

A Little Feedback Can Simplify Sensor Network Cooperation

Anand Sarwate, *Member, IEEE*, and Michael Gastpar, *Member, IEEE*,

Abstract—Shannon’s discovery of digital communication has shaped the architecture of virtually all communication systems in use today. The digital communication paradigm is built around the notion of bits and uses careful coding to deliver bits reliably end-to-end. It has been shown that this architectural principle can lead to a very significant performance penalty in wireless sensor networks. For a limited class of network scenarios, it was shown that optimal architectures are analog in nature, simple and scalable. In this paper, we show that more generally, simple analog architectures crucially depend on feedback to the sensors. Interesting questions then concern the amount of feedback needed and the resulting trade-off with performance. This paper provides rules-of-thumb for the selection of the number of feedback bits.

Index Terms—joint source-channel coding, fading, sensor networks, data fusion, feedback

I. INTRODUCTION

TO BE SCALABLE and robust, sensor network architectures must be simple. However, they also have to live off of very limited resources, and therefore cannot be too wasteful. Thus, the quest is for architectures that are simple yet have near-optimal performance guarantees. One way to investigate questions of optimality is through the lens of information theory, which can provide insights into the fundamental performance limits of distributed estimation systems such as sensor networks. In this paper we will take as a case study the problem of estimating an underlying source observed by sensors who must forward their observations to a fusion center. Information-theoretic studies of sensor networks have addressed issues such as distributed compression (see [3] for an overview) and channel capacity [4]; a general overview can also be found in [5]. Many of these works are focused on the *scaling law* performance of these systems. That is, how does the asymptotic distortion or capacity scale with the number of sensors in the network?

From an information-theoretic perspective, it is well-known that separately encoding sensor observations and then communicating them independently to a fusion center is suboptimal

Manuscript received 1 May 2009; revised 15 February 2010. Portions of this work appeared in [1], [2]. The work of A.D. Sarwate was supported by an NDSEG Graduate Research Fellowship which is sponsored by the U.S. Department of Defense. The work of M. Gastpar was supported by the National Science Foundation under award CCF-0347298 (CAREER) and NSF CNS-0627024.

A.D. Sarwate is with the Information Theory and Applications Center in the California Institute for Telecommunications and Information Technology (CALIT2) at the University of California, San Diego (e-mail: asarwate@ucsd.edu).

M. Gastpar is with the Department of Electrical Engineering and Computer Sciences at the University of California, Berkeley, and the Department of Electrical Engineering, Mathematics and Computer Science, Delft University of Technology, The Netherlands (e-mail: gastpar@eecs.berkeley.edu).

Digital Object Identifier 10.1109/JSAC.2010.100920.

in terms of the distortion in reconstructing the source. Because this *separation principle* does not hold, joint source-channel coding strategies can yield dramatic improvements over the best approach using separate encoding. In some cases, a very simple analog encoding scheme is optimal. A clear example of this was given in [6], where it was shown that *uncoded transmission* is exactly optimal for a Gaussian source observed by sensors that communicate over a Gaussian channel. More precisely, for this model, the asymptotic distortion scales as $\Theta(M^{-1})$ for the optimal scheme and like $\Omega(1/\log M)$ for a scheme that uses separate source and channel codes. Other works have also investigated linear processing schemes for forwarding information in sensor networks [7]–[9], which give some insights into the sensing performance achievable by simple protocols.

In practical applications, the true correlation structure between the sensors’ observations may not be known prior to deployment. There’s many a slip ’twixt design and deployment¹, and so the protocols we design should be able to adapt to different correlation structures. For example, in networks monitoring electromagnetic phenomena, the relative phases of the sensor’s observations may be quite sensitive to their placement. Because precise positioning may be difficult, the underlying sensor fusion protocol should be robust to this uncertainty. In monitoring or tracking applications the movement of a target or source can create time variation in the sensor field that requires periodic recalibration of the fusion protocol. In this paper we demonstrated how a type of observation uncertainty we call *source fading* can lead to catastrophic failure in the strategy of [6]. We then show that if very limited communication between sensors is allowed, a simple alignment protocol can partially recover the scaling performance. In particular, we show that K bits of feedback can *align* the observations of the sensors, yielding a distortion that scales as $O(M^{-K/(K+2)})$.

The study of feedback has a long and storied history in information theory [11], where it has been shown to lower error rates [12], [13], lower complexity [14]–[16], and provide a bridge between communication and control theory [17], [18]. It can sometimes increase capacity in networks or channels with uncertainty or memory [19]–[31]. In a sense, the method in this paper is a way of using *zero-rate feedback* to combat uncertainty in the observation model. We propose a simple two-phase protocol that uses limited feedback or communication between the sensors: in the first phase the sensors align their observations, and in the second they transmit their information to the fusion center.

¹Apologies to Pallasas [10, X.32]).

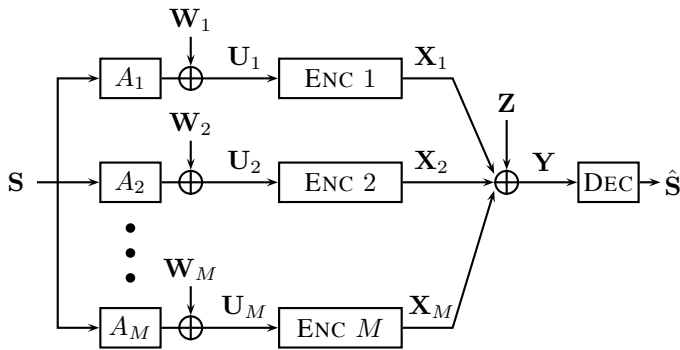


Fig. 1. Gaussian network with fading observations.

II. A MODEL FOR CORRELATION UNCERTAINTY

The basic model we will consider is shown in Figure 1. A source generates a vector of iid symbols $\mathbf{S} = (S[1], S[2], \dots, S[n])$. The M sensors observe the source through the *observation function* $A(\cdot)$:

$$\mathbf{U}_m = \mathbf{A}_m(\mathbf{S}) + \mathbf{W}_m, \quad (1)$$

where \mathbf{W}_m is iid Gaussian noise with variance σ_W^2 , independent across sensors m and time. The observation functions $\{A_m(\cdot)\}$ are random variables taking values in a set of functions \mathcal{A} according to some joint distribution $P_A(a)$. The choice of \mathcal{A} depends on the specifics of the sensors' design and reflects the relationship between the quantities of interest and the actual observed variables. We do not assume the sensors know the realization of $\{A_m(\cdot)\}$, but must estimate it. We will assume furthermore that the functions $\{A_m(\cdot)\}$ do not change over time.

