
TELECOM PARISTECH

Thesis Midterm Report

# Computing a Function of Correlated Sources

*By:*
Milad SEFIDGARAN

*Supervisor:*
Dr. Aslan TCHAMKERTEN

March 2011

## I. Problem Statement

Assume a sensor network with a directed tree configuration as in Figure 1(a), constituted of sensors that have local measurements and a root who wants to compute a function of these measurements. What is the minimum number of bits needed to be communicated from these sensors toward the root within the defined configuration? Obviously this amount may be less than the case where the root wants to recover all the measurements, but what is the minimum amount of bits to be transmitted for enabling the reliable function computation at root, given to the function, joint probability distribution of measurements and the node configuration?

In this thesis we consider the problem of finding the minimum number of bits needed to be transmitted for recovering a function of correlated sources with high probability. We consider different configurations. One we considered is the Slepian-Wolf configuration (Figure 1(b)) which is one of the sub-configurations that can be seen in networks, as in Figure 1(a). We have derived inner and outer bounds for the union region of rate[1] pairs that lead to reliable function computation at receiver in this configuration which depend on the function and joint probability distribution of random variables and we have shown that the achievable rate region is tight in some special cases.

This report is organized as following. In Section II, we review some related works. Section III states our results and prospective and future works are given in Section IV.

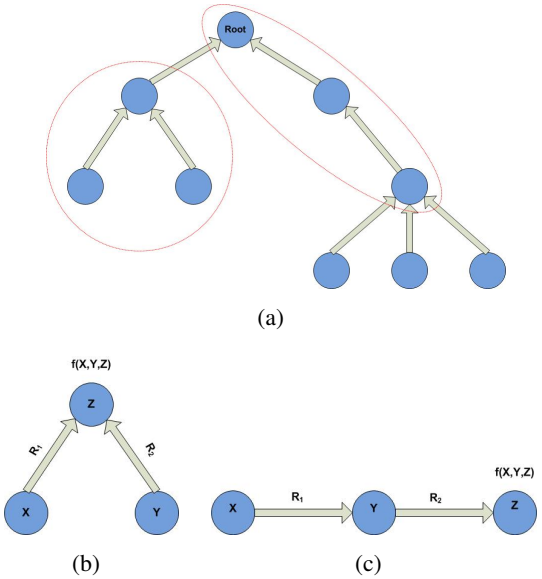



Fig. 1. (a) A wireless sensor network , (b) Slepian-Wolf Configuration and (c) Cascade Configuration

## II. Background

The problem of computing a function of some random variables has been studied for different configurations. For a Point-to-Point communication, the problem have been well-studied and for both one-way communication and interactive communication the minimum required bits for reliable function computation has been derived. But for a network, there are not any general results, even for simple networks such as Slepian-Wolf (Figure 1(b)) and Cascade (Figure 1(c)) networks. These networks are in particular interests since they are the basic networks that compose a directed tree network. We review the obtained results in the Point-to-Point, Slepian-Wolf, Cascade and a general configurations.

[1] number of bits for one time function computation

### A. Point-to-Point

For the Point-to-Point configuration, for interactive case the problem has been first studied by Yao [25] where each of the terminals has access to a random variable and the goal is to compute a function of these random variables at one terminal with the minimum number of transmitted bits. For this goal, terminals alternatively send message to each other until one of the nodes is able to compute the function. He derived upper and lower bound on the minimum number of transmitted bits for both deterministic and probabilistic algorithms. Then, Orlitsky in [14] and [15] investigated how much interaction can help in reducing the overall number of transmitted bits. He showed that in a problem of exchanging the sources in a point-to-point error free channel, one way communication may need exponentially more number of bits than the optimal one (i.e. optimal strategy with arbitrary number of interactions) and two round communication is almost optimal (It is less than four times of optimal number of bits) but not optimal (It may need more than twice of optimal number of bits).

Orlitsky and Roche [16] have considered the problem in a point to point communication with side information at the receiver who wants to compute a function of the random variables. For both one-way and two-way communication they derived an information-theoretic characterization on the minimum number of bits needed to be communicated. They showed that minimum number of bits depends on entropy of a conditional characteristic graph (introduced by Witsenhausen [22]) that differs according to the function and joint probability distribution of random variables. Ma et al. [12] have generalized their results for $K$ round communication.

### B. Slepian-Wolf

For Slepian-Wolf configuration there is no result for interactive case and for one-way communication the significant result is obtained by Körner and Marton [10]. They derived the rate region for a special case where the function is the sum modulo two of binary random variables with symmetric joint probability distributions. Later, Han and Kobayashi [8] generalized their results for sum modulo $p$ and symmetric distribution. Also they characterized the necessary and sufficient conditions that the rate region of the problem of reliably computing a function is the same as the rate region of the Slepian and Wolf source coding problem [17].

