Energy Efficient Estimation of Gaussian Sources Over Inhomogeneous Gaussian MAC Channels

Shuangqing Wei, Rajgopal Kannan, Sitharama Iyengar and Nageswara S. Rao

Abstract

It has been shown lately the optimality of uncoded transmission in estimating Gaussian sources over homogeneous/symmetric Gaussian multiple access channels (MAC) using multiple sensors. It remains, however, unclear whether it still holds for any arbitrary networks and/or with high channel signal-to-noise ratio (SNR) and high signal-to-measurement-noise ratio (SMNR). In this paper, we first provide a joint source and channel coding approach in estimating Gaussian sources over Gaussian MAC channels, as well as its sufficient and necessary condition in restoring Gaussian sources with a prescribed distortion value. An interesting relationship between our proposed joint approach with a more straightforward separate source and channel coding scheme is then established. Further comparison of these two schemes with the uncoded approach reveals a lacking of a consistent ordering of these three strategies in terms of the total transmission power consumption under a distortion constraint for arbitrary in-homogeneous networks. We then formulate constrained power minimization problems and transform them to relaxed convex geometric programming problems, whose numerical results exhibit that either separate or uncoded scheme becomes dominant over a linear topology network. In addition, we prove that the optimal decoding order to minimize the total transmission powers for both source and channel coding parts is solely subject to the ranking of MAC channel qualities, and has nothing to do with the ranking of measurement qualities. Finally, asymptotic results for homogeneous networks are obtained which not only confirm the existing optimality of the uncoded approach, but also show that the asymptotic SNR exponents of these three approaches are all the same. Moreover, the proposed joint approach share the same asymptotic ratio with respect to high SNR and high SMNR as the uncoded scheme.

1 Introduction

Recent years have witnessed a tremendous growth of interests in wireless ad hoc and sensor networks from both academia and industry, due to their ease of implementation, infrastructure-less nature,

[1]S. Wei is with the Department of ECE, Louisiana State University, Baton Rouge
[2]R. Kannan and S. Iyengar are with the Department of CS, Louisiana State University, Baton Rouge
[3]N. Rao is with the Computer Science and Mathematics Division, Oak Ridge National Laboratory
[4]This work is funded in part from DOE-ORNL (Sensornets Program Sept. 2006- 2008)

as well as the huge potentials in civil and military applications. In many instances, sensor nodes are deployed to serve a common purpose such as surveillance and monitoring environments. One of the major design metrics is to maximize the lifetime of a sensor net while meeting the constraints imposed by the quality of reconstruction, such as the resulting ultimate distortion measure when data collected by sensors are fused to construct an estimate of the monitored source. A critical factor affecting the lifetime of a sensor net is the amount of total power expenditure that senor nodes spend on transmitting their measurements to a fusion center. The power consumptions are closely related with the way sensors collecting and processing measurements, as well as the communication link quality between sensor nodes and the fusion center.

In this paper, assuming L sensor nodes send measurements of a Gaussian source to a fusion center via a one-hop interference limited wireless link, we investigate the issue of power allocations across sensors with and without local compression and channel coding. A similar system model for Gaussian sensor networks has also been adopted recently by [1, 2, 3] and [4]. In [1, 2], the authors investigated joint source-channel coding paradigm and analyzed how distortion scales with respect to the size of the network. They showed that uncoded transmission achieves the lower bound of the mean squared error distortion as the number of sensors grow to infinity in *symmetric* networks. However, no exact source and channel coding schemes are provided for general system settings other than the uncoded scheme. In [3], the "exact" optimality of uncoded transmission is proved even for the homogeneous Gaussian networks with *finite* number of sensor nodes. As pointed out in [3], it remains unclear though what approach is more favorable when a system becomes non-symmetric with a finite number of sensors.

The objectives of this paper are two folds. First, we will propose a joint source-channel coding approach and then establish its relationship with the separate source and channel coding strategy. Second, we will investigate the optimal rate and power allocation strategy in order to minimize the total transmission power under the constraint that the mean squared error value in estimating the Gaussian source remotely is no greater than a prescribed threshold. In particular, we will compare the resulting total power consumptions of three distinct processing schemes, namely, joint source and channel coding, separate source and channel coding and uncoded amplify-and-forward approaches for in-homogeneous networks, and demonstrate the well known result of the optimality of uncoded approach for estimating Gaussian sources in homogeneous networks does not always hold in inhomogeneous networks.

Our contributions in this paper can be summarized as follows:

- A joint source and channel coding approach is proposed, whose achievable rate region is obtained. An interesting relationship between this approach and separate source and channel coding approach is then established.

- Optimal decoding order for both joint and separate source channel coding is found which is only a function of MAC channel ranking order, and has nothing to do with the power level of source measurement noise, when we intend to minimize the total transmission power.

- Relaxed geometric programming problems are formulated in order to compare three schemes. Numerical results demonstrate the uncoded transmission is not always the best option. The ordering of the three schemes is highly dependent on relative channel qualities, measurement noise levels, as well as the distortion threshold.

- Asymptotic results for large size homogeneous networks with finite SNR and SMNR are obtained which show the optimality of uncoded transmission from another perspective. More importantly, a condition is found under which the scaling factor of received channel SNR versus signal-to-distortion-noise-ratio (SDNR), as SMNR grows to infinity, of joint approach is equal to that of the uncoded scheme. In addition, we prove the SNR exponents of all three schemes are the same.

The paper is organized as follows. System model is set up in Section 2. A joint source and channel coding scheme is proposed in Section 3, in which we establish its achievable rate region, as well as its relationship with the separate source and channel coding scheme. In order to compare joint and separate approaches, the formulated total power minimization problems are solved using geometric programming approach in Section 4, where we also obtain the optimal decoding order for non-homogeneous networks. Uncoded approach is revisited in Section 5 from the perspective of making comparisons with the former two approaches. In Section 6, we compare the aforementioned three schemes in asymptotic region for homogeneous networks. Finally, numerical results of our comparisons for arbitrary two-node networks are presented in Section 7.

2 System Model

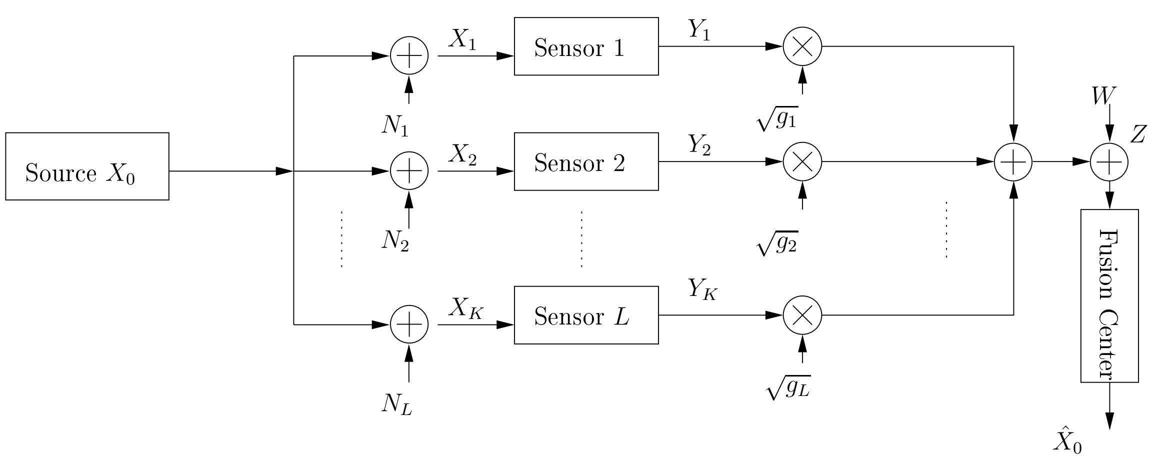


Figure 1: System model

Assume L sensor nodes observe a common Gaussian source $X_0[i], i = 1, \cdots, n$, where $X_0[i] \sim \mathcal{N}(0, \sigma_S^2)$ are identically and independently distributed Gaussian random variables with mean zero and variance σ_S^2. The measurements $X_j[i] = X_0[i] + N_j[i]$, $j = 1, \cdots, L$ from L sensors experience independent additive Gaussian measurement noise $N_j[i] \sim \mathcal{N}(0, \sigma_{N_j}^2)$, where independence is assumed to hold across both space and time. Let $Y_j[i]$ denote the transmitted signal from sensor j at time i, which satisfies an average power constraint:

$$\frac{1}{n}\sum_{i=1}^{n} |Y_j[i]|^2 \leq P_j,\ j = 1 \cdots, L. \tag{1}$$

.

The processed signals $\{Y_j[i]\}$ then go through a Gaussian multiple access channel and are superposed at a fusion center resulting in $Z[i] = \sum_{j=1}^{L} \sqrt{g_j} Y_j[i] + W[i]$, where $W[i] \sim \mathcal{N}(0, \sigma_W^2)$ are white Gaussian noise introduced at the fusion center and assumed independent with $N_j[i]$. Coefficients $g_j, j = 1, \cdots, L$ capture the underlying channel pathloss and fading from sensors to the fusion center. In this paper, we assume coherent fusion is conducted in the sense that g_j are assumed perfectly known by the fusion center. Upon receiving $\{Z[i]\}$, the fusion center constructs an estimate $\{\hat{X}_0[i]\}$ of $\{X_0[i]\}$ such that the average mean squared error $D_E \triangleq \lim_{n\to\infty} \frac{1}{n} \sum_{i=1}^{n} E\left|X_0[i] - \hat{X}_0[i]\right|^2$ of the estimation satisfies $D_E \leq D$, where D is a prescribed upper bound for estimation error.

What interests us in this paper is power efficient schemes to estimate the Gaussian source remotely with a prescribed mean squared error. Three approaches, namely, joint source and channel coding, separate source and channel coding, and uncoded amplify-and-forward schemes, will be investigated in the sequel.

3 Joint Source-Channel Based Fusion and Its Relationship with Separate Source and Channel Coding

In [5], a joint source and channel coding scheme is proposed to estimate two correlated Gaussian sources remotely at a fusion center where measurements from two sensors are received through a Gaussian MAC channel. Achievable rate region was obtained as a function of the required distortion tuple in restoring two correlated sources. Inspired by their work, we, in this section, will first develop an achievable rate region for our proposed joint source-channel coding (JSCC) approach for any arbitrary network with $L > 1$ sensor nodes and then demonstrate an interesting relationship of JSCC with a separate source and channel coding scheme (SSCC) which is a straightforward combination of the recent findings on CEO problem [6] and traditional MAC channel coding [7] with independent sources.

3.1 Achievable Rate Region for Distributed Joint Source-Channel Coding in Estimation of Gaussian Sources

Let $\tilde{R}_j, j = 1, \cdots, L$ denote the compression rate at the j-th sensor. There are total $2^{n\tilde{R}_j}$ source codewords $\mathbf{U}_j = \{\mathbf{U}_j^{(k)}, k = 1, \cdots, 2^{n\tilde{R}_j}\}$ from which sensor j selects $\mathbf{U}_j^{(m_j)} = \{U_j^{(m_j)}[i], i = 1, \cdots, n\}$ to represent $\mathbf{X_j} = \{X_j[i], i = 1, \cdots, n\}$. The joint approach we propose here is to let each sensor directly transmit a scaled version of a source codeword $\mathbf{U}_j^{(m_j)}$. The scaling factor introduced herein is to maintain the average transmission power P_j by sensor j, $j = 1, \cdots, L$. Since L sensors see the same Gaussian source with independent measurement noise, L quantization vectors $\{\mathbf{U}_j^{(m_j)}, j = 1, \cdots, L\}$ are correlated. As a result, the decoding at fusion center needs to take into account of such correlation when it performs joint decoding of these L codewords. The decoded source/channel codeword $\hat{\mathbf{U}}_j^{(m_j)}$ are then linearly combined to obtain an MMSE estimate $\{\hat{X}_0[i], i = 1, \cdots, n\}$ of the Gaussian source $\{X_0[i], i = 1, \cdots, n\}$.

