Energy-Distortion Tradeoffs in Gaussian Joint Source-Channel Coding Problems

Aman Jain, Deniz Gündüz, Sanjeev R. Kulkarni, *Fellow, IEEE*, H. Vincent Poor, *Fellow, IEEE*, and Sergio Verdú, *Fellow, IEEE*

***Abstract*—The information-theoretic notion of energy efficiency is studied in the context of various joint source-channel coding problems. The minimum transmission energy $E(D)$ required to communicate a source over a noisy channel so that it can be reconstructed within a target distortion D is analyzed. Unlike the traditional joint source-channel coding formalisms, no restrictions are imposed on the number of channel uses per source sample. For single-source memoryless point-to-point channels, $E(D)$ is shown to be equal to the product of the minimum energy per bit $E_{b\min}$ of the channel and the rate-distortion function $R(D)$ of the source, regardless of whether channel output feedback is available at the transmitter. The primary focus is on Gaussian sources and channels affected by additive white Gaussian noise under quadratic distortion criteria, with or without perfect channel output feedback. In particular, for two correlated Gaussian sources communicated over a Gaussian multiple-access channel, inner and outer bounds on the energy-distortion region are obtained, which coincide in special cases. For symmetric channels, the difference between the upper and lower bounds on energy is shown to be at most a constant even when the lower bound goes to infinity as $D \to 0$. It is also shown that simple uncoded transmission schemes perform better than the separation-based schemes in many different regimes, both with and without feedback.**

***Index Terms*—Energy efficiency, feedback, information theory, joint source-channel coding, multiple-access channel (MAC), separate source and channel coding, uncoded transmission.**

I. Introduction

A FUNDAMENTAL problem in communications is to transmit a message from a source terminal to a destination over a noisy channel such that the destination can reconstruct the source message with the highest fidelity. In general, we can associate a cost for using the channel and also define the fidelity of the reconstruction by a distortion function. Naturally, there is a tradeoff between the available budget for transmission and the achievable distortion at the destination. In classical models, it is assumed that there is an average budget per use of the channel as well as a fixed bandwidth ratio that specifies the number of channel uses per source sample. Then, the problem is to find the minimum distortion achievable for the given average budget and a specified bandwidth ratio which characterizes the *power-distortion tradeoff* of the given system.

In this paper, we introduce the notion of an *energy-distortion tradeoff*. The "energy" refers to the cost of using the communication channel per source observation. Thus, to properly capture the use of energy in this joint compression-communication framework, we relax the following two related restrictions: first, rather than constraining the cost of each channel use for a fixed bandwidth ratio, we constrain the total budget (per source sample) used over all the channel uses; second, we place no restriction on the number of channel uses allowed per source observation (*bandwidth ratio*). In this model, by removing the restrictions on bandwidth ratio, we identify the fundamental limit on the minimum energy requirements without any constraints on spectral efficiency.

The main objective of this paper is to explore the possibility of reducing the energy consumption in joint source-channel coding problems by allowing an unrestricted number of channel uses per observation. To do so, we first cast the problem of energy-distortion tradeoff within an information-theoretic framework. We show that, for point-to-point settings, separation holds for memoryless stationary sources and channels. However, our main focus is on the case in which Gaussian bivariate sources are to be communicated over an additive white Gaussian noise (AWGN) affected multiple-access channel (MAC) with or without feedback.

A potential application of our model is in wireless sensor networks where a physical phenomenon is observed at the sensor nodes and is to be reconstructed at a fusion center. Ultrawideband has been considered as a viable communication strategy for sensor networks because of several benefits including good performance in the low-power regime [6]. In most sensor network applications, the sensors are expected to be severely energy constrained while the required information rates are relatively low. As we show in this paper, in such networks, removing the constraint on the bandwidth ratio substantially reduces the energy requirements in many cases.

For a single-source point-to-point communication system, separate source and channel coding is known to be optimal in

Manuscript received August 13, 2010; revised September 28, 2011; accepted October 06, 2011. Date of publication January 31, 2012; date of current version April 17, 2012. This work was supported in part by the Center for Science of Information, the National Science Foundation (NSF) Science and Technology Center under Grant CCF-0939370, and by the NSF under Grant CNS-09-05398. The work of D. Gündüz was supported by the European Commission's Marie Curie International Reintegration Grant Fellowship with Reference 256410 under the Seventh Framework Programme, and by the Spanish Ministry of Science and Innovation under Project TEC2010-17816 (JUNTOS). The material in this paper was presented in part at the 2010 IEEE Information Theory Workshop.

A. Jain was with the Department of Electrical Engineering, Princeton University, Princeton, NJ 08544 USA. He is now with Goldman Sachs & Co., New York, NY 10282 USA (e-mail: amanjain@princeton.edu).

D. Gündüz is with the Centre Tecnològic de Telecomunicacions de Catalunya, Parc Mediterrani de la Tecnologia, E-08860 Barcelona, Spain (e-mail: deniz.gunduz@cttc.es).

S. R. Kulkarni, H. V. Poor, and S. Verdú are with the Department of Electrical Engineering, Princeton University, Princeton, NJ 08544 USA (e-mail: kulkarni@princeton.edu; poor@princeton.edu; verdu@princeton.edu).

Communicated by M. Gastpar, Associate Editor for Shannon Theory.

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TIT.2012.2184912

terms of the power-distortion tradeoff. Naturally, the optimality of separation applies to the energy-distortion tradeoff as well: for a given level of distortion D, the minimal value of the transmission energy $E(D)$ is achieved by lossy compression (at the rate $R(D)$ per source sample) followed by channel encoding in the most energy efficient manner, i.e., by operating the channel in the wideband regime such that the transmitter uses minimum energy per bit $E_{b\min}$. In fact, this analogy extends to a general cost function on channel use to yield the *cost-distortion* tradeoff for the source and channel pair. Similarly to the power-distortion tradeoff, the cost-distortion (and hence, the energy-distortion) tradeoff is unchanged in the presence of feedback when the channel is memoryless. The results for the single-source scenario are presented in Section II.

The situation is considerably more complicated for multiuser settings. It is well known that the optimality of source-channel separation does not extend to multiuser scenarios other than in a number of special cases [1], [7]. Taking the next natural step from the single-user scenario, in Section III we introduce the problem with two sources that are to be conveyed to a single destination through an additive memoryless Gaussian MAC. For the two-source model, we are interested in the set $\mathcal{E}(D_1, D_2)$ of energy consumption pairs (E_1, E_2) which can achieve the distortion pair (D_1, D_2) for the two sources. As we show in Section IV, there is a provable energy efficiency advantage in increasing the bandwidth ratio in some situations.

In addition to studying the simple setup where no channel output feedback is available at the encoders in Section IV, in Section V we consider the effects of the availability of perfect instantaneous channel output feedback. The model with feedback finds possible applications in sensor networks for which the fusion center (central receiver) has abundant power and bandwidth and can provide accurate feedback about its channel observations to the energy-limited sensor nodes. For the case of unit bandwidth ratio, these models have been studied in [10] and [11] with and without feedback, respectively (see also [16] and references therein). An interesting result in [10] and [11] is that uncoded transmission is optimal when the channel signal-to–noise ratio is below a certain threshold.

Exact characterization of the region $\mathcal{E}(D_1, D_2)$ is a difficult problem in the most general form. We provide outer (converse) bounds on $\mathcal{E}(D_1, D_2)$ with and without feedback. For the inner (achievability) bounds, in each case, we propose a separate source and channel coding scheme and an uncoded transmission scheme. In the proposed separate source and channel coding scheme, the observations are compressed into digital messages (see, e.g., [14] and [19]), which are then orthogonally transmitted to the receiver. When feedback is not available, a very simple uncoded transmission scheme in which both encoders transmit suitably scaled versions of their observations (see [11] and references therein) is more efficient than the separation-based scheme for large distortions. When feedback is available, we propose an uncoded transmission scheme which is motivated by the capacity achieving coding scheme for a Gaussian MAC [15]. The main idea of the scheme is for both transmitters to keep improving the estimates at the receiver using very low power uncoded transmissions of the "estimation error" at the receiver. The coding scheme in [15] is extended in [12] to a MAC with noisy feedback, proving that its effectiveness is not limited to the perfect feedback scenario. For the symmetric setup, we show that the energy-distortion tradeoff achieved by uncoded transmission is close to the lower bound. In fact, numerical experiments suggest that uncoded transmission outperforms separation for the symmetric case.

A related problem, where two or more sensors observe independent noisy versions of a single Gaussian source and communicate them to a central receiver over a Gaussian MAC with or without feedback, has been studied in [3] and [4] for a finite bandwidth ratio and in [8] from an energy-distortion perspective. In these cases, the uncoded transmission schemes are either exactly optimal or optimal in a scaling sense (for a large number of sensors).

II. Single-Source Scenario

We begin by studying the single source and point-to-point communication channel scenario, both with and without feedback. We define and study the cost-distortion tradeoff for such channels, a special case of which is the energy-distortion tradeoff.

A. System Model

Consider m independent and identically distributed (i.i.d.) realizations of a source, according to the common distribution P_S. We denote these m outcomes as a vector $S^m = (S_1, S_2, \ldots, S_m)$. The vector S^m is observed at an encoder (transmitter) which maps it onto a channel codeword $X^n = (X_1, \ldots, X_n)$ of length n. The channel input X_n undergoes a random transformation to the output Y_i observed at a decoder (receiver), for $i = 1, \ldots, n$. The transformation is characterized by the stationary, memoryless conditional distribution $P_{Y|X}$.

If there is no feedback, the channel input X_i is a function only of S^m, i.e.

$$X_i = f_i\left(S^m\right) \tag{1}$$

for some $f_i : \mathbb{R}^m \to \mathbb{R}$, for $i = 1, \ldots, n$. If feedback is available, we allow causal and perfect channel output feedback to the encoder, i.e.

$$X_i = f_i\left(S^m, Y^{i-1}\right) \tag{2}$$

for some $Y^{i-1} = (Y_1, \ldots, Y_{i-1})$ and $f_i : \mathbb{R}^{m+i-1} \to \mathbb{R}$, for $i = 1, \ldots, n$.

The task at the decoder is to generate an estimate $\hat{S}_j$ of each of the source realizations S_j, for $j = 1, \ldots, m$. These estimates are functions of the channel outputs at the receiver, i.e.

$$\hat{S}_j = g_j\left(Y^n\right) \tag{3}$$

where $g_j : \mathbb{R}^n \to \mathbb{R}$, for $j = 1, \ldots, m$.

The decoder needs to ensure that the average distortion (given by a function $d(\hat{S}, S) \in \mathbb{R}_+$) does not exceed a target level. At the same time, at the encoder, the total average cost of transmitting X^n (as given by the cost function $b(X) \in \mathbb{R}_+$), normalized by m, is restricted to be less than some cost constraint. Define a (D, E, m, n) code to be a collection of encoding functions $\{f_i\}_{i=1}^n$ and decoding functions $\{g_j\}_{j=1}^m$ that satisfy

$$\sum_{j=1}^{m} \mathbb{E}\left[d(\hat{S}_j, S_j)\right] \leq m\, D \tag{4}$$

and

$$\sum_{i=1}^{n} \mathbb{E}\left[b(X_i)\right] \leq m\, E \tag{5}$$

for $D, E \geq 0$. Note that the cost restriction in (5) scales linearly in the number of observations m rather than the number of channel uses n, which is unlike the usual formulation of classical joint source-channel coding problems. This allows us to remove the constraints on n for a given m, and study the cost per observation rather than in terms of channel uses.

Define the *bandwidth ratio* to be the ratio of channel uses and the number of observations, i.e., n/m. For a fixed distortion target D, we define the *cost-distortion tradeoff* function for the given setup as

$$E(D) = \min\Big\{E : \text{for all } \epsilon > 0,\ \text{a } (D+\epsilon, E+\epsilon, m, n) \text{ code exists for some } m, n\ \in \mathbb{N}\Big\}. \tag{6}$$

Note that the definition of $E(D)$ does not impose any requirement on the bandwidth ratio, and, therefore, truly reflects the ultimate fundamental limit on the transmission cost incurred for a given distortion. In this paper, we are interested in the Gaussian channels where the "cost" of using the channel is the energy expended in transmission, thus turning the cost-distortion tradeoff into energy-distortion tradeoff.

B. Characterization of the Cost-Distortion Tradeoff

The optimal cost-distortion tradeoff for a single source and point-to-point channel can be achieved by source-channel separation. In the source-channel separation scheme, the source is compressed into as few information bits as possible and then those bits are transmitted reliably to the receiver with as little cost incurred per bit as possible.

To state this result, we recall some well-known definitions. For the communication channel characterized by $P_{Y|X}$, the capacity per unit cost C is given by [18]

$$\mathsf{C} = \sup_{P>0} \frac{C(P)}{P} \tag{7}$$

where $C(P) = \sup_{P_X : \mathbb{E}[b(X)] \leq P} I(X;Y)$ is the capacity cost function for the channel. For the particular case where cost is the transmission energy, we define the minimum energy per bit $E_{b\min}$ to be

$$E_{b\min} = \mathsf{C}^{-1} = \inf_{P>0} \frac{P}{C(P)}. \tag{8}$$

Similarly, the rate-distortion function for the source P_S is given by

$$R(D) = \inf_{\substack{P_{\hat{S}|S}: \\ \mathbb{E}[d(\hat{S},S)] \leq D}} I(\hat{S}; S). \tag{9}$$

Theorem 1: The cost-distortion tradeoff function is equal to

$$E(D) = \frac{R(D)}{\mathsf{C}} \tag{10}$$

regardless of whether channel output feedback is available at the transmitter.