More recently, Doshi, Shah, and Médard [5] derived conditions under which a rate pair can be achieved for fixed code length and fixed error probability. This characterization is, however, not single-letter.

### C. Cascade

For the problem of function computation with some distortions in a cascade configuration (Figure 1(c)) without any side information at the receiver (i.e. $Z$ is a constatnt in Figure 1(c) ), Cuff et al. [4] proposed an inner and outer bound which are tight for the lossless case. Note that their result is for the case where there is not any side information at receiver, so it can not be generalized to the cascade configuration with multiple intermediate nodes nor to be used in larger configurations like a network in Figure 1(a). For interactive case, there is not any result.

### D. General Network

The obtained results for general network are for the following special cases.

a) Tsitsiklis [20] have considered the problem of decentralized decision in a network, i.e. there are some processor with some observations and all of them want to agree on a decision (which can be a function). He considered the problem if processors can make compatible decisions, locally or not. If yes, what the complexity of finding these compatible decisions is and if not, how many bits should be communicated to agree on a decision. They derived the complexity classes of finding compatible decisions and proposed a scheme which leads to convergence of decisions with the criteria of minimizing a cost function. Note that in their setting, all of the nodes want to have the same decision

and it is different from our problem where there is a node who wants to compute a function (make a decision).

b) Giridhar and Kumar [7] considered the problem of function computation in the wireless sensor networks and have derived the maximum possible refresh rate of function computation for symmetric, type-sensitive and type-threshold functions in the collocated and planar multihop network.

c) Distributed averaging of i.i.d sources within a prescribed mean square error distortion has been studied by Su and El Gamal [18].

d) Feizi and Médard [6] generalized the results of [5] for a tree network where the root wants to compute the function of the sources. (Which is not a single letter characterization).

e) Kowshik and Kumar [11] derived necessary and sufficient conditions on each encoder of nodes for reliably computing a function in a network, i.e. the encoder at each node should produce different outputs for inputs that may cause different function value.

f) The problem from the *Network Coding* view has been studied by Appuswamy et al. [1] for independent sources. They defined the modified concept of min-cut and showed that the modified min-cut is an upper bound for function computation. They also have shown that this upper bound is tight in some cases.

g) Ayaso et al. [2] considered a network with point-to-point memoryless noisy independent channels and have given a lower bound on the necessary computation time.

Lack of general results for basic networks such as Slepian-Wolf and Cascade configuration motivates us to consider these basic networks which may enable us to derive results for a general network. We finish the bibliography section by reviewing the results in rate distortion theory.

The rate distortion theory is closely related to the reliable function computation problem. In fact, for reliable function computation problem there is a stricter condition than the rate distortion problem, however for the proof of converse, the rate distortion theory is usually used.

The rate distortion region has been derived for point to point communication with side information at decoder by Wyner and Ziv [23]. Yamamoto [24] has shown that this result is valid in the case that the goal is to decode a function of source and side information with some distortions. Some upper and lower bounds on the rate region for the case of correlated sources have been proposed by Berger and Tung [21] and the rate region was fully derived for the case of correlated sources with one distortion criteria by Berger and Yeung [3].

## III. Results

The results we derived is for the problem of computing a function of correlated sources. More precisely, a receiver wants to compute a function $f$ of two correlated sources $X$ and $Y$ and side information $Z$. What is the minimum number of bits that needs to be communicated by each transmitter? This setting extends the Orlitsky-Roche setting [16] to multiple sources.

We first establish an outer bound to the rate region by simply applying the converse result of Point-to-Point communication with side information at receiver in [16]. Then we propose an inner bound. Although this inner bound is not tight in general, we show that it is tight for the case where $X$ is *inferable*, i.e., when $X$ is a function of $f(X, Y, Z)$ and $Z$, and for the case where $Y$ is constant. In the latter case, we recover Orlitsky and Roche's result.

Orlitsky and Roche showed that the minimum number of bits needed for computing $f(X, Z)$ is the solution of a mutual information minimization over maximal independent sets defined over a certain characteristic graph given by $X$, $Z$, and $f$. In contrast, our inner bound involves an optimization over (finite) *multisets* of maximal independent sets. We show, through an example where $X$ is inferable, that multisets may indeed increase the set of achievable rate pairs.

### A. Problem statement

Let $f$ be a function of three random variables, $X$, $Y$, and $Z$. A transmitter knows $X$, another knows $Y$, and a receiver knows $Z$. What are the minimum numbers of bits that needs to be communicated by each transmitter so that the receiver can compute $f(X,Y,Z)$ reliably? The precise problem formulation extends [16] to multiple transmitters.

Let $\mathcal{X}$, $\mathcal{Y}$, $\mathcal{Z}$, and $\mathcal{F}$ be finite sets, and $f : \mathcal{X}\times\mathcal{Y}\times\mathcal{Z} \to \mathcal{F}$. Let $\{(x_i, y_i, z_i)\}_{i=1}^{\infty}$ be independent instances of random variables $(X, Y, Z)$ taking values over $\mathcal{X} \times \mathcal{Y} \times \mathcal{Z}$ and distributed according to $p(x, y, z)$.

