Lossy Computing of Correlated Sources with Fractional Sampling

Xi Liu*, Osvaldo Simeone† and Elza Erkip*
*ECE Dept., Polytechnic Inst. of NYU, Brooklyn, NY, USA
†CWCSPR, ECE Dept., NJIT, Newark, NJ, USA
Email:xliu02@students.poly.edu, osvaldo.simeone@njit.edu, elza@poly.edu

***Abstract*—This paper considers the problem of lossy compression for the computation of a function of two correlated sources, both of which are observed at the encoder. Due to presence of observation costs, the encoder is allowed to observe only subsets of the samples from both sources, with a fraction of such sample pairs possibly overlapping. For both Gaussian and binary sources, the distortion-rate function, or rate-distortion function, is characterized for selected functions and with quadratic and Hamming distortion metrics, respectively. Based on these results, for both examples, the optimal measurement overlap fraction is shown to depend on the function to be computed by the decoder, on the source correlation and on the link rate. Special cases are discussed in which the optimal overlap fraction is the maximum or minimum possible value given the sampling budget, illustrating non-trivial performance trade-offs in the design of the sampling strategy.**

I. Introduction

Consider an encoder endowed with a sensor that is able to measure two correlated discrete memoryless source sequences $S_1^n = (S_{1,1}, ..., S_{1,n})$ and $S_2^n = (S_{2,1}, ..., S_{2,n})$, as shown in Fig. 1. Due to the energy cost of source acquisition, sampling, quantization and compression, it might not be possible for the sensor to fully measure the sources S_1 and S_2. To simplify, this limitation can be modelled by imposing that only $n\theta_k$ samples can be measured from each source S_k, $k = 1, 2$, with $0 \leq \theta_k \leq 1$. The encoder compresses the measured samples to nR bits, where R is the communication rate in bits per source symbol. Based on the received bits, the decoder reconstructs a lossy version of a target function $T^n = f^n(S_1^n, S_2^n)$ of source sequences S_1^n and S_2^n, which is such that $T_i = f(S_{1,i}, S_{2,i})$, $i = 1, ..., n$. We refer to the above problem as *lossy computing with fractional sampling*.

A key aspect of the problem of lossy computing with fractional sampling is that the encoder is allowed to choose *which* samples to measure given the sampling budget (θ_1, θ_2). To fix the ideas, assume that we have $(\theta_1 = 0.5, \theta_2 = 0.5)$, so that only half of the samples can be observed from both sources. As two extreme strategies, the encoder can either measure the same samples from both sources, say $S_{1,i}, S_{2,i}$ for $i = 1, ..., n/2$, or it can measure the first source S_1 for the first $n/2$ samples, namely $S_{1,i}$ for $i = 1, ..., n/2$, and the second source S_2 for the remaining $n/2$ samples, namely $S_{2,i}$ for $i = n/2 + 1, ..., n$. With the first sampling strategy, the encoder is able to directly calculate the desired function $T_i = f(S_{1,i}, S_{2,i})$ for $i = 1, ..., n/2$, while having no information (beside the prior distribution) about T_i for the remaining samples. With the second strategy, instead, the encoder collects partial information about T at all times in the form of samples from source S_1 or source S_2. As it will be discussed in this paper, the optimal sampling strategy depends critically on the function $f(\cdot, \cdot)$, on the correlation between $S_{1,i}$ and $S_{2,i}$, and on the link rate R.

A. Related Work and Contributions

With full sampling of both sources, i.e., $(\theta_1 = 1, \theta_2 = 1)$, the encoder can directly calculate the function $T^n = f^n(S_1^n, S_2^n)$ and the problem at hand reduces to the standard rate-distortion set-up (see, e.g., [1]). Instead, if the encoder can only measure one of the two sources, i.e., $(\theta_1 = 1, \theta_2 = 0)$ or $(\theta_1 = 0, \theta_2 = 1)$, the problem at hand becomes a special case of the indirect source coding set-up introduced in [2]. For a discussion on problems related to computing and compression in network scenarios, we refer to [3]. The framework of source coding with fractional sampling was introduced in our previous work [4] for a model in which an energy-constrained sensor measures independent Gaussian sources for optimized fraction of time and the receiver wishes to reconstruct all sources with given quadratic distortion constraints. The model is also related to that of compression with actions of [5].

This paper formulates the problem of lossy computing with fractional sampling of correlated sources (Section II). After providing a general expression for the distortion-rate and the rate-distortion functions (Section III), we focus on two specific examples that illustrate the trade-offs involved in the design of the sampling strategy. Specifically, we first consider correlated Gaussian sources and assume that linear functions of the form $T = w_1 S_1 + w_2 S_2$ are to be reconstructed at the decoder with quadratic distortion constraints (Section IV). We then consider correlated binary sources with arbitrary functions $T = f(S_1, S_2)$ and Hamming distortion (Section V). Various conclusions are drawn regarding conditions under which the optimal sampling strategy prescribes the maximum or the minimum possible overlap between the samples measured from the two sources.

II. System Model

In this section, we formally introduce the system model of interest. As shown in Fig. 1, the encoder has access to two discrete memoryless source sequences $S_1^n = (S_{1,1}, ..., S_{1,n})$ and

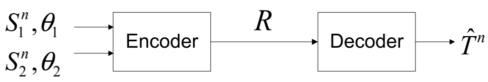

Figure 1. The encoder measures correlated sources S_1 and S_2 for a fraction of time θ_1 and θ_2, respectively, and the decoder estimates a function $T^n = f^n(S_1^n, S_2^n)$.

$S_2^n = (S_{2,1}, ..., S_{2,n})$ respectively, which consist of n independent and identically distributed (i.i.d.) samples $(S_{1,i}, S_{2,i})$ with $S_{1,i} \in \mathcal{S}_1$ and $S_{2,i} \in \mathcal{S}_2$, $i = 1, ..., n$, where $\mathcal{S}_1$ and $\mathcal{S}_2$ are the alphabet sets for S_1 and S_2 respectively. All alphabets are assumed to be finite unless otherwise stated. Due to presence of observation costs, we assume the encoder can only sample a fraction θ_k of the samples for source S_k, with $0 \leq \theta_k \leq 1$ for $k = 1, 2$. Given the i.i.d. nature of the sources, without loss of generality, we assume that the encoder measures the first θ_1 fraction of samples of source S_1 and measures the θ_2 fraction of samples of S_2 starting from sample $n(\theta_1 - \theta_{12}) + 1$[1], as shown in Fig. 2. The samples measured at the encoder from the two sources thus overlap for a fraction θ_{12}, with θ_{12} satisfying

$$\theta_{12,min} \leq \theta_{12} \leq \theta_{12,max}, \tag{1}$$

with $\theta_{12,min} = (\theta_1 + \theta_2 - 1)^+$ and $\theta_{12,max} = \min(\theta_1, \theta_2)$, where $(\cdot)^+$ denotes $\max(\cdot, 0)$. We refer to the triple $(\theta_1, \theta_2, \theta_{12})$ as a *sampling profile*, and to (θ_1, θ_2) as the *sampling budget*.