For simplicity, in this paper we will assume $A_m(\cdot)$ is a scalar gain, and with some abuse of notation we write $A_m(s) = A_m \cdot s$. For more examples where $A_m(\cdot)$ is a filter, see [32]. A simple example is the case where each $A_m = 1$ almost surely, in which case we recover the source observation model studied in [6] and many other works.

We call this random observation model *source fading* because the coefficients $\{A_m\}$ act like channel gains in a fading channel with the source as its input. We assume the following:

$$\{A_m\} \text{ are iid with distribution } P_A(a) \quad (2)$$

$$P_A(a) = p_A(-a) \quad (3)$$

$$|A_m| < \nu \quad \text{a.s.} \quad (4)$$

We call this *bounded real scalar fading*. For much of this paper we will further simplify to the case where $\{A_m\}$ are iid and equiprobable on the set $\{-1, +1\}$, which we will refer to as *sign fading*. We assume that the sensors and destination know the distributions of the gains $\{A_m\}$ but not their realization.

The first question to ask for sensors observing faded observations is to what extent they can determine the fading process. Consider the case where each sensor receives $A_m \mathbf{S} + \mathbf{W}_m$, where the A_m are drawn iid from some distribution over a finite set \mathcal{A} . The sensors can estimate their own marginal distribution locally. If two fading gain $a_1, a_2 \in \mathcal{A}$ induce different marginal distributions on U_m at sensor m , then the

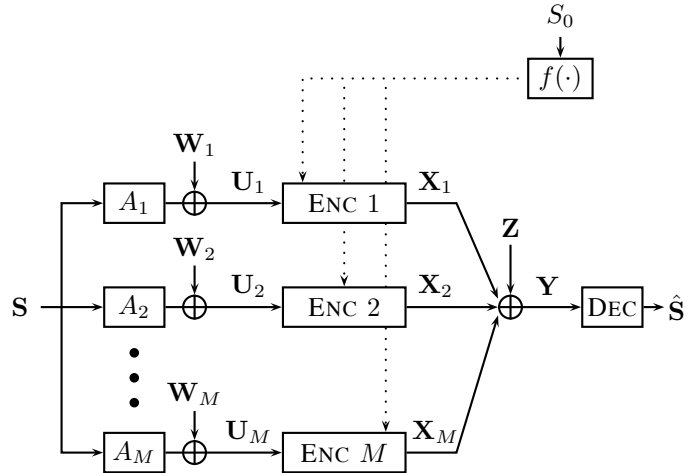


Fig. 2. Network with feedback signals.

sensor can in theory discriminate between them based on the empirical (marginal) statistics. If the $\{A_m\}$ are continuous then we can design estimators based on the empirical observations or estimate $\{A_m\}$ on a quantized set of values.

The problem introduced by fading observations in this setting is from *different* fading functions inducing the *same* marginal distribution on U_m . For point-to-point systems, this is similar to the rate-distortion problem considered in [33]. In order to collaborate, the sensors must disambiguate between the fading processes which could induce their local distribution. We call this problem one of *alignment*.

Sensor m converts its entire observation sequence S into a codeword $\mathbf{X}_m = (X_m[1], X_m[2], \dots, X_m[n])$. For simplicity we will assume that the length of the codeword is the same as the length of the source vector. For each m , the codeword must satisfy a power constraint P :

$$E \left[\|\mathbf{X}_m\|_2^2 \right] \leq P, \quad (5)$$

where $\|\cdot\|_2$ is the Euclidean norm. The codewords are transmitted across a Gaussian multiple-access channel:

$$\mathbf{Y} = \sum_{m=1}^M \mathbf{X}_m + \mathbf{Z}, \quad (6)$$

where \mathbf{Z} is iid and Gaussian with mean 0 and variance σ_Z^2 . The receiver uses \mathbf{Y} to form an estimate $\hat{\mathbf{S}}$ that minimizes the squared error:

$$D(M, P) = E_S \left[\|\mathbf{S} - \hat{\mathbf{S}}\|^2 \right]. \quad (7)$$

The goal of our coding scheme is to minimize the distortion $D(M, P)$.

A. Feedback model

In the protocol proposed in this paper we make use of a feedback signal to align the sensors, as shown in 2. We propose a simple two-phase protocol. The first phase is an *alignment phase* in which the sensors communicate among each other to learn the structure of their source fading. This is analogous to

the use of training sequences for channel estimation. In our simplified model we do not consider localized protocols for the sensors to align themselves. Distributed consensus procedures (see the forthcoming [34]) may also be useful in this context. In the *transmission phase* they communicate with the central receiver over the shared multiaccess channel. We define the time axis so that the discovery phase is at times $n \leq 0$ and the transmission phase is at times $n > 0$. In the discovery phase sensor m forms an estimate \hat{A}_m of its own observation gain A_m based on its own observations and some feedback.

Consider an alignment phase of a single step, so that during the alignment phase sensor m observes $A_m S[0] + W_m[0]$. For notational clarity, let $S_0 = S[0]$. During the alignment phase, a feedback signal $f(S_0)$ is given to all sensors. The function $f(\cdot)$ may be stochastic – for example, it may be a noisy observation of the source. Note that an estimator with access to $\{\mathbf{U}_m\}$ cannot distinguish between $\{\mathbf{U}_m\}$ and $\{-\mathbf{U}_m\}$ unless it has some additional information about \mathbf{S} . Hence, we will assume that the signal $f(S_0)$ comes from an observer not subject to source fading. We will consider several different examples for $f(S_0)$:

- *Perfect feedback* refers to the case where $f(S_0) = S_0$. This is clearly an unrealistic assumption but provides a bound on how well our protocol can perform even in the best case.
- *Single-bit noisy feedback* refers to the case where a single bit $\text{sgn}(A_b S_0 + W_b)$ from an extra *beacon* sensor is broadcast to sensors $1, 2, \dots, M$. In a real system this beacon sensor may be a chosen “leader” or may be an feedback signal from the fusion center itself.
- *Multi-bit feedback* refers to quantized perfect feedback from a longer alignment phase. The K bits ($\text{sgn}(S[-K+1]), \dots, \text{sgn}(S[-1]), \text{sgn}(S[0])$) are given to all sensors to help them align.

The techniques we use to analyze these feedback signals can be extended to more complex scenarios, but these simple examples highlight the benefits of a short alignment signal for the asymptotic distortion scaling.

B. A general protocol for alignment

We will analyze a simple protocol for alignment based on the feedback signal. In the discovery phase sensor m forms an estimate \hat{A}_m of its own observation gain A_m based on its own observations and the feedback. In the transmission phase they use an uncoded transmission protocol to communicate to the destination. Let

$$\eta_m = \sqrt{\frac{P}{\hat{A}_m^2 \sigma_S^2 + \sigma_W^2}}. \quad (8)$$

Sensor m forwards $\eta \hat{A}_m U_m$ to compensate for its observation gain. If a sufficient fraction of the sensors are aligned, the distortion will converge to 0 rapidly as the number of sensors increases.