Proof: It readily follows from established results, as shown in Appendix A. ∎

Theorem 1 along with (8) immediately implies the following result on energy-distortion tradeoff.

Corollary 1: The energy-distortion tradeoff function is equal to

$$E(D) = E_{b\min} \times R(D) \tag{11}$$

regardless of whether channel output feedback is available at the transmitter.

C. Gaussian Source and Channel Under Quadratic Cost and Distortion

For the AWGN channel and the memoryless Gaussian source, let the source variance be denoted as σ_S^2 and the communication channel be characterized by $Y_i = X_i + Z_i$ where the noise is i.i.d. Gaussian with variance σ_Z^2. Furthermore, we define the channel cost function as $b(x) = x^2$ and the distortion function as $d(\hat{s}, s) = (\hat{s} - s)^2$.

For this formulation, we have that

$$E_{b\min} = 2\sigma_Z^2 \log_e 2 \tag{12}$$

and

$$R(D) = \frac{1}{2} \log_2^+ \left(\frac{\sigma_S^2}{D}\right) \tag{13}$$

where $\log^+(x) = \log(x)$ if $x \geq 1$ and 0 otherwise. Therefore, Corollary 1 gives

$$E(D) = \sigma_Z^2 \log_e^+ \left(\frac{\sigma_S^2}{D}\right). \tag{14}$$

Note that in order to achieve (14) for any $D < \sigma_S^2$, we cannot use the uncoded scheme of Goblick [5] due to the restriction $m = n$. On the other hand, for an AWGN channel with perfect channel output feedback, the optimal tradeoff can be achieved by the simple uncoded Schalkwijk–Kailath (SK) scheme [9]. The SK scheme can also be adapted to joint source-channel coding for the transmission of a Gaussian source over an AWGN channel [17]. That modified joint source-channel coding SK scheme does not require the compression of the source, yet it achieves the optimal power-distortion tradeoff for any fixed bandwidth ratio [17]. By using the modified SK scheme [17] with high enough bandwidth ratio, we can approach (14) as closely as desired.

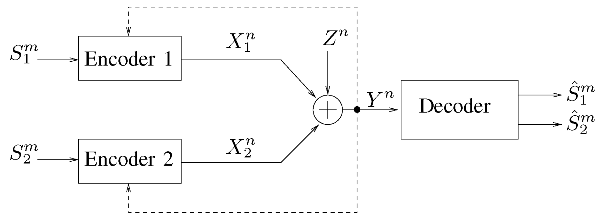

Fig. 1. Setup with two correlated memoryless Gaussian sources and an AWGN MAC.

III. Two-Source Scenario: Basic Setup

We proceed to study the case of two correlated Gaussian sources being communicated to a central receiver over a Gaussian MAC. For this purpose, we need to extend the definition of energy-distortion tradeoff to include the case of multiple sources. To do so, we first introduce the notion of energy-distortion tradeoff region in this section.

Consider a Gaussian MAC with two encoders and one decoder. The encoders observe m i.i.d. realizations of a correlated and jointly Gaussian source pair denoted by (S_1, S_2). Therefore, the first encoder observes $S_1^m = (S_{1,1}, S_{1,2}, \ldots, S_{1,m})$ and the second encoder observes $S_2^m = (S_{2,1}, S_{2,2}, \ldots, S_{2,m})$. We let $S_{k,j} \sim \mathcal{N}(0, \sigma_k^2)$ for $k = 1, 2$ and $\mathbb{E}[S_{1,j} S_{2,j}] = \rho\sigma_1\sigma_2$, for $j = 1, \ldots, m$, where ρ is the coefficient of correlation between the two source components.

We focus our attention on the AWGN MAC. Hence, the n channel outputs $Y^n = (Y_1, \ldots, Y_n)$ at the receiver are given by

$$Y_i = X_{1,i} + X_{2,i} + Z_i \tag{15}$$

where $Z_i \sim \mathcal{N}(0, \sigma_Z^2)$ are i.i.d., for $i = 1, \ldots, n$. The receiver (decoder) uses Y^n to generate estimates $\hat{S}_{k,j}$ of $S_{k,j}$:

$$\hat{S}_{k,j} = g_{k,j}(Y^n) \tag{16}$$

where $g_{k,j} : \mathbb{R}^n \to \mathbb{R}$, for $k = 1, 2$ and $j = 1, \ldots, m$. For the case of no feedback, the encoders map their observation vectors to n channel inputs $X_k^n = (X_{k,1}, \ldots, X_{k,n})$ through the encoding functions $f_{k,i} : \mathbb{R}^m \to \mathbb{R}$, i.e.

$$X_{k,i} = f_{k,i}(S_k^m) \tag{17}$$

for $k = 1, 2$ and $i = 1, \ldots, n$. When perfect, causal feedback is available at the encoders, the channel inputs $X_{k,i}$ are additionally dependent on the prior channel outputs $Y^{i-1} = (Y_1, \ldots, Y_{i-1})$, i.e.

$$X_{k,i} = f_{k,i}\left(S_k^m, Y^{i-1}\right) \tag{18}$$

for some $f_{k,i} : \mathbb{R}^{m+i-1} \to \mathbb{R}$ for $k = 1, 2$ and $i = 1, \ldots, n$ (see Fig. 1).

Given σ_1^2, σ_2^2, σ_Z^2, and ρ, define a $(D_1, D_2, E_1, E_2, m, n)$ code to be a collection of encoding and decoding functions that satisfy

$$\sum_{j=1}^{m} \mathbb{E}[(\hat{S}_{k,j} - S_{k,j})^2] \leq m\, D_k \tag{19}$$

and

$$\sum_{i=1}^{n} \mathbb{E}[(X_{k,i})^2] \leq m\, E_k \tag{20}$$

for $k = 1, 2$. We further assume that $D_k \leq \sigma_k^2$ for $k = 1, 2$.

For a fixed target distortion pair (D_1, D_2), we define (E_1, E_2) to be an achievable energy consumption point if m, n and a $(D_1 + \epsilon, D_2 + \epsilon, E_1 + \epsilon, E_2 + \epsilon, m, n)$ code exist for all $\epsilon > 0$. The energy-distortion tradeoff region (denoted by $\mathcal{E}(D_1, D_2)$) is defined to be the collection of all achievable energy consumption points. We note that the set $\mathcal{E}(D_1, D_2)$ is closed and convex.

In the symmetric case in which we set $\sigma_1 = \sigma_2 = \sigma$, $E_1 = E_2 = E$, and $D_1 = D_2 = D$, the energy-distortion region is completely characterized by

$$E_{\text{sym}}(D) = \min\left\{E : (E, E) \in \mathcal{E}(D, D)\right\}. \tag{21}$$

IV. Two-Source Scenario: No Feedback

In this section, we study the case in which no feedback is available. In particular, we provide an outer bound (converse result) and two inner bounds (achievability results) on the energy-distortion tradeoff region.

A. Converse

The following theorem provides a converse on the energy requirements in the setup with no feedback.

Theorem 2: For the setup with no feedback, any $(E_1, E_2) \in \mathcal{E}(D_1, D_2)$ must satisfy

$$E_k \geq \frac{\sigma_Z^2}{(1-\hat{\rho})^2} \log_e\left(\frac{\sigma_k^2}{D_k}(1-\rho^2)\right) \tag{22}$$

for $k = 1, 2$, and

$$E_1 + E_2 + 2\hat{\rho}\sqrt{E_1 E_2} \geq 2\sigma_Z^2 (\log_e 2)\, R_{S_1,S_2}(D_1, D_2) \tag{23}$$

for some $0 \leq \hat{\rho} \leq |\rho|$, where $R_{S_1,S_2}(D_1, D_2)$ is the minimum sum rate needed to achieve both D_1 and D_2 at the receiver when the encoders cooperate [see (76)–(77)].

Proof: See Appendix B. ■

Theorem 2 immediately implies the following corollary, by setting $\sigma_1 = \sigma_2 = \sigma$, $E_1 = E_2 = E$, and $D_1 = D_2 = D$.

Corollary 2: For the symmetric setting, we have a lower bound on $E_{\text{sym}}(D)$ given by

$$E_{\text{lb}}(D) = \begin{cases} \min_{0 \leq \hat{\rho} \leq |\rho|} \max\left\{\frac{\sigma_Z^2}{(1-\hat{\rho})^2} \log_e^+\left(\frac{\sigma^2}{D}(1-\rho^2)\right), \frac{\sigma_Z^2}{2(1+\hat{\rho})} \log_e^+\left((1-\rho^2)\frac{\sigma^4}{D^2}\right)\right\} & \text{if } |\rho| \leq 1 - \frac{D}{\sigma^2} \\ \min_{0 \leq \hat{\rho} \leq |\rho|} \max\left\{\frac{\sigma_Z^2}{(1-\hat{\rho})^2} \log_e^+\left(\frac{\sigma^2}{D}(1-\rho^2)\right), \frac{\sigma_Z^2}{2(1+\hat{\rho})} \log_e^+\left(\frac{1+|\rho|}{\frac{2D}{\sigma^2}+|\rho|-1}\right)\right\} & \text{if } |\rho| > 1 - \frac{D}{\sigma^2} \end{cases} \tag{24}$$

B. Achievability

For the achievability part, we analyze two different schemes. The first one is separate source and channel coding. In this scheme, the source coding part relies on the Gaussian two-terminal source coding problem which has been considered earlier in [14] and [19]. In the first step, encoder k encodes its observations using an average of R_k bits per observation. In the next step, these bits are transmitted to the receiver with minimum energy expenditure ($E_{b\min}$) per encoded bit. Furthermore, we let both encoders use the MAC orthogonally such that they do not interfere with each other. Apart from the practical reasons due to the modularity it provides, separate source and channel coding is also motivated by its theoretical optimality in the point-to-point scenario.

Theorem 3: Without feedback, any (E_1, E_2) pair satisfying the following conditions belongs to $\mathcal{E}(D_1, D_2)$:

$$E_1 \geq \sigma_Z^2 \log_e^+ \left(\frac{\sigma_1^2}{D_1} \left(1 - \rho^2 \left(1 - e^{-E_2/\sigma_Z^2}\right)\right)\right) \quad (25)$$

$$E_2 \geq \sigma_Z^2 \log_e^+ \left(\frac{\sigma_2^2}{D_2} \left(1 - \rho^2 \left(1 - e^{-E_1/\sigma_Z^2}\right)\right)\right) \quad (26)$$

and

$$E_1+E_2 \geq \sigma_Z^2 \log_e^+ \left(\frac{(1-\rho^2)\sigma_1^2\sigma_2^2}{2D_1D_2} \left(1+\sqrt{1+\frac{4\rho^2 D_1 D_2}{(1-\rho^2)^2\sigma_1^2\sigma_2^2}}\right)\right). \quad (27)$$

Proof: See Appendix C. ■

Theorem 3 immediately implies the following corollary, by setting $\sigma_1 = \sigma_2 = \sigma$, $E_1 = E_2 = E$, and $D_1 = D_2 = D$.

Corollary 3: For the symmetric setting, we have an upper bound on $E_{\text{sym}}(D)$ given by

$$E_{\text{sep}}(D) = \max\left\{ \sigma_Z^2 \log_e^+ \left(\frac{(1-\rho^2)\sigma^2}{2D} \left(1+\sqrt{1+\frac{4\rho^2 D}{(1-\rho^2)^2\sigma^2}}\right)\right), \frac{\sigma_Z^2}{2} \log_e^+ \left(\frac{(1-\rho^2)\sigma^4}{2D^2} \left(1+\sqrt{1+\frac{4\rho^2 D^2}{(1-\rho^2)^2\sigma^4}}\right)\right)\right\}. \quad (28)$$

Remark 1: There is a finite gap between the curves $E_{\text{sep}}(D)$ and $E_{\text{lb}}(D)$ even as $D \to 0$, given by

$$\lim_{D\to 0} E_{\text{sep}}(D) - E_{\text{lb}}(D) = \frac{\sigma_Z^2}{2} \log_e \left(\frac{1}{1-\rho^2}\right) \quad (29)$$

whereas both $E_{\text{sep}}(D)$ and $E_{\text{lb}}(D)$ go to infinity as $D \to 0$.

Next, we turn our attention to another transmission scheme in which the transmitters simply transmit scaled versions of their observations (and thus, have a bandwidth ratio of unity). The primary motivation for considering an uncoded scheme is its optimality in related settings (see, e.g., [3], [5], and [10]). Since the bandwidth ratio of the transmission scheme proposed in the proof of Theorem 4 is unity, the results of Theorem 4 are also directly available from [10] and [16] by replacing power constraints with energy constraints.

Theorem 4 ([10], [16]): Without feedback, any (E_1, E_2) pair satisfying the following conditions belongs to $\mathcal{E}(D_1, D_2)$:

$$\frac{D_1}{\sigma_1^2} \geq \frac{(1-\rho^2)E_2 + \sigma_Z^2}{E_1 + E_2 + 2|\rho|\sqrt{E_1E_2} + \sigma_Z^2} \quad (30)$$

and

$$\frac{D_2}{\sigma_2^2} \geq \frac{(1-\rho^2)E_1 + \sigma_Z^2}{E_1 + E_2 + 2|\rho|\sqrt{E_1E_2} + \sigma_Z^2}. \quad (31)$$

Proof: Available in [10, Th. IV. 3]. ■

We note that unlike the separation-based achievability result in which the bandwidth ratio approaches infinity, uncoded transmission has unit bandwidth ratio. It is known that for the setting in Fig. 1 for a unit bandwidth ratio, uncoded transmission is optimal in terms of power-distortion tradeoff at low enough powers [10].