**Definition 1** (Code)**.** *A $(n, R_1, R_2)$ code consists of two encoding functions*

$$\varphi_X : \mathcal{X}^n \to \{1, 2, .., 2^{nR_1}\},$$
$$\varphi_Y : \mathcal{Y}^n \to \{1, 2, .., 2^{nR_2}\},$$

*and a decoding function*

$$\psi : \{1, 2, .., 2^{nR_1}\} \times \{1, 2, .., 2^{nR_2}\} \times \mathcal{Z}^n \to \mathcal{F}^n .$$

*The error proability of a code is defined as*

$$P(\psi(\varphi_X(\mathbf{X}), \varphi_Y(\mathbf{Y}), \mathbf{Z}) \neq f(\mathbf{X}, \mathbf{Y}, \mathbf{Z})),$$

*where $\mathbf{X} \stackrel{def}{=} X_1, \ldots, X_n$ and*

$$f(\mathbf{X}, \mathbf{Y}, \mathbf{Z}) \stackrel{def}{=} \{f(X_1, Y_1, Z_1), ..., f(X_n, Y_n, Z_n)\} .$$

**Definition 2** (Rate Region)**.** *A rate pair $(R_1, R_2)$ is achievable if, for any $\epsilon > 0$ and all $n$ large enough, there exists a $(n, R_1, R_2)$ code whose error probability is no larger than $\varepsilon$. The rate region is the closure of the set of achievable $(R_1, R_2)$.*

The problem we consider here is to characterize the rate region for given $p(x, y, z)$ and $f$.

### B. Preliminaries

Conditional characteristic graph [22], [9] plays a key role in coding for computing:

**Definition 3** (Conditional Characteristic Graph )**.** *Given $(X, Y) \sim p(x, y)$ and $f(X, Y)$, the conditional characteristic graph $G_{X|Y}$ of $X$ given $Y$ is the (undirected) graph whose vertex set is $\mathcal{X}$ and such that $x_i$ and $x_j$ are connected whenever there exists a $y \in \mathcal{Y}$ with*

i. $p(x_i, y)p(x_j, y) > 0$,
ii. $f(x_i, y) \neq f(x_j, y)$.

**Notation.** *Given two random variables $X$ and $W$, where $X$ ranges over $\mathcal{X}$ and $W$ over* subsets *of $\mathcal{X}$,*[2] *we write $X \in W$ whenever $P(X \in W) = 1$.*

**Definition 4** (Conditional Graph Entropy [16])**.** *The conditional entropy of a graph is defined as*

$$H_{G_{X|Y}}(X|Y) = \min_{\substack{W-X-Y \\ X\in W\in\Gamma(G_{X|Y})}} I(W; X|Y). \tag{1}$$

We now extend the definition of conditional characteristic graph to allow conditioning on variables that take values over independent sets and to allow side information.

Recall that an independent set of a graph $G$ is a subset of vertices, no two of which are connected. The set of independent sets of $G$ and the set of maximal independent sets of $G$ are denoted by $\Gamma(G)$ and $\Gamma^*(G)$, respectively.

[2]I.e., a sample of $W$ is a subset of $\mathcal{X}$.

**Definition 5** (Generalized Conditional Characteristic Graph)**.** *Given* $(X, Y, W) \sim p(x, y, w)$ *and* $f(X, Y)$ *such that* $Y \in W \in \Gamma(G_{Y|X})$*, let* $\tilde{f}(x, w) \stackrel{def}{=} f(x, y)$ *for* $x \in \mathcal{X}$*,* $y \in w \in \Gamma(G_{Y|X})$*, with* $p(x, y, w) > 0$*. The conditional characteristic graph of* $X$ *given* $W$*, denoted by* $G_{X|W}$*, is the conditional characteristic graph of* $X$ *given* $W$ *with respect to* $p(x, w)$ *and* $\tilde{f}(X, W)$*.*

**Definition 6** (Generalized Conditional Characteristic Graph with Side Information)**.** *Given* $(X, Y, Z, W) \sim p(x, y, z, w)$ *and* $f(X, Y, Z)$ *such that* $Y \in W \in \Gamma(G_{Y|X,Z})$*, let* $\tilde{f}(x, w, z) = f(x, y, z)$ *for* $x \in \mathcal{X}$*,* $z \in \mathcal{Z}$*,* $y \in w \in \Gamma(G_{Y|X,Z})$*,*[3] *and* $p(x, y, z, w) > 0$*. The conditional characteristic graph of* $X$ *given* $W$ *and* $Z$*, denoted by* $G_{X|W,Z}$*, is the conditional characteristic graph of* $X$ *given* $(W, Z)$ *with respect to* $p(x, w, z)$ *and* $\tilde{f}(X, W, Z)$*.*

**Definition 7** (Inferable Random Variable)**.** *Given* $(X, Y, Z) \sim p(x, y, z)$ *and* $f(X, Y, Z)$*,* $X$ *is said to be inferable if it is a function of* $f(X, Y, Z)$ *and* $Z$*.*

The following lemma can be deduced from Definitions 3 to 7.