The decoder wishes to estimate a function $T^n = f^n(S_1^n, S_2^n)$, where $T_i = f(S_{1,i}, S_{2,i})$ for $i = 1, ..., n$. We let $d: \mathcal{T} \times \hat{\mathcal{T}} \to [0, +\infty)$ be a distortion measure, where $\mathcal{T}$ and $\hat{\mathcal{T}}$ are the alphabet sets of the variables T and $\hat{T}$ respectively. We assume, without loss of generality, that for each $t \in \mathcal{T}$ there exists a $\hat{t} \in \mathcal{T}$ such that $d(t, \hat{t}) = 0$. The link between the encoder and the decoder can support a rate of R bits/sample. Formal definitions follow.

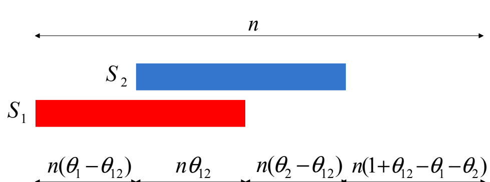

Figure 2. Sampling profile $(\theta_1, \theta_2, \theta_{12})$ at the encoder: a fraction, $\theta_1 - \theta_{12}$, of samples is measured only from source S_1; a fraction, θ_{12}, of samples is measured from both sources; a fraction, $\theta_2 - \theta_{12}$, of samples is measured only from source S_2; and the remaining fraction, $1 + \theta_{12} - \theta_1 - \theta_2$, of samples is not measured for either source ($0 \leq \theta_1, \theta_2 \leq 1$, and θ_{12} as in (1)).

Definition 1: A $(n, R, D, \theta_1, \theta_2, \theta_{12})$ code for the problem of lossy computing of two memoryless sources with fractional sampling consists of an encoder $h : \mathcal{S}_1^{n\theta_1} \times \mathcal{S}_2^{n\theta_2} \to \{1, ..., 2^{nR}\}$, which maps the measured θ_1-fraction of source S_1, i.e., $(S_{1,1}, ..., S_{1,n\theta_1})$, and the measured θ_2-fraction of source S_2, i.e., $(S_{2,n(\theta_1-\theta_{12})+1}, ..., S_{2,n(\theta_1+\theta_2-\theta_{12})})$, into a message of rate R bits per source sample (where the normalization is with respect to the overall number of samples n); and a decoder $g : \{1, ..., 2^{nR}\} \to \hat{\mathcal{T}}^n$, which maps the message from the encoder into an estimate $\hat{T}^n$, such that distortion constraint D is satisfied, i.e.,

$$\frac{1}{n} E\left[\sum_{i=1}^{n} d(T_i, \hat{T}_i)\right] \leq D. \tag{2}$$

Given any sampling profile $(\theta_1, \theta_2, \theta_{12})$, a tuple $(R, D, \theta_1, \theta_2, \theta_{12})$ is said to be achievable, if for any $\epsilon > 0$, and sufficiently large n, there exists a $(n, R, D+\epsilon, \theta_1, \theta_2, \theta_{12})$ code. The *distortion-rate function* for a given sampling profile $D(R, \theta_1, \theta_2, \theta_{12})$ is defined as $D(R, \theta_1, \theta_2, \theta_{12}) = \inf\{D$: the tuple $(R, D, \theta_1, \theta_2, \theta_{12})$ is achievable$\}$. The *distortion-rate function* with sampling budget (θ_1, θ_2), $D(R, \theta_1, \theta_2)$, is defined as $D(R, \theta_1, \theta_2) = \min_{\theta_{12}} D(R, \theta_1, \theta_2, \theta_{12})$ where the minimum is taken over all θ_{12} satisfying (1). Similar definitions are used for the rate-distortion function. Specifically, the *rate-distortion function* given a sampling profile $(\theta_1, \theta_2, \theta_{12})$ and distortion D is defined as $R(D, \theta_1, \theta_2, \theta_{12}) = \inf\{R$: the tuple $(R, D, \theta_1, \theta_2, \theta_{12})$ is achievable$\}$, and the rate-distortion function with sampling budget (θ_1, θ_2) as $R(D, \theta_1, \theta_2) = \min_{\theta_{12}} R(D, \theta_1, \theta_2, \theta_{12})$ where the minimum is taken over all θ_{12} satisfying (1).

III. Rate-Distortion Trade-Off with Fractional Sampling

In this section, we characterize the distortion-rate functions $D(R, \theta_1, \theta_2, \theta_{12})$ and $D(R, \theta_1, \theta_2)$ defined above as well as their rate-distortion counterparts. To elaborate, we first define the standard distortion-rate function for the memoryless source T as $D_{12}(R) = \min_{p(\hat{t}|t):\, I(T;\hat{T}) \leq R} E[d(T, \hat{T})]$ [1]. We similarly define the corresponding rate-distortion function with full sampling as $R_{12}(D) = \min_{p(\hat{t}|t):\, E[d(T,\hat{T})] \leq D} I(T; \hat{T})$. Moreover, we define the indirect distortion-rate function for compression of T when only S_k is observed at the encoder, for $k = 1, 2$, as $D_k(R) = \min_{p(\hat{t}|s_k):\, I(S_k;\hat{T}) \leq R} E[d(T, \hat{T})]$. We similarly define the corresponding rate-distortion function $R_k(D) = \min_{p(\hat{t}|s_k):\, E[d(T,\hat{T})] \leq D} I(S_k; \hat{T})$. Finally, we define $D_{k,min}$ as $D_{k,min} = \lim_{R \to \infty} D_k(R) = \min_{g_k(\cdot)} E(d(g_k(S_k), T))$, for $k = 1, 2$, where function $g_k(\cdot)$ is defined as $g_k : \mathcal{S}_k \to \hat{\mathcal{T}}$, which maps S_k to an estimate $\hat{T}$.

Lemma 1: For any given sampling profile $(\theta_1, \theta_2, \theta_{12})$ and link rate R, the distortion-rate function for computing T is given by[2]

$$\begin{aligned} D(R, \theta_1, \theta_2, \theta_{12}) = \min_{R_1, R_{12}, R_2 \geq 0} &(\theta_1 - \theta_{12}) D_1\left(\frac{R_1}{\theta_1 - \theta_{12}}\right) \\ &+ \theta_{12} D_{12}\left(\frac{R_{12}}{\theta_{12}}\right) + (\theta_2 - \theta_{12}) D_2\left(\frac{R_2}{\theta_2 - \theta_{12}}\right) \\ &+ (1 + \theta_{12} - \theta_1 - \theta_2) D_{max}, \end{aligned} \tag{3}$$

[1] Throughout the paper, quantities such as $n\theta_1$, $n\theta_2$ and $n(\theta_1 + \theta_2 - \theta_{12})$ are implicitly assumed to be rounded to the largest smaller integer.