The protocol is as follows:

- 1) Each sensor receives a common feedback signal $f(S_0)$.
- 2) Sensors estimate their observation gain $\hat{A}_m = g(f(S_0), U_m)$ using a MAP estimation rule g .

- 3) For the remaining time, the sensors compensate using \hat{A}_m and forward their observations uncoded:

$$\mathbf{X}_m = \eta \hat{A}_m \mathbf{U}_m. \quad (9)$$

- 4) The receiver performs an MMSE estimate on the received signal vector.

III. FULLY DIGITAL AND FULLY ANALOG ARCHITECTURES ARE INEFFICIENT

Information theory provides a mathematical framework for the analysis of communication systems. One of the key result of Shannon’s landmark paper [35] is the *separation principle*, which says that optimal end-to-end performance in estimation over a single link can be obtained by concatenating an optimal quantizer and an optimal error-correcting code. This result nicely decouples the overall system design problem into the separate problems of quantizer design and error-correction. This architectural primitive is the backbone of most communication systems in use today. Indeed, the separation principle can be extended to many other scenarios; an overview can be found in [36].

Sensor networks present several new challenges for communications engineers. The most interesting such networks operate via wireless links that are severely resource-constrained (e.g. in terms of power or bandwidth) and unreliable. A key challenge, therefore, is to design simple and lightweight protocols that can enable cooperation between the sensors through local communication and data processing. Shannon’s separation approach leads to some architectural simplifications: the overall sensor network problem is split into two problems, that of distributed data compression, and that of networked error-correction coding. Both problems have received recent research attention. One distributed compression problem that has a clear connection to sensor networks, the so-called “CEO problem,” has been studied in [37]–[40]. However, there are several significant drawbacks to employing such an approach, including, (i), the “separation theorem” does not hold for typical sensor network topologies, thus using the separation approach can and does result in severe performance penalties (we will briefly comment on this in the sequel); and (ii), separation-based architectures usually require high coding complexity and delay, making them of limited value for sensor networks.

Although many natural communication systems are inherently digital (such as written language) many early engineered systems did not employ sophisticated encoding. These point-to-point systems were, not surprisingly, analog and to a good extent linear [41]. The digital communication paradigm has all but made this obsolete, but for wireless sensor networks, the situation is fundamentally different, and indeed, no similar separation theorem holds. Therefore, it is worth the effort to explore other architectural paradigms, and the simplest appears to be that of linear analog architectures. The immediate advantage of such a simple architecture is scalability. What is perhaps more surprising is that such architectures manage to perform well for a number of topologies that are reminiscent of wireless sensor networks. In [42] it was shown that a simple *linear analog* scheme can achieve the optimal distortion-power tradeoff for the example given in the previous section.

A. Shannon's separation principle

In a separation-based architecture, the sensors first perform an optimal distributed source code to convert their observations $\{\mathbf{U}_m\}$ into strings of bits such that a decoder with access to all the bits can reconstruct a source estimate $\hat{\mathbf{S}}$. We can lower bound the optimal distortion by assuming that the realizations of $\{A_m\}$ are given to the sensors. For a target distortion D , let $R_S(D)$ be the rate-distortion function for the source \mathbf{S} with distortion limit D . Note that $R(D)$ is vector-valued and represents a rate tuple (R_1, R_2, \dots, R_M) . The bits generated by this source code are then treated as *independent* messages and are communicated over the multiple-access channel with an optimal channel code. Let $C(P)$ be the capacity function for the multiple-access channel under the cost constraint P . Note that $C(P)$ is also a tuple of achievable rates $(R_1^{(c)}, \dots, R_M^{(c)})$ for reliable communication across the multiple-access channel.

Suppose we have a distributed source code for \mathbf{S} with rates $r = (r_1, \dots, r_M)$. If $R(D) = r$ and $r < C(P)$ then we can compress the source and transmit the compressed messages reliably across the channel. Thus if $R(D) < C(P)$ component-wise, we can achieve distortion D across the channel. If $r > C(P)$ in any component, then the rate tuple generated by the source code cannot be communicated reliably across the channel.

Let $R_{tot} = \sum R_m$. The source coding problem is known as the CEO problem in the information theory literature [37]. Using a CEO source code we find that the distortion as a function of the sum rate is given by [38, Equation (6)]:

$$D(R_{tot}, M) = \frac{\sigma_S^2}{\frac{\sigma_S^2}{\sigma_W^2} M (1 - \exp(-2R_{tot}/M)) + 1}. \quad (10)$$

The sum capacity of the Gaussian multiple-access channel is upper-bounded by the case when all of the sensors can collaborate. The total power is MP , so:

$$R_{tot} \leq \frac{1}{2} \log \left(1 + \frac{M^2 P}{\sigma_Z^2} \right), \quad (11)$$

Substituting, the achievable distortion is lower bounded by

$$D(M) \geq \frac{\sigma_S^2 \sigma_W^2}{\sigma_W^2 + \sigma_S^2 M \left(1 - \left(\frac{\sigma_S^2}{\sigma_S^2 + M^2 P} \right)^{1/M} \right)}. \quad (12)$$

Taking the limit as $M \rightarrow \infty$ shows that $D(M) = \Omega(1/\log M)$ behavior described in the introduction.

B. Uncoded transmission

In the uncoded transmission scheme, each sensor scales its own observation to meet the power constraint of the channel. Define the constant η as

$$\eta = \sqrt{\frac{P}{\sigma_S^2 + \sigma_W^2}}. \quad (13)$$

Then

$$\mathbf{X}_m = \eta \mathbf{U}_m [i] = \eta (A_m \mathbf{S} + \mathbf{W}_m). \quad (14)$$

In the case without sign fading ($A_m = 1$ almost surely) it was shown in [6] that uncoded transmission attains the optimal performance exactly. Specifically, the performance of uncoded transmission is found to be

$$D \geq \frac{\sigma_S^2 \sigma_W^2}{M \sigma_S^2 + \sigma_W^2} \left(1 + \frac{M(\sigma_S^2 \sigma_Z^2 / \sigma_W^2)}{\frac{M \sigma_S^2 + \sigma_W^2}{\sigma_S^2 + \sigma_W^2} M P + \sigma_Z^2} \right). \quad (15)$$

The optimality of this performance was shown by explicitly bounding correlations via an argument due to Witsenhausen [43]. For general parameters A_m it can be shown that uncoded transmission attains the best possible distortion scaling [42]. Our first result shows that bounded scalar fading results in misalignment and renders uncoded transmission useless.