**Lemma 1.** *Given* $(X, Y, Z, W) \sim p(x, y, z, w)$ *and* $f(X, Y, Z)$*, we have*

$$G_{Y|W,Z} = G_{Y|X,Z},$$

*for all* $W$ *such that* $X \in W \in \Gamma(G_{X|Y,Z})$*, in each of the following cases:*

a. $p(x, y, z) > 0$ *for all* $(x, y, z) \in \mathcal{X} \times \mathcal{Y} \times \mathcal{Z}$*;*
b. $\Gamma(G_{X|Y,Z})$ *consists only of singletons;*
c. $X$ *is inferable given* $p(x, y, z)$ *and* $f(X, Y, Z)$*.*

### *C. Main Results*

First result proposes an outer bound to the rate region, which is obtained by using the converse result in [16].

**Theorem 1** (Outer bound)**.** [4] *If* $(R_1, R_2)$ *is achievable then*

$$\begin{aligned} R_1 &\geq H_{G_{X|Y,Z}}(X|Y, Z), \\ R_2 &\geq H_{G_{Y|X,Z}}(Y|X, Z), \\ R_1 + R_2 &\geq H_{G_{X,Y|Z}}(X, Y|Z). \end{aligned}$$

Our next result provides an inner bound to the rate region.

Recall that a multiset $\mathrm{m}(\mathcal{S})$ of a set $\mathcal{S}$ is a set of elements from $\mathcal{S}$ possibly with repetitions (for instance, if $\mathcal{S} = \{0, 1\}$, then $\{0, 1, 1\}$ is a multiset). We denote the collection of all multisets of $\mathcal{S}$ by $\mathrm{M}(\mathcal{S})$.

**Theorem 2** (Inner bound)**.** $(R_1, R_2)$ *is achievable whenever*

$$\begin{aligned} R_1 &\geq I(X; W_1|W_2, Z), \\ R_2 &\geq I(Y; W_2|W_1, Z), \\ R_1 + R_2 &\geq I(X; W_1|Z) + I(Y; W_2|W_1, Z), \end{aligned}$$

*for some* $W_1$ *and* $W_2$ *that satisfy*

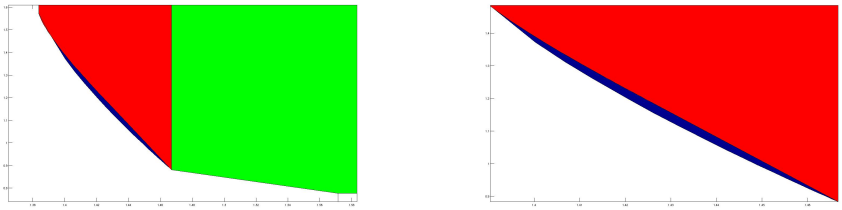
$$\begin{aligned} W_1 - X - (Y, Z, W_2), \\ W_2 - Y - (X, Z, W_1), \end{aligned}$$

*and*

$$X \in W_1 \in M(\Gamma^*(G_{X|Y,Z})),$$

[3] By definition $\Gamma(G_{Y|X,Z}) = \Gamma(G_{Y|(X,Z)})$.

[4] The proofs of the theorems can be found at Appendix.



(a) (b)

Fig. 2. Example of rate region for the case where one random variable is inferable.

$$Y \in W_2 \in M(\Gamma^*(G_{Y|W_1,Z}))\,.$$

When there is no side information at the decoder, i.e., when $Z$ is a constant, it can be shown that the two Markov chains in the theorem are equivalent to the single long Markov chain $W_1 - X - Y - W_2$, and that the sum rate inequality becomes as $R_1 + R_2 \geq I(X, Y; W_1, W_2)$.

For a single transmitter, i.e., in the case where one of the sources, say $Y$, is constant, $W_2$ is independent of $X$, $Z$ and $W_1$, and the achievable rate region reduces to $R_1 \geq I(X; W_1|Z)$ for some $W_1$ that satisfies $W_1 - X - Z$ and $X \in W_1 \in M(\Gamma^*(G_{X|Z}))$. Orlitsky and Roche [16] showed that the smallest value of $I(X; W_1|Z)$ is achieved by some $W_1$ taking values over the maximal independent sets $\Gamma^*(G_{X|Z})$. For multiple transmitters, choosing $W_1$ and $W_2$ over the maximal independent sets $\Gamma^*(G_{X|Y,Z})$ and $\Gamma^*(G_{Y|W_1,Z})$ need not be optimal. In fact, we have

$$\mathcal{R}(\Gamma^*) \subseteq \mathcal{R}(\Gamma) \subseteq \mathcal{R}(M(\Gamma^*)),$$

where $\mathcal{R}(\Gamma^*)$, $\mathcal{R}(\Gamma)$, and $\mathcal{R}(\Gamma)$ denote the achievable rate regions obtained by restricting $W_1$ and $W_2$ to take values over maximally independent sets, all independent sets, and multisets of maximally independent sets. An example where we have strict inclusion of these regions is illustrated in the example after Theorem 3. However, at the moment we do not have any analytical proof or intuition for this.