Corollary 4: For the symmetric setting, we have an upper bound on $E_{\text{sym}}(D)$ given by

$$E_{\text{unc}}(D) = \frac{\sigma_Z^2\left(1-\frac{D}{\sigma^2}\right)}{2(1+|\rho|)\frac{D}{\sigma^2} - (1-\rho^2)} \quad (32)$$

for $D > \sigma^2(1-|\rho|)/2$. If $D \leq \sigma^2(1-|\rho|)/2$, then D cannot be achieved using the uncoded transmission scheme proposed in the proof of Theorem 4.

Remark 2: We note that all the upper and lower bounds (viz., bounds given in Corollaries 2, 3, and 4) on $E_{\text{sym}}(D)$ presented in this section decrease with $|\rho|$. An intuitive explanation of this fact is that less information needs to be transmitted from the two encoders to the receiver when the correlation between the observations is higher.

C. Numerical Examples

Using numerical examples, we first examine the reduction in the energy consumption due to bandwidth expansion. Fig. 2 compares the lower bound on the energy requirements when the bandwidth ratio is 1 [10, Corollary 4.1], and the upper bound obtained from the separation-based scheme without any constraint on the bandwidth ratio. As is clear from Fig. 2, for low distortion, significant energy savings are possible by expanding the bandwidth of transmission.

Next, we compare the two achievability schemes and the converse bound. In Fig. 3, we show the lower bound on $E_{\text{sym}}(D)$, i.e., $E_{\text{lb}}(D)$, obtained from Corollary 2, for the source correlation values of $\rho = 0.2$ and 0.8, under the assumption that σ^2 and σ_Z^2 have unit variance. Also plotted are the upper bounds $E_{\text{sep}}(D)$ and $E_{\text{unc}}(D)$ obtained from Corollary 3 and Corollary 4, respectively. The x-axis represents the distortion D and the y-axis represents the energy requirement $E_{\text{sym}}(D)$.

A couple of observations are worth pointing out.

1) At low correlation values ($\rho = 0.2$), the performance of separation-based coding is very close to the lower bound for all target distortion values. The gap is larger at high correlation values ($\rho = 0.8$).

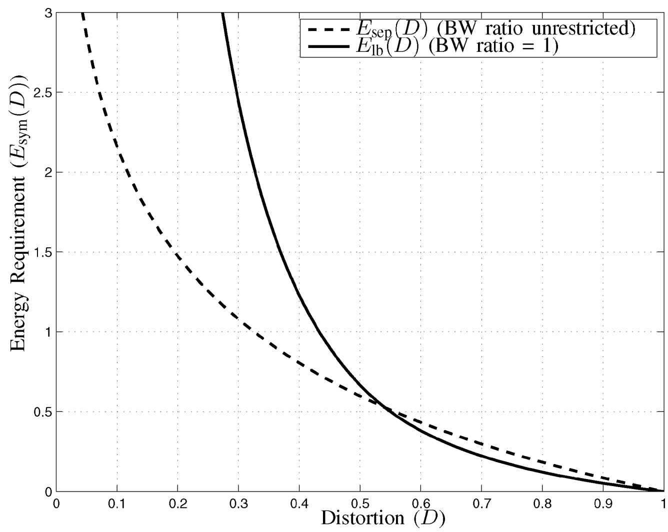

Fig. 2. Upper bound (based on separate source and channel coding, bandwidth ratio unrestricted) and lower bound (for $\text{bandwidth ratio} = 1$) on $E_{\text{sym}}(D)$ for $\rho = 0.5$ and no feedback.

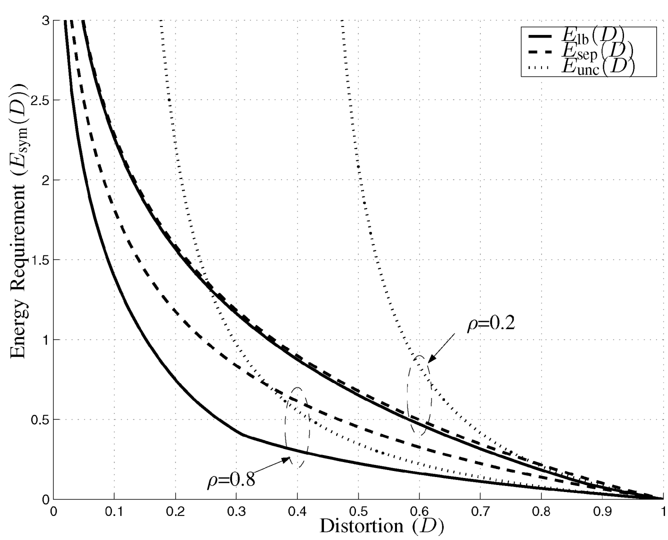

Fig. 3. Upper and lower bounds on $E_{\text{sym}}(D)$ for $\rho = 0.2, 0.8$, and no feedback.

2) For any correlation, there are large enough distortion values such that the uncoded transmission has lower energy requirements than the separation-based scheme. This demonstrates the suboptimality of separate source and channel coding (as proposed in the proof of Theorem 3) in terms of the energy-distortion tradeoff.

V. Two-Source Scenario: Feedback

In this section, we study the Gaussian MAC with noiseless, causal feedback. We propose a converse as well as a new uncoded transmission scheme which, unlike the one in Section IV, makes use of the feedback link. We also note that separate source-channel coding as proposed for the setup without feedback also carries over to the feedback case.

A. Converse

Theorem 5: For the case when feedback is present, any $(E_1, E_2) \in \mathcal{E}(D_1, D_2)$ satisfies

$$E_k \geq \frac{\sigma_Z^2}{(1-\hat{\rho})^2} \log_e \left(\frac{\sigma_k^2}{D_k}(1-\rho^2) \right) \tag{33}$$

for $k = 1, 2$, and

$$E_1 + E_2 + 2\hat{\rho}\sqrt{E_1 E_2} \geq 2\sigma_Z^2 \left(\log_e 2\right) R_{S_1,S_2}(D_1, D_2) \tag{34}$$

for some $0 \leq \hat{\rho} \leq 1$.

Proof: The proof is similar to the proof of Theorem 2 as given in Appendix B, except that now there are no restrictions on the correlations $\hat{\rho}_i$ between the transmissions $X_{1,i}$ and $X_{2,i}$. Thus, Lemma 2 does not hold in the presence of feedback. ■

Remark 3: The only difference between the converses with feedback (Theorem 5) and without feedback (Theorem 2) is that the correlation $\hat{\rho}$ between the transmissions is bounded by $|\rho|$ in the case where feedback is absent.

Theorem 5 immediately implies the following corollary.

Corollary 5: For the symmetric setting, we have a lower bound on $E_{\text{sym}}(D)$ given by

$$E_{\text{lb}}(D) = \begin{cases} \min_{0\leq\hat{\rho}\leq 1} \max\left\{ \frac{\sigma_Z^2}{(1-\hat{\rho})^2} \log_e^+ \left(\frac{\sigma^2}{D}(1-\rho^2) \right), \frac{\sigma_Z^2}{2(1+\hat{\rho})} \log_e^+ \left((1-\rho^2)\frac{\sigma^4}{D^2} \right) \right\} \\ \qquad \text{if } |\rho| \leq 1 - \frac{D}{\sigma^2} \\ \min_{0\leq\hat{\rho}\leq 1} \max\left\{ \frac{\sigma_Z^2}{(1-\hat{\rho})^2} \log_e^+ \left(\frac{\sigma^2}{D}(1-\rho^2) \right), \frac{\sigma_Z^2}{2(1+\hat{\rho})} \log_e^+ \left(\frac{1+|\rho|}{\frac{2D}{\sigma^2}+|\rho|-1} \right) \right\} \\ \qquad \text{if } |\rho| > 1 - \frac{D}{\sigma^2} \end{cases} \tag{35}$$

B. Achievability

Similarly to the setup with no feedback, we study two different achievability schemes. However, since the separate source and channel coding scheme proposed in Section IV-B does not use the feedback link, it also works for the setting with feedback. So, for the feedback case, whenever we mention a separate source and channel coding scheme, it refers to the scheme discussed in Section IV-B. Also proposed in this section is an uncoded transmission scheme that makes use of the feedback link, similar to the SK scheme for the single-user case [2], [9], [17].

The basic idea of the uncoded transmission scheme is similar to the SK scheme for a point-to-point channel. In every step, using the perfect channel output feedback, each transmitter calculates the "error" for its own source, i.e., the difference between the minimum mean-square error (MMSE) estimate at the receiver and the actual source realization. These errors are then scaled and transmitted simultaneously by both transmitters over the MAC. The transmission power for every channel use is taken to be fixed and very small (approaching zero). Based on the received signals, the receiver updates its estimates for both the sources, which is known at the transmitters as well. The scheme is terminated as soon as the target distortions for both sources are achieved at the receiver. We note that the scheme proposed here is similar to the channel-coding scheme proposed in [15], with the main difference being the elimination of "quantization" and "mapping" steps.

ALGORITHM Uncoded Transmission (P, λ)

1) Define: $\epsilon_{k,0} = S_k$, $\mathbb{E}[\epsilon_{k,0}^2] = \sigma_k^2$, $\hat{S}_{k,0} = 0$ for $k = 1$, 2, and $\hat{\rho}_0 = |\rho|$.
2) Execute the following steps for every time $t \in \mathbb{N}$ until

$$\mathbb{E}[\epsilon_{k,t}^2] \leq D_k \tag{36}$$

for $k = 1, 2$:

1) Encoder 1 transmits

$$X_{1,t} = \sqrt{\frac{P}{\mathbb{E}[\epsilon_{1,t-1}^2]}}\,\epsilon_{1,t-1} \tag{37}$$

and encoder 2 transmits

$$X_{2,t} = \sqrt{\frac{\lambda P}{\mathbb{E}[\epsilon_{2,t-1}^2]}}\,\epsilon_{2,t-1}\mathrm{sgn}(\hat{\rho}_{t-1}) \tag{38}$$

where $\mathrm{sgn}(x)$ denotes the sign of x and is taken to be -1 for $x = 0$;

2) The received signal at the receiver is

$$Y_t = X_{1,t} + X_{2,t} + Z_t \tag{39}$$

where $Z_t \sim \mathcal{N}(0, \sigma_Z^2)$; and

3) The receiver (and transmitters) update

$$\hat{S}_{k,t} = \hat{S}_{k,t-1} - \frac{\mathbb{E}[Y_t \epsilon_{k,t-1}]}{\mathbb{E}[Y_t^2]} Y_t \tag{40}$$

and

$$\epsilon_{k,t} = \epsilon_{k,t-1} - \frac{\mathbb{E}[Y_t \epsilon_{k,t-1}]}{\mathbb{E}[Y_t^2]} Y_t \tag{41}$$

where

$$\mathbb{E}[Y_t^2] = P(1+\lambda) + 2P\sqrt{\lambda}|\hat{\rho}_{t-1}| + \sigma_Z^2 \tag{42}$$

$$\mathbb{E}[Y_t \epsilon_{1,t-1}] = \sqrt{\mathbb{E}[\epsilon_{1,t-1}^2]}\sqrt{P}\left(1 + \sqrt{\lambda}|\hat{\rho}_{t-1}|\right) \tag{43}$$

$$\mathbb{E}[Y_t \epsilon_{2,t-1}] = \sqrt{\mathbb{E}[\epsilon_{2,t-1}^2]}\sqrt{P}\left(\sqrt{\lambda} + |\hat{\rho}_{t-1}|\right)\mathrm{sgn}(\hat{\rho}_{t-1}) \tag{44}$$

$$\mathbb{E}[\epsilon_{1,t}^2] = \mathbb{E}[\epsilon_{1,t-1}^2]\frac{\lambda P(1-\hat{\rho}_{t-1}^2) + \sigma_Z^2}{P(1+\lambda+2\sqrt{\lambda}|\hat{\rho}_{t-1}|) + \sigma_Z^2} \tag{45}$$

$$\mathbb{E}[\epsilon_{2,t}^2] = \mathbb{E}[\epsilon_{2,t-1}^2]\frac{P(1-\hat{\rho}_{t-1}^2) + \sigma_Z^2}{P(1+\lambda+2\sqrt{\lambda}|\hat{\rho}_{t-1}|) + \sigma_Z^2} \tag{46}$$

and

$$\hat{\rho}_t = \frac{\hat{\rho}_{t-1}\sigma_Z^2 - \mathrm{sgn}(\hat{\rho}_{t-1})\sqrt{\lambda}P(1-\hat{\rho}_{t-1}^2)}{\sqrt{P(1-\hat{\rho}_{t-1}^2) + \sigma_Z^2}\sqrt{\lambda P(1-\hat{\rho}_{t-1}^2) + \sigma_Z^2}}. \tag{47}$$

The algorithm operates on individual source pairs (S_1, S_2), and aims to achieve a distortion of D_1 and D_2 in their respective reconstructions at the receiver. The algorithm takes as parameters the values of P and λ, such that each transmission by encoder 1 has energy P and each transmission by encoder 2 has energy λP. The internal variables/parameters are $\hat{S}_{k,t}$, $\epsilon_{k,t}$ and $\hat{\rho}_t$. The variable $\hat{S}_{k,t}$ tracks the best estimate of S_k at the receiver based on all the information (i.e., $\hat{S}_{k,t-1}$ and Y_t) available at the receiver by time t. The variable $\epsilon_{k,t} = \hat{S}_{k,t-1} - S_{k,t-1}$ is the "error" in the reconstruction at the receiver, and is what actually is transmitted by the encoders (up to a scaling factor). The quantity $\hat{\rho}_t$ evolves deterministically over time and denotes the correlation between the two errors at time t (and thus, between the two transmissions at time $t+1$). Throughout the rest of the discussion in this section, we treat $\epsilon_{k,t}$ as the distortion achieved at the receiver at time t. All the notation is kept as consistent as possible with [15].