We are interested in deriving the achievable region of rate tuples $\{\tilde{R}_j, j = 1, \cdots, L\}$ such that $2^{n\tilde{R}_j}, j = 1, \cdots, L$ source/channel codewords can be decoded with asymptotic zero error and the mean squared error D_E satisfies $D_E \leq D$.

Theorem 1. *To make $D_E \leq D$, $\tilde{R}_i$ satisfy*

$$\tilde{R}_i = I(X_i; U_i) = r_i + \frac{1}{2}\log\left[1 + \frac{\sigma_S^2}{\sigma_{N_i}^2}\left(1 - 2^{-2r_i}\right)\right], \ i = 1, \cdots, L \tag{2}$$

where $r_i \geq 0, i = 1, \cdots, L$ are chosen based on

$$\frac{1}{D_E} = \frac{1}{\sigma_S^2} + \sum_{k=1}^{L} \frac{1 - 2^{-2r_k}}{\sigma_{N_k}^2} \geq \frac{1}{D} \tag{3}$$

and $I(X_j; U_j)$ denotes the mutual information between X_j and a Gaussian random variable U_j, which is associated with X_j by

$$U_j = X_j + V_j, j = 1, \cdots, L \tag{4}$$

where V_j, independent of X_j, are independent Gaussian random variables with mean 0 and variance $\sigma_{V_i}^2 = \sigma_{N_i}^2/(2^{r_i} - 1)$.

Proof. The proof is a straightforward application of the techniques used in proving Lemma 10 in [6]. For brevity, we only provide an outline here.

We quantize $\{X_j[i]\}$ with $2^{n\tilde{R}_j}$ Gaussian vectors $\{\hat{U}_j[i]\}$ such that the source symbol $X_j[i]$ can be constructed from the quantized symbol through a test channel [7]: $X_j[i] = \hat{U}_j[i] + \hat{V}_j[i]$, where $\hat{V}_j[i]$ is a Gaussian random variable with mean zero and variance $2^{-2\tilde{R}_j}\sigma_{X_j}^2$, which is independent of $\hat{U}_j[i] \sim \mathcal{N}\left(0, (1 - 2^{-2\tilde{R}_j})\sigma_{X_j}^2\right)$. Equivalently, we can also represent $\hat{U}_j[i]$ as $\hat{U}_j[i] = \alpha X_j[i] + \tilde{V}_j[i]$, where α is the linear-MMSE estimate coefficient and $\tilde{V}_j[i]$ is the resultant estimation error. By orthogonal principle, we have

$$\alpha = \frac{\sigma^2_{\hat{U}_j}}{\sigma^2_{X_j}} = \left(1 - 2^{-2\tilde{R}_j}\right)$$

and $\tilde{V}_j[i]$ is a Gaussian variable independent of $X_j[i]$ with mean zero and variance $2^{-2\tilde{R}_j}\left(1 - 2^{-2\tilde{R}_j}\right)\sigma^2_{X_j}$. Therefore, after normalization, we obtain

$$U_j = \frac{1}{\alpha}\hat{U}_j = X_j + \frac{1}{\alpha}\tilde{V}_j = X_j + V_j \tag{5}$$

where $V_j \sim \mathcal{N}\left(0, \sigma^2_{X_j}/(2^{2\tilde{R}_j} - 1)\right)$. We introduce variables r_j such that

$$2^{2\tilde{R}_j} - 1 = \left(2^{2r_j} - 1\right)\frac{\sigma^2_{X_j}}{\sigma^2_{N_j}} \tag{6}$$

which proves (4). We can also see that r_j is actually the conditional mutual information between X_j and U_j given X_0, i.e. $r_j = I\left(X_j; U_j|X_0\right)$. Since $I(X_j;U_j) = H(U_j) - H(V_j)$, it is then straightforward to show that (2) holds.

Given $U_j = X_j + V_j$ and $X_j = X_0 + N_j$, where N_j and V_j are independent, we can construct the LMMSE estimate of X_0 by $\hat{X}_0 = \sum_{j=1}^{L}\beta_j U_j$, where coefficients β_j can be determined again using Orthogonal Principle. Based on Equations (95) and (96) in [6], we obtain the desired result for the mean squared error in (3).

□

From the proof of Theorem 1, it can be seen that U_i and U_j are correlated due to the correlation between X_i and X_j, whose correlation can be captured by $\rho_{i,j}$, the covariance coefficient between X_i and X_j, which can be computed as

$$\rho_{i,j} = \frac{E[X_i X_j]}{\sqrt{E|X_i|^2 E|X_j|^2}} = \frac{\sigma^2_S}{\sqrt{(\sigma^2_S + \sigma^2_{N_i})(\sigma^2_S + \sigma^2_{N_j})}} \tag{7}$$

The covariance coefficient $\tilde{\rho}_{i,j}$ between U_i and U_j can be obtained accordingly as:

$$\tilde{\rho}_{i,j} = \rho_{i,j}\sqrt{(1 - 2^{-2\tilde{R}_i})(1 - 2^{-2\tilde{R}_j})}. \tag{8}$$

After substituting $\tilde{R}_i$ determined in Theorem 1 into it, we obtain

$$\tilde{\rho}_{i,j} = \sqrt{\frac{q_i q_j}{(1 + q_i)(1 + q_j)}}, \tag{9}$$

where $q_i = \frac{\sigma_S^2}{\sigma_{N_i}^2}(1-2^{-2r_i})$.

For any given subset $S \subseteq \{1,2,\cdots,L\}$, define vectors $\mathbf{U}(\mathbf{S}) = \left[U_{\pi_1},\cdots,U_{\pi_{|S|}}\right]$ and $\mathbf{U}(\mathbf{S^c}) = \left[U_{\pi_{|S|+1}},\cdots,U_{\pi_L}\right]$, where π is an arbitrary ordering of the L indexes. The covariance matrix of $\mathbf{U} = [\mathbf{U}(\mathbf{S}),\mathbf{U}(\mathbf{S^c})]^T$ can thus be decomposed as

$$\mathbf{\Sigma_U} = E\left[\mathbf{U}\mathbf{U}^T\right] = \begin{bmatrix} \mathbf{\Sigma_S} & \mathbf{\Sigma_{S,S^c}} \\ \mathbf{\Sigma_{S^c,S}} & \mathbf{\Sigma_{S^c}} \end{bmatrix}, \tag{10}$$

where $\mathbf{\Sigma_S}$, $\mathbf{\Sigma_{S^c}}$, $\mathbf{\Sigma_{S,S^c}}$ denote the auto- and cross-covariance matrices of $\mathbf{U}(\mathbf{S})$ and $\mathbf{U}(\mathbf{S^c})$. The entries of $\mathbf{\Sigma_U}$ are $(\mathbf{\Sigma_U})_{i,j} = \tilde{\rho}_{i,j}\sqrt{P_iP_j}$ for $i \neq j$ and $(\mathbf{\Sigma_U})_{i,i} = P_i$, $i,j \in \{1,\cdots,L\}$, where $\tilde{\rho}_{i,j}$ is obtained in (9).

After each sensor maps the observation vector to U_j, an additional scaling factor $\gamma_j = \sqrt{\frac{P_j}{\sigma_{U_j}^2}}$ is imposed on U_j, where $\sigma_{U_j}^2 = \sigma_{X_j}^2 / \left(1-2^{-2\tilde{R}_j}\right)$ in order to keep the average transmission power of $Y_j[i] = \gamma_j U_j[i]$ as P_j. The received signal at the fusion center can thus be written as

$$Z[i] = \sum_{j=1}^{L} \gamma_j U_j[i]\sqrt{g_j} + W[i] \tag{11}$$

Theorem 2. *Given the received signal $Z[i], i = 1,\cdots,n$ in (11), the quantization rate $\tilde{R}_i, i = 1,\cdots,L$ obtained in Theorem 1 satisfies the following inequalities in order to restore X_0 at the fusion center with distortion no less than D:*

$$\begin{aligned}\tilde{R}(S) \leq & \sum_{i=1}^{|S|-1} I\left(U_{\pi_i};U_{\pi_{i+1}}^{\pi_{|S|}}\right) + I\left(U(S);U(S^c)\right) \\ & + I\left(U(S);Z|U(S^c)\right), \forall S \subseteq \{1,\cdots,L\},\end{aligned} \tag{12}$$

where $\tilde{R}(S) = \sum_{i\in S}\tilde{R}_i$, $U(S) = \{U_i, i \in S\}$, S^c is the complementary set of S, $\{\pi_1,\cdots,\pi_{|S|}\}$ is an arbitrary permutation of S, and $U_{\pi_{i+1}}^{\pi_{|S|}} = \left[U_{\pi_{i+1}},\cdots,U_{\pi_{|S|}}\right]$.

Proof. The proof follows the footsteps of the one proving achievability of the capacity region for regular MAC channel with independent channel codewords. The difference here is that we need to take into account the correlations of the channel inputs from each user when the joint typical sequence technique is used to compute the upper bound of the probability of various error events. The details are deferred to the Appendix A. □

It can be easily seen that when inputs to the channel are independent, the first and second terms in (12) vanish and consequently the inequality reduces to the one characterizing the capacity region for MAC channels with independent inputs [7, Chap.15.3].

Next, we prove a sequence of lemmas in order to establish a connection between the JSCC and SSCC approaches.

Lemma 1. *Given $U_j = X_j + V_j, j = 1, 2$ as in (4), $U_2 \to X_2 \to X_1 \to U_1$ forms a Markov chain. As a result, we have*

$$I(U_2; X_2|U_1) = I(U_2; X_2) - I(U_1; U_2) \quad (13)$$

$$I(U_1; X_1|U_2) = I(U_1; X_1) - I(U_2; U_1) \quad (14)$$

$$I(U_1, U_2; X_1, X_2) = I(X_1; U_1) + I(X_2; U_2) - I(U_1; U_2) \quad (15)$$

Proof. See Appendix B. □

Lemma 2. *For $U_j = X_j + V_j$, $j = 1, \cdots, L$, the following relation of mutual information holds*

$$I\left(U(S); X(S)\right) = \sum_{i \in S} I(U_i; X_i) - \sum_{i=1}^{|S|-1} I\left(U_{\pi_i}; U_{\pi_{i+1}}^{\pi_{|S|}}\right), \forall S \subseteq \{1, \cdots, L\}. \quad (16)$$

Proof. WLOG, consider $S = \{1, 2, \cdots, s\}$. Define $\tilde{\mathbf{U}}_{\mathbf{2}} = [U_2, U_3, \cdots, U_s]$ and $\tilde{\mathbf{X}}_{\mathbf{2}} = [X_2, X_3, \cdots, X_s]$. Apparently, $\tilde{\mathbf{U}}_{\mathbf{2}} \to \tilde{\mathbf{X}}_{\mathbf{2}} \to X_1 \to U_1$ forms a Markov chain. From (15), we immediately obtain:

$$I\left[U(S); X(S)\right] = I(X_1; U_1) + I(\tilde{\mathbf{U}}_{\mathbf{2}}; \tilde{\mathbf{X}}_{\mathbf{2}}) - I(U_1; \tilde{\mathbf{U}}_{\mathbf{2}}) \quad (17)$$

Using same idea, it can be shown that

$$I(\tilde{\mathbf{U}}_{\mathbf{2}}; \tilde{\mathbf{X}}_{\mathbf{2}}) = I(U_2; X_2) + I(U_3 \cdots, U_s; X_3, \cdots, X_s) - I(U_2; U_3^s), \quad (18)$$

where $I(U_3 \cdots, U_s; X_3, \cdots, X_s)$ can be decomposed in a similar manner. Such decomposition can be conducted iteratively until we reach

$$I(X_{s-1}, X_s; U_{s-1}, U_s) = I(X_{s-1}; U_{s-1}) + I(U_s; X_s) - I(U_s; U_{s-1}) \quad (19)$$

Combining all iterations yields the desired result:

$$I\left(U(S); X(S)\right) = \sum_{i=1}^{s} I(U_i; X_i) - \sum_{i=1}^{s-1} I\left(U_i; U_{i+1}^s\right) \quad (20)$$