It can be shown that the inner bound is not always tight, e.g., for the mod 2 sum computation problem discussed in [10], [8]. However, the inner bound is tight for the case that one of the random variables is constant ([16]) and also for the case where one of the random variables is inferable:

**Theorem 3** (Rate Region - Inferable Random Variable)**.** *If $X$ is inferable given $(X, Y, Z) \sim p(x, y, z)$ and $f(X, Y, Z)$, then the rate region is the closure of rate pairs $(R_1, R_2)$ such that*

$$\begin{aligned} R_1 &\geq H(X|W, Z), \\ R_2 &\geq I(Y; W|X, Z), \\ R_1 + R_2 &\geq H(X|Z) + I(Y; W|X, Z), \end{aligned}$$

*for some $W$ that satisfies*

$$W - Y - (X, Z),$$

$$Y \in W \in \mathcal{W} \subset M(\Gamma^*(G_{Y|X,Z})),$$

*with*

$$|\mathcal{W}| \leq |\mathcal{Y}| + 2.$$

**Example 1.** *Consider the situation with no side information given by $f(X, Y) = Y$(mod 2) $+ 3X$, $\mathcal{X} = \{0, 1, 2\}$, $\mathcal{Y} = \{0, 1, 2\}$, and*

$$p(x, y) = \begin{bmatrix} .21 & .03 & .12 \\ .06 & .15 & .16 \\ .03 & .12 & .12 \end{bmatrix}.$$

*It can be checked that* $\Gamma(G_{X|Y}) = \{\{0\},\{1\},\{2\}\}$*, Furthermore, since* $X$ *is inferable, we have*

$$\Gamma(G_{Y|W_1}) = \Gamma(G_{Y|X}) = \{\{0\},\{1\},\{2\},\{0,2\}\},$$

*where the first equality holds by Lemma 1.c.*

*The upper Figure 1 represents the rate region provided by Theorem 3. The green area represents* $\mathcal{R}(\Gamma^*)$*, the union of the green and the red areas represents* $\mathcal{R}(\Gamma)$*, and the union of the green, red, and blue areas represents the* $\mathcal{R}(M(\Gamma^*))$ *with* $|M(\Gamma^*)| \leq 5$*. The lower figure emphasizes the difference between* $\mathcal{R}(\Gamma)$ *and* $\mathcal{R}(M(\Gamma^*))$*.*

## IV. Prospective and Future works

For our future works, we propose to consider the following aspects,

1) Is there any analytical proof or intuition explaining why using multi-sets instead of sets may improve the achievable rate region?
2) In the coding scheme proposed by Körner and Marton [10], they jointly apply function compression and Slepian-Wolf compression (they do it in one step by multiplying a matrix by input). The question to consider is how to implement the joint compression in general case, which might improve the achievable rate region, or to prove that separation is optimal and propose a separation based coding scheme for their problem. Also, we will try to generalize their result for the case of non-symmetric distribution or for other boolean functions which may help in determining the exact rate region of our studied problem.
3) In the problem of function computation in a cascade configuration (Figure 1(c)), what is the rate region when there is side information available at receiver? By deriving the rate region for this problem, one could extend the result for cascade network with multiple intermediate nodes. However, the problem will become different and more difficult than [**?**] because the intermediate node may not be able to compute the function.
4) Assume a network of nodes in a directed tree configuration (Figure 1(a)) where each node may have access to some sources and the goal is computing a function of these sources at the root. Since this network is constituted of Slepian-Wolf and Cascade sub-networks (Figure 1(a)), by using the results for these sub-networks, one may be able to find an appropriate achievable rate region for this network.
5) The problem we considered, Slepian-Wolf configuration, is the special case of MAC channel, in the sense that transmitters send their messages in separate channels or in other words $p(y|x_1,x_2) = 1$ for $y = (x_1,x_2)$ and zero otherwise. One prospective may be considering the MAC channel in general case. Nazer and Gastpar [13] have considered this problem and derived the computation capacity for the special case where the channel is symmetric and where both channel and function are linear. They showed that even for correlated sources, separation is not optimal. We try to generalize their results for arbitrary function and MAC channel.