[2] For any given convex function $Q(x)$ for $x \geq 0$, we define $0 \cdot Q(x/0) = 0$, for $x \geq 0$, if $\lim_{x \to 0} x \cdot Q(1/x) = 0$.

with $D_{max} = \min_{\hat{t}\in\hat{\mathcal{T}}} E[d(T,\hat{t})]$, and where the minimization is taken under the constraint

$$R_1 + R_2 + R_{12} \leq R. \tag{4}$$

For convenience, we let

$$\begin{aligned} D_{min}(\theta_1,\theta_2,\theta_{12}) =& \lim_{R\to\infty} D(R,\theta_1,\theta_2,\theta_{12}) \\ =&(\theta_1-\theta_{12})D_{1,min} + (\theta_2-\theta_{12})D_{2,min} \\ &+ (1+\theta_{12}-\theta_1-\theta_2)D_{max}. \end{aligned} \tag{5}$$

Similarly, for any given sampling profile $(\theta_1,\theta_2,\theta_{12})$ and distortion level $D \geq D_{min}(\theta_1,\theta_2,\theta_{12})$, the rate-distortion function for computing T is given by

$$\begin{aligned} R(D,&\theta_1,\theta_2,\theta_{12}) = \min_{D_1,D_{12},D_2} (\theta_1-\theta_{12})R_1\left(\frac{D_1}{\theta_1-\theta_{12}}\right) \\ &+\theta_{12}R_{12}\left(\frac{D_{12}}{\theta_{12}}\right) + (\theta_2-\theta_{12})R_2\left(\frac{D_2}{\theta_2-\theta_{12}}\right), \end{aligned} \tag{6}$$

where the minimization is taken over all choices of D_1, D_2 and D_{12} satisfying $D_{12} \geq 0$,

$$D_1 \geq (\theta_1-\theta_{12})D_{1,min}, \tag{7a}$$

$$D_2 \geq (\theta_2-\theta_{12})D_{2,min}, \tag{7b}$$

$$D_1+D_2+D_{12}+(1+\theta_{12}-\theta_1-\theta_2)D_{max} \leq D, \tag{7c}$$

In the lemma above, rate R_k is assigned for the description of the $(\theta_k-\theta_{12})$-fraction of samples in which only source S_k is measured, $k=1,2$, while rate R_{12} is assigned for the description of the θ_{12}-fraction of samples in which both sources are measured (recall Fig. 2). Distortions D_1, D_2 and D_{12} are the corresponding average per-symbol distortions in the reconstruction of T at the decoder. The proof follows immediately from the independence of the samples measured from the different fraction of samples, and it is thus omitted. The following property is a consequence of the operational definitions given above.

Lemma 2: $D(R,\theta_1,\theta_2)$ is continuous and convex in R, for $R \geq 0$. Similarly, $R(D,\theta_1,\theta_2)$ is continuous and convex in D, for $D \geq D_{min}(\theta_1,\theta_2)$, where $D_{min}(\theta_1,\theta_2) = \lim_{R\to\infty} D(R,\theta_1,\theta_2) = \min_{\theta_{12}} D_{min}(\theta_1,\theta_2,\theta_{12})$.

IV. Gaussian Sources

In this section, we focus on the case in which sources S_1 and S_2 are jointly Gaussian, zero-mean, unit-variance and correlated with coefficient ρ, with $\rho \in [-1,1]$. The decoder wishes to compute a weighted sum function $T = f(S_1,S_2) = w_1S_1 + w_2S_2$, with $w_1,w_2 \in \mathbb{R}$, under the mean square error (MSE) distortion measure $d(t,\hat{t}) = (t-\hat{t})^2$. In the following, we study two specific choices for the weights $w_1=1, w_2=0$ and $w_1=w_2=1$, resulting in the weighted sum functions $T=S_1$ and $T=S_1+S_2$, respectively. These two cases are selected in order to illustrate the impact of the choice of the function $f(S_1,S_2)$ on the optimal sampling strategy. The discussion can be extended with appropriate modifications to arbitrary choices of weights (w_1,w_2).

A. Computation of $T = S_1$

Proposition 1: For a given sampling budget (θ_1,θ_2), the distortion-rate function for computing $T=S_1$ is

$$D(R,\theta_1,\theta_2) = \begin{cases} 1-\theta_1+\theta_1 2^{-\frac{2R}{\theta_1}}, & \text{if } R \leq \frac{\theta_1}{2}\log_2\left(\frac{1}{\rho^2}\right), \\ 1-\theta_1-\rho^2(\theta_2-\theta_{12}^*) & \\ \quad +(\theta_1+\theta_2-\theta_{12}^*)2^{-\frac{2R}{\theta_1+\theta_2-\theta_{12}^*}} & \\ \quad \cdot\left(\rho^2\right)^{\frac{\theta_2-\theta_{12}^*}{\theta_1+\theta_2-\theta_{12}^*}}, & \text{otherwise}, \end{cases} \tag{8}$$

where $\theta_{12}^* = \theta_{12,min}$ is the optimal fraction of samples to be measured by both the encoder and the decoder. The rate-distortion function $R(D,\theta_1,\theta_2)$ can be obtained by inverting function (8) with respect to variable D.

Proof: See Appendix A. ■

Proposition 1 confirms the intuition that if the receiver is interested in source 1 only, i.e., $T=S_1$, the encoder should simultaneously measure both sources S_1 and S_2 for a fraction of time to be kept as small as possible. Moreover, if $R \leq \theta_1/2\log_2(1/\rho^2)$, the entire rate R is used to describe only the θ_1-fraction of samples measured from source S_1 only; otherwise, both the θ_1-fraction of source S_1 and the $(\theta_2-\theta_{12}^*)$-fraction of source S_2 that is not overlapped are described at positive rates. Note that, for rate $R \leq \theta_1/2\log_2(1/\rho^2)$, since only source S_1 is described, the choice of the overlapping fraction, in fact, does not matter, i.e., any θ_{12} satisfying $\theta_{12,min} < \theta_{12} \leq \theta_{12,max}$ is also optimal in this case.