Proposition 1. *For the Gaussian network with fading observations satisfying (2)–(4), under the uncoded transmission scheme the distortion scales with M as $\Omega(1)$.*

Proof: We can only improve the performance of uncoded transmission by assuming that each sensor m knows $|A_m|$. For sensor m , the density of U_m is

$$p_{U_m|A_m}(u_m|A_m) = \frac{1}{\sqrt{2\pi(A_m^2 \sigma_S^2 + \sigma_W^2)}} \exp\left(-\frac{u_m^2}{2(A_m^2 \sigma_S^2 + \sigma_W^2)}\right).$$

This density is identical for $A_m = \pm a_m$. The sensors apply the gains

$$\eta_m = \sqrt{\frac{P}{A_m^2 \sigma_S^2 + \sigma_W^2}}$$

Note that

$$\sum_{m=1}^M \frac{P}{\sigma_S^2 A_m^2 + \sigma_W^2} + \sigma_Z^2 \geq M \frac{P}{\sigma_S^2 \nu^2 + \sigma_W^2} + \sigma_Z^2 \geq M \mu^2$$

for some $\mu > 0$.

Thus we can lower bound the distortion by

$$D(M) \geq \frac{\sigma_S^2}{\frac{B}{\sqrt{M}} \mu + 1}. \quad (16)$$

Now note that $B = \sum_{m=1}^M \eta_m A_m$, and

$$E[\eta_m A_m] = E\left[\sqrt{\frac{P A_m^2}{A_m^2 \sigma_S^2 + \sigma_W^2}} \operatorname{sgn} A_m\right] = 0$$

$$E[\eta_m^2 A_m^2] = E\left[\frac{P A_m^2}{A_m^2 \sigma_S^2 + \sigma_W^2}\right] = \sigma_B^2 < \infty.$$

So by the central limit theorem [44], $B M^{-1/2}$ converges in distribution to a Gaussian random variable with mean 0 and variance σ_B^2 .

We can now write the expected distortion in the limit:

$$\begin{aligned} \lim_{M \rightarrow \infty} E[D(M)] &\geq \lim_{M \rightarrow \infty} E\left[\frac{\sigma_S^2}{(B M^{-1/2} \mu)^2 + 1}\right] \\ &= E\left[\lim_{M \rightarrow \infty} \frac{\sigma_S^2}{(B M^{-1/2} \mu)^2 + 1}\right] \\ &= \int \frac{\sigma_S^2}{\xi^2 + 1} \frac{1}{\sqrt{2\pi\sigma_B^2\mu^2}} e^{-(1/2\sigma_B^2\mu^2)\xi^2} \\ &> 0, \end{aligned}$$

Thus the expected distortion does not converge to 0 as $M \rightarrow \infty$, so $D(M) = \Omega(1)$. ■

C. Remarks

The preceding arguments showed that under scalar source fading, the uncoded transmission scheme fails because the expected gain of the scheme does not grow with the number of sensors. Therefore a fully analog scheme that does not use the feedback cannot succeed because the signal \mathbf{Y} at the receiver is not strong enough to overcome the observation and channel noises. By contrast, a fully digital architecture may still achieve a distortion scaling like $O(1/\log M)$ by modifying the CEO code appropriately. We shall see in the next section that appropriate alignment of the sensor observations can partially recover the performance of uncoded transmission and gives better distortion scaling than the logarithmic scaling of the fully digital system.

IV. ANALOG ARCHITECTURES WITH FEEDBACK

We now turn to the analysis of our protocol for handling source fading with feedback. The overall system is shown in Figure 2. For the following discussion we will focus on sign fading; extensions to scalar fading are straightforward and given later. After receiving the common feedback $f(S_0)$, each sensor m forms an estimate $\hat{A}_m = g_m(f(S_0), U_m)$ of its observation gain. Conditioned on a value $S_0 = s_0$, the sensor observations are independent and identically distributed, so the events of successful estimation of the observation gains are independent from sensor to sensor. By the exchangeability of the gains A_m , each sensor should attempt to maximize their probability of success, and will adopt the same decision rule $g_m(\cdot) = g(\cdot)$. Let $v(s_0)$ denote the probability that a sensor correctly estimates its own observation gain. Let $\hat{A}_m = g(f(S_0), U_m)$.

The MMSE estimate of S given Y will depend on the random variable $B = \sum_{m=1}^M A_m \hat{A}_m$. For a fixed B define

$$L(B, M) = \frac{\sigma_S^2 \sigma_W^2}{\frac{B^2}{M + (\sigma_Z^2 / \sigma_W^2) \eta^{-2}} \sigma_S^2 + \sigma_W^2}, \quad (17)$$

where η is given by (8). The expected distortion is then

$$D_{unc}(MP) = E_B [L(B, M)] . \quad (18)$$

Let Γ denote the number of sensors for which $\hat{A}_m = A_m$. Conditioned on the value of S_0 , the event of each sensor being aligned correctly depends only on the noise value at that sensor, and hence is an independent Bernoulli random variable with parameter $v(s_0)$. Therefore Γ is a binomial random variable with parameters $(M, v(s_0))$.

The distortion achievable after the feedback is given by

$$E[D | f(S_0)] = E_{S_0} \left[\sum_{k=0}^M L(M - 2k, M) P(\Gamma = k) \right], \quad (19)$$

where the expectation is taken over S_0 and

$$P(\Gamma = k) = \binom{M}{k} v(s_0)^k (1 - v(s_0))^{M-k}. \quad (20)$$

For equation (19) to converge to 0, each term in the summation must converge to 0 as $M \rightarrow \infty$. The convergence is in turn dependent on S_0 and the decision rule $g(\cdot)$ by which the sensors estimate their alignment. Intuitively, if S_0 is close to zero, it will be difficult to determine A_m and thus the probability $v(S_0)$ that sensor m aligns correctly will be close to $1/2$. We therefore wish to find a decision rule $g(\cdot)$ that minimizes the probability of error for each sensor, or, alternately, that maximizes the probability of correct alignment.

In order for $D(M) = O(M^{-1})$, the fraction Γ/M of sensors that are aligned must be strictly greater than half. However, since the Γ depends on the realization of the feedback, we cannot have $\Gamma/M > 1/2 + \epsilon$ almost surely. Instead, we allow ϵ to scale with M and balance the achievable distortion with the probability of getting $\Gamma/M > 1/2 + \epsilon$. This balancing approach is given by the following proposition.