## Publication

M. Sefidgaran, A. Tchamkerten, Computing a function of correlated sources: a rate region. *Submitted to ISIT 2011.*

## References


[1] R. Appuswamy, M. Franceschetti, N. Karamchandani, and K. Zeger. Network coding for computing: Cut-set bounds. *Information Theory, IEEE Transactions on*, 57(2):1015 –1030, February 2011.

[2] O. Ayaso, D. Shah, and M. A. Dahleh. Information theoretic bounds for distributed computation over networks of point-to-point channels. *Information Theory, IEEE Transactions on*, 56:6020–6039, December 2010.

[3] T. Berger and R. W. Yeung. Multiterminal source encoding with one distortion criterion. *Information Theory, IEEE Transactions on*, 35(2):228 –236, March 1989.

[4] P. Cuff, Han-I Su, and A. El Gamal. Cascade multiterminal source coding. In *Information Theory, 2009. ISIT 2009. IEEE International Symposium on*, pages 1199 –1203, 2009.

[5] V. Doshi, D. Shah, M. Médard, and S. Jaggi. Distributed functional compression through graph coloring. In *Proceedings of the Data Compression Conference*, pages 93–102, 2007.

[6] S. Feizi and M. Médard. When do only sources need to compute? on functional compression in tree networks. In *Proceedings of the 47th annual Allerton conference on Communication, control, and computing*, pages 447–454, 2009.

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

[8] T. S. Han and K. Kobayashi. A dichotomy of functions f(x,y) of correlated sources (x,y) from the viewpoint of the achievable rate region. *Information Theory, IEEE Transactions on*, 33:69–76, January 1987.

[9] J. Körner. Coding of an information source having ambiguous alphabet and the entropy of graphs. In *Transactions, 6th Prague Conference on Information Theory*, 1973.

[10] J. Körner and K. Marton. How to encode the modulo-two sum of binary sources (corresp.). *Information Theory, IEEE Transactions on*, 25(2):219 – 221, March 1979.

[11] H. Kowshik and P. R. Kumar. Zero-error function computation in sensor networks. In *Proceedings of the 48th IEEE Conference on Decision and Control*, 2009.

[12] N. Ma, P. Ishwar, and P. Gupta. Information-theoretic bounds for multiround function computation in collocated networks. In *Proceedings of ISIT - Volume 4*, pages 2306–2310, 2009.

[13] B. Nazer and M. Gastpar. Computation over multiple-access channels. *Information Theory, IEEE Transactions on*, 53(10):3498–3516, October 2007.

[14] A. Orlitsky. Worst-case interactive communication. i. two messages are almost optimal. *Information Theory, IEEE Transactions on*, 36(5):1111 –1126, September 1990.

[15] A. Orlitsky. Worst-case interactive communication. ii. two messages are not optimal. *Information Theory, IEEE Transactions on*, 37(4):995 –1005, July 1991.

[16] A. Orlitsky and J. R. Roche. Coding for computing. *36th IEEE Symposium on Foundations of Computer Science*, pages 502 –511, 1995.

[17] D. Slepian and J. Wolf. Noiseless coding of correlated information sources. *Information Theory, IEEE Transactions on*, 19(4):471 – 480, July 1973.

[18] Han-I Su and A. El Gamal. Distributed lossy averaging. In *Information Theory, 2009. ISIT 2009. IEEE International Symposium on*, 28 2009.

[19] Sundar Subramanian, Piyush Gupta, and Sanjay Shakkottai. Scaling bounds for function computation over large networks. In *Information Theory, 2007. ISIT 2007. IEEE International Symposium on*, pages 136 –140, 2007.

[20] J. N. Tsitsiklis. Problems in decentralized decision making and computation. *Ph.D. Thesis, Department of EECS, MIT*, November 1984.

[21] S. Tung. Multiterminal source coding (ph.d. thesis abstr.). *Information Theory, IEEE Transactions on*, 24(6):787, November 1978.

[22] H. Witsenhausen. The zero-error side information problem and chromatic numbers (corresp.). *Information Theory, IEEE Transactions on*, 22(5):592 – 593, September 1976.

[23] A. D. Wyner and J. Ziv. The rate-distortion function for source coding with side information at the decoder. *Information Theory, IEEE Transactions on*, 22:1–10, Januray 1976.

[24] H. Yamamoto. Correction to 'wyner-ziv theory for a general function of the correlated sources'. *Information Theory, IEEE Transactions on*, 29(2):803 – 807, March 1982.

[25] A. C. Yao. Some complexity questions related to distributive computing(preliminary report). In *Proceedings of the eleventh annual ACM symposium on Theory of computing*, STOC '79, pages 209–213, 1979.


## APPENDIX

*Sketch of the Proof of Theorem 1:* The first (second) inequality can be derived by assuming that receiver has access to $Y$ ($X$), so the problem reduces to the Point-to-Point problem and now we use the converse result in [16]. For the third inequality one can assume that two transmitters have access to both random variable and again we use the converse result in [16]. ■

*Sketch of the Proof of Theorem 2:* We consider a two-step coding procedure; a compression phase followed by a Slepian-Wolf coding [17] of the compressed sequences.