B. Computation of $T = S_1 + S_2$

We now consider the case in which the desired function is $T=S_1+S_2$. Note that T is a Gaussian random variable with zero mean and variance $D_{max} = 2(1+\rho)$, and that T and S_1 (or S_2) are jointly Gaussian with correlation coefficient $\tilde{\rho} = \sqrt{(1+\rho)/2}$. Moreover, since $T=0$ for $\rho=-1$, it is enough to focus on $\rho \in (-1,1]$. We observe that the distortion-rate function for $T=S_1+S_2$ is given by $D_{12}(R) = 2(1+\rho)2^{-2R}$, for $R \geq 0$ [1]. Moreover, the indirect distortion-rate function is given as $D_k(R) = 2(1+\rho)(1-\tilde{\rho}^2+\tilde{\rho}^2 2^{-2R})$, for $R \geq 0$ and $k=1,2$ [6].

Proposition 2: Given sampling budget (θ_1,θ_2), the distortion-rate function for computing $T=S_1+S_2$ is

$$\begin{aligned} D(R,\theta_1,\theta_2) =& \min_{\theta_{12},R_{12}} (1+\rho)^2(\theta_1+\theta_2-2\theta_{12})2^{-\frac{2(R-R_{12})}{\theta_1+\theta_2-2\theta_{12}}} \\ &+2(1+\rho)\left(1+\rho\theta_{12}+\theta_{12}2^{-\frac{2R_{12}}{\theta_{12}}}\right) - (1+\rho)^2(\theta_1+\theta_2), \end{aligned} \tag{9}$$

where the minimization in (9) is taken over all θ_{12} satisfying (1) and all R_{12} satisfying $0 \leq R_{12} \leq R$.

Proof: See Appendix B. ■

In order to obtain further analytical insight into the optimal sampling strategy, we now consider some special cases of interest.

Corollary 1: For $R \to \infty$, we have

$$D_{min}(\theta_1,\theta_2) = 2(1+\rho)(1+\rho\theta_{12}^*) - (1+\rho)^2(\theta_1+\theta_2), \tag{10}$$

where $\theta_{12}^* = \theta_{12,min}$ if $\rho > 0$, $\theta_{12}^* = \theta_{12,max}$ if $\rho < 0$, and θ_{12}^* is arbitrary if $\rho = 0$.

This corollary is easily obtained from Proposition 2. It says that, if the sources (S_1, S_2) have positive correlation, i.e., $\rho > 0$, and there are no rate limitations ($R \to \infty$), the MSE distortion increases linearly with θ_{12}, and it is thus optimal to set θ_{12} to be the smallest possible value $\theta_{12}^* = \theta_{12,min}$. In contrast, if $\rho < 0$, the MSE distortion decreases linearly with θ_{12}, and thus the optimal θ_{12}^* is the largest possible value, $\theta_{12}^* = \theta_{12,max}$. This shows the relevance of the source correlation in designing the optimal sampling strategy.

The general conclusions about the optimal sampling strategies discussed above for infinite rate can be extended to finite rates R in certain regimes. Corollary 2 below states that if $\rho \leq 0$, then, just as in the case of infinite rate R of Corollary 1, the encoder should set θ_{12} to be as large as possible, i.e., $\theta_{12}^* = \theta_{12,max}$, irrespective of the value of R. Furthermore, Corollary 3 below suggests that for sufficiently small rates, the optimal overlap θ_{12}^* tends to be maximum, i.e., $\theta_{12}^* = \theta_{12,max}$, for a larger range of correlation coefficients ρ than $\rho \leq 0$. This is mainly because when rate R is small enough, it is generally more efficient to use the available rate to describe T directly during the overlapping θ_{12}-fraction, rather than indirectly describing T based on observations of S_1 or S_2 alone.

Corollary 2: For $\rho \leq 0$, the distortion-rate function is

$$D(R,\theta_1,\theta_2) = \begin{cases} (\theta_1+\theta_2-\theta_{12}^*)(1+\rho)^2 2^{-\frac{2R}{\theta_1+\theta_2-\theta_{12}^*}} \\ \cdot \left(\frac{2}{1+\rho}\right)^{\frac{\theta_{12}^*}{\theta_1+\theta_2-\theta_{12}^*}} + 2(1+\rho)(1+\rho\theta_{12}^*) \\ \quad -(1+\rho)^2(\theta_1+\theta_2), \text{ if } R > \frac{\theta_{12}^*}{2}\log_2\left(\frac{2}{1+\rho}\right) \\ 2(1+\rho)\left(1-\theta_{12}^*+\theta_{12}^* 2^{-\frac{2R}{\theta_{12}^*}}\right), \quad \text{otherwise}, \end{cases} \tag{11}$$

where $\theta_{12}^* = \theta_{12,max}$ is the optimal overlapping fraction.

Proof: The proof is obtained by solving (9) for $\rho \leq 0$. For $\rho \leq 0$, we can show that it is optimal to have $\theta_{12}^* = \theta_{12,max}$ by simply considering the monotonicity of function $D(R, \theta_1, \theta_2, \theta_{12})$, written as a function of θ_{12}, with respect to θ_{12}. Details of this step are omitted here. With θ_{12}^* known, the corresponding optimal R_{12}^* and the resulting minimum distortion can be computed. ■

Corollary 3: For any $0 < \rho \leq 1$, if $R \leq (\theta_{12,min}/2)\log_2(2/(1+\rho))$, the distortion-rate function is given as

$$D(R,\theta_1,\theta_2) = 2(1+\rho)(1-\theta_{12}^*) + 2(1+\rho)\theta_{12}^* 2^{-\frac{2R}{\theta_{12}^*}}, \tag{12}$$

where $\theta_{12}^* = \theta_{12,max}$.

Proof: Given $R \leq (\theta_{12,min}/2)\log_2(2/(1+\rho))$, for any feasible θ_{12} satisfying (1), we always have $R \leq (\theta_{12}/2)\log_2(2/(1+\rho))$. In this case, for any given θ_{12}, applying the standard Lagrangian method to (9), we obtain $R_{12}^* = R$. Substituting into (9) and considering the monotonicity of function $D(R, \theta_1, \theta_2, \theta_{12})$ with respect to θ_{12}, we can show that the optimal overlap fraction is given by $\theta_{12}^* = \theta_{12,max}$, leading to the distortion-rate function as stated in the corollary. ■

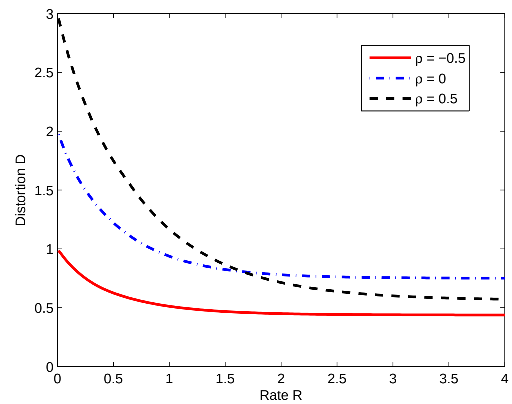

Figure 3. Distortion-rate function when $\theta_1 = 0.5$ and $\theta_2 = 0.75$, with correlation coefficient ρ chosen to be $\rho = -0.5, 0, 0.5$, respectively.