As the whole derivation does not rely on the exact order of $\{1, \cdots, s\}$, we thus complete the proof of Lemma 2. □

Lemma 3. *For the same $U(S)$ and $X(S)$ as in Lemma 1, we have*

$$\begin{aligned} I\left(U(S);X(S)\right)-I\left(U(S);U(S^c)\right) &= I\left[U(S);X(S)|U(S^c)\right] \\ &= I\left[U(S);X_0|U(S^c)\right]+\sum_{i\in S} I\left[U_i;X_i|X_0\right] \end{aligned} \tag{21}$$

Proof. See appendix □

Theorem 3. *When each sensor performs independent vector quantization and subsequently transmits the resulting scaled quantization vector through a Gaussian MAC channel, to reconstruct the Gaussian source at fusion center with distortion no greater than D, the necessary and sufficient condition is for any subset $S \subseteq \{1,2,\cdots,L\}$, the following inequality holds*

$$I\left[U(S);X_0|U(S^c)\right]+\sum_{i\in S} I\left[U_i;X_i|X_0\right] \le I\left[U(S);Z|U(S^c)\right] \tag{22}$$

where

$$LHS = -\frac{1}{2}\log\left[\frac{D_E}{\sigma_S^2}+\sum_{i\in S^c}\frac{D_E}{\sigma_{N_i}^2}\left(1-2^{-2r_i}\right)\right]+\sum_{i\in S} r_i \tag{23}$$

and

$$RHS = \frac{1}{2}\log\left\{1+\frac{1}{\sigma_W^2}\sqrt{\mathbf{g}}(S)^T\mathbf{Q}_{\mathbf{\Sigma_S}}\sqrt{\mathbf{g}}(S)\right\} \tag{24}$$

with $\sqrt{\mathbf{g}}(S)^T = \left[\sqrt{g_i}, i\in S\right]$ *and* $\mathbf{Q}_{\mathbf{\Sigma_S}} = \mathbf{\Sigma_S} - \mathbf{\Sigma_{S,S^c}}\mathbf{\Sigma_{S^c}}^{-1}\mathbf{\Sigma_{S^c,S}}$. *The auto- and cross-covariance matrices $\mathbf{\Sigma_S}$, $\mathbf{\Sigma_{S^c}}$, $\mathbf{\Sigma_{S,S^c}}$ and $\mathbf{\Sigma_{S^c,S}}$ are defined as in (10).*

Proof. To construct an estimate of X_0 at a fusion center with distortion no greater than D is equivalent to requiring that the minimum compression rate $\tilde{R}_i$ satisfies $\tilde{R}_i = I(X_i;U_i)$, as required by local vector quantization, and that $\tilde{R}_i, i=1,\cdots,L$ are in the region determined in Theorem 2. Consequently, the conditions are translated to

$$\begin{aligned} \sum_{i\in S} I(X_i;U_i) \le \sum_{i=1}^{|S|-1} I\left(U_{\pi_i};U_{\pi_{i+1}}^{\pi_{|S|}}\right) \\ +I\left(U(S);U(S^c)\right)+I\left(U(S);Z|U(S^c)\right) \end{aligned} \tag{25}$$

From Lemma 2 and Lemma 3, this condition is equivalent to

$$I\left[U(S);X_0|U(S^c)\right]+\sum_{i\in S} I\left[U_i;X_i|X_0\right] \le I\left[U(S);Z|U(S^c)\right]. \tag{26}$$

Define $r_i = I\left(X_i; U_i | X_0\right)$. Then it is straightforward to show that the LHS of (26) is equal to that in (23) by computing the mean squared error of estimating X_0 using $U(S)$ or $[U(S), U(S^c)]$ [6], which is

$$E\left[|X_0|^2 | U(S)\right] = \left[\frac{1}{\sigma_S^2} + \sum_{i \in S} \frac{1}{\sigma_{N_i}^2}\left(1 - 2^{-2r_i}\right)\right]^{-1} \tag{27}$$

Given $I\left[U(S); Z | U(S^c)\right] = H\left[Z | U(S^c)\right] - H\left[Z | U(S), U(S^c)\right]$ and $\mathbf{U}$ and Z are Gaussian random vector/variables, it is sufficient to get the conditional variance of Z given the vector $U(S^c)$. This can be boiled down to finding the conditional variance of $\sum_{i \in S} \sqrt{g_i}\gamma_i U_i$ given $U(S^c)$ as $Z = \sum_{i=1}^{L} \sqrt{g_i}\gamma_i U_i + W$.

Based on Theorem 3 in [8], we have

$$\mathrm{Cov}\left[U(S) | U(S^c)\right] = \mathbf{\Sigma_S} - \mathbf{\Sigma_{S,S^c}} \mathbf{\Sigma_{S^c}}^{-1} \mathbf{\Sigma_{S^c,S}} \tag{28}$$

Therefore,

$$\mathrm{Var}\left(\sum_{i \in S} \sqrt{g_i}\gamma_i U_i\right) = \sqrt{\mathbf{g}}(S)^T \mathbf{Q}_{\mathbf{\Sigma_S}} \sqrt{\mathbf{g}}(S). \tag{29}$$

The entropy can thus be computed accordingly yielding

$$\begin{aligned} H\left[Z | U(S^c)\right] &= \frac{1}{2}\log\left[2\pi e\left(\sigma_W^2 + \sqrt{\mathbf{g}}(S)^T \mathbf{Q}_{\mathbf{\Sigma_S}} \sqrt{\mathbf{g}}(S)\right)\right] \\ H\left[Z | U(S^c), U(S)\right] &= \frac{1}{2}\log\left(2\pi e \sigma_W^2\right) \end{aligned} \tag{30}$$

which leads to (24), and hence completes the proof. □

3.2 Relationship With Separate Source-Channel Coding Approach

If we look closely at (22) and (23), we can easily see that the LHS of the achievable rate region for the JSCC approach actually characterizes the rate-distortion region for Gaussian sources with conditionally independent (CI) condition [6].

Under the CI assumption, distributed source coding at sensors includes two steps. The first step is the same as in JSCC, in which an independent vector quantization for Gaussian source at each sensor is conducted with respect to the observed signal $\mathbf{X_j} = \{X_j[i], i = 1, \cdots, n\}$, which generates a vector $\mathbf{U_j^k} = \{U_j^k[i], i = 1, \cdots, n\}$, $k \in \{1, \cdots, 2^{n\tilde{R}_j}\}$. In the second step, those indexes of k_j are further compressed using Slepian-Wolf's random binning approach [6, 9]. Consequently, there are 2^{nR_j} bins for sensor j, which contain all representation vectors $\mathbf{U_j^k}$ of measurements $\mathbf{X_j}$. It was shown in [6] that R_j satisfy: $\sum_{j \in S} R_j \geq I\left[U(S); X_0 | U(S^c)\right] + \sum_{i \in S} I\left[U_i; X_i | X_0\right]$, for all $S \subseteq \{1, 2, \cdots, L\}$ in order to restore X remotely with distortion no greater than D.

For SSCC, to send indexes of bins correctly to the fusion center, independent Gaussian codewords $\{Y_j[i] \sim \mathcal{N}(0, P_j), i = 1, \cdots, n\}$ for $j = 1, \cdots, L$ for each bin index are generated at L sensors. To ensure indexes are correctly decoded at the fusion center, the rate tuple $\{R_i, i = 1, \cdots, L\}$ should also be contained in the capacity region of Gaussian MAC channel with independent channel inputs under power constraints $\{P_j, j = 1, \cdots, L\}$. The region is characterized by $\sum_{i \in S} R_i \leq \frac{1}{2}\log\left[1 + \sum_{j \in S} \frac{P_j g_j}{\sigma_W^2}\right]$, for all $S \subseteq \{1, 2, \cdots, L\}$.

The data processing at the fusion center consists of three phases. In the first phase, channel decoding is performed to recover the indexes of bins containing $\{\mathbf{U_j^k}, j = 1, \cdots, L\}$. In the second phase, joint typical sequences $\{\mathbf{U_j^k}\}$ are obtained from L bins whose indexes are restored. In the last phase, $\{\mathbf{U_j^k}\}$ are linearly combined to estimate the source vector $\{X_0[i]\}$ under the minimum mean squared error (MMSE) criterion.

Under SSCC, we can therefore obtain the sufficient and necessary condition for restoring X_0 with MSE no greater than D:

$$
\begin{aligned}
& I\left[U(S); X_0 | U(S^c)\right] + \sum_{i \in S} I\left[U_i; X_i | X_0\right] \\
& \leq \frac{1}{2}\log\left[1 + \sum_{j \in S} \frac{P_j g_j}{\sigma_W^2}\right], \forall S \subseteq \{1, 2, \cdots, L\}.
\end{aligned} \tag{31}
$$

In general, we cannot say which approach, JSCC or SSCC, is better in terms of the size of rate region. This can be seen more clearly when we look at a particular case for $L = 2$. When there are only two sensors, to reconstruct $\{X_0[i]\}$ with a distortion no greater than D using JSCC or SSCC proposed as above, the transmission powers P_1 and P_2, as well as r_1 and r_2 satisfy:

$$
\begin{aligned}
& r_1 - \frac{1}{2}\log\left\{\frac{D_E}{\sigma_S^2} + \frac{D_E}{\sigma_{N_2}^2}\left(1 - 2^{-2r_2}\right)\right\} \\
& \leq \frac{1}{2}\log\left(1 + \frac{P_1 g_1 (1 - \tilde{\rho}_{1,2}^2)}{\sigma_W^2}\right)
\end{aligned} \tag{32}
$$

$$
\begin{aligned}
& r_2 - \frac{1}{2}\log\left\{\frac{D_E}{\sigma_S^2} + \frac{D_E}{\sigma_{N_1}^2}\left(1 - 2^{-2r_1}\right)\right\} \\
& \leq \frac{1}{2}\log\left(1 + \frac{P_2 g_2 (1 - \tilde{\rho}_{1,2}^2)}{\sigma_W^2}\right)
\end{aligned} \tag{33}
$$

$$
\begin{aligned}
& r_1 + r_2 + \frac{1}{2}\log\left(\frac{\sigma_S^2}{D_E}\right) \\
& \leq \frac{1}{2}\log\left(1 + \frac{P_2 g_2 + P_1 g_1 + 2\tilde{\rho}_{1,2}\sqrt{P_1 g_1 P_2 g_2}}{\sigma_W^2}\right)
\end{aligned} \tag{34}
$$

where $\tilde{\rho}_{1,2}$ denotes the covariance coefficient between U_i and U_j, which is zero for SSCC and

$$\tilde{\rho}_{1,2} = \sqrt{\frac{q_1 q_2}{(1+q_1)(1+q_2)}}, \tag{35}$$

for JSCC, as obtained in (9) with r_i satisfying

$$1/D_E = \frac{1}{\sigma_S^2} + \sum_{k=1}^{2} \frac{1-2^{-2r_k}}{\sigma_{N_k}^2} \geq \frac{1}{D}. \tag{36}$$

It can be easily seen from (32)-(34) that inequalities of (32) and (33) under JSCC are dominated by those under SSCC, i.e. $\{P_j, r_j\}$ satisfying (32) and (33) under JSCC also satisfies the corresponding inequalities under SSCC, while the inequality (34) under JSCC dominates that under SSCC.

To compare SSCC and JSCC, we next formulate a constrained optimization problem in which the objective is to minimize the total transmission power of L sensors with a constraint that the distortion in restoring X is no greater than D. For $L = 2$, the problem can be stated as

$$\min_{P_i, r_i, i=1,2} P_1 + P_2, \text{ subject to (32)-(34) and (36).} \tag{37}$$

which becomes power/rate allocations for SSCC and JSCC, respectively, for different correlation coefficients $\tilde{\rho}$. The optimization results for SSCC and JSCC under different channel and measurement parameters will reveal to us the relative efficiency of SSCC and JSCC, which will be further compared with that for an uncoded scheme, as investigated in the next few sections.