Pick $W_1$ and $W_2$ as in the theorem. These random variables together with $X, Y, Z$ are distributed according to some $p(x, w_1, y, w_2, z)$.

For $w_1 \in \Gamma(G_{X|Y,Z})$ and $w_2 \in \Gamma(G_{Y|W_1,Z})$, define $\tilde{f}(w_1, w_2, z)$ to be equal to $f(x, y, z)$ for all $x \in w_1$ and $y \in w_2$ such that $p(x, y, z) > 0$. Further, for $\boldsymbol{w}_1 = (w_{1,1}, \ldots, w_{1,n})$ and $\boldsymbol{w}_2 = (w_{2,1}, \ldots, w_{2,n})$ let

$$\tilde{f}(\boldsymbol{w}_1, \boldsymbol{w}_2, \boldsymbol{z}) \stackrel{\text{def}}{=} \left\{\tilde{f}(w_{1,1}, w_{2,1}, z_1), \ldots, \tilde{f}(w_{1,n}, w_{2,n}, z_n)\right\}.$$

Generate $2^{nI(X;W_1)}$ sequences

$$\boldsymbol{w}_1^{(i)} = \left(w_{1,1}^{(i)}, w_{1,2}^{(i)}, \ldots, w_{1,n}^{(i)}\right),$$

$i \in \{1, 2, \ldots, 2^{nI(X;W_1)}\}$, i.i.d. according to the marginal distribution $p(w_1)$, and randomly bin these sequences uniformly into $2^{nR_1}$ bins. Similarly, generate $2^{nI(Y;W_2)}$ sequences

$$\boldsymbol{w}_2^{(i)} = (w_{2,1}^{(i)}, w_{2,2}^{(i)}, \ldots, w_{2,n}^{(i)}),$$

i.i.d. according to $p(w_2)$, and randomly bin them uniformly into $2^{nR_2}$ bins. Reveal the bin assignments $\phi_1$ and $\phi_2$ to the encoders and to the decoder.

*Encoding:* The $\mathbf{X}$-transmitter finds a sequence $\boldsymbol{w}_1$ that is robust jointly typical with $\boldsymbol{x}$, and sends the index of the bin that $\boldsymbol{w}_1$ belongs to, i.e., $\phi_1(\boldsymbol{w}_1)$. The $\mathbf{Y}$-transmitter proceeds similarly and sends $\phi_2(\boldsymbol{w}_2)$. If a transmitter doesn't find such an index it declares an errors, and if there are more than one indices, the transmitter selects one of them randomly and uniformly.

*Decoding:* Given $\mathbf{z}$ and the index pair $(t_1, t_2)$, declare $\tilde{f}(\hat{\boldsymbol{w}}_1, \hat{\boldsymbol{w}}_2, \boldsymbol{z})$ if there exists a unique $(\hat{\boldsymbol{w}}_1, \hat{\boldsymbol{w}}_2, \boldsymbol{z})$ that is jointly robust typical and such that $\phi_1(\hat{\boldsymbol{w}}_1) = t_1$ and $\phi_2(\hat{\boldsymbol{w}}_2) = t_2$, and if $\tilde{f}(\hat{\boldsymbol{w}}_1, \hat{\boldsymbol{w}}_2, \boldsymbol{z})$ is defined. Otherwise declare an error.

*Probability of Error:* There are two types of error. The first type of error occurs when no $\boldsymbol{w}_1$'s, respectively $\boldsymbol{w}_2$'s, is robust jointly typical with $\boldsymbol{x}$, respectively with $\boldsymbol{y}$. The probability of each of these two errors is shown to be negligible in [16]. Hence, the probability of the first type of error is negligible.

The second type of error refers to the Slepian-Wolf coding procedure. By symmetry of the encoding and decoding procedures, the probability of error of the Slepian-Wolf coding procedure, averaged over sources outcomes, over $\boldsymbol{w}_1$'s and $\boldsymbol{w}_2$'s, and over the binning assignments, is the same as the average error probability conditioned that the transmitters select $\boldsymbol{W}_1^{(1)}$ and $\boldsymbol{W}_2^{(1)}$. Note that whenever $(\hat{\boldsymbol{W}}_1, \hat{\boldsymbol{W}}_2) = (\boldsymbol{W}_1^{(1)}, \boldsymbol{W}_2^{(1)})$, there is no error, i.e., $f(\boldsymbol{X}, \boldsymbol{Y}, \boldsymbol{Z}) = \tilde{f}(\boldsymbol{W}_1^{(1)}, \boldsymbol{W}_2^{(1)}, \boldsymbol{Z})$ by definition of robust typicality and by the definitions of $W_1$ and $W_2$. We now compute the probability of the event $(\hat{\boldsymbol{W}}_1, \hat{\boldsymbol{W}}_2) \neq (\boldsymbol{W}_1^{(1)}, \boldsymbol{W}_2^{(1)})$.