Proposition 2. *Suppose the feedback function $f(\cdot)$ and decision rule $g(\cdot)$ are chosen such that there exist constants $\epsilon_1 > 0$, $\epsilon_2 > 0$, and functions $\alpha(M)$ and $\beta(M)$ such that*

$$\lim_{M \rightarrow \infty} \alpha(M) = 0 \quad (21)$$

$$\lim_{M \rightarrow \infty} \beta(M) = 0 \quad (22)$$

$$\lim_{M \rightarrow \infty} M\beta(M)^2 = \infty, \quad (23)$$

and if $|S_0| \geq \epsilon_1 \alpha(M)$ then $v(S_0) \geq \frac{1}{2} + \epsilon_2 \beta(M)$. Then as $M \rightarrow \infty$, the feedback/decision pair (f, g) achieves a distortion $D(M)$ in the transmission phase satisfying

$$D(M) = O \left(\max \left\{ \alpha(M), \frac{1}{M\beta(M)^2} \right\} \right) \quad (24)$$

for the network with sign fading.

Remark. Equation (23) says that if the source sample S_0 on which we base our feedback is large enough, i.e. at least $\epsilon_1 \alpha(M)$, then the probability of successful alignment is larger than $1/2 + \epsilon_2 \beta(M)$.

Proof: Suppose that (21) – (23) hold. Let H be the event $(|S_0| \geq \epsilon_1 \alpha(M))$. Let Γ be a binomial random variable with parameters $(M, v(s_0))$ so that $E[\Gamma] = Mv(s_0)$. We have the following Chernoff bound:

$$P \left(|\Gamma M v(s_0)| > \frac{M\beta(M)\epsilon_2}{\sqrt{2}} \right) \leq 2 \exp(-\epsilon_2^2 M\beta(M)^2). \quad (25)$$

We can also write the following bounds, using the assumptions in (23):

$$P(H^c) = \frac{1}{\sqrt{2\pi\sigma_S^2}} \int_{-\epsilon_1 \alpha(M)}^{\epsilon_1 \alpha(M)} e^{-x^2/2\sigma_S^2} dx \quad (26)$$

$$\leq \sqrt{\frac{2}{\pi\sigma_S^2}} \epsilon_1 \alpha(M) \quad (27)$$

$$v(S_0 | H) \geq \frac{1}{2} + \epsilon_2 \beta(M). \quad (28)$$

We evaluate the distortion separately on the events

$$F_1 = H^c$$

$$F_2 = H \cap (\Gamma > (\epsilon_2/2)M\beta(M))$$

$$F_3 = H \cap (\Gamma \leq (\epsilon_2/2)M\beta(M)).$$

On H^c we upper bound the distortion by letting $\Gamma = M/2$. On H , we have equation (28), so we can let $\Gamma = M/2$ on F_2 and $\Gamma = \frac{1}{2} + (\epsilon_2/2)M\beta(M)$ on F_3 . Putting this together and noting that the distortion must be less than σ_S^2 , we rewrite (19) as :

$$\begin{aligned} E_{S_0} [L(M - 2\Gamma, M)(1_{F_1} + 1_{F_2} + 1_{F_3})] \\ \leq \sigma_S^2 (P(H^c) + P(\Gamma > (\epsilon_2/2) | H)) \\ + L(M\epsilon_2\beta(M)/2, M)P(\Gamma \leq (\epsilon_2/2) | H) \\ \leq \sigma_S^2 \left(\sqrt{\frac{2}{\pi\sigma_S^2}}\epsilon_1\alpha(M) + 2\exp(-\epsilon^2 M\beta(M)^2) \right) \\ + L(M\epsilon_2\beta(M)/2, M). \end{aligned}$$

The first term converges to zero with the slower of $\alpha(M)$ and $\exp(-M\beta(M)^2)$. The second term converges to zero as $O(M^{-1}\beta(M)^{-2})$. Since $\beta(M)^{-2}$ is sub-linear, we can ignore this latter term, so the distortion is $O(\max\{\alpha(M), M^{-1}\beta(M)^{-2}\})$. ■

The bounds for this proof depend only on the relationship between the values of S_0 and the success probability $v(S_0)$. The latter depends only on the noise distribution, and thus it is possible to treat non-Gaussian noises, although in this case the destination's linear estimator will not necessarily be the MMSE estimator.

1) *Example: perfect feedback:* To gain further insight, let us consider an idealized case in which $f(\cdot)$ is the identity function, so that the sensors get to know the source sample exactly; we call this *perfect feedback*. We emphasize that this feedback is only available during the discovery phase and not for all time, and that we are assuming that the discovery phase is one sample long.

Conditioned on the value of $S_0 = s_0$, each sensor is left with the problem of detecting antipodal signals $\pm s_0$ in the presence of Gaussian noise with a uniform prior. The maximum a priori probability (MAP) rule for this problem is a threshold test at 0; for $s_0 > 0$, if $U_m > 0$ then $\hat{A}_m = 1$, and for $s_0 < 0$ if $U_m > 0$ then $\hat{A}_m = -1$. The probability of success for this rule is

$$v_p(s_0) = 1 - Q\left(\frac{|s_0|}{\sigma_W}\right), \quad (29)$$

where

$$Q(x) = \frac{1}{\sqrt{2\pi}} \int_x^\infty e^{-y^2/2} dy. \quad (30)$$

The success probability, conditioned on $H = |S_0| \geq \epsilon_M$, is

$$v_p(S_0 | H) \geq \frac{1}{2} + \frac{\epsilon_M}{2\sqrt{2\pi\sigma_W^2}}. \quad (31)$$

Proposition 2 tells us that the distortion scales faster than $\max\{\epsilon_M, M^{-1}\epsilon_M^{-2}\}$. Equating these two scaling rates, we can set $\epsilon_M = O(M^{-1/3})$ to yield a scaling rate of $O(M^{-1/3})$.

Suppose instead that the sensors do not receive S_0 , but instead a one bit quantization of S_0 , or $f(S_0) = \text{sgn}(S_0)$. Since the the MAP rule with full knowledge of S_0 was a threshold test at 0 for all values of S_0 , the MAP rule for this case is the same. Another way to phrase the decision rule is if $f(S_0) = \text{sgn}(U_m)$ then $\hat{A}_m = 1$, otherwise $\hat{A}_m = -1$. We again condition on the value of S_0 , which gives the the

same success probabilities as (29) and (31). Therefore the distortion with this one bit of feedback is also $O(M^{-1/3})$. It is interesting to note that in this case the achievable distortion scaling does not depend on the ‘‘richness’’ of the actual available feedback.