4 Optimal Power and Rate Allocations to Minimize the Total Transmission Power

4.1 Geometric Programming Solution to Power/Rate Allocations

The constrained optimization problems in (37) are non-convex. They can, however, be solved efficiently using standard techniques in convex optimization by transforming the original problems into relaxed convex geometric programming problems [10].

In this section, we take SSCC as an example to demonstrate how it works. For SSCC with $\tilde{\rho} = 0$, the rate tuple (R_1, R_2) should be taken from the boundary of the capacity region for two-user

Gaussian MAC channels to minimize P_1+P_2. Consequently,

$$R_1 = \frac{\alpha}{2}\log\left(\frac{\sigma_W^2+P_1g_1}{\sigma_W^2}\right)+\frac{1-\alpha}{2}\log\left(1+\frac{P_1g_1}{g_2P_2+\sigma_W^2}\right)$$
$$R_2 = \frac{\alpha}{2}\log\left(1+\frac{P_2g_2}{g_1P_1+\sigma_W^2}\right)+\frac{1-\alpha}{2}\log\left(\frac{\sigma_W^2+P_2g_2}{\sigma_W^2}\right) \quad (38)$$

where $\alpha\in[0,1]$ is a time sharing factor.

Define $y_j = 2^{2R_j}$, $z_j = 2^{2r_j}$ for $j=1,2$. We can transform this total power minimization problem for SSCC with $L=2$ to an equivalent generalized Signomial Programming problem [11]:

$$\min_{P_j,y_j,z_j,j=1,2} P_1+P_2, \text{ subject to:} \quad (39)$$
$$\left(g_2P_2+\sigma_W^2\right)^{(1-\alpha)} y_1 \le$$
$$\left(1+\frac{g_1P_1}{\sigma_W^2}\right)^{\alpha}\left(g_1P_1+g_2P_2+\sigma_W^2\right)^{(1-\alpha)} \quad (40)$$
$$\left(g_1P_1+\sigma_W^2\right)^{\alpha} y_2 \le$$
$$\left(1+\frac{g_2P_2}{\sigma_W^2}\right)^{(1-\alpha)}\left(g_1P_1+g_2P_2+\sigma_W^2\right)^{\alpha} \quad (41)$$
$$\frac{\sigma_s^2}{\sigma_{N_2}^2+\sigma_s^2}z_2^{-1}+D^{-1}\frac{\sigma_s^2\sigma_{N_2}^2}{\sigma_{N_2}^2+\sigma_s^2}y_1^{-1}z_1 \le 1 \quad (42)$$
$$\frac{\sigma_s^2}{\sigma_{N_1}^2+\sigma_s^2}z_1^{-1}+D^{-1}\frac{\sigma_s^2\sigma_{N_1}^2}{\sigma_{N_1}^2+\sigma_s^2}y_2^{-1}z_2 \le 1 \quad (43)$$
$$y_1^{-1}y_2^{-1}z_1z_2\sigma_s^2/D \le 1 \quad (44)$$
$$D^{-1}+\sigma_{N_1}^{-2}z_1^{-1}+\sigma_{N_2}^{-2}z_2^{-1} \le \sigma_s^{-2}+\sigma_{N_1}^{-2}+\sigma_{N_2}^{-2} \quad (45)$$

where constraints (40) and (41) are obtained by relaxing equality constraints in (38), and constraints (42)-(45) result from the transformation of (32)-(34), which are in the form of $f(x)\le 1$, where $f(x)$ is a posynomial function of n variables [10]: $f(x)=\sum_{k=1}^{K} c_k x_1^{a_{1k}} x_2^{a_{2k}}\cdots x_n^{a_{nk}}$, where $c_k \ge 0$ and $x_j > 0$ for $j=1,\cdots,n$ and $a_{ij}\in\mathcal{R}$. In addition, constraints (40) and (41) are in the form of generalized signomial functions [11, 12] with fractional powers.

Single condensation technique [11, 12] can then be applied to convert this Signomial programming problem to a standard geometric programming (GP) problem. In this method, we replace $(1+P_1g_1/\sigma_W^2)$ in the RHS of (40) by its geometric mean $\beta_{11}^{\beta_{11}}\left(\frac{g_1P_1}{\sigma_W^2\beta_{12}}\right)^{\beta_{12}}$, and similarly $(1+P_2g_2/\sigma_W^2)$ in the RHS of (41) by $\beta_{21}^{\beta_{21}}\left(\frac{g_2P_2}{\sigma_W^2\beta_{22}}\right)^{\beta_{22}}$, where $\beta_{i,j}\ge 0$ and $\beta_{i,1}+\beta_{i,2}=1$ for $i=1,2$. In addition, we also replace $(g_1P_1+g_2P_2+\sigma_W^2)$ by its geometric mean: $\left(\frac{\sigma^2}{\gamma_1}\right)^{\gamma_1}\left(\frac{g_1P_1}{\gamma_2}\right)^{\gamma_2}\left(\frac{g_2P_2}{\gamma_3}\right)^{\gamma_3}$, where $\gamma_i\ge 0$ and $\sum_{i=1}^{3}\gamma_i=1$. Finally, to handle fractional powers in the LHS of (40) and (41), we introduce two auxiliary variables t_1 and t_2 to replace $g_1P_1+\sigma_W^2$ and $g_2P_2+\sigma_W^2$, respectively, in the LHS of

(40) and (41). Accordingly, two additional posynomials are introduced on the list of constraints: $g_i P_i + \sigma_W^2 \le t_i$ for $i = 1, 2$.

The resulting standard geometric programming problem can thus be solved in an iterative manner by repeatedly updating normalization coefficients β_{ij} and γ_i, and applying interior point method for a given vector of these coefficients [11, 12].

Using the similar method, we can also transform the optimization problem for the JSCC approach to a convex Geometric programming problem. The details are skipped here [13].

4.2 Optimal Source/Channel Decoding Order for Non-Symmetric Channels

Although the optimization problems formulated in (37) can only be solved algorithmically, we can still manage to obtain some insights by scrutinizing the problem structures. In this section, we will reveal some relationships between the optimal decoding order and channel attenuation factors for non-symmetric networks.

4.2.1 Separate Source and Channel Coding

We first show the optimal source encoding/decoding order, as well as channel decoder order for SSCC is uniquely determined by the ordering of channel attenuation factors $\{g_i, i = 1, \cdots, L\}$, and has nothing to do with the ranking of sensor measurement noise power $\{\sigma_{N_i}^2, i = 1, \cdots, L\}$.

Theorem 4. *For SSCC, let π^* denote any permutation of $\{1, \cdots, L\}$ such that $g_{\pi^*(1)} \le g_{\pi^*(2)} \le \cdots \le g_{\pi^*(L)}$. To minimize the total transmission power, the optimal decoding order for channel codes at receiver is in the reversed order of π^*, i.e. interference cancellation is in the order $\pi^*(L), \pi^*(L-1), \cdots, \pi^*(1)$, which is also the decoding order of distributed source codewords.*

Proof. The proof consists of two steps. First, we will determine the channel decoding order for a given vector of source encoding rates $\{R_i, i = 1, \cdots, L\}$. For SSCC, the rate tuple $\{R_i\}$ satisfies $\sum_{i\in S} R_i \le \frac{1}{2}\log\left(1 + \sum_{i\in S} P_i g_i/\sigma_W^2\right)$, which is equivalent to

$$\sum_{i\in S} X_i \ge f(S) \triangleq \prod_{i\in S} 2^{2R_i} - 1, \forall S \subseteq \{1, \cdots, L\}. \tag{46}$$

where $X_i = P_i g_i/\sigma_W^2$ and $f : 2^E \to \mathcal{R}_+$ is a set function with $E \triangleq \{1, \cdots, L\}$. Given $\{R_i\}$, the optimization problem then becomes $\min \sum_{i=1}^{L} \sigma_W^2 X_i/g_i$, subject to (46).

Based on the Corollary 3.13 in [14], the set of power vectors $\{X_i\}$ satisfying (46) is a contra-polymatroid $\mathcal{G}(f)$, as f satisfies (1) $f(\phi) = 0$ (2) $f(S) \le f(T)$ if $S \subset T$ (3) $f(S) + f(T) \le f(S \cup T) + f(S \cap T)$. From Lemma 3.3 in [14], the minimizing vector $\{X_i, i \in E\}$ for $\min \sum_{i=1}^{L} \sigma_W^2 X_i/g_i$ is a vertex point $\{X_{\pi^*(i)}\}$ of $\mathcal{G}(f)$, where π^* is a permutation on the set E such that $1/g_{\pi^*(1)} \ge \cdots \ge$

$1/g_{\pi^*(L)}$ and

$$\begin{aligned} X_{\pi^*(i)} &= f\left(\{\pi^*(1),\cdots,\pi^*(i)\}\right) - f\left(\{\pi^*(1),\cdots,\pi^*(i-1)\}\right) \\ &= \prod_{j=1}^{i} 2^{2R_{\pi^*(j)}} - \prod_{j=1}^{i-1} 2^{2R_{\pi^*(j)}} \end{aligned} \tag{47}$$

which thus proves the first part of this Theorem.

We next use (47) to transform the objective function to

$$\sum_{i=1}^{L} \sigma_W^2 X_i/g_i = \sum_{i=1}^{L-1} \left(\frac{1}{g_{\pi^*(i)}} - \frac{1}{g_{\pi^*(i+1)}}\right) \prod_{j=1}^{i} 2^{2*R_{\pi^*(j)}} \tag{48}$$

For SSCC, rate tuples $\{R_i, i=1,\cdots,L\}$ satisfy $\sum_{i\in S} R_i \geq I\left(X(S);U(S)|U(S^c)\right)$. It is therefore quite straightforward to show that in order to minimize the total transmission power in (48), we need to have $\sum_{j=1}^{i} R_{\pi^*(j)}$ achieve the lower bound, i.e.

$$\sum_{j=1}^{i} R_{\pi^*(j)} = I\left(X_{\pi^*(1)}^{\pi^*(i)}; U_{\pi^*(1)}^{\pi^*(i)} | U_{\pi^*(i+1)}^{\pi^*(L)}\right) \tag{49}$$

which implies that the decoding order for source codewords is $\pi^*(L),\cdots,\pi^*(1)$, independent of the ordering of variances $\{\sigma_{N_i}, i=1,\cdots,L\}$ of measurement noise.

□

Theorem 4 implies that sensor $\pi^*(L)$ does not conduct random binning and its quantization vector $\{U_{\pi^*(L)}[n]\}$ is restored first. $\{U_{\pi^*(L)}[n]\}$ is then used as side information to restore sensor $\pi^*(L-1)$'s source codeword $\{U_{\pi^*(L-1)}[n]\}$ from this node's first stage Gaussian source vector quantization, which resides in the bin whose index is decoded from the channel decoding step. This process continues until sensor $\pi^*(1)$'s first stage quantization vector $\{U_{\pi^*(1)}[n]\}$is restored by using all other sensors' quantization as side information.

In the next section, we will see a similar conclusion can be reached for JSCC.

4.2.2 Joint Source and Channel Coding

Unlike in the SSCC case where we have a nice geometric (contra-polymatroid) structure which enables us to reach a conclusion valid for any arbitrary asymmetric networks, JSCC in general lacks such a feature for us to exploit. We will instead , in this section, focus on a case with only $L=2$ sensor nodes and establish a similar result as in Section 4.2.1 for optimal channel decoding orders. WLOG, we assume $g_1 > g_2$ in the subsequent analysis.

Theorem 5. *Given a pair of quantization rates $\tilde{R}_1$ and $\tilde{R}_2$, when $g_1 > g_2$, the optimal decoding order to minimize the total transmission power $P_1 + P_2$ is to decoder node 1's signal first, and then node 2's information after removing the decoded node 1's signal from the received signal, i.e.*

$$\tilde{R}_1 = I(U_1; Z) = \frac{1}{2}\log\left(1 + \frac{\left(\sqrt{P_1 g_1} + \tilde{\rho}\sqrt{P_2 g_2}\right)^2}{\sigma_W^2 + P_2 g_2 (1-\tilde{\rho}^2)}\right)$$
$$\tilde{R}_2 = I(U_2; Z, U_1) = \frac{1}{2}\log\left(\frac{P_1(1-\tilde{\rho}^2) + \sigma_W^2}{\sigma_W^2 (1-\tilde{\rho}^2)}\right) \tag{50}$$

where $\tilde{\rho}$ is the same as in 35.