For a given target distortion pair (D_1, D_2), it can be shown that the uncoded transmission algorithm terminates for some choice of P and any $\lambda > 0$. Furthermore, the following result provides an upper bound on the energy consumption of the algorithm.

Theorem 6: For the setting with feedback, choose any $\delta, \lambda > 0$. Then, for $0 < P < P_0$ for some P_0, the uncoded transmission scheme terminates within time

$$T = \left\lceil \frac{(1+\delta)\,\sigma_Z^2}{P} \max\left\{\log_e\left(\frac{\sigma_1^2}{D_1}\right), \frac{1}{\lambda}\log_e\left(\frac{\sigma_2^2}{D_2}\right)\right\}\right\rceil. \tag{48}$$

Furthermore, the energy consumption point

$$(E, \lambda E) \tag{49}$$

where

$$E = \sigma_Z^2 \max\left\{\log_e\left(\frac{\sigma_1^2}{D_1}\right), \frac{1}{\lambda}\log_e\left(\frac{\sigma_2^2}{D_2}\right)\right\} \tag{50}$$

is achievable.

Proof: See Appendix D. ■

Remark 4: Note that by setting $\lambda = \log(\sigma_2^2/D_2)/\log(\sigma_1^2/D_1)$, we get the achievable energy consumption pair

$$\left(\sigma_Z^2\log_e\left(\frac{\sigma_1^2}{D_1}\right), \sigma_Z^2\log_e\left(\frac{\sigma_2^2}{D_2}\right)\right) \tag{51}$$

which can also be achieved with orthogonal transmissions, i.e., by treating the system as two separate single source point-to-point channels. However, also note that the achievability point (49) in Theorem 6 is just an upper bound on the actual energy consumption of the uncoded transmission scheme. An accurate estimate of the energy incurred by uncoded transmission is difficult to obtain in the general case. Theorem 7 in the following gives an analytical result concerning the energy consumption of uncoded transmission for the symmetric setting.

Theorem 7: For the symmetric setting with feedback, we have an upper bound on $E_{\rm sym}(D)$ given by

$$E_{\rm unc}(D) = \begin{cases} \frac{\sigma_Z^2}{4}\log_e\left(\frac{(1+|\rho|)\sigma^2}{2D-(1-|\rho|)\sigma^2}\right)+\frac{\sigma_Z^2}{2}\left(\frac{D}{2D-(1-|\rho|)\sigma^2}-\frac{1}{1+|\rho|}\right) & \text{if } D\geq\sigma^2(1-|\rho|) \\ \frac{\sigma_Z^2}{4}\log_e\left(\frac{1+|\rho|}{1-|\rho|}\right)+\frac{\sigma_Z^2}{2}\left(\frac{|\rho|}{1+|\rho|}\right)+\sigma_Z^2\log_e\left(\frac{(1-|\rho|)\sigma^2}{D}\right) & \text{if } 0\leq D<\sigma^2(1-|\rho|) \end{cases} \quad (52)$$

Proof: See Appendix E. The main idea of the proof is to approximate the time evolution of $\hat{\rho}_t$, distortion $D_t = \mathbb{E}[\epsilon_{k,t}^2]$, and energy $E_t = tP$, by the following set of differential equations obtained from (40)–(47) and letting $P \to 0$

$$\frac{1}{D_t}\frac{dD_t}{dt} = -(1+|\hat{\rho}_t|)\left(\frac{1}{\sigma_Z^2}\frac{dE_t}{dt}\right) \quad (53)$$

$$\frac{d\hat{\rho}_t}{dt} = -\left(\hat{\rho}_t+\operatorname{sgn}(\hat{\rho}_t)\right)^2(1-\hat{\rho}_t)\left(\frac{1}{\sigma_Z^2}\frac{dE_t}{dt}\right) \quad (54)$$

where $D_0 = \sigma^2$, $\hat{\rho}_0 = |\rho|$, and $E_0 = 0$. ■

Remark 5: For the symmetric setup, the uncoded transmission scheme is exactly optimal when $\rho = 0$ or $|\rho| = 1$. Furthermore, when $\rho = 0$, the separation-based scheme (as proposed in the proof of Theorem 3) is also optimal though it has exactly twice the energy consumption of the lower bound when $|\rho| = 1$. On the other hand, when $|\rho| = 1$, another trivial separation-based scheme in which both the encoders use exactly the same code (within a factor of $+1$ or -1, according to whether $\rho = +1$ or -1) and transmit synchronously to the decoder is optimal. In this case, the MAC effectively reduces to a point-to-point channel with two transmit and one receive antennas.

Remark 6: Using expression (52), it can be shown that there is a finite gap between the curves $E_{\rm unc}(D)$ and $E_{\rm lb}(D)$ even as $D \to 0$, i.e.

$$\lim_{D\to 0} E_{\rm unc}(D) - E_{\rm lb}(D) = \sigma_Z^2\left(\frac{|\rho|}{2(1+|\rho|)}-\frac{1}{4}\log_e\left((1-|\rho|)(1+|\rho|)^3\right)\right) \quad (55)$$

whereas both $E_{\rm unc}(D)$ and $E_{\rm lb}(D)$ go to infinity as $D \to 0$.

On the other hand, note that $E_{\rm lb}(D)$ is different with and without feedback, while $E_{\rm sep}(D)$ is the same. However, the asymptotic gap (as $D \to 0$) between the two curves is still the same, and is given by (29).

We also note that the asymptotic gap (55) of the uncoded transmission scheme is smaller than the asymptotic gap (29) of the separation-based scheme.

C. Numerical Examples

We now compare the two achievability schemes and the converse obtained in Sections V-A and V-B. Fig. 4 shows $E_{\rm lb}(D)$ obtained from Corollary 5, $E_{\rm sep}(D)$ from Corollary 3, and $E_{\rm unc}(D)$ from Theorem 7. The two cases considered are low correlation ($\rho = 0.2$) and high correlation ($\rho = 0.8$). As earlier, the x-axis represents the distortion D, and the y-axis represents the energy requirement $E_{\rm sym}(D)$, under the assumption that $\sigma^2 = \sigma_Z^2 = 1$.

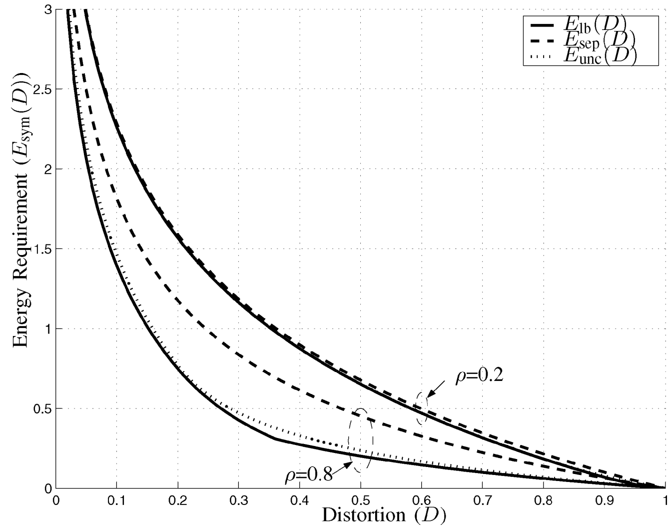

Fig. 4. Upper and lower bounds on $E_{\rm sym}(D)$ for $\rho = 0.2, 0.8$ when feedback is present.

At low correlation values (e.g., $\rho = 0.2$), all the bounds are close to each other. In particular, the gap in the energy requirements of the uncoded transmission scheme and the lower bound is almost indistinguishable except at higher distortion values. However, the bounds are not as tight for $\rho = 0.8$ for which the uncoded transmission scheme has a clear advantage over the separation-based scheme. Comparing the figures with and without feedback for the same correlation coefficient, we note that the lower bound decreases slightly in the presence of feedback, while the separation scheme cannot benefit from the feedback. On the other hand, the uncoded scheme benefits greatly from the availability of feedback which enables it to take advantage of the available bandwidth, and its performance approaches the lower bound for all correlation coefficient values.

While we have closed-form expressions for both $E_{\rm unc}(D)$ and $E_{\rm sep}(D)$, it is difficult to determine analytically whether uncoded transmission always outperforms separate source and channel coding. Numerical simulations suggest that this is indeed the case. For example, Fig. 5 shows the difference in energy requirements (i.e., $E_{\rm sep}(D) - E_{\rm unc}(D)$) for all values of distortion and $|\rho|$.

VI. Conclusion

We have considered the issue of minimal transmission energy requirements in joint source-channel systems. In particular, we have studied an information-theoretic notion of energy efficiency for systems in which observations are communicated from sensors to a central receiver over a wireless medium. We have imposed no restrictions on the kind of signaling schemes that can be employed or the amount of wireless resources (bandwidth) available. In particular, we have defined and studied the energy requirements in two different Gaussian settings: a single source point-to-point channel, and two correlated Gaussian

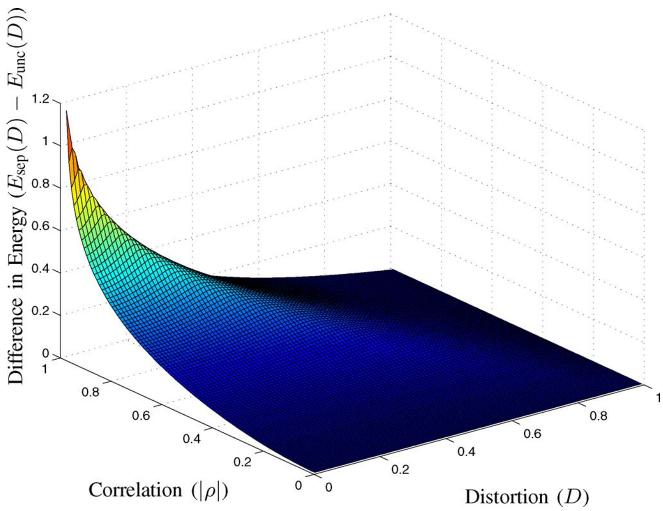

Fig. 5. Excess energy requirement of the separation-based scheme over the uncoded transmission scheme.

sources communicating over a Gaussian MAC. Additionally, for both single-source and two-source cases, we have studied the setting in which noiseless, causal channel output feedback is available at the transmitters.

For the single source point-to-point channel case, we have exactly characterized the minimum transmission energy required per source observation, for a wide class of sources and channels. The minimum energy is given by the product of the minimum energy per bit for the channel part and the rate-distortion function for the source part. As expected, separation is shown to be optimal and the availability of feedback is shown not to decrease the energy requirements.

For the case of two transmitters observing a bivariate memoryless Gaussian source and transmitting over a memoryless Gaussian MAC, we have provided upper and lower bounds on the minimum energy requirement. The upper bounds are obtained by analyzing a separate source and channel coding scheme, and a multiaccess generalization of the SK scheme. With feedback, numerical results suggest that uncoded transmission always has lower energy consumption than separate source and channel coding. We note that when the sources are independent, the upper and lower bounds coincide, both with and without feedback.

For the two-source case with channel feedback, the proposed uncoded transmission scheme is motivated by the achievability part in [15]. Its analysis, however, is complicated due to the fact that the time evolution of the internal variables of the scheme happens in a complex and mutually dependent fashion. We have simplified the analysis by making approximations using a system of differential equations. The solution to this system of differential equations results in the energy-distortion tradeoff achieved by uncoded transmission when the transmission power vanishes.

One of the main points illustrated by this study is that simple uncoded transmission schemes might be attractive in multiuser systems from an energy efficiency perspective, extending similar observations in [3] and [11] to the wideband regime. Furthermore, besides lower computational complexity, uncoded transmission schemes also benefit from their operation on a per symbol basis, drastically reducing both coding delays and storage requirements.

Appendix A
Proof of Theorem 1

Proof: The achievability part is a direct application of separate source and channel coding. The main idea is to first compress the observation vector at a rate of $R(D)$ information bits per observation using a rate-distortion optimal source coding scheme. Next, given large enough block lengths, each of the $R(D)$ bits can be transmitted at an average cost of C^{-1} units per bit by employing an appropriate channel code that achieves the maximum capacity per unit cost C (see [18] and references therein).

We now focus on the converse part. Fixing a distortion target D, for any $\epsilon > 0$, a $(D, E+\epsilon = E(D)+\epsilon, m, n)$ code exists for some $m, n \in \mathbb{N}$. For any such code, we note that

$$I(\hat{S}^m; S^m) \leq I(X^n; Y^n) \tag{56}$$

from the data-processing inequality.