Define event $E(i, j)$ as

$$\begin{aligned} E(i,j) = \Big\{ &(\boldsymbol{W}_1^{(i)}, \boldsymbol{W}_2^{(j)}, \boldsymbol{Z}) \in \mathcal{T}, \\ &\phi_1(\boldsymbol{W}_1^{(i)}) = \phi_1(\boldsymbol{W}_1^{(1)}), \phi_2(\boldsymbol{W}_2^{(j)}) = \phi_2(\boldsymbol{W}_2^{(1)}) \Big\} \end{aligned}$$

where $\mathcal{T}$ denotes the ($\varepsilon$-) jointly robust typical set with respect to distribution $p(w_1, w_2, z)$. The probability of the second type of error is upper bounded as

$$\begin{aligned} P((\hat{\boldsymbol{W}}_1, \hat{\boldsymbol{W}}_2) &\neq (\boldsymbol{W}_1^{(1)}, \boldsymbol{W}_2^{(1)})) \\ &= P(E^c(1,1) \cup (\cup_{(i,j)\neq(1,1)} E(i,j))) \\ &\leq P(E^c(1,1)) + \sum_{\substack{(i,1) \\ i\neq 1}} P(E(i,1)) \\ &\quad + \sum_{\substack{(1,j) \\ j\neq 1}} P(E(1,j)) + \sum_{\substack{(i,j) \\ i\neq 1 \\ j\neq 1}} P(E(i,j)) \end{aligned} \tag{2}$$

Now, one can show that

$$\begin{aligned} P(E^c(1,1)) &\leq \epsilon \\ P(E(i,1)) &\overset{\approx}{\leq} 2^{-n(R_1 + I(W_1;W_2,Z))} \\ P(E(1,j)) &\overset{\approx}{\leq} 2^{-n(R_2 + I(W_2;W_1,Z))} \\ P(E(i,j)) &\overset{\approx}{\leq} 2^{-n(R_1+R_2+I(W_1;W_2)+I(W_1,W_2;Z))}. \end{aligned} \tag{3}$$

Hence, from (2) and (3)

$$\begin{aligned} & P((\hat{\boldsymbol{W}}_1, \hat{\boldsymbol{W}}_2) \neq (\boldsymbol{W}_1^{(1)}, \boldsymbol{W}_2^{(1)})) \\ & \overset{\approx}{\leq} \varepsilon + 2^{nI(X;W_1)} 2^{-n(R_1 + I(W_1;W_2,Z))} \\ & + 2^{nI(Y;W_2)} 2^{-n(R_2 + I(W_2;W_1,Z))} \\ & + 2^{n(I(X;W_1)+I(Y;W_2))} 2^{-n(R_1+R_2+I(W_1;W_2)+I(W_1,W_2;Z))} . \end{aligned}$$

The error probability of the second type is thus negligible whenever the theorem conditions are satisfied. (Note that $I(X;W_1) + I(Y;W_2) - I(W_1;W_2) - I(W_1,W_2;Z) = I(X;W_1|Z) + I(Y;W_2|W_1,Z)$.) ■

*Sketch of the Proof of Theorem 3:* The achievability of this theorem is a special case of Theorem 2 with $W_1 = X$.

Now for the converse, in the single transmitter case, Orlitsky and Roche used Wyner and Ziv's converse arguments. For the multiple transmitters case, we proceed similarly, and use the Berger and Yeung's converse arguments of [3, Theorem 1], which considers the rate distortion problem with one single distortion criterion.[5] To apply this rate distortion result, note that, since $X$ is inferable, if $f(\boldsymbol{X}, \boldsymbol{Y}, \boldsymbol{Z})$ can be computed with high probability, then $(\boldsymbol{X}, f(\boldsymbol{X}, \boldsymbol{Y}, \boldsymbol{Z}))$ can be recovered within small Hamming average distortion.

There is one small caveat in applying the converse arguments of [3, Theorem 1]. In our case we need the distortion measures to be defined over functions of the sources. To extend [3, Theorem 1] to the case where the distortion measures are defined over functions of the sources, one proceeds as in [24] that showed Wyner and Ziv's result [23] can be extended to the case where the distortion measure is defined over a function of the source and the side information.

Note that the rate distortion achievability results do not, in general, provide a direct way for establishing achievability results for computing problems (whether for single or multiple sources). The reason is that in the rate distortion problems [23] and [3] one usually considers average distortion between the source and the reconstruction block whereas in the computation problem we consider the more stringent block distortion criterion. ■

[5]Source $\boldsymbol{X}$ should be perfectly reconstructed while source $\boldsymbol{Y}$ should be reconstructed within some prescribed average distortion.