Proof. See appendix D. □

5 Uncoded Sensor Transmission in Fusion

For Gaussian sensor networks as modeled in Section 2, it has been shown recently [1, 2] that uncoded transmission, i.e. each sensor only forwards a scaled version of its measurements to the fusion center, asymptotically achieves the lower bound on distortion when the number of sensors grow to infinity and system is symmetric. In the context of the theme of this paper, we, in this section, investigate the optimal power allocation strategy when a finite number of sensors deploy the uncoded scheme under more general channel conditions.

For uncoded transmission, the transmitted signal by node j is $Y_j[i] = \alpha_j X_j[i]$, where $\alpha_j = \sqrt{\frac{P_j}{\sigma_S^2 + \sigma_{N_j}^2}}$ is a scaling factor to make the transmission power $E|Y_j[i]|^2 = P_j$. The received signal at the fusion center is therefore $Z[i] = \sum_{j=1}^{L} Y_j[i]\sqrt{g_j} + W[i]$. The linear MMSE estimate of $X_0[i]$ is: $\hat{X}_0[i] = \gamma Z[i]$, where the coefficient γ can be obtained using Orthogonal principle: $E\left[\left(X_0[i] - \hat{X}_0[i]\right) Z[i]\right] = 0$. The resultant MSE is

$$E\left|X_0[i] - \hat{X}_0[i]\right|^2 = \sigma_S^2 \frac{\sum_{j=1}^{L} \frac{P_j g_j \sigma_{N_j}^2}{\sigma_S^2 + \sigma_{N_j}^2} + \sigma_W^2}{\sum_{j=1}^{L} P_j g_j + B + \sigma_W^2} \tag{51}$$

where $B = \sum_{i=1}^{L} \sum_{i \neq j, j=1}^{L} \rho_{i,j} \sqrt{P_i g_i}\sqrt{P_j g_j}$ and the covariance coefficients $\rho_{i,j}$ is the same as in (7).

When $L = 2$, the power control problem under a distortion constraint $E\left|X_0[i] - \hat{X}_0[i]\right|^2 \leq D$ for the uncoded scheme can be formulated as:

$$
\begin{aligned}
&\min P_1 + P_2 \ \text{ subject to:} \\
&\frac{1}{2}\log\left[\frac{\sigma_S^2}{D}\left(1+\frac{\sigma_{N_1}^2}{\sigma_S^2+\sigma_{N_1}^2}\frac{P_1 g_1}{\sigma_W^2}+\frac{\sigma_{N_2}^2}{\sigma_S^2+\sigma_{N_2}^2}\frac{P_2 g_2}{\sigma_W^2}\right)\right] \\
&\leq \frac{1}{2}\log\left[1+\frac{P_1 g_1 + P_2 g_2 + 2\rho_{1,2}\sqrt{P_1 g_1 P_2 g_2}}{\sigma_W^2}\right]
\end{aligned}
\tag{52}
$$

This problem can again be transformed to a GP problem using the condensation technique applied in Section 4.1. We skip the details here.

What deserves our attention is that when we compare the constraint in (52) with (34), there is a striking similarity when we substitute r_i with $\tilde{R}_i$, whose relationship was introduced in (6). After substitution, (34) becomes

$$
\begin{aligned}
&\frac{1}{2}\log\left\{\frac{\sigma_S^2}{D}\left[1+A_1\right]\left[1+A_2\right]\right\} \\
&\leq \frac{1}{2}\log\left[1+\frac{P_1 g_1 + P_2 g_2 + 2\tilde{\rho}_{1,2}\sqrt{P_1 g_1 P_2 g_2}}{\sigma_W^2}\right]
\end{aligned}
\tag{53}
$$

where $A_i = \frac{\sigma_{N_i}^2}{\sigma_S^2+\sigma_{N_i}^2}(2^{2\tilde{R}_i}-1)$ for $i = 1, 2$, and $\tilde{\rho}_{1,2}$ and $\rho_{1,2}$ are associated as in (8).

For JSCC, we have $\tilde{R}_i \leq I(U_i; Y|U_j) + I(U_i; U_j)$, which is equivalent to

$$
2^{\tilde{R}_i} - \frac{1}{1-\tilde{\rho}_{1,2}^2} \leq \frac{P_i g_i}{\sigma_W^2} \tag{54}
$$

We can infer from (32)-(34), as well as (52) that it is in general hard to argue which approach is the most energy efficient in terms of the total power consumption under a common distortion constraint D, which will be further exemplified in our simulation results in Section 7. There, we will see the most energy efficient approach depends on exact values of $\sigma_{N_j}^2$, as well as g_i and D.

However, when a system becomes homogeneous and symmetric in the sense that $\sigma_{N_i}^2 = \sigma_{N_j}^2$ and $g_i = g_j$ for all $i, j \in \{1, \cdots, L\}$, we have consistent results for both finite number of L and asymptotically large L, as revealed in the next section, when we compare these three approaches.

6 Energy Consumption Comparison for Homogeneous Networks: Finite and Asymptotic Results

In this section, we provide analytical results on comparisons between different transmission strategies proposed thus far, including JSCC, SSCC and uncoded schemes in terms of their total transmission power consumptions when system becomes homogeneous.

6.1 Comparison under finite L and finite SNR

Theorem 6. *When a system of $L < \infty$ sensors becomes symmetric, the total power consumption for the separate, joint and uncoded schemes proposed previously follow:*

$$\left(\frac{Pg}{\sigma_W^2}\right)_{LoB} < \left(\frac{Pg}{\sigma_W^2}\right)_A < \left(\frac{Pg}{\sigma_W^2}\right)_J < \left(\frac{Pg}{\sigma_W^2}\right)_S \tag{55}$$

where $\left(\frac{Pg}{\sigma_W^2}\right)_{LoB}$ is the lower-bound on transmission power, and g is the common channel gain from each sensor to the fusion center. Indexes A, J and S represent the uncoded, joint and separate encoding schemes, respectively.

Proof. The proof hinges upon the analysis of rate and power allocations for symmetric networks for different schemes.

Separate Coding:

We first look at the separate source and channel coding approach. In symmetric networks, for source coding part, each node employs identical compression rate R, which satisfies $LR = Lr + \frac{1}{2}\log\frac{\sigma_S^2}{D}$, where r is the solution to

$$\frac{1}{\sigma_S^2} + \frac{L}{\sigma_{N_1}^2}(1 - 2^{-2r}) = \frac{1}{D}. \tag{56}$$

As a result [6],

$$LR = -\frac{L}{2}\log\left(1 - \frac{\sigma_{N_1}^2}{L}\left(\frac{1}{D} - \frac{1}{\sigma_S^2}\right)\right) + \frac{1}{2}\log\frac{\sigma_S^2}{D} \tag{57}$$

For channel coding part, to minimize the total transmission power, it is optimal to let each sensor transmit at the same power P and same rate R which can be achieved by jointly decoding all node's information at the fusion center. Therefore, the total compression rate also satisfies

$$LR = \frac{1}{2}\log\left(1 + \frac{LPg}{\sigma_W^2}\right) \tag{58}$$

Combining (57) and (58) yields

$$\begin{aligned}\left(\frac{Pg}{\sigma_W^2}\right)_S &= -\frac{1}{L} + \frac{\sigma_S^2}{LD}\left(2^{2r}\right)^L \\ &= -\frac{1}{L} + \frac{\sigma_S^2}{LD}\left[1 - \left(\frac{1}{D} - \frac{1}{\sigma_S^2}\right)\frac{\sigma_{N_1}^2}{L}\right]^{-L}\end{aligned} \tag{59}$$

Joint Coding:

For the joint source-channel coding scheme, the vector quantization rate for each sensor is equal to $\tilde{R} = I(X_1; U_1)$, which is associated with r as shown by (2), where r can be further obtained using

(56).

By applying the techniques used in driving the constraints in (32)-(34) to the $L > 2$ case when channel is symmetric, the quantization rate and the resultant L-user multiple access channel with L correlated inputs $\{U_1, \cdots, U_L\}$ are associated by

$$\begin{aligned} Lr + \frac{1}{2}\log\frac{\sigma_S^2}{D} &= I(U_1, \cdots, U_L; Z) \\ &= H(Z) - H(Z|U_1, \cdots, U_L), \end{aligned} \tag{60}$$

where the conditional entropy $H(Z|U_1, \cdots, U_L) = \frac{1}{2}\log\left(2\pi e\sigma_W^2\right)$, and the entropy of $Z = \sum_{j=1}^{L}\sqrt{g}\gamma U_j + W$ is $H(Z) = \frac{1}{2}\log\left[2\pi e\left(\sigma_W^2 + g\mathrm{Var}(\gamma\sum_j U_j)\right)\right]$. Given the covariance coefficient between U_i and U_j: $\tilde{\rho} = \frac{\sigma_S^2}{\sigma_S^2+\sigma_{N_1}^2}(1 - 2^{-2\tilde{R}})$, we have

$$E|\gamma\sum_{j=1}^{L}U_j|^2 = LP + (L^2 - L)\tilde{\rho}P \tag{61}$$

From (56) and (2), we obtain

$$\tilde{\rho} = \left(\frac{\sigma_S^2}{D} - 1\right)\left(L + \frac{\sigma_S^2}{D} - 1\right)^{-1}. \tag{62}$$

The transmission power P for joint coded scheme in symmetric networks can thus be computed using (60) and (61):

$$\left(\frac{Pg}{\sigma_W^2}\right)_J = \left[-\frac{1}{L} + \frac{\sigma_S^2}{LD}\left(2^{2r}\right)^L\right]\left(1 + \left(L - \frac{1}{L}\right)\tilde{\rho}\right)^{-1} \tag{63}$$

Comparing (63) with (59), it is apparent that $\left(\frac{Pg}{\sigma_W^2}\right)_J < \left(\frac{Pg}{\sigma_W^2}\right)_S$.

Uncoded Scheme

When sensor transmits scaled measurements in a symmetric network, we can obtain the minimum transmission power P by making the mean squared error obtained in (51) equal to D and substituting P_j, g_j and $\sigma_{N_j}^2$ by P, g and $\sigma_{N_1}^2$, respectively. As a result, we obtain

$$\left(\frac{Pg}{\sigma_W^2}\right)_A = \left[\frac{L^2\sigma_S^2}{\sigma_S^2 + \sigma_{N_1}^2} - L\left(\frac{\sigma_S^2}{D} - 1\right)\frac{\sigma_{N_1}^2}{\sigma_S^2 + \sigma_{N_1}^2}\right]^{-1}\left(\frac{\sigma_S^2}{D} - 1\right) \tag{64}$$

To compare $\left(\frac{Pg}{\sigma_W^2}\right)_A$ with $\left(\frac{Pg}{\sigma_W^2}\right)_J$, we need to introduce an auxiliary variable. Define

$$\tilde{Q}_L = 1 - \frac{\sigma_S^2 + \sigma_{N_1}^2}{\sigma_S^2}\tilde{\rho} < 1 - \tilde{\rho}. \tag{65}$$

We can therefore re-derive the minimum power for uncoded scheme:

$$\left(\frac{Pg}{\sigma_W^2}\right)_A = \frac{1}{L}\left(\frac{1}{\tilde{Q}_L} - 1\right). \tag{66}$$

In addition, we can express $2^{2r} = \left[1 - \left(\frac{1}{D} - \frac{1}{\sigma_S^2}\right)\frac{\sigma_{N_1}^2}{L}\right]$ as

$$2^{2r} = \left[\frac{\tilde{Q}_L}{L}\left(L + \frac{\sigma_S^2}{D} - 1\right)\right]^{-1} = \frac{1-\tilde{\rho}}{\tilde{Q}_L} > 1 \tag{67}$$

which is used to transform $\left(\frac{Pg}{\sigma_W^2}\right)_J$ as

$$\left(\frac{Pg}{\sigma_W^2}\right)_J = \frac{1}{L}\left\{\left(\frac{1-\tilde{\rho}}{\tilde{Q}_L}\right)^{L-1}\frac{1}{\tilde{Q}_L} - \frac{L-1+\sigma_S^2/D}{L\sigma_S^2/D}\right\} \tag{68}$$

where the second term $\frac{L-1+\sigma_S^2/D}{L\sigma_S^2/D} < 1$ due to $L > 1$ and $\sigma_S^2 > D$. Since $1 - \tilde{\rho} > \tilde{Q}_L$, after comparing (68) and (66), it is straightforward to show $\left(\frac{Pg}{\sigma_W^2}\right)_J > \left(\frac{Pg}{\sigma_W^2}\right)_A$.