Next, we lower bound the left-hand side of (56)

$$\begin{aligned}
&I(\hat{S}^m; S^m)\\
&\geq \inf_{P_{\hat{S}^m|S^m}: \sum_{j=1}^m \mathbb{E}\left[d(\hat{S}_j, S_j)\right] \leq m\,(D+\epsilon)} I(\hat{S}^m; S^m) && (57)\\
&= \inf_{P_{\hat{S}^m|S^m}: \sum_{j=1}^m \mathbb{E}\left[d(\hat{S}_j, S_j)\right] \leq m\,(D+\epsilon)} \sum_{j=1}^m I(\hat{S}^m; S_j | S^{j-1}) && (58)\\
&\geq \inf_{P_{\hat{S}^m|S^m}: \sum_{j=1}^m \mathbb{E}\left[d(\hat{S}_j, S_j)\right] \leq m\,(D+\epsilon)} \sum_{j=1}^m I(\hat{S}_j; S_j | S^{j-1}) && (59)\\
&= \inf_{P_{\hat{S}^m|S^m}: \sum_{j=1}^m \mathbb{E}\left[d(\hat{S}_j, S_j)\right] \leq m(D+\epsilon)} \sum_{j=1}^m I(\hat{S}_j; S_j) - I(S_j; S^{j-1}) && (60)\\
&= \inf_{P_{\hat{S}^m|S^m}: \sum_{j=1}^m \mathbb{E}\left[d(\hat{S}_j, S_j)\right] \leq m\,(D+\epsilon)} \sum_{j=1}^m I(\hat{S}_j; S_j) && (61)\\
&= m \inf_{P_{\hat{S}|S}: \mathbb{E}[d(\hat{S}, S)] \leq D+\epsilon} I(\hat{S}; S) && (62)\\
&= m\, R(D+\epsilon) && (63)
\end{aligned}$$

where (57) holds since the right-hand side is the minimization of the mutual information over all possible distributions of $\hat{S}^m$ so that the total distortion criterion (4) is satisfied; (58) and (59) follow from the chain rule and nonnegativity of mutual information; (60) follows because S_j–$\hat{S}_j$–S^{j-1} forms a Markov chain; (61) follows from the memoryless source assumption; (62) follows from the convexity of mutual information $I(\hat{S}; S)$ in the conditional distribution $P_{\hat{S}|S}$; and finally, (63) follows from the definition of the rate-distortion function.

We can also upper bound the right-hand side of (56):

$$I(X^n; Y^n) \leq \sup_{P_{X^n}: \sum_{i=1}^n \mathbb{E}[b(X_i)] \leq m\,(E+\epsilon)} I(X^n; Y^n) \tag{64}$$

$$= \sup_{P_{X^n}: \sum_{i=1}^n \mathbb{E}[b(X_i)] \le m\,(E+\epsilon)} \sum_{i=1}^n I(X^n; Y_i | Y^{i-1}) \tag{65}$$

$$\le \sup_{P_{X^n}: \sum_{i=1}^n \mathbb{E}[b(X_i)] \le m\,(E+\epsilon)} \sum_{i=1}^n I(X^n; Y_i) \tag{66}$$

$$= \sup_{P_{X^n}: \frac{1}{n}\sum_{i=1}^n \mathbb{E}[b(X_i)] \le \frac{m}{n}\,(E+\epsilon)} \sum_{i=1}^n I(X_i; Y_i) \tag{67}$$

$$\le n \sup_{P_X: \mathbb{E}[b(X)] \le \frac{m}{n}\,(E+\epsilon)} I(X;Y) \tag{68}$$

$$= n\, C\left(\frac{m}{n}(E+\epsilon)\right) \tag{69}$$

$$\le m(E+\epsilon) \sup_{P>0} \frac{C\,(P)}{P} \tag{70}$$

$$= m(E+\epsilon)\, \mathsf{C} \tag{71}$$

where (64) holds since the right-hand side is the maximization of the mutual information over all distributions of X^n; (66) follows from the fact that Y_i–X^n–Y^{i-1} forms a Markov chain; (67) holds since Y_i depends on X^n only through X_i; (68) follows from the concavity of mutual information $I(X;Y)$ in the distribution of P_X; (69) follows from the definition of channel capacity;(70) is obtained by setting $P = m(E+\epsilon)/n$; and finally, (71) follows from [18, Th. 2]. Note that the arguments for (64)–(71) hold regardless of whether feedback is available at the encoder.

Substituting (63) and (69) into (56), we get

$$R(D+\epsilon) \le (E+\epsilon)\, \mathsf{C}. \tag{72}$$

However, since (72) should hold for all $\epsilon > 0$ and $R(D)$ is a continuous function in D whenever $R(D)$ is finite (see, e.g., [13]), we get that

$$R(D) \le E\, \mathsf{C} \tag{73}$$

immediately establishing the converse. ■

Appendix B
Proof of Theorem 2

Define $R_{S_1|S_2}(D_1)$ to be the minimum rate needed to achieve distortion D_1 at the receiver when S_2 is available at both the first encoder and the receiver. Similarly, we define $R_{S_2|S_1}(D_2)$. It is known that

$$\begin{aligned} R_{S_1|S_2}(D_1) &= \inf_{\substack{P_{\hat{S}_1|S_1,S_2}: \\ \mathbb{E}[(\hat{S}_1 - S_1)^2] \le D_1}} I(\hat{S}_1; S_1 | S_2) \\ &= \frac{1}{2}\log_2^+\left(\frac{\sigma_1^2(1-\rho^2)}{D_1}\right) \end{aligned} \tag{74}$$

and similarly

$$R_{S_2|S_1}(D_2) = \frac{1}{2}\log_2^+\left(\frac{\sigma_2^2(1-\rho^2)}{D_2}\right). \tag{75}$$

Next, we define $R_{S_1,S_2}(D_1,D_2)$ to be the minimum sum rate needed to achieve both D_1 and D_2 at the receiver when the encoders cooperate to encode their observations. It is straightforward to show that (e.g., [21, Th. 6] and [10, Th. 3.1])

$$R_{S_1,S_2}(D_1,D_2) = \inf_{\substack{P_{\hat{S}_1,\hat{S}_2|S_1,S_2}: \\ \mathbb{E}[(\hat{S}_1 - S_1)^2] \le D_1 \\ \mathbb{E}[(\hat{S}_2 - S_2)^2] \le D_2.}} I(\hat{S}_1, \hat{S}_2; S_1, S_2) \tag{76}$$

$$= \begin{cases} \frac{1}{2}\log_2^+\left(\frac{\sigma_1^2}{D_1}\right) & \text{if } \rho^2 \ge \frac{1-\frac{D_2}{\sigma_2^2}}{1-\frac{D_1}{\sigma_1^2}} \\ \frac{1}{2}\log_2^+\left(\frac{\sigma_1^2\sigma_2^2}{D_1 D_2}(1-\rho^2)\right) & \text{if } \rho^2 \le \left(1-\frac{D_2}{\sigma_2^2}\right)\left(1-\frac{D_1}{\sigma_1^2}\right) \\ \frac{1}{2}\log_2^+\left(\frac{1-\rho^2}{\frac{D_1 D_2}{\sigma_1^2\sigma_2^2} - \left(|\rho| - \sqrt{\left(1-\frac{D_1}{\sigma_1^1}\right)\left(1-\frac{D_2}{\sigma_2^1}\right)}\right)^2}\right) & \text{otherwise} \end{cases} \tag{77}$$

under the assumption that $D_1/\sigma_1^2 \le D_2/\sigma_2^2$.

Before providing the proof of Theorem 2, we need a few lemmas.

Lemma 1: If a $(D_1+\epsilon, D_2+\epsilon, E_1+\epsilon, E_2+\epsilon, m, n)$ code exists, then it satisfies

$$m\, R_{S_1|S_2}(D_1+\epsilon) \le \sum_{i=1}^n I(X_{1,i}; Y_i | X_{2,i}) \tag{78}$$

$$m\, R_{S_2|S_1}(D_2+\epsilon) \le \sum_{i=1}^n I(X_{2,i}; Y_i | X_{1,i}) \tag{79}$$

and

$$m\, R_{S_1,S_2}(D_1+\epsilon, D_2+\epsilon) \le \sum_{i=1}^n I(X_{1,i}, X_{2,i}; Y_i) \tag{80}$$

regardless of whether channel feedback is available or not at the transmitters, where $X_{k,i}$ and Y_i, for $k = 1, 2$ and $i = 1, 2, \ldots, n$, are the transmissions from encoder k and the received signals at the decoder, respectively, at time i.

Proof: The proof relies on considering different cut-sets that separate at least one encoder with the decoder. Thus, each cut-set then reduces the setting to a point-to-point source-channel coding problem which admits the use of source-channel separation.

First, consider (80). We note that

$$I(S_1^m, S_2^m; \hat{S}_1^m, \hat{S}_2^m) \le I(S_1^m, S_2^m; Y^n) \tag{81}$$

$$\le I(S_1^m, S_2^m, X_1^n, X_2^n; Y^n) \tag{82}$$

$$= I(X_1^n, X_2^n; Y^n) + I(S_1^m, S_2^m; Y^n | X_1^n, X_2^n) \tag{83}$$

$$= I(X_1^n, X_2^n; Y^n) \tag{84}$$

$$= \sum_{i=1}^n I(X_1^n, X_2^n; Y_i | Y^{i-1}) \tag{85}$$

$$\leq \sum_{i=1}^{n} I(X_1^n, X_2^n; Y_i) \quad (86)$$

$$= \sum_{i=1}^{n} I(X_{1,i}, X_{2,i}; Y_i) \quad (87)$$

where (81) follows from the data-processing inequality; (84) follows by noting that, conditioned on channel inputs X_1^n and X_2^n, the channel output Y^n is independent of S_1^m and S_2^m; (86) follows from the fact that Y_i–(X_1^n, X_2^n)–Y^{i-1} is a Markov chain; and (87) follows by noting that Y_i depends on the pair (X_1^n, X_2^n) only through $(X_{1,i}, X_{2,i})$. Note also that (81)–(87) hold with or without feedback.

At the same time, we can also lower bound the left-hand side of (81) in a manner similar to (57)–(63):

$$\begin{aligned}
&I(S_1^m, S_2^m; \hat{S}_1^m, \hat{S}_2^m) \\
&\geq \inf_{\substack{P_{\hat{S}_1^m, \hat{S}_2^m | S_1^m, S_2^m}: \\ \sum_{j=1}^m \mathbb{E}\left[(\hat{S}_{k,j} - S_{k,j})^2\right] \leq m(D_k + \epsilon), \text{ for } k=1,2}} I(S_1^m, S_2^m; \hat{S}_1^m, \hat{S}_2^m) && (88) \\
&\geq \inf_{\substack{P_{\hat{S}_1^m, \hat{S}_2^m | S_1^m, S_2^m}: \\ \sum_{j=1}^m \mathbb{E}\left[(\hat{S}_{k,j} - S_{k,j})^2\right] \leq m(D_k + \epsilon), \text{ for } k=1,2}} \sum_{j=1}^{m} I(S_{1,j}, S_{2,j}; \hat{S}_{1,j}, \hat{S}_{2,j}) && (89) \\
&\geq m \inf_{\substack{P_{\hat{S}_1, \hat{S}_2 | S_1, S_2}: \\ \mathbb{E}\left[(\hat{S}_k - S_k)^2\right] \leq (D_k + \epsilon), \text{for } k=1,2}} I(S_1, S_2; \hat{S}_1, \hat{S}_2) && (90) \\
&= m\, R_{S_1, S_2}(D_1 + \epsilon, D_2 + \epsilon) && (91)
\end{aligned}$$

where (90) is due to convexity of the mutual information term in the conditional distribution $P_{\hat{S}_1, \hat{S}_2 | S_1, S_2}$. Therefore, (81)–(87) and (88)–(91) together imply (80).

Next, let us consider (78). As earlier

$$\begin{aligned}
I(\hat{S}_1^m; S_1^m | S_2^m) &\leq I(\hat{S}_1^m Y^n; S_1^m | S_2^m) && (92) \\
&= I(Y^n; S_1^m | S_2^m) + I(\hat{S}_1^m; S_1^m | S_2^m, Y^n) && (93) \\
&= I(Y^n; S_1^m | S_2^m) && (94) \\
&\leq I(Y^n; X_1^n | S_2^m) && (95) \\
&= \sum_{i=1}^{n} I(Y_i; X_1^n | S_2^m, Y^{i-1}) && (96) \\
&= \sum_{i=1}^{n} I(Y_i; X_1^n | S_2^m, Y^{i-1}, X_{2,i}) && (97) \\
&= \sum_{i=1}^{n} I(Y_i; X_{1,i} | X_{2,i}) && (98)
\end{aligned}$$

where (94) follows by noting that, conditioned on S_2^m, S_1^m–Y^n–$\hat{S}_1^m$ is a Markov chain; (95) follows from the data-processing inequality; (97) follows since $X_{2,i}$ is a function of S_2^m and possibly, Y^{i-1}; and (98) follows from the fact that Y_i depends on X_1^n, S_2^m, and Y^{i-1} only through $X_{1,i}$ and $X_{2,i}$.

