The capacity region of the discrete memoryless interference channel with strong interference (Corresp.)," *IEEE Trans. Inf. Theory*, vol. 33, no. 5, pp. 710–711, Sep. 1987.

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

[8] T. Cover, A. El Gamal, and M. Salehi, "Multiple access channels with arbitrarily correlated sources," *IEEE Trans. Inf. Theory*, vol. 26, no. 6, pp. 648–657, Nov. 1980.

[9] R. Dougherty, C. Freiling, and K. Zeger, "Networks, matroids, and non-Shannon information inequalities," *IEEE Trans. Inf. Theory*, vol. 53, no. 6, pp. 1949–1969, Jun. 2007.

[10] P. Gács and J. Körner, "Common information is far less than mutual information," *Problems Control Inf. Theory*, vol. 150, no. 2, pp. 149–162, 1973.

[11] M. Gastpar, "Cut-set arguments for source-channel networks," in *Proc. Int. Symp. Inf. Theory (ISIT)*, Jun./Jul. 2004, p. 34.

[12] Y. Geng, A. Gohari, C. Nair, and Y. Yu, "On Marton's inner bound for two receiver broadcast channels," in *Proc. ITA Workshop*, 2011. [Online]. Available: http://ita.ucsd.edu/workshop/11/files/paper/paper_688.pdf

[13] A. Giridhar and P. R. Kumar, "Computing and communicating functions over sensor networks," *IEEE J. Sel. Areas Commun.*, vol. 23, no. 4, pp. 755–764, Apr. 2005.

[14] A. A. Gohari and V. Anantharam, "Information-theoretic key agreement of multiple terminals—Part I: Source model," *IEEE Trans. Inf. Theory*, vol. 56, no. 8, pp. 3973–3996, Aug. 2010.

[15] A. A. Gohari and V. Anantharam, "Information-theoretic key agreement of multiple terminals—Part II: Channel model," *IEEE Trans. Inf. Theory*, vol. 56, no. 8, pp. 3997–4010, Aug. 2010.

[16] A. Gohari and V. Anantharam, "An outer bound to the admissible source region of broadcast channels with arbitrarily correlated sources and channel variations," in *Proc. 46th Annu. Allerton Conf. Commun., Control, Comput.*, Urbana, IL, USA, Sep. 2008, pp. 301–308.

[17] A. A. Gohari and V. Anantharam, "A generalized cut-set bound," in *Proc. IEEE Int. Symp. Inf. Theory (ISIT)*, Seoul, Korea, Jun./Jul. 2009, pp. 99–103.

[18] D. Gündüz, E. Erkip, A. Goldsmith, and H. V. Poor, "Source and channel coding for correlated sources over multiuser channels," *IEEE Trans. Inf. Theory*, vol. 55, no. 9, pp. 3927–3944, Sep. 2009.

[19] W. Kang and S. Ulukus, "A new data processing inequality and its applications in distributed source and channel coding," *IEEE Trans. Inf. Theory*, vol. 57, no. 1, pp. 56–69, Jan. 2011.

[20] G. Kramer and S. A. Savari, "Cut sets and information flow in networks of two-way channels," in *Proc. Int. Symp. Inf. Theory (ISIT)*, Jun./Jul. 2004, p. 33.

[21] G. Kramer, Y. Liang, and S. Shamai, "Outer bounds on the admissible source region for broadcast channels with dependent sources," in *Proc. Inf. Theory Appl. Workshop*, Feb. 2009, pp. 162–179.

[22] S. Lim, Y.-H. Kim, A. El Gamal, and S.-Y. Chung, "Noisy network coding," *IEEE Trans. Inf. Theory*, vol. 57, no. 5, pp. 3132–3152, May 2011.