Lower Bound of Transmission Power

Since $X_0 \to \{X_1, \cdots, X_L\} \to Z \to \hat{X}_0$ forms a Markov chain, by Data Processing Inequality [7], we have $I(X_0; \hat{X}_0) \leq I(X_1, \cdots, X_L; Z)$. On one hand, to ensure $E|X_0 - \hat{X}_0|^2 \leq D$, it can be shown $I(X_0; \hat{X}_0) \geq \frac{1}{2}\log\frac{\sigma_S^2}{D}$ using rate distortion results [7]. On the other hand, the mutual information $I(X_1, \cdots, X_L; Z)$ is upper-bounded by the mutual information of an additive noise Gaussian channel with channel gain $\sqrt{g}$ and total transmission power upper-bounded by $E|\sum_{j=1}^{L} X_j|^2 = LP + (L^2 - L)\rho P$, where $\rho = \frac{\sigma_S^2}{\sigma_S^2 + \sigma_{N_1}^2}$ is the covariance coefficient between X_i and X_j for $i \neq j$. Consequently,

$$I(X_1, \cdots, X_L; Z) \leq \frac{1}{2}\log\left[1 + \frac{Pg}{\sigma_W^2}\left(L + (L^2 - L)\rho\right)\right] \tag{69}$$

From $\frac{1}{2}\log\frac{\sigma_S^2}{D} \leq I(X_0; \hat{X}_0) \leq I(X_1, \cdots, X_L; Z)$, we obtain the lower-bound of transmission power:

$$\left(\frac{Pg}{\sigma_W^2}\right)_{LoB} = \frac{\frac{\sigma_S^2}{D} - 1}{L + (L^2 - L)\rho} \tag{70}$$

Compare this lower bound with (64), we have $\left(\frac{Pg}{\sigma_W^2}\right)_{LoB} < \left(\frac{Pg}{\sigma_W^2}\right)_A$.

Therefore, we have shown that the order in (55) holds and thus completed the proof.

□

6.2 Comparison under $L \to \infty$ and finite SNR

In this section, we provide the scaling behaviors of the total transmission power of various schemes studied so far. In particular, we are interested in how $\frac{Pg}{\sigma_W^2}$ scales with respect to the number of sensors L in a symmetric system under a common constraint on distortion no greater than D. The analysis is quite straightforward based on the results for a finite L number of sensors that we have obtained in (59), (68), (66) and (70), for separate, joint, uncoded schemes and the lower bound, respectively.

Theorem 7.

$$\begin{aligned}
\lim_{L\to\infty}\left(\frac{LgP}{\sigma_W^2}\right)_S &= \frac{\sigma_S^2}{D}\exp\left[\sigma_{N_1}^2\left(\frac{1}{D}-\frac{1}{\sigma_S^2}\right)\right]-1 \\
\lim_{L\to\infty}\left(\frac{LgP}{\sigma_W^2}\right)_J &= \exp\left[\sigma_{N_1}^2\left(\frac{1}{D}-\frac{1}{\sigma_S^2}\right)\right]-\frac{D}{\sigma_S^2} \\
\lim_{L\to\infty}\left(\frac{L^2gP}{\sigma_W^2}\right)_A &= \lim_{L\to\infty}\left(\frac{L^2gP}{\sigma_W^2}\right)_{LoB} \\
&= \left(\frac{1}{D}-\frac{1}{\sigma_S^2}\right)\left(\sigma_S^2+\sigma_{N_1}^2\right) \qquad (71)
\end{aligned}$$

Proof. The proof of these convergence results is quite straightforward based upon the results for finite L as above, and is skipped here.

□

It is obvious that the transmission power of the uncoded scheme shares the same scaling factor as the lower-bound, which is in the order of $1/L^2$. The asymptotic optimality of the uncoded scheme in symmetric Gaussian sensor networks is not a new result, which has been attained previously in [1] [15]. In [1, 15], the authors assumed a fixed transmission power and showed that the distortion achieved using uncoded approach has the same asymptotic scaling law as that obtained via a lower bound. Here, we provide a different perspective in assessing its optimality for the uncoded scheme in terms of the total transmission power while meeting a fixed distortion constraint.

Both joint and separate coding schemes have the scaling factor in the order of $1/L$. Asymptotically, joint coding scheme saves in total transmission power by a factor of σ_S^2/D as compared with the separated approach.

6.3 Comparison under finite L and SNR $\to \infty$

Given the number of sensors $L < \infty$, we are interested in the scaling factors associated with transmission SNR P/σ_W^2, mean squared error D and measurement noise variance σ_N^2 in homogeneous networks. In particular, we need to investigate how the following asymptotic factors are related, $\sigma_N^2 \to 0$, $D \to 0$ and $P/\sigma_W^2 \to \infty$.

We first need to identify the limit imposed upon the scaling factor related with σ_N^2 and D. As can be seen from both (56) and (64) under SSCC, JSCC and uncoded approaches, it is required that

$$\lambda\left(\sigma_N^2\right) \triangleq \frac{\sigma_N^2}{L}\left(\frac{1}{D}-\frac{1}{\sigma_S^2}\right) \leq 1. \tag{72}$$

Denote $\gamma^* = \lim_{\sigma_N^2 \to 0} \frac{\ln D}{\ln \sigma_N^2}$. It can be seen that γ^* has to satisfy $\gamma^* \in [0,1]$ in order to have the inequality in (72) hold. Consequently,

$$\lambda^* = \lim_{\sigma_N^2 \to 0} \lambda\left(\sigma_N^2\right) = \begin{cases} 0, & 0 \leq \gamma^* < 1 \\ \frac{1}{L} & \gamma^* = 1 \end{cases} \tag{73}$$

Theorem 8. *The asymptotic ratios associated with D, σ_N^2 and P/σ_W^2 have the following relationship:*

$$\lim_{\sigma_N^2 \to 0} \frac{\left(Pg/\sigma_W^2\right)_S}{\sigma_S^2/D} = \frac{1}{L(1-\lambda^*)^L} \tag{74}$$

$$\lim_{\sigma_N^2 \to 0} \frac{\left(Pg/\sigma_W^2\right)_J}{\sigma_S^2/D} = \frac{1}{L^2(1-\lambda^*)^L} \tag{75}$$

$$\lim_{\sigma_N^2 \to 0} \frac{\left(Pg/\sigma_W^2\right)_A}{\sigma_S^2/D} = \frac{1}{L^2(1-\lambda^*)} \tag{76}$$

Proof. The proof follows straightforwardly with (59), (63) and (64) for SSCC, JSCC and uncoded schemes, respectively, as we let $\sigma_N^2 \to 0$ and $\frac{\sigma_N^2}{D} \approx L\lambda^*$. □

We can see from (74)-(76) that when $\gamma^* \in [0,1)$, i.e. $\lambda^* = 0$, the JSCC and uncoded schemes have the same aysmptotic ratio between the received SNR and S-MSE-Ratio, which is smaller than that for the SSCC approach by a factor of $1/L$. If $\gamma^* = 1$, i.e. $\lambda^* = 1/L$, uncoded approach has the smallest ratio among all schemes. However, if we introduce the SNR exponent as defined in [16],

$$\eta \triangleq -\lim_{\sigma_N^2 \to 0} \frac{\ln D}{\ln\left(Pg/\sigma_W^2\right)}, \tag{77}$$

All three approaches share the same ratio $\eta = 1$ for finite L.

Theorem 8 therefore provides us another perspective to compare these remote estimation approaches. It demonstrates the proposed joint source and channel coding scheme has potentially the same asymptotic performance as the uncoded one in high channel SNR and high measurement SNR regions for all ratio exponents $\gamma^* \in [0,1)$. In addition, speaking of SNR exponent of distortion measure in the high SNR region, all three schemes investigated in this paper share the same asymptotic ratio $\eta = 1$.

An additional remark we next make is about the limitation as to adopting SNR exponent η as a metric to characterize the asymptotic performance in large SNR regions [16]. It can be seen clearly from the above analysis that the SNR exponent obscures the asymptotic difference between SSCC and JSCC, as well as the uncoded approaches, which have different linear ratios as shown in Theorem 8. These differences are gone once log-scale is imposed, however.

Note also that we have implicitly assumed that the spectral efficiency of the system model in this paper, which is the ratio the source bandwidth over channel bandwidth [16], is one.

7 Numerical Results

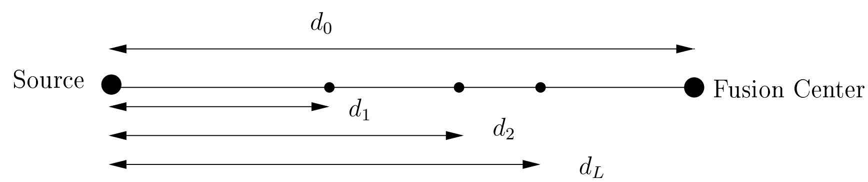


Figure 2: 1-D location Model

In this section, the three approaches proposed in this paper are examined and compared with each other by looking at each of their optimal total transmission powers under the constraint of restoring X_0 with MSE no greater than D, some prescribed threshold. Particularly, we consider a linear network topology where a source, fusion center and $L = 2$ sensor nodes are located on a same line as illustrated in Figure 2. To associate positions of sensor nodes with channel gains and measurement noise, we assume a path-loss model with coefficient β_s and β_c for g_i and $\sigma_{N_i}^2$, respectively: $g_i \propto 1/(d_0 - d_i)^{\beta_c}$ and $\sigma_{N_i}^2 \propto d_j^{\beta_s}$, for $i = 1, \cdots, L$, where d_0 is the distance between source and fusion center, and d_i is the distance between the i-th sensor and source. Given a distortion upper-bound $D < \sigma_S^2$ and $\beta_c = \beta_s = \beta$, the distance between the source and fusion center has to satisfy the following inequality, $d_0 < (L)^{1/\beta} \left[\frac{1}{D} - \frac{1}{\sigma_S^2}\right]^{-1/\beta}$, which is obtained by making the MSE using $\{X_i\}$ to estimate X_0 no greater than D.

We then run the geometrical programming based optimization algorithm to determine the minimum total transmission powers for various approaches. We consider 9 spots uniformly distributed between the source and fusion center for possible locations of two sensors, which are indexed by integers 1 through 9. The smaller the index value is, the closer the sensor is located to the source. Figure 3(a), Figure 3(b) and Figure 3(c) demonstrate the total minimum power consumption $P_1 + P_2$ as a function of nodes' locations for three sensor processing schemes, from which we have following observations:

- As proved in Theorem 6, when network is symmetric, $P_{total,A} < P_{total,J} < P_{total,S}$.
- Under a relatively large distortion constraint (e.g. $D = 0.5$), uncoded scheme is the most energy efficient among the three candidates for all sensor locations, as shown by Figure 3(a).

- Under relatively small distortion constraints (e.g. $D = 0.1$ and $D = 0.01$), separate coding approach becomes the most energy efficient when the relative position of two sensors becomes more asymmetric. For example, in both Figure 3(b) and Figure 3(c), at a location with an index pair $(1, 9)$, i.e. the first sensor is closest to the source and the second sensor is closest to the fusion center, we have $P_{total,A} > P_{total,J} > P_{total,S}$.