Also, we lower bound the left-hand side of (92) as follows:

$$\begin{aligned}
&I(\hat{S}_1^m; S_1^m | S_2^m) \\
&\geq \inf_{\substack{P_{\hat{S}_1^m | S_1^m, S_2^m}: \\ \sum_{j=1}^m \mathbb{E}\left[(\hat{S}_{1,j} - S_{1,j})^2\right] \leq m\,(D_1 + \epsilon)}} I(\hat{S}_1^m; S_1^m | S_2^m) && (99) \\
&= \inf_{\substack{P_{\hat{S}_1^m | S_1^m, S_2^m}: \\ \sum_{j=1}^m \mathbb{E}\left[(\hat{S}_{1,j} - S_{1,j})^2\right] \leq m\,(D_1 + \epsilon)}} \sum_{j=1}^{m} I(\hat{S}_1^m; S_{1,j} | S_2^m, S_1^{j-1}) && (100) \\
&\geq \inf_{\substack{P_{\hat{S}_1^m | S_1^m, S_2^m}: \\ \sum_{j=1}^m \mathbb{E}\left[(\hat{S}_{1,j} - S_{1,j})^2\right] \leq m\,(D_1 + \epsilon)}} \sum_{j=1}^{m} I(\hat{S}_{1,j}; S_{1,j} | S_2^m, S_1^{j-1}) && (101) \\
&= \inf_{\substack{P_{\hat{S}_1^m | S_1^m, S_2^m}: \\ \sum_{j=1}^m \mathbb{E}\left[(\hat{S}_{1,j} - S_{1,j})^2\right] \leq m\,(D_1 + \epsilon)}} \sum_{j=1}^{m} \Big(I(\hat{S}_{1,j} A; S_{1,j} | S_{2,j}) \\
&\qquad - I(A; S_{1,j} | S_{2,j}) \Big) && (102) \\
&\geq \inf_{\substack{P_{\hat{S}_1^m | S_1^m, S_2^m}: \\ \sum_{j=1}^m \mathbb{E}\left[(\hat{S}_{1,j} - S_{1,j})^2\right] \leq m\,(D_1 + \epsilon)}} \sum_{j=1}^{m} I(\hat{S}_{1,j}; S_{1,j} | S_{2,j}) && (103) \\
&\geq m \inf_{\substack{P_{\hat{S}_1 | S_1, S_2}: \\ \mathbb{E}\left[(\hat{S}_1 - S_1)^2\right] \leq (D_1 + \epsilon)}} I(\hat{S}_1; S_1 | S_2) && (104) \\
&= m\, R_{S_1 | S_2}(D_1 + \epsilon) && (105)
\end{aligned}$$

where (101) is obtained by reducing the set of random variables; in (102), we set $A = \{S_{2,j}^c, S_1^{j-1}\}$ where $S_{2,j}^c = S_2^m \setminus S_{2,j}$; (103) follows since the pair $(S_{1,j}, S_{2,j})$ is independent of A; and (104) is due to convexity of mutual information in conditional distribution.

The inequalities (92)–(98) and (99)–(105) immediately imply (79). The relation (80) can be obtained similarly. ■

We need another lemma which, given the correlated information at the two encoders, puts a limit on the maximum correlation that can be achieved among the transmissions from the two encoders. This result, with Lemma 1, could then be used to provide a limit on the maximum information the two encoders can convey to the receiver.

Lemma 2: For the given system model without feedback, for any encoder pair, we have

$$\mathrm{corr}\,(X_{1,i}, X_{2,i}) \leq \rho \quad (106)$$

where $\mathrm{corr}(X, Y)$ is the correlation between the random variables X and Y.

Proof: The main idea of the proof is along the lines of the proof of [10, Lemma C.1] and uses the following two lemmas.

Lemma 3 ([20, Th. 1]): For a sequence of pairs of independent random variables $(W_{1,i}, W_{2,i})_{i=1}^n$, we have

$$\begin{aligned}
&\sup_{f_1^n, f_2^n} \mathbb{E}\left[f_1^n\left(W_1^n\right) f_2^n\left(W_2^n\right)\right] \\
&\qquad \leq \sup_{i \in \{1,\ldots,n\}, f_{1,i}, f_{2,i}} \mathbb{E}\left[f_{1,i}\left(W_{1,i}\right) f_{2,i}\left(W_{2,i}\right)\right] && (107)
\end{aligned}$$

where $W_k^n = \{W_{k,1}, \ldots, W_{k,n}\}$ for $k = 1, 2$. Also, the supremum in (107) is over the functions f_k^n and $f_{k,i}$ for $k = 1$, 2 and $i = 1, \ldots, n$ satisfying

$$\mathbb{E}\left[f_k^n\left(W_k^n\right)\right] = 0 \tag{108}$$
$$\mathbb{E}\left[\left(f_k^n\left(W_k^n\right)\right)^2\right] = 1 \tag{109}$$
$$\mathbb{E}\left[f_{k,i}(W_{k,i})\right] = 0 \tag{110}$$

and

$$\mathbb{E}\left[\left(f_{k,i}(W_{k,i})\right)^2\right] = 1 \tag{111}$$

for $k = 1$, 2 and $i = 1, \ldots, n$.

The other lemma employs the Hirschfield–Gebelein–Rényi maximal correlation to upper bound the maximal correlation between the transmissions by the two encoders.

Lemma 4 ([22, Sec. IV, Lemma 10.2]): For jointly Gaussian random variables W_1 and W_2 with coefficient of correlation ρ, we have

$$\sup_{f_1, f_2} \mathbb{E}\left[f_1(W_1) f_2(W_2)\right] = |\rho| \tag{112}$$

where the supremum is over all functions f_1 and f_2 satisfying

$$\mathbb{E}[f_k(W_k)] = 0 \tag{113}$$

and

$$\mathbb{E}[(f_k(W_k))^2] = 1 \tag{114}$$

for $k = 1$, 2.

Finally, the proof of Lemma 2 is by noting that the transmissions $X_{1,i}$ and $X_{2,i}$ are functions ($f_{1,i}$ and $f_{2,i}$) of the observation vectors S_1^m and S_2^m, respectively. Notice that $(X_{k,i} - \mathbb{E}[X_{k,i}])/\sqrt{\mathrm{var}(X_{k,i})}$ has zero mean and unit variance. Therefore, for every $i = 1, \ldots, n$

$$\begin{aligned}
&\mathbb{E}\left[(X_{1,i} - \mathbb{E}[X_{1,i}])(X_{2,i} - \mathbb{E}[X_{2,i}])\right] \\
&\quad \leq \sqrt{\mathrm{var}(X_{1,i})\mathrm{var}(X_{2,i})} \sup_{i, f_1, f_2} \mathbb{E}\left[f_1(S_{1,i}) f_2(S_{1,i})\right] && (115) \\
&\quad \leq |\rho| \sqrt{\mathrm{var}(X_{1,i})\mathrm{var}(X_{2,i})} && (116)
\end{aligned}$$

where (115) is directly from Lemma 3 and (116) is from Lemma 4 by noting that the observations $S_{1,i}$ and $S_{2,i}$ are correlated with the coefficient ρ for every $i = 1, \ldots, n$. The inequality (116) immediately implies the statement of Lemma 2. ■

Proof of Theorem 2: Let the correlation between $X_{1,i}$ and $X_{2,i}$ be $\hat{\rho}_i$, for $i = 1, \ldots, n$. We can upper bound the variance of $X_{1,i}$ conditioned on $X_{2,i}$, since the variance of $X_{1,i}$ cannot exceed the MMSE of the linear estimate

$$\mathbb{E}[X_{1,i}] + \hat{\rho}_i \sqrt{\frac{\mathrm{var}(X_{1,i})}{\mathrm{var}(X_{2,i})}} \left(X_{2,i} - \mathbb{E}[X_{2,i}]\right) \tag{117}$$

of $X_{1,i}$. This consideration immediately gives us that

$$\mathrm{var}(X_{1,i}|X_{2,i}) \leq (1 - \hat{\rho}_i^2)\,\mathrm{var}(X_1). \tag{118}$$

We have a similar inequality, for $\mathrm{var}(X_{2,i}|X_{1,i})$. Furthermore

$$\begin{aligned}
&\mathrm{var}\left(X_{1,i} + X_{2,i}\right) \\
&\quad = \mathrm{var}(X_{1,i}) + \mathrm{var}(X_{2,i}) + 2\mathrm{cov}(X_{1,i}, X_{2,i}) && (119) \\
&\quad \leq \mathrm{var}(X_{1,i}) + \mathrm{var}(X_{2,i}) + 2\hat{\rho}_i \sqrt{\mathrm{var}(X_{1,i})\mathrm{var}(X_{2,i})}. && (120)
\end{aligned}$$

Also, define

$$\hat{\rho} = \frac{\sum_{i=1}^n \hat{\rho}_i \sqrt{\mathrm{var}(X_{1,i})\mathrm{var}(X_{2,i})}}{\sqrt{\sum_{i=1}^n \mathrm{var}(X_{1,i})}\sqrt{\sum_{i=1}^n \mathrm{var}(X_{2,i})}} \tag{121}$$

and

$$\mathrm{var}(X_k) = \frac{1}{m} \sum_{i=1}^n \mathrm{var}(X_{k,i}) \tag{122}$$

for $k = 1, 2$. Note that

$$\mathrm{var}(X_k) \leq E_k + \epsilon \tag{123}$$

since $\mathrm{var}(X_{k,i}) \leq \mathbb{E}[X_{k,i}^2]$ for $k = 1$, 2 and $i = 1, \ldots, n$, and due to the restriction (20).

Let us first prove (22) for $k = 1$. Continuing from (78)

$$\begin{aligned}
m\, R_{S_1|S_2}(D_1 + \epsilon) &\leq \sum_{i=1}^n I(X_{1,i}; Y_i | X_{2,i}) && (124) \\
&\leq \sum_{i=1}^n \frac{1}{2} \log_2 \left(1 + \frac{\mathrm{var}(X_{1,i}|X_{2,i})}{\sigma_Z^2}\right) && (125) \\
&\leq \sum_{i=1}^n \frac{1}{2} \frac{(1 - \hat{\rho}_i^2)\mathrm{var}(X_{1,i})}{\sigma_Z^2} \log_e 2 && (126) \\
&= \frac{\sum_{i=1}^n \mathrm{var}(X_{1,i}) - \sum_{i=1}^n \hat{\rho}_i^2 \mathrm{var}(X_{1,i})}{2\sigma_Z^2} \log_e 2 && (127) \\
&\leq \frac{m\,(1 - \hat{\rho}^2)\mathrm{var}(X_1)}{2\sigma_Z^2} \log_e 2 && (128) \\
&\leq \frac{m\,(1 - \hat{\rho}^2)(E_1 + \epsilon)}{2\sigma_Z^2} \log_e 2 && (129)
\end{aligned}$$

where (125) follows from the capacity of an AWGN channel under the constraints on the variance of the channel input; (126) follows by noting that $\log_e(1 + x) \leq x$ for all $x \geq 0$, and from (118); (128) follows by noting that

$$\begin{aligned}
&\sum_{i=1}^n \left(\sqrt{\hat{\rho}_i^2 \mathrm{var}(X_{1,i})}\sqrt{\mathrm{var}(X_{2,i})}\right) \\
&\qquad \leq \sqrt{\sum_{i=1}^n \hat{\rho}_i^2 \mathrm{var}(X_{1,i})}\sqrt{\sum_{i=1}^n \mathrm{var}(X_{2,i})} && (130)
\end{aligned}$$

due to the Cauchy–Schwarz inequality, which immediately implies

$$\sum_{i=1}^{n} \hat{\rho}_i^2 \mathrm{var}(X_{1,i}) \geq \hat{\rho}^2 \sum_{i=1}^{n} \mathrm{var}(X_{1,i}) \tag{131}$$

and hence, (128); and (129) is from (123). The relation (22) is now immediately implied by (129), since (129) should hold for all values of $\epsilon > 0$. Relation (22) for $k = 2$ is also proved similarly.

Next, for (23), consider the following series of manipulations, starting from (80):

$$\begin{aligned} &m\, R_{S_1,S_2}(D_1+\epsilon, D_2+\epsilon) \\ &\leq \sum_{i=1}^{n} I(X_{1,i}, X_{2,i}; Y_i) &(132)\\ &\leq \sum_{i=1}^{n} \frac{1}{2}\log_2\left(1+\frac{\mathrm{var}(X_{1,i}+X_{2,i})}{\sigma_Z^2}\right) &(133)\\ &\leq \sum_{i=1}^{n}\left(\mathrm{var}(X_{1,i})+\mathrm{var}(X_{2,i})+2\hat{\rho}_i\sqrt{\mathrm{var}(X_{1,i})\mathrm{var}(X_{2,i})}\right)\frac{\log_e 2}{2\sigma_Z^2} &(134)\\ &= \left(\sum_{i=1}^{n}\mathrm{var}(X_{1,i})+\sum_{i=1}^{n}\mathrm{var}(X_{2,i}) +2\sum_{i=1}^{n}\hat{\rho}_i\sqrt{\mathrm{var}(X_{1,i})\mathrm{var}(X_{2,i})}\right)\frac{\log_e 2}{2\sigma_Z^2} &(135)\\ &\leq m\left((E_1+\epsilon)+(E_2+\epsilon)+2\hat{\rho}\sqrt{(E_1+\epsilon)(E_2+\epsilon)}\right)\frac{\log_e 2}{2\sigma_Z^2} &(136) \end{aligned}$$

where most of the arguments are similar to those used in (124)–(129), while (134) follows from (120), and (136) follows from the definition of $\hat{\rho}$ in (121) and from (123). Since (136) should hold for all $\epsilon > 0$, (23) follows immediately.