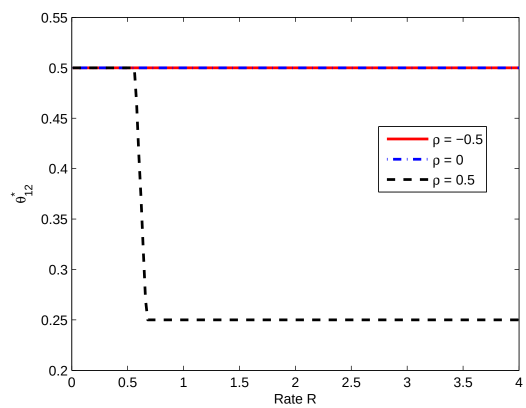

Figure 4. Optimal overlap fraction θ_{12}^* as a function of rate R when $\theta_1 = 0.5$ and $\theta_2 = 0.75$, with correlation coefficient ρ chosen to be $\rho = -0.5, 0, 0.5$, respectively.

C. Numerical Results

In this subsection, we numerically evaluate the distortion-rate tradeoff for computation of function $T = S_1 + S_2$. Recall that for $T = S_1$, the optimal overlap fraction is always $\theta_{12,min}$. Fig. 3 and Fig. 4 show the minimum MSE distortion D and the optimal overlap fraction θ_{12}^* versus rate R, respectively, for $\theta_1 = 0.5$, $\theta_2 = 0.75$, and $\rho = -0.5, 0, 0.5$. The curves are obtained by numerically solving the optimization in (9). It can be seen from Fig. 4 that, as predicted by Corollary 2, the optimal overlap fraction θ_{12}^* is equal to the maximum possible fraction $\theta_{12,max} = 0.5$, for $\rho = -0.5 < 0$ and $\rho = 0$. Moreover, for $\rho = 0.5 > 0$, with sufficiently small rates R, as described in Corollary 3, the optimal overlap fraction θ_{12}^* equals to the maximum overlap $\theta_{12,max} = 0.5$. However, as R grows beyond some threshold, θ_{12}^* drops to the minimum value $\theta_{12,min} = 0.25$, which is consistent with Corollary 1.

V. Binary Sources

In this section, we consider binary sources so that $\mathcal{S}_1 = \mathcal{S}_2 = \mathcal{T} = \hat{\mathcal{T}} = \{0,1\}$, and (S_1, S_2) is a doubly symmetric binary source (DSBS) characterized by probability $\Pr[S_1 \neq S_2] = p$, $0 \leq p \leq 1/2$. We take the Hamming distortion as the distortion measure, i.e., $d(t, \hat{t}) = 1 - \delta_{t\hat{t}}$, where $\delta_{t\hat{t}} = 1$ if $t = \hat{t}$ and $\delta_{t\hat{t}} = 0$ otherwise. Since all non-trivial binary functions are equivalent, up to relabeling, to either the exclusive OR or the AND [7], it suffices to consider only these two options for function $T = f(S_1, S_2)$: (*i*) the exclusive OR or binary sum, i.e., $T = S_1 \oplus S_2$; (*ii*) the AND or binary product, i.e., $T = S_1 \otimes S_2$. In the following, we focus on deriving the rate-distortion $R(D, \theta_1, \theta_2)$ for convenience, since in general it takes a simpler analytical form as compared to the distortion-rate function $D(R, \theta_1, \theta_2)$.

A. Computation of $T = S_1 \oplus S_2$

Proposition 3: For given sampling budget (θ_1, θ_2), the rate-distortion function for computing $T = S_1 \oplus S_2$ is given by

$$R(D,\theta_1,\theta_2) = \begin{cases} h(p) - h\left(\frac{D-(1-\theta_{12}^*)p}{\theta_{12}^*}\right), & \\ \quad \text{if } (1-\theta_{12}^*)p \leq D < p, & \\ 0, \quad \text{if } D \geq p, & \end{cases} \tag{13}$$

where $h(x) = -x\log_2(x) - (1-x)\log_2(1-x)$ is the binary entropy function, and $\theta_{12}^* = \theta_{12,max}$ is the optimal overlap fraction, for $(1-\theta_{12}^*)p \leq D < p$.

The above proposition can be proved by using the fact that $T = S_1 \oplus S_2$ is a Bernoulli(p) random variable, and is independent of S_1 and S_2. Therefore, the observation of either S_1 or S_2 is not useful for computing T, and thus one should choose the overlap fraction to be as large as possible, i.e., $\theta_{12}^* = \theta_{12,max}$. The rate-distortion function (13) then follows immediately from the rate-distortion function of the binary random variable T [1].

B. Computation of $T = S_1 \otimes S_2$

In this subsection, we focus on the binary product $T = S_1 \otimes S_2$, which is Bernoulli distributed with probability $(1-p)/2$. For convenience, we start by finding the minimum possible distortion at the decoder given (θ_1, θ_2), i.e., $D_{min}(\theta_1, \theta_2)$ as defined in Lemma 2, and the minimum required rate to achieve it. Then, we proceed to derive the rate-distortion function.

Proposition 4: For given sampling budget (θ_1, θ_2), the minimum achievable distortion for computing $T = S_1 \otimes S_2$ is given by

$$D_{min}(\theta_1,\theta_2) = \frac{1-p}{2} + \left(p - \frac{1}{2}\right)(\theta_1 + \theta_2) + \left(\frac{1-3p}{2}\right)\theta_{12}^*, \tag{14}$$

where $\theta_{12}^* = \theta_{12,max}$ if $1/3 \leq p \leq 1/2$ and $\theta_{12}^* = \theta_{12,min}$ if $0 \leq p < 1/3$. Moreover, distortion $D_{min}(\theta_1, \theta_2)$ can be achieved as long as $R \geq R_{min}(\theta_1,\theta_2) = \theta_1 + \theta_2 - \left(2 - h\left(\frac{1-p}{2}\right)\right)\theta_{12}^*$.

Proof: See Appendix C. ∎

The results in Proposition 4 can be seen as the counterpart of Corollary 1 for binary sources. In fact, they show that, for sufficiently large R, if $0 \leq p < 1/3$, the average Hamming distortion increases linearly with θ_{12} and thus we should set θ_{12} to the smallest possible value $\theta_{12,min}$; instead, if $1/3 \leq p \leq 1/2$, the optimal value of θ_{12} is the largest possible, namely, $\theta_{12,max}$.