2) *Example: one bit of noisy feedback:* Suppose instead that each sensor is given access to the sign of the signal received at an extra sensor b :

$$f(S_0) = G_b = \text{sgn}(A_b S_0 + W_b). \quad (32)$$

Sensor m must then decide if $A_m = A_b$ or $A_m \neq A_b$. Again, we condition on $S_0 = s_0$. The probability that $G_b = A_b$ is given by

$$P(G_b = A_b | S_0 = s_0) = 1 - Q\left(\frac{|s_0|}{\sigma_W}\right). \quad (33)$$

Since we cannot exactly know the signs A_b and A_m , we assume without loss of generality that $A_b = 1$ and attempt to distinguish between the two hypotheses $A_m = A_b$ and $A_m = -A_b$.

Suppose we have full knowledge of $U_b = A_b S_0 + W_b$. Under the hypothesis $A_m = A_b$, the pair $(A_b s_0 + W_0, A_m s_0 + W_m)$ is jointly Gaussian with mean (s_0, s_0) . Under the hypothesis $A_m = -A_b$ they are jointly Gaussian with mean $(s_0, -s_0)$. The decision rule in this case is again a threshold test on the line $U_m = 0$. If U_m and U_b have the same sign, then sensor m guesses $\hat{A}_m = \hat{A}_b = 1$, and if they have different sign it guesses $\hat{A}_m = -\hat{A}_b = -1$. This rule is again indifferent to the value of S_0 , as in the perfect feedback case.

The decision rule for A_m that maximizes a posteriori probability of the observations is therefore given by

$$\hat{A}_m = \begin{cases} g(G_b, U_m) = 1 & \text{if } G_b = \text{sgn}(U_m) \\ g(G_b, U_m) = -1 & \text{if } G_b \neq \text{sgn}(U_m), \end{cases} \quad (34)$$

regardless of the value of S_0 . The probability of success is given by

$$v_n(s_0) = 1 - 2Q\left(\frac{|s_0|}{\sigma_W}\right)Q\left(\frac{-|s_0|}{\sigma_W}\right). \quad (35)$$

So the conditional probability of success is:

$$v_n\left(S_0 \mid |S_0| > \frac{\epsilon_M}{2}\right) \geq \frac{1}{2} + \frac{\epsilon_M}{\sqrt{2\pi\sigma_W^2}} + \frac{\epsilon_M^2}{2\pi\sigma_W^2}. \quad (36)$$

The ϵ_M term is dominant as $M \rightarrow \infty$ in this conditional probability, the same as in the perfect feedback case. From our previous analysis, we can see that $D(M) = O(M^{-1/3})$.

The noisy feedback example models a situation in which one sensor acts as a ‘‘beacon’’ by broadcasting the sign of its observation to the other sensors. In a scaling sense, the sign of the noisy observation is ‘‘as good’’ as knowing the sign of the source sample exactly, although the constants in the convergence become worse as the noise becomes more severe.

A. Many bits of feedback

We now consider the effect of lengthening the discovery phase by allowing the feedback involve more than one sample of the source. Let $S_0 = S[0]$ and $S_{-1} = S[-1]$ be the two source samples on which we base our feedback $f(S_0, S_{-1})$.

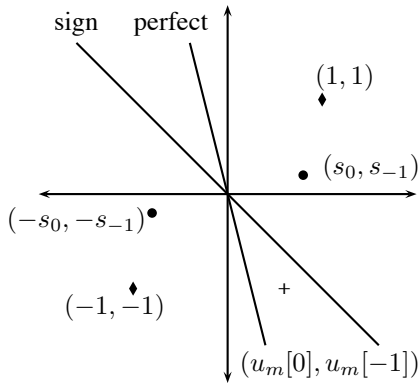


Fig. 3. MAP rules for perfect feedback and perfect sign feedback. The plus is the noisy observation $(u_m[0], u_m[-1])$. Under perfect information, $\hat{A}_m = 1$, whereas with only sign information the expected probability of success is maximized when $\hat{A}_m = -1$.

Conditioned on a realization (s_0, s_{-1}) of these two samples, the sensor observations are again independent, so each sensor should seek to maximize its own probability of success. To illustrate the effect of adding more feedback, we consider the perfect feedback of Section IV-1 to simplify the expressions. A similar analysis can be carried out for the noisy feedback case.

Suppose our feedback is the pair (S_0, S_{-1}) . Conditioned on the values of S_0 and S_{-1} , sensor m 's observations $(U_m[0], U_m[-1])$ are jointly Gaussian with mean (s_0, s_{-1}) under the hypothesis $A_m = 1$ and mean $(-s_0, -s_{-1})$ under the hypothesis $A_m = -1$. The problem is again the same as that of detecting antipodal signals in the presence of Gaussian noise, and the MAP rule is a threshold test shown in Figure 3. The probability of success for this rule is

$$v_p(s_0, s_{-1}) = 1 - Q\left(\frac{\sqrt{s_0^2 + s_{-1}^2}}{\sigma_W}\right). \quad (37)$$

Let H be the event $(|S_0| \leq \epsilon\alpha(M), |S_{-1}| \leq \epsilon\alpha(M))$, and H^c its complement. Then we have:

$$v_p(S_0, S_{-1} | H^c) \geq 1 - Q\left(\frac{\sqrt{2}\epsilon\alpha(M)}{\sigma_w}\right) \quad (38)$$

$$\geq \frac{1}{2} + \epsilon\alpha(M) \frac{1}{2\sqrt{\pi\sigma_W^2}}. \quad (39)$$

Since S_0 and S_{-1} are independent, $P(H)$ is proportional to $\alpha(M)^2$. The analysis in Proposition 2 implies that $D(M) = O(\max\{\alpha(M)^2, M\alpha(M)^{-2}\})$. Setting these two terms equal gives $\alpha(M) = M^{-4}$ so $D(M) = O(M^{-1/2})$.

Extending the above analysis in a standard way shows that with K samples of feedback we get distortion $D(M) = O(M^{-K/(K+2)})$. By choosing K arbitrarily large, we get closer to the optimal rate of $O(M^{-1})$.

Proposition 3. Consider the observation network with sign fading. Then the K -bit feedback scheme outlined above achieves a distortion that scales like $O(M^{-K/(K+2)})$.

Suppose we only get the signs of S_0 and S_{-1} , so that $f(S_0, S_{-1}) = (\text{sgn}(S_0), \text{sgn}(S_{-1}))$. The threshold test in the

MAP rule for perfect feedback depends on the actual values of (s_0, s_{-1}) , as opposed to the threshold test when $K = 1$. Thus the scaling result above does not immediately follow. Sensor m must then determine, based on the observation pair $(U_m[0], U_m[-1])$, whether $A_m = 1$ or $A_m = -1$ in a way that maximizes its probability of making a correct decision. We would like for the probability of success to be greater than $1/2$.