The proof of Theorem 2 can now be concluded by proving that $\hat{\rho} \in [0, |\rho|]$. To do so, we first note that

$$\begin{aligned} 0 &\leq \sum_{i=1}^{n}\sqrt{\mathrm{var}(X_{1,i})\mathrm{var}(X_{2,i})} \\ &\leq \sqrt{\sum_{i=1}^{n}\mathrm{var}(X_{1,i})}\sqrt{\sum_{i=1}^{n}\mathrm{var}(X_{2,i})} \end{aligned} \tag{137}$$

by the Cauchy–Schwarz inequality, which implies $\hat{\rho} \in [0, |\rho|]$ from (121) and the fact that $\hat{\rho}_i \in [0, |\rho|]$ from Lemma 2. ■

APPENDIX C
PROOF OF THEOREM 3

Proof: We prove here that (E_1, E_2) pairs satisfying the conditions in (25)–(27) can be achieved by separate source and channel coding. In the first step, both encoders separately encode their observations (at rates R_1 and R_2, respectively) such that the distortion targets D_1 and D_2 for the two sources are achieved at the receiver. The conditions on R_1 and R_2 for the achievability of (D_1, D_2) are [14], [19]

$$R_1 \geq \frac{1}{2}\log_2^+\left(\frac{\sigma_1^2}{D_1}\left(1-\rho^2\left(1-2^{-2R_2}\right)\right)\right) \tag{138}$$

$$R_2 \geq \frac{1}{2}\log_2^+\left(\frac{\sigma_2^2}{D_2}\left(1-\rho^2\left(1-2^{-2R_1}\right)\right)\right) \tag{139}$$

and

$$R_1+R_2 \geq \frac{1}{2}\log_2^+\left(\frac{(1-\rho^2)\sigma_1^2\sigma_2^2}{2D_1D_2}\left(1+\sqrt{1+\frac{4\rho^2 D_1 D_2}{(1-\rho^2)^2\sigma_1^2\sigma_2^2}}\right)\right). \tag{140}$$

Thereafter, the encoded information bits are communicated to the receiver in separate time-slots by the encoders. Note that this separate (orthogonal) operation reduces the MAC to a point-to-point AWGN channel for each of the transmitters. Thus, the energy requirement for transmitting each information bit (by either of the transmitters) to the receiver is $E_{b\min} = 2\sigma_Z^2 \log_e 2$. Thus, if (R_1, R_2) is an achievable rate pair for the source coding problem (see [14] and [19]), then $(E_1, E_2) = (E_{b\min}R_1, E_{b\min}R_2) \in \mathcal{E}(D_1, D_2)$ is an achievable energy consumption pair. Therefore, the conditions (25)–(27) on (E_1, E_2) can be obtained by replacing R_i with $E_i/E_{b\min}$ in (138)–(140). ■

APPENDIX D
PROOF OF THEOREM 6

Proof: For T given by (48), from (45) and the fact that $\mathbb{E}[\epsilon_{k,0}^2] = \sigma_k^2$, we get

$$\log_e\left(\frac{\mathbb{E}[\epsilon_{1,T}^2]}{\sigma_k^2}\right) = \sum_{t=1}^{T-1}\log_e\left(\frac{\lambda P(1-\hat{\rho}_{t-1}^2)+\sigma_Z^2}{P(1+\lambda+2\sqrt{\lambda}|\hat{\rho}_{t-1}|)+\sigma_Z^2}\right). \tag{141}$$

We can further bound the right-hand side of (141) as follows:

$$\begin{aligned} \frac{\lambda P(1-\hat{\rho}_{t-1}^2)+\sigma_Z^2}{P(1+\lambda+2\sqrt{\lambda}|\hat{\rho}_{t-1}|)+\sigma_Z^2} &\leq \frac{\lambda P+\sigma_Z^2}{P(1+\lambda)+\sigma_Z^2} &(142)\\ &\leq 1-\frac{P}{(1+\delta)\sigma_Z^2} &(143) \end{aligned}$$

for all sufficiently small $P > 0$, where (143) follows by noticing that $P(1+\lambda) \leq \delta\sigma_Z^2$ for all sufficiently small P. From (141), (143), and the fact that $\log_e(1-x) \leq -x$ for $x < 1$, we get

$$\log_e\left(\frac{\mathbb{E}[\epsilon_{1,T}^2]}{\sigma_k^2}\right) \leq -\frac{TP}{(1+\delta)\sigma_Z^2} \tag{144}$$

for all sufficiently small P. Since $\mathbb{E}[\epsilon_{1,T}^2]$ represents the distortion in the reconstruction of S_1 at the receiver at time T, (48) immediately implies that $\mathbb{E}[\epsilon_{1,T}^2] \leq D_1$. A similar result can be proven for the achievability of distortion D_2 for source S_2.

The result for the achievable energy consumption point (49) is straightforward after noting that the energy consumption at encoder/transmitter 1 is TP and at encoder/transmitter 2 is λTP, and that the choice of $\delta > 0$ in (48) is arbitrary. ■

APPENDIX E
PROOF OF THEOREM 7

Proof: We analyze the uncoded transmission algorithm for some $P > 0$ (to be decided later) and $\lambda = 1$. Furthermore, without loss of generality, let us restrict $0 \le \rho \le 1$, since if $\rho < 0$ we could replace S_1 with $-S_1$ without affecting the joint distribution (except changing the sign of ρ), the energy requirements, or the distortion constraints.

Let the algorithm terminate in T time slots. From Theorem 6, it can be deduced that

$$T \le \frac{2\sigma_Z^2 \log_e(\sigma^2/D)}{P}. \tag{145}$$

Let us first focus on the analysis of the behavior of $\hat{\rho}_t$. Note that $\hat{\rho}_0 = \rho \ge 0$. For the time being, let us additionally assume that $\rho < 1$. Let

$$T_0 = \min\{t : \hat{\rho}_{t+1} < 0\} \tag{146}$$

be the time (possibly infinity) before $\hat{\rho}$ hits negative values. In the definition of T_0, we require the uncoded transmission algorithm to keep operating regardless of the stopping condition (36). Next, we show that $\hat{\rho}$ decreases till time T_0 and then settles at a value of (almost) zero.

From (47), $\hat{\rho}_t$ satisfies

$$\hat{\rho}_{t+1} - \hat{\rho}_t = -\frac{(\hat{\rho}_t + \mathrm{sgn}(\hat{\rho}_t))P(1-\hat{\rho}_t^2)}{P(1-\hat{\rho}_t^2) + \sigma_Z^2}. \tag{147}$$

From (147), a Maclaurin series expansion of $\hat{\rho}_{t+1} - \hat{\rho}_t$ in terms of the parameter P leads to

$$\hat{\rho}_{t+1} - \hat{\rho}_t = -(\hat{\rho}_t + \mathrm{sgn}(\hat{\rho}_t))(1-\hat{\rho}_t^2)\frac{P}{\sigma_Z^2} + O(P^2) \tag{148}$$

where $O(f(P))$ represents any term $g(P)$ such that

$$\left|\lim_{P\to 0}\frac{g(P)}{f(P)}\right| < \infty. \tag{149}$$

Note that when $x \in [0, c_1]$ for some $0 \le c_1 < 1$

$$(x+\mathrm{sgn}(x))(1-x^2) \in \left[\min\{1, (1+\rho)(1-\rho^2)\}, \frac{32}{27}\right] \subseteq \left[c_2, \frac{32}{27}\right] \tag{150}$$

for some $c_2 > 0$. Hence, from (148), for all sufficiently small $P > 0$ and $t = 1, \ldots, T_0$, $\hat{\rho}_{t+1} < \hat{\rho}_t$. This, along with (150), implies that

$$\frac{|\hat{\rho}_{t+1} - \hat{\rho}_t|}{P} \in [c_3, c_4] \tag{151}$$

for some constants $0 < c_3 \le c_4$ and all sufficiently small P, for $t = 1, \ldots, T_0$. Therefore

$$T_0 \le \frac{\rho}{c_3 P} \tag{152}$$

since the change in the value of $\hat{\rho}_t$ is at least $c_3 P$ in every time step.

Also, note that the function $\left((1+x)(1-x^2)\right)^{-1}$ is uniformly differentiable over the interval $[0, \rho]$. Thus

$$\max_{x\in[\hat{\rho}_{t+1},\hat{\rho}_t]}\left|\frac{1}{(1+x)(1-x^2)} - \frac{1}{(1+\hat{\rho}_t)(1-\hat{\rho}_t^2)}\right| \le c_5|\hat{\rho}_{t+1} - \hat{\rho}_t| \le c_6 P \tag{153}$$

for some constants c_5 and c_6 and all $t = 1, \ldots, T$, where we have used (151) in obtaining (153).

From (148), for any $T_1 \le T_0$, we get that

$$\sum_{t=0}^{T_1}\frac{1}{(1+\hat{\rho}_t)(1-\hat{\rho}_t^2)}\int_{\hat{\rho}_t}^{\hat{\rho}_{t+1}} d\hat{\rho} = \sum_{t=0}^{T}\frac{\hat{\rho}_{t+1} - \hat{\rho}_t}{(1+\hat{\rho}_t)(1-\hat{\rho}_t^2)} = \sum_{t=0}^{T_1}\left(-\frac{P}{\sigma_Z^2} + O(P^2)\right) \tag{154}$$

which implies

$$\sum_{t=0}^{T_1}\left(\int_{\hat{\rho}_t}^{\hat{\rho}_{t+1}}\left(\frac{1}{(1+\hat{\rho})(1-\hat{\rho}^2)} + O(P)\right)d\hat{\rho}\right) = (T_1+1)\left(-\frac{P}{\sigma_Z^2} + O(P^2)\right) \tag{155}$$

from (153). Since $|\hat{\rho}_{T_1} - \hat{\rho}_0| \le 1$ and $T_1 = O(P)$ [from (152)], (155) yields the following relation:

$$\int_{\rho}^{\hat{\rho}_{T_1}}\frac{1}{(1+\hat{\rho})(1-\hat{\rho}^2)}d\hat{\rho} = -\frac{T_1 P}{\sigma_Z^2} + O(P) \tag{156}$$

which, by noting that $T_1 P$ is the energy expended (by each transmitter) till time T_1, gives us that

$$E_1 = \frac{\sigma_Z^2}{2}\frac{(\rho-\rho_1)}{(1+\rho)(1+\rho_1)} + \frac{\sigma_Z^2}{4}\log_e\left(\frac{1-\rho_1}{1-\rho}\frac{1+\rho}{1+\rho_1}\right) + O(P). \tag{157}$$

Here, $E_1 = T_1 P$ is the energy requirement of taking $\hat{\rho}$ from ρ to $0 < \rho_1 < \rho$. Let $E_0 = T_0 P$ be the energy spent in taking $\hat{\rho}$ from an initial value of ρ to 0. Then

$$E_0 = \frac{\sigma_Z^2}{2}\left(\frac{\rho}{1+\rho} + \frac{1}{2}\log_e\left(\frac{1+\rho}{1-\rho}\right)\right) + O(P). \tag{158}$$

Next, we show that $|\hat{\rho}_t| < O(P)$ for every $t > T_0$. First, we recall that $\hat{\rho}_{T_0+1} < 0$ but $\hat{\rho}_{T_0} \ge 0$. Furthermore, from (148), it can be obtained that

$$|\hat{\rho}_{t+1} - \hat{\rho}_t| < \frac{2P}{\sigma_Z^2} \tag{159}$$

for sufficiently small P, regardless of the value of $\hat{\rho}_t \in [-1, +1]$. Now, suppose that for some $t > T_0$, $|\hat{\rho}_t| > 2P/\sigma_Z^2$ while $|\hat{\rho}_{t-1}| \le 2P/\sigma_Z^2$. This implies, from (159), that $\hat{\rho}_{t-1}$ has the same sign as $\hat{\rho}_t$ since two consecutive time values of $\hat{\rho}$ cannot differ by more that $2P/\sigma_Z^2$. However, from (148), notice that if $|\hat{\rho}_{t-1}| \le 2P/\sigma_Z^2$ then the sign of $(\hat{\rho}_t - \hat{\rho}_{t+1})$ should be the opposite of $\hat{\rho}_{t-1}$ for P small. This, along with (159), implies that

$$|\hat{\rho}_t| \le \frac{2P}{\sigma_Z^2} \tag{160}$$

for all $t \geq T_0$ which contradicts our assumption that $|\hat{\rho}_t| \leq 2P/\sigma_Z^2$. Thus, if $|\hat{\rho}_{t-1}| \leq 2P/\sigma_Z^2$, then $|\hat{\rho}_{t'}| \leq 2P/\sigma_Z^2$ for all $t' \geq t$.

Having understood a little bit about the behavior of $\hat{\rho}$ over time, let us turn our attention to the behavior of the distortion of the estimates (i.e., $D_t = \mathbb{E}[\epsilon_{k,t}^2]$ for $k = 1, 2$) over time t.

From (45) and the fact that $D_0 = \mathbb{E}[\epsilon_{k,0}^2] = \sigma^2$, we immediately obtain

$$\log_e\left(\frac{D_T}{\sigma^2}\right) = \sum_{t=0}^{T} \log_e\left(\frac{P(1-\hat{\rho}_t^2)+\sigma_Z^2}{2P(1+|\hat{\rho}_t|)+\sigma_Z^2}\right) \tag{161}$$

for $k = 1, 2$. However, since $\mathbb{E}[\epsilon_{k,t}^2]$ is also the average distortion in the estimate of S_k at time t, we set $\mathbb{E}[\epsilon_{k,T}^2] = D$ to get

$$\log_e\left(\frac{D}{\sigma^2}\right) = \sum_{t=0}^{T} \log_e\left(\frac{P(1-\hat{\rho}_t^2)+\sigma_Z^2}{2P(1+|\hat{\rho}_t|)+\sigma_Z^2}\right) \tag{162}$$

$$= \sum_{t=0}^{T} \left(-\frac{P}{\sigma_Z^2}(1+|\hat{\rho}_t|)^2 + O(P^2)\right) \tag{163}$$

where (162) follows from (161), and (163) follows from the Maclaurin series expansion of each of the summands in (162) (in the parameter P around $P = 0$).