Before we proceed to investigate the general rate-distortion function $R(D, \theta_1, \theta_2)$, we first derive the indirect rate-distortion function $R_1(D)$ for $T = S_1 \otimes S_2$ when only S_1 is observed at the encoder.

Lemma 3: The indirect rate-distortion function for $T = S_1 \otimes S_2$ is given by

$$R_1(D) = \begin{cases} \min\limits_{\frac{1-p-2D}{1-2p} \leq y \leq 1} h\left(D + y(1-p) + \frac{p-1}{2}\right) - \frac{1}{2}h(y) & \\ \quad -\frac{1}{2}h(2d + y(1-2p) + p - 1), \quad \frac{p}{2} < D \leq \frac{1-p}{2}, & \\ 0, \quad D \geq \frac{1-p}{2}, & \end{cases} \tag{15}$$

Proof: See Appendix D. ∎

By symmetry, the indirect rate-distortion function $R_2(D)$ for T when S_2 is observed at the encoder is also given by Lemma 3. The rate-distortion function $R_{12}(D)$ for variable T is instead given from standard results [1] as $R_{12}(D) = h((1-p)/2) - h(D)$ if $0 \leq D \leq (1-p)/2$, and $R_{12}(D) = 0$ if $D > (1-p)/2$.

Proposition 5: For a given sampling budget (θ_1, θ_2), the rate-distortion function for computing $T = S_1 \otimes S_2$ is given as

$$R(D,\theta_1,\theta_2) = \begin{cases} \min\limits_{\theta_{12}, D_3, D_{12}} \theta_{12}\left(h\left(\frac{1-p}{2}\right) - h\left(\frac{D_{12}}{\theta_{12}}\right)\right) & \\ \quad +(\theta_1+\theta_2-2\theta_{12})R_1\left(\frac{D_3}{\theta_1+\theta_2-2\theta_{12}}\right), & \\ \quad \text{if } D_{min}(\theta_1,\theta_2) \leq D < \frac{1-p}{2}, & \\ 0, \quad \text{if } D \geq \frac{1-p}{2}, & \end{cases} \tag{16}$$

where $D_{min}(\theta_1, \theta_2)$ is as given in Proposition 4 and the minimization is taken over all choices of θ_{12}, D_3 and D_{12} such that (1) is satisfied, $p(\theta_1 + \theta_2 - 2\theta_{12})/2 \leq D_3 \leq (1-p)(\theta_1+\theta_2-2\theta_{12})/2$, $p\theta_{12}/2 \leq D_{12} \leq (1-p)\theta_{12}/2$, and

$$D_3 + D_{12} + \left(\frac{1-p}{2}\right)(1 + \theta_{12} - \theta_1 - \theta_2) = D. \tag{17}$$

Proof: See Appendix E. ∎

C. Numerical Results

In this subsection, we numerically evaluate the distortion-rate tradeoff for computation of function $T = S_1 \otimes S_2$. Recall that for $T = S_1 \oplus S_2$, the optimal overlap fraction is always $\theta_{12,max}$. Fig. 5 and Fig. 6 plot the minimum average Hamming distortion D and the optimal overlap fraction θ_{12}^* for $\theta_1 = 0.5$, $\theta_2 = 0.75$, and $p = 0.1, 0.2, 0.4$. In Fig. 5, as predicted by Proposition 4, the minimum rate $R_{min}(\theta_1, \theta_2)$ that achieves distortion $D_{min}(\theta_1, \theta_2)$, is given by 0.9982, 0.9927, 0.69 for $p = 0.1, 0.2, 0.4$, respectively. It can be observed from Fig. 6, for $p = 0.4 > 1/3$, the optimal overlap fraction θ_{12}^* is equal to the maximum possible value $\theta_{12,max} = 0.5$, for any $0 \leq$

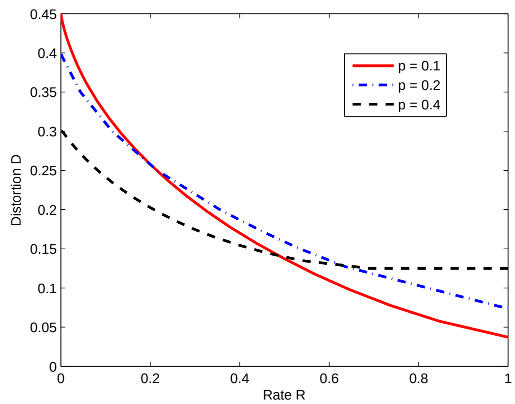

Figure 5. Distortion-rate function when $\theta_1 = 0.5$ and $\theta_2 = 0.75$, with p chosen to be $p = 0.1, 0.2, 0.4$, respectively.

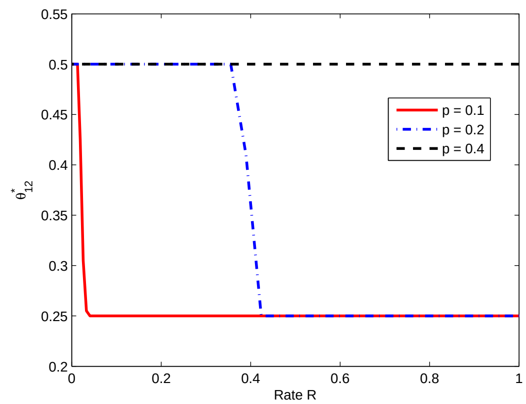

Figure 6. Optimal overlap fraction θ^*_{12} as a function of R when $\theta_1 = 0.5$ and $\theta_2 = 0.75$, with p chosen to be $p = 0.1, 0.2, 0.4$, respectively.

$R \leq 1$. However, for smaller probabilities $p = 0.1, 0.2$, the optimal overlap fraction equals to the maximum possible value $\theta_{12,max} = 0.5$ for sufficiently smaller rates and then drops to the minimum possible value $\theta_{12,min} = 0.25$ after R grows beyond a threshold. Moreover, the smaller the probability p is, the larger range of rates R over which the optimal overlap fraction θ^*_{12} is $\theta_{12,min} = 0.25$.

We note that with a larger p, it is easier to describe T directly, since $T \sim$ Bernoulli$((1-p)/2)$, but the indirect description of T based on S_1 or S_2 becomes more difficult since T becomes less correlated with S_1 or S_2.[3] This explains why the optimal overlap fraction should be chosen as the maximum possible value $\theta_{12,max} = 0.5$ when p is larger than $1/3$ (see the curve $p = 0.4$). In this sense, the regime $p \geq 1/3$ may be considered as the binary counterpart of the regime $\rho \leq 0$ for the Gaussian sum case in Section IV-B. For probabilities $p < 1/3$, the numerical results above imply that the optimal overlap depends on the link rate R. Similar to the Gaussian sum case when $0 < \rho \leq 1$ (Corollary 3), when R is sufficiently small, it remains optimal to choose the overlap fraction to be the maximum possible; however, as R grows sufficiently large, it is more advantageous to have the overlap fraction as small as possible, which is consistent with Proposition 4.