[23] N. Ma and P. Ishwar, "Infinite-message distributed source coding for two-terminal interactive computing," *in Proc. 47th Annual Allerton Conf. Commun., Control, and Comput.*, Monticello, IL, USA, 2009, pp. 1510–1517.
[24] I. Maric, R. Yates, and G. Kramer, "The capacity region of the strong interference channel with common information," in *Proc. Asilomar Conf. Signals, Syst., Comput.*, Pacific Grove, CA, USA, 2005, pp. 1737–1741.
[25] P. Minero and Y.-H. Kim, "Correlated sources over broadcast channels," in *Proc. IEEE Int. Symp. Inf. Theory (ISIT)*, Seoul, Korea, Jun./Jul. 2009, pp. 2780–2784.
[26] C. Nair. *A Note on Outer Bounds for Broadcast Channel*. [Online]. Available: http://arxiv.org/abs/1101.0640, accessed Aug. 22, 2014.
[27] C. Nair and A. El Gamal, "An outer bound to the capacity region of the broadcast channel," *IEEE Trans. Inf. Theory*, vol. 53, no. 1, pp. 350–355, Jan. 2007.
[28] B. Nazer and M. Gastpar, "Computation over multiple-access channels," *IEEE Trans. Inf. Theory*, vol. 53, no. 10, pp. 3498–3516, Oct. 2007.
[29] A. Orlitsky and J. R. Roche, "Coding for computing," *IEEE Trans. Inf. Theory*, vol. 47, no. 3, pp. 903–917, Mar. 2001.
[30] S. S. Pradhan, S. Choi, and K. Ramchandran, "A graph-based framework for transmission of correlated sources over multiple-access channels," *IEEE Trans. Inf. Theory*, vol. 53, no. 12, pp. 4583–4604, Dec. 2007.
[31] T. Han and M. H. M. Costa, "Broadcast channels with arbitrarily correlated sources," *IEEE Trans. Inf. Theory*, vol. 33, no. 5, pp. 641–650, Sep. 1987.
[32] H. Yamamoto, "Wyner–Ziv theory for a general function of the correlated sources (Corresp.)," *IEEE Trans. Inf. Theory*, vol. 28, no. 5, pp. 803–807, Sep. 1982.
[33] S. Wolf and J. Wullschleger, "New monotones and lower bounds in unconditional two-party computation," *IEEE Trans. Inf. Theory*, vol. 54, no. 6, pp. 2792–2797, Jun. 2008.
[34] V. Prabhakaran and M. Prabhakaran, "Assisted common information with applications to secure two-party computation," in *Proc. IEEE Int. Symp. Inf. Theory (ISIT)*, Feb. 2010, pp. 2602–2606.
[35] G. Kramer and C. Nair, "Comments on 'broadcast channels with arbitrarily correlated sources'," in *Proc. IEEE Int. Symp. Inf. Theory*, Jun./Jul. 2009, pp. 2777–2779.
[36] S. Kamath and V. Anantharam, "Non-interactive simulation of joint distributions: The Hirschfeld–Gebelein–Rényi maximal correlation and the hypercontractivity ribbon," in *Proc. 50th Annu. Allerton Conf. Commun., Control, Comput.*, Monticello, IL, USA, Oct. 2012, pp. 1057–1064.
[37] K. Khezeli and J. Chen, "An improved outer bound on the admissible source region for broadcast channels with correlated sources," in *Proc. IEEE Int. Symp. Inf. Theory*, Jun./Jul. 2014, pp. 466–470.

Amin Gohari is an Assistant Professor at Sharif University of Technology, Tehran, Iran. Dr. Gohari received his M.Sc. and Ph.D. degree in electrical engineering in 2010 from the University of California, Berkeley, and his B.Sc. degree in 2004 from Sharif University of Technology, Iran. He received the 2010 Eli Jury Award from UC Berkeley, Department of Electrical Engineering, for "outstanding achievement in the area of communication networks," and the 2009-2010 Bernard Friedman Memorial Prize in Applied Mathematics from UC Berkeley, Department of Mathematics, for "demonstrated ability to do research in applied mathematics." He also received the Gold Medal from the 41st International Mathematical Olympiad (IMO 2000) and the First Prize from the 9th International Mathematical Competition for University Students (IMC 2002).

Venkat Anantharam is on the faculty of the EECS Department at the University of California at Berkeley. He received the Philips India Medal and the President of India Gold Medal from IIT Madras in 1980 and an NSF Presidential Young Investigator award in 1988. He is a co-recipient of the 1998 Prize Paper Award of the IEEE Information Theory Society, and a co-recipient of the 2000 Stephen O. Rice Prize Paper Award of the IEEE Communications Theory Society. He received the Distinguished Alumnus Award from IIT Madras in 2008.