Next, let us assume that $T \leq T_0$, i.e., the target distortion is attained before $\hat{\rho}_t$ falls below 0. From (163)

$$\log_e\left(\frac{D}{\sigma^2}\right) = \sum_{t=0}^{T} \left(\frac{\hat{\rho}_{t+1}-\hat{\rho}_t}{1-\hat{\rho}_t} + O(P^2)\right) \tag{164}$$

$$= \sum_{t=0}^{T} \left(\int_{\hat{\rho}_{t+1}}^{\hat{\rho}_t} \frac{1}{1-\hat{\rho}}\, d\hat{\rho} + O(P^2)\right) \tag{165}$$

$$= \int_{\rho}^{\hat{\rho}_T} \frac{1}{1-\hat{\rho}}\, d\hat{\rho} + O(P) \tag{166}$$

where we have used (148) and the fact that $\hat{\rho}_t \geq 0$ in obtaining (164); and similar arguments as in (153)–(157) to obtain (165). The equation (166) implies that

$$\hat{\rho}_T = 1 - \frac{\sigma^2}{D} e^{O(P)} \tag{167}$$

which along with (157) and by letting $P \to 0$ yields that

$$E = \frac{\sigma_Z^2}{4} \log_e\left(\frac{(1+\rho)\sigma^2}{2D-(1-\rho)\sigma^2}\right) + \frac{\sigma_Z^2}{2}\left(\frac{D}{2D-(1-\rho)\sigma^2} - \frac{1}{1+\rho}\right) \tag{168}$$

is achievable, when $\hat{\rho}_T \geq 0$. The condition of $\hat{\rho}_T \geq 0$ can alternately be written as

$$D \geq (1-\rho)\sigma^2 \tag{169}$$

as $P \to 0$, from (167). This demonstrates the first part of (52).

To show the second part of (52), let us assume that the termination time $T \geq T_0$. Yet again, from (45) we have

$$\log_e\left(\frac{D}{\sigma^2}\right) = \sum_{t=0}^{T_0} \left(\frac{\hat{\rho}_{t+1}-\hat{\rho}_t}{1-\hat{\rho}_t} + O(P^2)\right) + \sum_{t=T_0+1}^{T} \left(-\frac{P}{\sigma_Z^2}(1+|\hat{\rho}_t|)^2 + O(P^2)\right) \tag{170}$$

$$= \int_{\rho}^{O(P)} \frac{1}{1-\hat{\rho}}\, d\hat{\rho} + \sum_{t=T_0+1}^{T} \left(-\frac{P}{\sigma_Z^2}\right) + O(P) \tag{171}$$

$$= \log_e(1-\rho) - \frac{(E-E_0)}{\sigma_Z^2} + O(P) \tag{172}$$

where we have used (163) and (148) to obtain (170) and used (160) and the Maclaurin expansion of $(1+\hat{\rho}_t)^2$ in the parameter $\hat{\rho}_t$ to obtain the first and second terms in (171). Using (158) with (172) and letting $P \to 0$ gives that

$$E = \frac{\sigma_Z^2}{4} \log_e\left(\frac{1+\rho}{1-\rho}\right) + \frac{\sigma_Z^2}{2}\left(\frac{\rho}{1+\rho}\right) + \sigma_Z^2 \log_e\left(\frac{(1-\rho)\sigma^2}{D}\right) \tag{173}$$

is achievable for $D < (1-\rho)\sigma^2$ as $P \to 0$. This demonstrates the second part of (52) for all $\rho \in [0, 1)$.

Finally, for the case in which $\rho = 1$, note that $S_{1,i} = S_{2,i}$ almost surely, for $i = 1, \ldots, n$. Therefore, each encoder knows the pair $(S_{1,i}, S_{2,i})$ for $i = 1, \ldots, n$, and, hence, could cooperate with the other encoder. This reduces the model to a two transmit antennas and one receive antenna point-to-point system with one source. From Section II, the energy-distortion tradeoff function is given by

$$E_{\text{sym}}(D) = \frac{\sigma_Z^2}{4} \log_e\left(\frac{\sigma^2}{D}\right) \tag{174}$$

taking into account that $E(D)$ given by Corollary 1 needs to be divided by two to account for energy consumed at each transmitter. Note that the expression (174) matches the expression (52) evaluated at $\rho = 1$, and the expression (35) for the lower bound. This establishes the second part of (52) for $\rho = 1$. ■

REFERENCES

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

[2] T. J. Cruise, "Achievement of rate-distortion bound over additive white noise channel utilizing a noiseless feedback channel," *Proc. IEEE*, vol. 55, no. 4, pp. 583–584, Apr. 1967.

[3] M. Gastpar, "Uncoded transmission is exactly optimal for a simple Gaussian "sensor" network," *IEEE Trans. Inf. Theory*, vol. 54, no. 11, pp. 5247–5251, Nov. 2008.

[4] M. Gastpar, "On the role of feedback in large sensor networks," in *Proc. Int. Zurich Semin. Commun.*, Feb. 2004, pp. 188–191.

[5] T. J. Goblick, "Theoretical limitations on the transmission of data from analog sources," *IEEE Trans. Inf. Theory*, vol. 11, no. 4, pp. 558–567, Oct. 1965.

[6] F. Granelli and M.-G. D. Benedetto, "MONET special issue editorial," *ACM/Springer J. Mobile Netw. Appl., Spec. Issue Ultra Wide Band Sens. Netw.*, vol. 11, no. 4, pp. 487–499, Aug. 2006.

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

[8] A. Jain, D. Gündüz, S. Kulkarni, H. V. Poor, and S. Verdú, "Energy efficient lossy transmission over sensor networks with feedback," in *Proc. 35th IEEE Int. Conf. Acoust., Speech, Signal Process.*, Dallas, TX, Mar. 2010, pp. 5558–5561.

[9] T. Kailath, "An application of Shannon's rate-distortion theory to analog communication over feedback channels," *Proc. IEEE*, vol. 55, no. 6, pp. 1102–1103, Jun. 1967.

[10] A. Lapidoth and S. Tinguely, "Sending a bivariate Gaussian over a Gaussian MAC," *IEEE Trans. Inf. Theory*, vol. 56, no. 6, pp. 2714–2752, Jun. 2010.

[11] A. Lapidoth and S. Tinguely, "Sending a bivariate Gaussian source over a Gaussian MAC with feedback," *IEEE Trans. Inf. Theory*, vol. 56, no. 4, pp. 1852–1864, Apr. 2010.

[12] A. Lapidth and M. Wigger, "On the AWGN MAC with imperfect feedback," *IEEE Trans. Inf. Theory*, vol. 56, no. 11, pp. 5432–5476, Nov. 2010.

[13] R. J. McEliece*, The Theory of Information and Coding*. Reading, MA: Addison-Wesley, 1977.

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

[15] L. H. Ozarow, "The capacity of the white Gaussian multiple access channel with feedback," *IEEE Trans. Inf. Theory*, vol. 30, no. 4, pp. 623–628, Jul. 1984.

[16] R. Rajesh and V. Sharma, "Source-channel coding for Gaussian sources over a Gaussian multiple access channel," presented at the 45th Annu. Allerton Conf. Commun., Control Comput., Monticello, IL, Sep. 2007.

[17] J. P. M. Schalkwijk and L. I. Bluestein, "Transmission of analog waveforms through channels with feedback," *IEEE Trans. Inf. Theory*, vol. 13, no. 4, pp. 617–619, Oct. 1967.

[18] S. Verdú, "On channel capacity per unit cost," *IEEE Trans. Inf. Theory*, vol. 36, no. 5, pp. 1019–1030, Sep. 1990.

[19] A. B. Wagner, S. Tavildar, and P. Viswanath, "Rate region of the quadratic Gaussian two-encoder source-coding problem," *IEEE Trans. Inf. Theory*, vol. 54, no. 5, pp. 1938–1961, May 2008.

[20] H. S. Witsenhausen, "On sequences of pairs of dependent random variables," *SIAM J. Appl. Math.*, vol. 28, no. 1, pp. 100–113, Jan. 1975.

[21] J.-J. Xiao and Z.-Q. Luo, "Compression of correlated Gaussian sources under individual distortion criteria," presented at the 43rd Annu. Allerton Conf. Commun., Control, Comput., Monticello, IL, Sep. 2005.

[22] Y. A. Rozanov*, Stationary Random Processes*. San Francisco, CA: Holden-Day, 1967.

Aman Jain received the B.Tech. degree from Indian Institute of Technology – Kanpur, India, in 2005, the M.A. degree from Princeton University in 2007, and the Ph.D. degree from Princeton University in 2010, all in Electrical Engineering.

In 2009, he was a Graduate Fellow with the McGraw Center for Teaching and Learning at Princeton University. Since 2010, he has been with Goldman, Sachs & Co.

Aman Jain received the Outstanding Teaching Assistant Award from the Dept. of E.E., Princeton University. In 2000, he attended the International Mathematics Olympiad Training Camp, India. His research interests include information-theory, wireless communications, networks and signal processing.

Deniz Gündüz received the B.S. degree in electrical and electronics engineering from the Middle East Technical University in 2002, and the M.S. and Ph.D. degrees in electrical engineering from Polytechnic Institute of New York University (formerly Polytechnic University), Brooklyn, NY in 2004 and 2007, respectively.

He is currently a Research Associate at CTTC in Barcelona, Spain. He also held a visiting researcher position at Princeton University. Previously he was a consulting Assistant Professor at the Department of Electrical Engineering, Stanford University and a postdoctoral Research Associate at the Department of Electrical Engineering, Princeton University.

He is the recipient of a Marie Curie Reintegration Grant funded by the European Union's Seventh Framework Programme (FP7), the 2008 Alexander Hessel Award of Polytechnic Institute of New York University given to the best PhD Dissertation, and a recipient of the Best Student Paper Award at the 2007 IEEE International Symposium on Information Theory (ISIT). His research interests lie in the areas of communication theory and information theory with special emphasis on joint source-channel coding, cooperative communications and energy efficiency.

Sanjeev R. Kulkarni (M'91–SM'96–F'04) received the B.S. in Mathematics, B.S. in E.E., M.S. in Mathematics from Clarkson University in 1983, 1984, and 1985, respectively, the M.S. degree in E.E. from Stanford University in 1985, and the Ph.D. in E.E. from M.I.T. in 1991.

From 1985 to 1991, he was a Member of the Technical Staff at M.I.T. Lincoln Laboratory. Since 1991, he has been with Princeton University where he is currently Professor of Electrical Engineering, and an affiliated faculty member in the Department of Operations Research and Financial Engineering and the Department of Philosophy. He spent January 1996 as a research fellow at the Australian National University, 1998 with Susquehanna International Group, and Summer 2001 with Flarion Technologies.

Prof. Kulkarni received an ARO Young Investigator Award in 1992, an NSF Young Investigator Award in 1994. He is a Fellow of the IEEE and has served as an Associate Editor for the IEEE Transactions on Information Theory. Prof. Kulkarni's research interests include statistical pattern recognition, nonparametric statistics, learning and adaptive systems, information theory, wireless networks, and image/video processing.

H. Vincent Poor (S'72–M'77–SM'82–F'87) received the Ph.D. degree in electrical engineering and computer science from Princeton University in 1977. From 1977 until 1990, he was on the faculty of the University of Illinois at Urbana-Champaign. Since 1990 he has been on the faculty at Princeton, where he is the Dean of Engineering and Applied Science, and the Michael Henry Strater University Professor of Electrical Engineering. Dr. Poor's research interests are in the areas of stochastic analysis, statistical signal processing and information theory, and their applications in wireless networks and related fields. Among his publications in these areas are *Quickest Detection* (Cambridge University Press, 2009), co-authored with Olympia Hadjiliadis, and *Information Theoretic Security* (Now Publishers, 2009), co-authored with Yingbin Liang and Shlomo Shamai.

Dr. Poor is a member of the National Academy of Engineering and the National Academy of Sciences, a Fellow of the American Academy of Arts and Sciences, and an International Fellow of the Royal Academy of Engineering (U.K.). He is also a Fellow of the Institute of Mathematical Statistics, the Optical Society of America, and other organizations. In 1990, he served as President of the IEEE Information Theory Society, in 2004–07 as the Editor-in-Chief of these *Transactions*, and in 2009 as General Co-chair of the IEEE International Symposium on Information Theory, held in Seoul, South Korea. He received a Guggenheim Fellowship in 2002 and the IEEE Education Medal in 2005. Recent recognition of his work includes the 2010 IET Ambrose Fleming Medal for Achievement in Communications, the 2011 IEEE Eric E. Sumner Award, the 2011 IEEE Information Theory Paper Award, and the degree of D.Sc. *honoris causa* from the University of Edinburgh, conferred in 2011.

Sergio Verdú (M'84–F'92) received the Telecommunications Engineering degree from the Universitat Politècnica de Barcelona in 1980, and the Ph.D. degree in Electrical Engineering from the University of Illinois at Urbana-Champaign in 1984. Since 1984 he has been a member of the faculty of Princeton University, where he is the Eugene Higgins Professor of Electrical Engineering.

A member of the National Academy of Engineering, Verdú is the recipient of the 2007 Claude Shannon Award and the 2008 IEEE Richard Hamming Medal. He was awarded a Doctorate Honoris Causa from the Universitat Politècnica de Catalunya, Barcelona in 2005.

Verdú's research has received several paper awards including the 1998 and 2011 Information Theory Paper Awards, the Information Theory Golden Jubilee Paper Award, and the 2006 Joint Communications/Information Theory Paper Award.

Sergio Verdú is currently Editor-in-Chief of Foundations and Trends in Communications and Information Theory.