VI. Conclusions

In this paper, we have considered the problem of lossy compression for computing a function of correlated sources. Motivated by the fact that acquiring the information necessary for computation may be costly in sensor networks, we assumed that the encoder can only observe a fraction of the samples from each source according to a sampling strategy that is subject to design. The results highlight the dependence of the optimal sampling strategy on the function to be computed by the decoder, on the source correlation and on the link rate. Interesting future work includes investigation of related scenarios with side information or distributed source coding.

[3] The correlation between T and S_1 or S_2 is given by $\sqrt{(1-p)/(1+p)}$.

Appendix A
Proof of Proposition 1

Given $T = S_1$, we have the distortion rate functions $D_1(R) = D_{12}(R) = 2^{-2R}$ and $D_2(R) = 1 - \rho^2 + \rho^2 2^{-2R}$ [6]. In this case, applying Lemma 1, we obtain

$$\begin{aligned} & D(R, \theta_1, \theta_2, \theta_{12}) \\ &= \min_{R_1, R_2, R_{12} \geq 0} (\theta_1 - \theta_{12}) 2^{-\frac{2R_1}{\theta_1 - \theta_{12}}} + \theta_{12} 2^{-\frac{2R_{12}}{\theta_{12}}} + (\theta_2 - \theta_{12}) \\ &\qquad \cdot \left(1 - \rho^2 + \rho^2 2^{-\frac{2R_2}{\theta_2 - \theta_{12}}}\right) + (1 + \theta_{12} - \theta_1 - \theta_2) \qquad (18) \\ &= \min_{0 \leq R_2 \leq R} \theta_1 2^{-\frac{2(R-R_2)}{\theta_1}} + (\theta_2 - \theta_{12}) \rho^2 2^{-\frac{2R_2}{\theta_2 - \theta_{12}}} \\ &\qquad + 1 - \theta_1 - \rho^2 (\theta_2 - \theta_{12}), \qquad (19) \end{aligned}$$

where the minimization in (18) is under the constraint (4). Note that the optimization in (18) is equivalent to that in (19), since in any optimal solution, we have $R_{12}/\theta_{12} = R_1/(\theta_1 - \theta_{12})$ by the convexity of function 2^{-2r} for $r \geq 0$, and the condition (4) must be met with equality. It can be easily seen that function $D(R, \theta_1, \theta_2, \theta_{12})$ above is monotonically non-decreasing with respect to θ_{12}. Therefore, the optimal overlap is the minimum possible, which equals $\theta^*_{12} = \theta_{12,min}$. Moreover, the optimal rate R^*_2 that minimizes (19) can be obtained using standard Lagrangian methods similar to [8] as:

$$R^*_2 = \frac{\theta_2 - \theta^*_{12}}{\theta_1 + \theta_2 - \theta^*_{12}} \left(R - \frac{\theta_1}{2} \log_2 \frac{1}{\rho^2}\right)^+ . \qquad (20)$$

With the so obtained R^*_2, the results in Proposition 1 follows immediately.

Appendix B
Proof of Proposition 2

Applying Lemma 1 to this case, for a given sampling profile $(\theta_1, \theta_2, \theta_{12})$, we obtain the distortion-rate function as

$$\begin{aligned}
& D(R, \theta_1, \theta_2, \theta_{12}) \\
= & \min_{R_1, R_{12}, R_2} 2(\theta_1 - \theta_{12})(1+\rho)\left(1 - \tilde{\rho}^2 + \tilde{\rho}^2 2^{-\frac{2R_1}{\theta_1 - \theta_{12}}}\right) \\
& + 2(\theta_2 - \theta_{12})(1+\rho)\left(1 - \tilde{\rho}^2 + \tilde{\rho}^2 2^{-\frac{2R_2}{\theta_1 - \theta_{12}}}\right) \\
& + 2\theta_{12}(1+\rho)2^{-\frac{2R_{12}}{\theta_{12}}} + 2(1 + \theta_{12} - \theta_1 - \theta_2)(1+\rho)
\end{aligned} \tag{21}$$

$$\begin{aligned}
= & \min_{R_1, R_{12}, R_2} 2\theta_{12}(1+\rho)2^{-\frac{2R_{12}}{\theta_{12}}} + 2(1 + \theta_{12} - \theta_1 - \theta_2)(1+\rho) \\
& + 2(\theta_1 + \theta_2 - 2\theta_{12})(1+\rho)\left(1 - \tilde{\rho}^2 + \tilde{\rho}^2 2^{-\frac{2(R_1+R_2)}{\theta_1+\theta_2-2\theta_{12}}}\right),
\end{aligned} \tag{22}$$

where the constraint on R_1, R_2 and R_{12} is as in (4). Note that in the above, the problem in (21) is reduced to an equivalent problem in (22) since $R_1/(\theta_1 - \theta_{12}) = R_2/(\theta_2 - \theta_{12})$ holds in any optimal solution by the convexity of 2^{-2r} for $r \geq 0$. Moreover, it can be easily seen that the condition (4) must hold with equality, i.e., we have $R_1 + R_2 = R - R_{12}$. Substituting this and $\tilde{\rho} = \sqrt{(1+\rho)/2}$ to (22) and taking the minimum over all θ_{12} satisfying (1), we can obtain $D(R, \theta_1, \theta_2)$ as stated in the proposition.

Appendix C
Proof of Proposition 4

For any given sampling profile $(\theta_1, \theta_2, \theta_{12})$, in order to minimize the distortion with respect to R, we can take R to be arbitrarily large (in fact, given the binary alphabets, $R = 1$ suffices). With no rate limitations, it is easy to see that, during the θ_{12}-fraction, T can be computed at the encoder and described to the decoder losslessly with a rate equal to the entropy of T, $h((1-p)/2)$. During the $(\theta_1 - \theta_{12})$-fraction, only source S_1 is observed and can be described to the decoder losslessly with a rate $h(1/2) = 1$. Based on source S_1, the best estimate at the decoder is as follows: $\hat{T} = 0$ if $S_1 = 0$, and $\hat{T} = 1$ if $S_1 = 1$, leading to average Hamming distortion $p/2$. Similarly, during the $(\theta_2 - \theta_{12})$-fraction, the average Hamming distortion is also $p/2$. During the $(1+\theta_{12}-\theta_1-\theta_2)$-fraction, neither source S_1 nor source S_2 is observed. Since T is Bernoulli distributed with $(1-p)/2$, the best estimate is given by $\hat{t} = 0$, leading to average Hamming distortion $(1-p)/2$. Therefore, when there is no constraint on rate R, we have

$$\begin{aligned}
& D(R, \theta_1, \theta_2, \theta_{12}) \\
= & \frac{p}{2}(\theta_1 + \theta_2 - 2\theta_{12}) + \frac{1-p}{2}(1 + \theta_{12} - \theta_1 - \theta_2) \\
= & \frac{1-p}{2} + \left(p - \frac{1}{2}\right)(\theta_1 + \theta_2) + \frac{1-3p}{2}\theta_{12}.
\end{aligned} \tag{23}$$