Under the two hypotheses, the pair $(U_m[0], U_m[-1])$ is jointly Gaussian with mean (s_0, s_{-1}) for $A_m = 1$, mean $(-s_0, -s_{-1})$ for $A_m = -1$ and covariance $\sigma_W^2 I$. The likelihood of observing $(U_m[0], U_m[-1])$ is the expectation of the conditional likelihood over all source pairs (s_0, s_{-1}) that could have generated $f(s_0, s_{-1})$. Because of the symmetry in the distribution of (S_0, S_{-1}) , the likelihoods under the two hypotheses are equal on the line orthogonal to the vector $f(s_0, s_{-1})$ so the MAP estimate is a threshold on that line.

To illustrate the difference between the MAP estimate for the case of perfect feedback versus the case of sign feedback, consider Figure 3. For a fixed (s_0, s_{-1}) , the probability of error is given by the probability that the noise exceeds the distance from the point (s_0, s_{-1}) to the threshold in the direction orthogonal to the decision boundary:

$$v_p(s_0, s_{-1}) = 1 - Q\left(\frac{|s_0| + |s_{-1}|}{\sqrt{2}\sigma_W}\right). \quad (40)$$

This differs from equation (37) by a shift in the norm inside the $Q(\cdot)$ function; with perfect knowledge of the samples we have an \mathcal{L}_2 norm, and with only the sign we have an \mathcal{L}_1 norm. This explains why the tests and errors were the same in the case when $K = 1$.

Let H be the event that $|S_0|, |S_{-1}| \leq \epsilon\alpha(M)$ and H^c its complement. Then we can bound the probability of success:

$$v_p(s_0, s_{-1} | H^c) \geq \frac{1}{2} + \frac{\epsilon}{\sqrt{\pi\sigma_W^2}}\alpha(M). \quad (41)$$

Thus from our previous analysis, $D(M) = O(M^{-1/2})$ as in the case when we have perfect knowledge of the source samples.

To see clearly the benefits of extending the scheme to multiple bits of feedback, consider the simulation shown in Figure 4. Here $\sigma_S^2 = 5$, $\sigma_W^2 = 1$, $\sigma_Z^2 = 10$, and $P = 10$. The figure shows the asymptotic distortion for the separation and optimal uncoded transmission schemes with no sign fading. The three remaining curves are simulations of the protocol with alignment phases of lengths 1, 2, and 3 under perfect feedback. With only a single sample of feedback, the asymptotic decay in the distortion does not appear unless the network is very large, but with three samples the distortion scaling is nearly optimal.

B. Feedback for bounded scalar fading

As an extension to these results on sign fading, we can modify the scheme for bounded real scalar fading. If sensor m estimates $|A_m|$ locally and finds that $|\hat{A}_m| < \epsilon'$ for some threshold ϵ' , it will not transmit. Since the gains are iid this affects at a constant fraction $P(|A_m| < \epsilon')$ of the sensors almost surely as $M \rightarrow \infty$. Consider the case of 1-bit feedback.

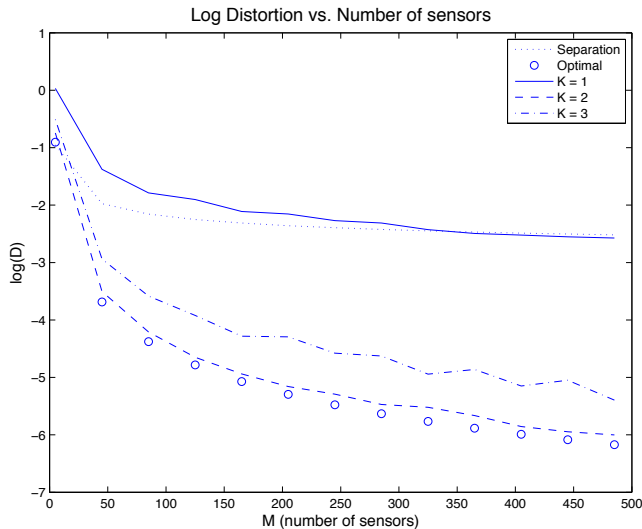


Fig. 4. Plot of $\log(D)$ as a function of M for separation-based coding (dotted), uncoded transmission with no sign fading (circles), and alignment phases of lengths 1, 2, and 3 under perfect feedback.

Suppose a beacon sensor broadcasts the sign of its observation G_b at time 0 to all the other sensors:

$$G_b = \text{sgn}(U_b[0]) = \text{sgn}(A_b S_0 + W_b[0]) \quad (42)$$

We call this signal the feedback function $f(U_b)$

Sensor m then checks if $G_b = G_m$ (defined analogously). If so, it estimates $\text{sgn}(A_m) = \text{sgn}(A_b)$, otherwise it estimates $\text{sgn}(A_m) = -\text{sgn}(A_b)$. Call this decision rule $g(G_b, U_m)$. This rule is the maximum a posteriori probability (MAP) rule for detecting the sign of A_m . Denote by $v_m(S_0)$ the probability of successfully estimating $\text{sgn}(A_m)$ at sensor m using this rule. Conditioned on S_0 , all the sensor observations are independent. We have already assumed that $A_m > \epsilon$, so $v_m(S_0) > 1/2 + \delta$ for all m . We have the same proposition relating the scaling rate to the success probability:

Proposition 4. *Suppose the feedback function $f(\cdot)$ and decision rule $g(\cdot)$ are chosen such that there exist constants $\epsilon_1 > 0$, $\epsilon_2 > 0$, and functions $\alpha(M)$ and $\beta(M)$ such that*

$$\lim_{M \rightarrow \infty} \alpha(M) = 0 \quad (43)$$

$$\lim_{M \rightarrow \infty} \beta(M) = 0 \quad (44)$$

$$\lim_{M \rightarrow \infty} M\beta(M)^2 = \infty, \quad (45)$$

and if $|S_0| \geq \epsilon_1 \alpha(M)$ then $v(S_0) \geq 1/2 + \epsilon_2 \beta(M)$. Then as $M \rightarrow \infty$, the feedback/decision pair (f, g) achieves a distortion $D(M)$ in the transmission phase satisfying

$$D(M) = O\left(\max\left\{\alpha(M), \frac{1}{M\beta(M)^2}\right\}\right). \quad (46)$$

for the Gaussian network with bounded real scalar fading.

We then have, by application of the previous proposition, that a single bit of feedback in the mode we have described biases the sum of the forwarded observations into a regime that scales faster than \sqrt{M} , which then avoids the central-limit behavior described in Proposition 1. It is clear from these examples that the situations in which the fading process $\{A_m\}$

is problematic are when A_m is zero-mean and symmetric. In these cases the sign of A_m plays a key role. This can be thought of as a phase uncertainty introduced by the fading observations.