- Overall, to minimize the total power expenditure, we should choose either uncoded transmission or separate coding scheme for a given pair of locations. This is a bit surprising as joint coded approach is often advocated more efficient (rate wise) than the separate one. It thus exemplifies that exact values of channel conditions and the level of measurement noise are crucial to concluding which scheme is the most power efficient in non-symmetric Gaussian networks with a finite number of sensors.

A Proof of Theorem 2

Proof. From Theorem 1, we know that each sensor finds from $2^{n\tilde{R}_j}$ codewords the closest one $U_j^{(k)} = \{U_j^{(k)}[i]\}$ to the observation vector $\{X_j[i]\}$ and then amplify-and-forwards $\{Y_j[i]\}$ to the fusion center. The decoder applies jointly typical sequence decoding [7] to seek $\{U_j^{(k)}, j = 1, \cdots, L\}$ from L codebooks which are jointly typical with the received vector $\{Z[i]\}$.

WLOG, re-shuffle $2^{n\tilde{R}_j}$ vectors such that $U_j^{(1)}$ is the vector selected by sensor $j \in \{1, \cdots, L\}$. We assume that a subset $U^{(1)}(S^c) = \{U_j^{(1)}, j \in S^c\}$ has been decoded correctly, while its complementary set $U^{(1)}(S) = \{U_j^{(1)}, j \in S\}$ is in error, which implies that the channel decoder at fusion center is in favor of a set of vectors $U^{(k)}(S) = \{U_j^{(k_j)}, k_j \neq 1, j \in S\}$, instead. Next, We will find the upper bound of the probability that $\left(U^{(1)}(S^c), Z\right)$ and $U^{(k)}(S)$ are jointly typical.

The technique to upper-bond this probability is quite similar as the one for MAC channels with independent channel inputs [7, Chap 15.3]. The major difference here is that the channel inputs from L sensors are correlated because of the testing channel model used in independent source coding, i.e. $U_j = X_j + V_j = X_0 + N_j + V_j$, for $j = 1, \cdots, L$.

The upper bound of the probability that $\left(U^{(1)}(S^c), Z\right)$ and $U^{(k)}(S)$ are jointly typical is therefore

$$2^{n(H(U(S),U(S^c),Z)+\epsilon)}2^{-n(H(U(S^c),Z)-\epsilon)}2^{-n\sum_{i\in S}(H(U_i)-\epsilon)} \tag{78}$$

$$= \exp_2\left(-n\left(H(U(S^c), Z) + \sum_{i\in S} H(U_i)\right.\right.$$
$$\left.\left.-H(U(S), U(S^c), Z) - (|S| + 2)\epsilon\right)\right) \tag{79}$$

where the first term in (78) is the upper bound for the number of jointly typical sequences of $(U(S), U(S^c), Z)$, the second term in (78) is the upper bound of the probability $P\left(U^{(1)}(S^c), Z\right)$ and the last term in (78) is the upper bound of the probability $P\left(U^{(k)}(S)\right)$. The summation in the last term in (78) is due to the independence of codebooks generated by each sensor and the assumption that decoder is in favor of some $U_j^{(k_j)}$ for $k_j \neq 1$ and $j \in S$, which are independent of $U^{(1)}(S)$.

Since we have at most $2^{n\sum_{j\in S}\tilde{R}_j}$ number of sequences to be confused with $U_j^{(1)}, j \in S$, we need

$$\begin{aligned}\sum_{j\in S}\tilde{R}_j &< H(U(S^c), Z) \\ &+ \sum_{i\in S} H(U_i) - H(U(S), U(S^c), Z) - (|S|+2)\epsilon \\ &= \sum_{i=1}^{|S|-1} I\left(U_{\pi_i}; U_{\pi_{i+1}}^{\pi_{|S|}}\right) + I\left(U(S); U(S^c), Z\right) - (|S|+2)\epsilon \end{aligned} \tag{80}$$

for all $S \subseteq \{1, \cdots, L\}$ and any arbitrarily small ϵ in order to achieve the asymptotic zero error probability as $n \to \infty$, which thus completes the proof.

□

B Proof of Lemma 1

Proof. As $U_2 \to X_2 \to X_1 \to U_1$ forms a Markov chain, we have

$$\begin{aligned} I(U_2; X_2, X_1|U_1) &\overset{(a)}{=} I(U_2; X_1, X_2, U_1) - I(U_2; U_1) \\ &\overset{(b)}{=} I(U_2; X_2) + I(U_2; X_1, U_1|X_2) - I(U_2; U_1) \\ &\overset{(c)}{=} I(U_2; X_2) - I(U_2; U_1) \end{aligned} \tag{81}$$

where equations (a) and (b) are due to the chain rule on conditional mutual information [7]. Given Markov chain of $U_2 \to X_2 \to X_1 \to U_1$, U_2 and (X_1, U_1) are conditionally independent given X_2 and consequently $I(U_2; X_1, U_1|X_2) = 0$ leading to equation (c).

On the other hand, following equations also hold under similar arguments:

$$\begin{aligned} I(U_2; X_2, X_1|U_1) &\overset{(1)}{=} I(U_2; X_2|U_1) + I(U_2; X_1|X_2, U_1) \\ &\overset{(2)}{=} I(U_2; X_2|U_1) + I(U_2; X_1, U_1|X_2) - I(U_2; U_1|X_2) \\ &\overset{(3)}{=} I(U_2; X_2|U_1) \end{aligned} \tag{82}$$

Therefore, combining (81) and (82) yields:

$$I(U_2; X_2|U_1) = I(U_2; X_2) - I(U_1; U_2) \tag{83}$$

and similarly,

$$I(U_1;X_1|U_2) = I(U_1;X_1) - I(U_1;U_2) \tag{84}$$

which thus proves (13) and (14).

We can prove (15) by firstly showing that

$$I(U_1,U_2;X_1,X_2) = I(X_1,X_2;U_1) + I(U_2;X_2,X_1|U_1) \tag{85}$$

under the chain rule, where

$$I(X_1,X_2;U_1) = I(X_1;U_1) + I(X_2;U_1|X_1) = I(X_1;U_1) \tag{86}$$

because of $U_2 \to X_2 \to X_1 \to U_1$. Since we have already proved (81), it is straightforward to show that (15) holds.

□

C Proof of Lemma 3

Proof. Due to the independence of measurement noise, $U(S) \to X(S) \to X(S^c) \to U(S^c)$ forms a Markov chain. The first equation in (21) is a direct application of (13).

The proof of the second equation is based upon another Markov chain by adding the source random variable X_0 into the former one: $U(S) \to X(S) \to X_0 \to X(S^c) \to U(S^c)$. From this Markov chain, we can deduce

$$\begin{aligned} I\left[U(S);X_0|U(S^c)\right] &= I\left[U(S^c),X_0;U(S)\right] - I\left[U(S^c);U(S)\right] \\ &= I\left[X_0;U(S)\right] + I\left[U(S^c);U(S)|X_0\right] - I\left[U(S);U(S^c)\right] \\ &= I\left[X_0;U(S)\right] - I\left[U(S);U(S^c)\right] \end{aligned} \tag{87}$$

and

$$\begin{aligned} \sum_{i\in S} I\left[U_i;X_i|X_0\right] &= I\left[U(S);X(S)|X_0\right] \\ &= I\left[U(S);X(S)\right] + I\left[U(S);X_0|X(S)\right] - I\left[X_0;U(S)\right] \\ &= I\left[U(S);X(S)\right] - I\left[X_0;U(S)\right] \end{aligned} \tag{88}$$

where the first equality is because of the conditional independence of (X_i,U_i) given X_0.

It can be seen that (87) and (88) yields

$$I\left[U(S);X_0|U(S^c)\right]+\sum_{i\in S}I\left[U_i;X_i|X_0\right]$$
$$=I\left(U(S);X(S)\right)-I\left(U(S);U(S^c)\right), \tag{89}$$

which completes the proof for Lemma 3.

□

D Proof of Theorem 5

Proof. Define $Z_i = 2^{2\tilde{R}_i}\left(1-\tilde{\rho}^2\right)$, for $i=1,2$. It can be shown that $Z_i>1$ by using $\tilde{\rho}^2=\rho^2(1-2^{-2\tilde{R}_1})(1-2^{-2\tilde{R}_2})$ and $0<\rho<1$ as shown in (7).

Given a pair of quantization rates $(\tilde{R}_1,\tilde{R}_2)$, the original optimization problem becomes

$$\begin{aligned}&\min P_1+P_2,\ \text{subject to:}\\ &P_1\geq\frac{(Z_1-1)\sigma_W^2}{(1-\tilde{\rho}^2)g_1}\triangleq b_1,\ P_2\geq\frac{(Z_2-1)\sigma_W^2}{(1-\tilde{\rho}^2)g_2}\triangleq b_2\\ &P_1g_1+P_2g_2+2\tilde{\rho}\sqrt{P_1g_1P_2g_2}\\ &\geq\sigma_W^2\left(2^{2\tilde{R}_1+2\tilde{R}_2}(1-\tilde{\rho}^2)-1\right)\triangleq b_3\end{aligned} \tag{90}$$

Define a matrix

$$\mathbf{A}=\begin{bmatrix} g_1 & \tilde{\rho}\sqrt{g_1g_2}\\ \tilde{\rho}\sqrt{g_1g_2} & g_2\end{bmatrix} \tag{91}$$

whose eigenvalue decomposition is: $\mathbf{A}=\mathbf{Q}\Lambda\mathbf{Q}^T$, where the diagonal matrix $\Lambda=\text{diag}\{\lambda_1,\lambda_2\}$ has eigenvalues λ_i of $\mathbf{A}$ and the column vectors of the matrix

$$\mathbf{Q}=\begin{bmatrix} q_{1,1} & q_{1,2}\\ q_{2,1} & q_{2,2}\end{bmatrix} \tag{92}$$

are normalized eigenvectors associated with λ_1 and λ_2 respectively, which satisfy:

$$\mathbf{Q}^T\mathbf{Q}=\begin{bmatrix} 1 & 0\\ 0 & 1\end{bmatrix}\triangleq\mathbf{I_2} \tag{93}$$

Notice that the last constraint in (90) is a quadratic form of variables $\sqrt{P_1}$ and $\sqrt{P_2}$, which essentially determines an ellipse, we can perform an unitary transformation by introducing two new variables Y_1 and Y_2:

$$\mathbf{Y}=[Y_1,Y_2]^T=\mathbf{Q}^T[\sqrt{P_1},\sqrt{P_2}]^T \tag{94}$$

such that the original optimization problem in (90) is transformed to an equivalent one:

$$\begin{aligned}&\min ||Y||^2 = Y_1^2 + Y_2^2, \text{ subject to:}\\&\mathbf{Q}[Y_1, Y_2]^T \geq [b_1, b_2]\\&\lambda_1 Y_1^2 + \lambda_2 Y_2^2 \geq b_3.\end{aligned} \tag{95}$$

We show next that $Y_1^2 + Y_2^2$ is minimized at the point where the second line $q_{2,1}Y_1 + q_{2,2}Y_2 = b_2$ intersects with the ellipse $\lambda_1 Y_1^2 + \lambda_2 Y_2^2 = b_3$ as shown in Figure 4. In order to prove this, we need to first prove that the intersection of two lines $q_{i,1}Y_1 + q_{i,2}Y_2 = b_i, i = 1, 2$ is inside the ellipse. Let $[Y_1^*, Y_2^*]$ denote the crossing point of the two lines, which can be determined as $[Y_1^*, Y_2^*] = \mathbf{Q}^T[b_1, b_2]^T$. It is sufficient to prove that $\lambda_1(Y_1^*)^2 + \lambda_2(Y_2^*)^2 < b_3$, which is equivalent to having

$$[b_1, b_2]\mathbf{A}[b_1, b_2]^T < b_3. \tag{96}$$

This holds as

$$\begin{aligned}&[b_1, b_2]\mathbf{A}[b_1, b_2]^T - b_3\\&= -\frac{\sigma_W^2}{1-\tilde{\rho}^2}\left[\sqrt{(Z_1 - 1)(Z_2 - 1)} - \tilde{\rho}\right]^2 < 0\end{aligned} \tag{97}$$

where Z_i were defined right below (50), and hence obtain the desired result.