It can be easily seen that, if $p \geq 1/3$, it is optimal to choose the maximum overlap fraction $\theta_{12}^* = \theta_{12,max}$; otherwise if $p < 1/3$, it is optimal to choose the minimum overlap fraction $\theta_{12}^* = \theta_{12,min}$. Substituting θ_{12}^* in (23), we obtain $D_{min}(\theta_1, \theta_2)$ as stated in the proposition. Finally, from the discussion above, it follows that, for any $R \geq R_{min}(\theta_1, \theta_2) = \theta_1 + \theta_2 - (2 - h((1-p)/2))\,\theta_{12}^*$, distortion $D_{min}(\theta_1, \theta_2)$ can be achieved at the decoder.

Appendix D
Proof of Lemma 3

The indirect rate-distortion function $R_1(D)$ for T is given by [2]

$$R_1(D) = \min_{p(\hat{t}|s_1): Ed(T,\hat{T}) \leq D} I(S_1; \hat{T}). \tag{24}$$

Let $p(\hat{t} = 1|s_1 = 0) = x$ and $p(\hat{t} = 1|s_1 = 1) = y$, where $0 \leq x \leq 1$ and $0 \leq y \leq 1$. Note that if we select $x = y = 0$, i.e., $\hat{T} = 0$ with probability 1, the average distortion $D = (1-p)/2$ is achievable at the decoder. Thus, for $D \geq (1-p)/2$, we have $R_1(D) = 0$. Moreover, from the proof of Proposition 4, it follows that $D \geq p/2$ must hold. For the nontrivial case $p/2 \leq D < (1-p)/2$, the expected distortion constraint can be written as

$$\begin{aligned}
E(d(T, \hat{T})) = & \frac{x}{2} + \frac{1-y}{2}(1-p) + \frac{y}{2}p \\
= & \frac{x + (2p-1)y + 1 - p}{2} \leq D,
\end{aligned} \tag{25}$$

and the mutual information $I(S_1; \hat{T})$ can be written as

$$\begin{aligned}
I(S_1; \hat{T}) &= H(\hat{T}) - H(\hat{T}|S_1) \\
&= h\left(\frac{x+y}{2}\right) - \frac{1}{2}h(x) - \frac{1}{2}h(y).
\end{aligned} \tag{26}$$

For any given y, considering the monotonicity of (26) with respect to x for $0 \leq x \leq 2d-(1-p)+(1-2p)y$, we can easily show that (26) is minimized at $x = 2d-(1-p)+(1-2p)y$, i.e., (25) is met with equality. Therefore, for $p/2 \leq D < (1-p)/2$, we can rewrite (24) as in (15) of the lemma.

Appendix E
Proof of Proposition 5

If we set $\hat{T} = 0$ at the decoder, the resulting Hamming distortion is $(1-p)/2$. Hence, for $D \geq (1-p)/2$, zero rate is required for description, i.e., $R(D, \theta_1, \theta_2) = 0$. For $D_{min}(\theta_1, \theta_2) \leq D < \frac{1-p}{2}$, for any given sampling profile $(\theta_1, \theta_2, \theta_{12})$, we can use Lemma 1 by setting $D_{1,min} = D_{2,min} = p/2$, $D_{12,min} = 0$ and $D_{max} = (1-p)/2$. Due to the convexity of $R_1(D)$, it is optimal to have $D_1/(\theta_1 - \theta_{12}) = D_2/(\theta_2 - \theta_{12})$ in any optimal solution. Moreover, with $D_{min}(\theta_1, \theta_2) \leq D < \frac{1-p}{2}$, for optimality, (7c) must be met with equality, i.e.,

$$D_1 + D_2 + D_{12} + \frac{(1 + \theta_{12} - \theta_1 - \theta_2)(1-p)}{2} = D, \tag{27}$$

and D_1, D_{12} and D_2 must be such that $D_1/(\theta_1 - \theta_{12})$, D_{12}/θ_{12} and $D_2/(\theta_2 - \theta_{12})$ are all less than or equal to $D_{max} = (1-p)/2$. If we let $D_3 = D_1 + D_2$, then D_3 satisfies

$$\frac{p(\theta_1 + \theta_2 - 2\theta_{12})}{2} \leq D_3 \leq \frac{(1-p)(\theta_1 + \theta_2 - 2\theta_{12})}{2}. \tag{28}$$

Finally, taking the minimum of $R(D, \theta_1, \theta_2, \theta_{12})$ over all θ_{12} satisfying (1), we obtain $R(D, \theta_1, \theta_2)$ as in the proposition for $D_{min}(\theta_1, \theta_2) \leq D < \frac{1-p}{2}$.

REFERENCES

[1] T. M. Cover and J. A. Thomas, *Elements of Information Theory 2nd Edition*. Wiley-Interscience, 2006.

[2] H. Witsenhausen, "Indirect rate distortion problems," *IEEE Trans. on Information Theory*, vol. 26, no. 5, pp. 518–521, 1980.

[3] A. E. Gamal and Y. H. Kim, *Network Information Theorey*. Cambridge University Press, 2012.

[4] X. Liu, O. Simeone, and E. Erkip, "Energy-efficient sensing and communication of parallel Gaussian sources," in *Proc. of IEEE ISIT*, Cambridge, MA, 2012, to appear.

[5] L. Zhao, Y. K. Chia, and T. Weissman, "Compression with actions," 2012. [Online]. Available: http://arxiv.org/abs/1204.2331

[6] Y. Oohama, "Gaussian multiterminal source coding," *IEEE Trans. on Information Theory*, vol. 43, no. 6, pp. 1912–1923, Nov. 1997.

[7] H. Yamamoto, "Wyner-Ziv theory for a general function of the correlated sources," *IEEE Trans. on Information Theory*, vol. 28, no. 5, pp. 803–807, Sep. 1982.

[8] X. Liu, O. Simeone, and E. Erkip, "Energy-efficient sensing and communication of parallel Gaussian sources," 2012. [Online]. Available: http://arxiv.org/pdf/1202.4098.pdf