Assume $\lambda_1 > \lambda_2$. Solving quadratic function of $|\lambda\mathbf{I_2} - \mathbf{A}| = 0$, we obtain eigenvalues λ_1 and λ_2:

$$\lambda_{1,2} = \frac{1}{2}(g_1 + g_2)\left[1 \pm \sqrt{1-\Delta}\right] \tag{98}$$

where $\Delta = \frac{4g_1g_2(1-\tilde{\rho}^2)}{(g_1+g_2)^2}$. The entries of eigenvectors can be computed accordingly:

$$\begin{aligned}q_{1,1} &= \sqrt{\frac{\lambda_1 - g_2}{2\lambda_1 - g_1 - g_2}},\ q_{1,2} = -\sqrt{\frac{\lambda_2 - g_2}{2\lambda_2 - g_1 - g_2}}\\q_{2,1} &= \sqrt{\frac{\lambda_1 - g_1}{2\lambda_1 - g_1 - g_2}},\ q_{2,2} = \sqrt{\frac{\lambda_2 - g_1}{2\lambda_1 - g_1 - g_2}}.\end{aligned} \tag{99}$$

Based on (98), we have $2\lambda_1 - g_1 - g_2 = -(2\lambda_2 - g_1 - g_2) = (g_1 + g_2)\sqrt{1-\Delta}$. Due to the non-negativeness of the ratios involved in (99), it can be shown that $\lambda_1 > g_1 > g_2 > \lambda_2$. In addition, because of $\lambda_1 + \lambda_2 = g_1 + g_2$, $q_{i,j}$'s satisfy:

$$q_{1,1} > |q_{1,2}|,\ q_{2,1} < q_{2,2}. \tag{100}$$

We can therefore conclude that the lengths of the semi-axis $1/\sqrt{\lambda_1}$ and $1/\sqrt{\lambda_2}$ of the ellipse in the direction of Y_1 and Y_2, respectively, satisfy $1/\sqrt{\lambda_1} < 1/\sqrt{\lambda_2}$.

Also, since the slopes of the lines $q_{2,1}Y_1+q_{2,2}Y_2 = b_2$ and $q_{1,1}Y_1+q_{1,2}Y_2 = b_1$ have the relationship of $q_{2,1}/q_{2,2} < 1 < q_{1,1}/|q_{1,2}|$, in addition, $[Y_1^*, Y_2^*]$ is inside the ellipse, the minimum distance in (95) is attained at the point where the line with smaller slope intersects with the ellipse, as illustrated by Figure 4, which implies to minimize the total transmission power $P_1 + P_2$, the second constraint on P_2 in (90), as well as the third one, should be active. This is equivalent to having: $\tilde{R}_2 = I(U_2; Z, U_1)$ and $\tilde{R}_1 + \tilde{R}_2 = I(U_1, U_2; Z) + I(U_2; U_1)$, and $\tilde{R}_1 = I(U_1; Z)$.

Consequently, when $g_1 > g_2$, the decoding at the fusion center follows exactly as that described in Theorem 5.

□

References

[1] M .Gastpar and M. Vetterli, "On the capacity of large Gaussian relay networks," *IEEE Transactions on Information Theory*, vol. 51, no. 3, pp. 765–779, March 2005.

[2] M .Gastpar and M. Vetterli, "Power, spatio-temporal bandwidth, and distortion in large sensor networks," *IEEE Journal on Selected Areas in Communications*, vol. 23, no. 4, pp. 745–754, April 2005.

[3] M. Gastpar, "Uncoded transmission is exactly optimal for a simple Gaussian "sensor" network," in *Information Theory and Applications Workshop*, UCSD, Jan. 2007.

[4] H. Behroozi and M. R. Soleymani, "Power-distortion performance of successive coding strategy in Gaussian CEO problem," in *ICASSP*, Nov. 2006, vol. 4, pp. IV–277–IV280.

[5] Amos Lapidoth and Stephan Tinguely, "Sending a bi-variate Gaussian source over a Gaussian mac," in *Proceedings International Symposium on Information Theory*, 2006.

[6] Y. Oohama, "Rate-distortion theory for Gaussian multiterminal source coding systems with several side information at the decoder," *IEEE Transactions on Information Theory*, vol. 51, no. 7, pp. 2577–2593, July 2005.

[7] T. M. Cover and J. A. Thomas, *Elements of Information Theory*, Wiley, New York, 2006, 2nd Edition.

[8] T. W. Anderson, *An Introduction to Multivariate Statistical Analysis*, Wiley Inter-Science, New York, 2003, 3rd Edition.

[9] V. Prabhakaran, D. Tse, and K. Ramchandran, "Rate region of the quadratic Gaussian CEO problem," in *Proceedings of ISIT*, 2004.

[10] S. Boyd and L. Vandenberghe, *Convex Optimization*, Cambridge University Press, New York, 2004.

[11] Stephen P. Boyd, Seung Jean Kimy, Lieven Vandenberghez, and Arash Hassibix, "A tutorial on geometric programming," *Optimization and Engineering*, 2006, To appear.

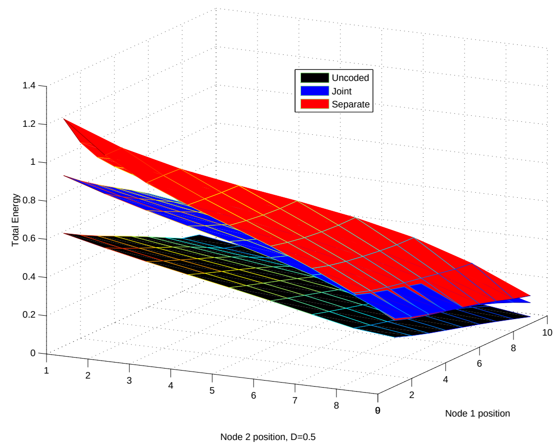
[12] M. Chiang, "Geometric programming for communication systems," *Short monograph in Foundations and Trends in Communications and Information Theory*, vol. 2, no. 1-2, pp. 1–154, July 2005.

[13] S. Wei, R. Kannan, S. Iyengar, and N. Rao, "Energy efficient data fusion in estimating Gaussian sources with sensor networks," *Tech. Report, Dept. ECE LSU*, 2007.

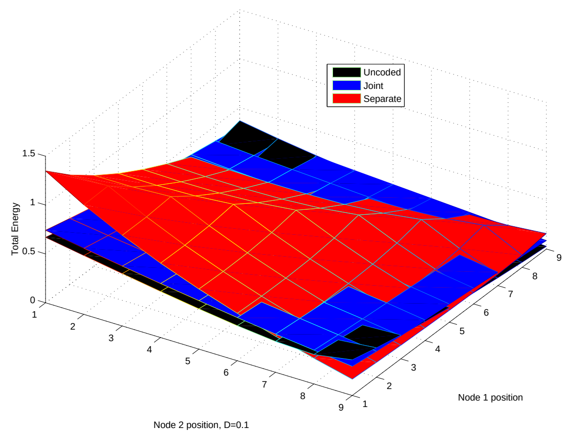
[14] D. Tse and S. Hanly, "Multi-access fading channels: Part I: Polymatroid structure, optimal resource allocation and throughput capacities," *IEEE Transactions on Information Theory*, vol. 44, no. 7, pp. 2796–2815, Nov. 1998.

[15] M .Gastpar and M. Vetterli, "Source-channel communication in sensor networks," in *Proc. IPSN 2003*, 2003.

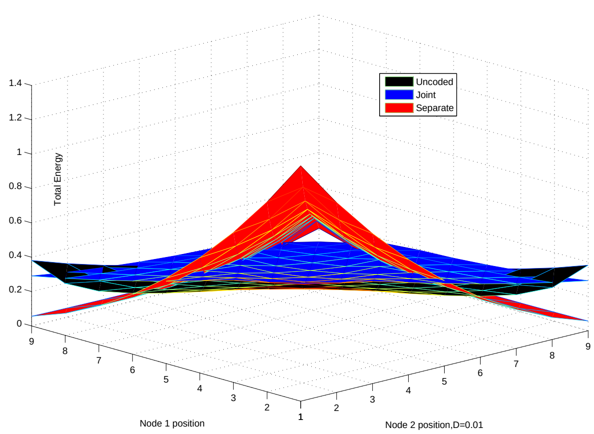
[16] G. Caire and K. Narayanan, "On the distortion snr exponent of hybrid digital-analog space-time coding," *IEEE Transactions on Information Theory*, vol. 53, no. 8, pp. 2867–2878, Aug. 2007.



(a) Result 1: $D = 0.5$, $\sigma_S^2 = \sigma_W^2 = 1$, $\beta_c = \beta_s = 2$.



(b) Results 2: $D = 0.1$, $\sigma_S^2 = \sigma_W^2 = 1$, $\beta_c = \beta_s = 2$
.



(c) Results 3: $D = 0.01$, $\sigma_S^2 = \sigma_W^2 = 1$, $\beta_c = \beta_s = 2$
.

Figure 3: Total power consumption for separate source-channel coding (Red), joint source-channel coding (Blue) and uncoded (Black) schemes

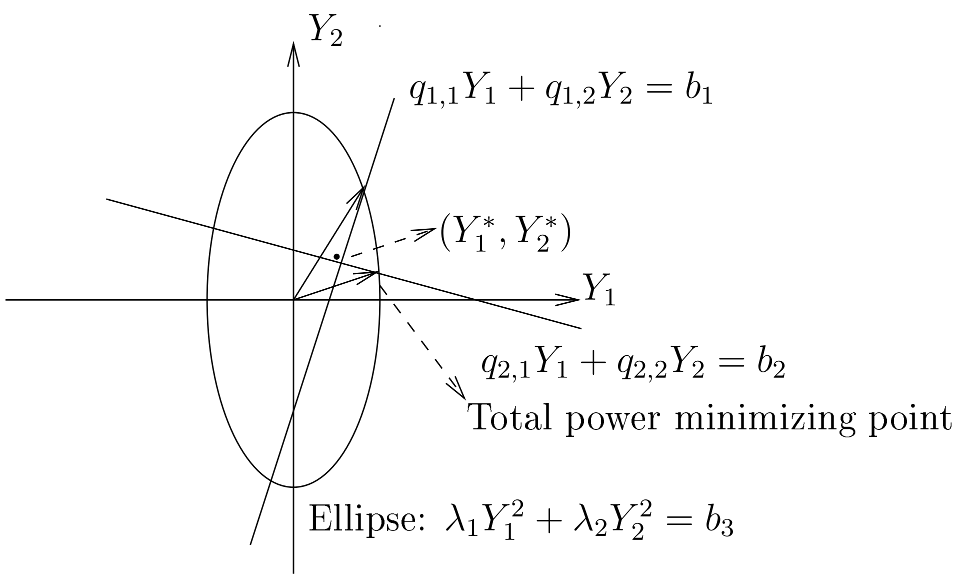
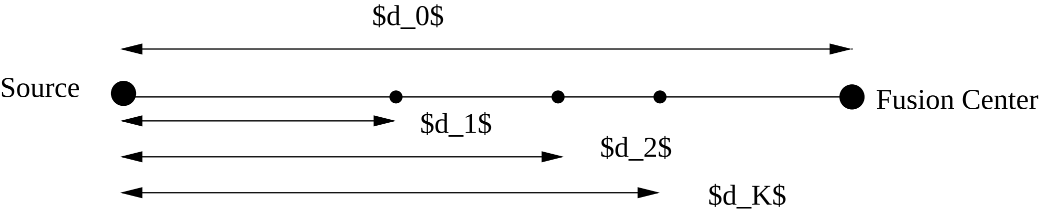


Figure 4: Total power minimization


d_0
Source
Fusion Center
d_1
d_2
d_K