V. DISCUSSION

We have seen how uncertainty in the correlation between sensors, or *misalignment*, could severely impact the performance of a simple, information-theoretically optimal sensor fusion strategy. We proposed a simple two-phase protocol in which a small number of bits is broadcast to all the sensors. This protocol partially resolves the uncertainty and regains much of the performance. This same framework can be applied to more complex problems that may appear in real sensor networks. In particular, it can be modified to account for other types of alignment uncertainty for sensor data fusion. By using limited feedback, the sensors can apply simple tests to align their observations. This in turn allows them to exploit the simple and near-optimal collaboration via linear analog communication. In this sense, a limited amount of feedback can greatly simplify sensor data fusion in many applications.

The example given here can be extended to more sophisticated mathematical models for sources and more complex sensing scenarios. One simple extension to the protocol would be to handle complex sources. In this case the underlying source is a complex circularly symmetric Gaussian random variable and the source fading could be a uniform phase shift. The sensors can use the feedback to partially align their phase vectors for the uncoded transmission phase. By using the same techniques as in the real scaling case, we can find the scaling law for the distortion. It may also be possible to adapt the recent work on distributed beamforming [45]–[49].

Another mathematical extension to the model is to include memory in the source observations by modeling the source fading as FIR filters. For ensembles of real filters, the ambiguity is at most a sign shift, as in the scalar fading case. To help align the sensors with a given filter, we can use the same beacon strategy described earlier. This will allow a sufficient fraction of the sensors to align themselves with high probability, allowing for further processing to enable coherent communication. If the sensor correlations change dynamically (but on a slower time scale), we can do periodic re-alignment to maintain system performance, similar to the use of pilot signals for slow fading wireless channels.

More challenging source misalignment problems arise when there are multiple sources or when the source fading takes the form of a delay. For multiple sources we may have a problem of *bandwidth mismatch* which has to be addressed in the uncoded transmission framework. There is also the interaction of the sensor network topology with the problem of alignment. In our case study, all sensors could communicate directly with a centralized fusion center and the distortion scaling was directly measured in terms of the number of sensors. Many networks may not have these direct connections, so the linear analog architecture described here must be modified for new scenarios. Misalignment in an general network can be modeled via delayed observations. As long as a fraction of sensors are misaligned, their contributions will result in a coherent

interference that scales as fast as the correctly aligned sensors. One way around this problem may be to allow the amount of feedback to scale with the number of sensors, but we leave this for future work.

The sensor network system designer often faces challenges in modeling the underlying phenomena to be sensed. The standard approach of separation-based estimation and communication leads to gross inefficiencies that may limit the lifetime of the sensor network. The statistical structure of the sensor's correlations may depend on the actual deployment, so simple schemes that exploit known structures may also have poor performance. The capability for feedback is already engineered into many communication systems for control and synchronization purposes. We can use a small amount of feedback to align the sensor observations and simultaneously achieve better scaling performance by enabling simplified collaborative protocols.

ACKNOWLEDGMENTS

The authors would like to acknowledge the useful suggestions of the reviewers which helped improve the clarity of this manuscript.

REFERENCES

- [1] A. D. Sarwate and M. Gastpar, "Estimation from misaligned observations with limited feedback," in *39th Conference on Information Sciences and Systems (CISS 2005)*, (Baltimore, MD), Mar. 2005.
- [2] A. D. Sarwate and M. Gastpar, "Fading observation alignment via feedback," in *4th International Symposium on Information Processing in Sensor Networks (IPSN 2005)*, (Los Angeles, CA), Apr. 2005.
- [3] K. Eswaran and M. Gastpar, "Foundations of distributed source coding," in *Distributed Source Coding: Theory, Algorithms, and Applications* (P. L. Dragotti and M. Gastpar, eds.), Academic Press, 2009.
- [4] P. Gupta and P. Kumar, "The capacity of wireless networks," *IEEE Trans. Inf. Theory*, vol. 46, pp. 388–404, March 2000.
- [5] M. Gastpar, "Information-theoretic bounds on sensor network performance," in *Wireless Sensor Networks: Signal Processing and Communications* (A. Swami, Q. Zhao, Y.-W. Hong, and L. Tong, eds.), Wiley, 2007.
- [6] M. Gastpar, "Uncoded transmission is exactly optimal for a simple gaussian sensor network," *IEEE Trans. Inf. Theory*, vol. 54, pp. 5427–5251, November 2008.
- [7] J. J. Xiao, A. Ribeiro, Z. Q. Luo, and G. B. Giannakis, "Distributed compression-estimation using wireless sensor networks," *IEEE Signal Processing Mag.*, pp. 27–41, July 2006.
- [8] W. Bajwa, J. Haupt, A. Sayeed, and R. Nowak, "Joint source-channel communication for distributed estimation in sensor networks," *IEEE Trans. Inf. Theory*, vol. 53, pp. 3629–3653, October 2007.
- [9] J.-J. Xiao, S. Cui, Z.-Q. Luo, and A. Goldsmith, "Linear coherent decentralized estimation," *IEEE Trans. Signal Process.*, vol. 56, pp. 757–770, February 2008.
- [10] W. R. Paton, *The Greek anthology*, vol. 4. Harvard University Press, 1918.
- [11] C. Shannon, "The zero error capacity of a noisy channel," *IRE Trans. Inf. Theory*, vol. IT-2, no. 6, pp. 8–19, 1956.
- [12] J. Schalkwijk and T. Kailath, "A coding scheme for additive noise channels with feedback - I : No bandwidth constraint," *IEEE Trans. Inf. Theory*, vol. 12, no. 2, pp. 172–182, 1966.
- [13] A. Sahai, S. Draper, and M. Gastpar, "Boosting reliability over awgn networks with average power constraints and noiseless feedback," in *Proc. 2005 IEEE International Symposium on Information Theory*, (Adelaide, Australia), 2005.
- [14] M. Horstein, "Sequential transmission using noiseless feedback," *IEEE Trans. Inf. Theory*, vol. 9, pp. 136–143, July 1963.
- [15] G. D. Forney, "Exponential error bounds for erasures, list, and decision feedback schemes," *IEEE Trans. Inf. Theory*, vol. 14, pp. 206–220, March 1968.
- [16] M. Burnashev, "Data transmission over a discrete channel with feedback, random transmission time," *Problems of Information Transmission*, vol. 12, October–December 1976.
- [17] N. Elia, "When Bode meets Shannon: control-oriented feedback communication schemes," *IEEE Trans. Autom. Control*, vol. 49, pp. 1477–1488, September 2004.
- [18] A. Sahai and S. Mitter, "The necessity and sufficiency of anytime capacity for stabilization of a linear system over a noisy communication link, part 1 : Scalar systems," *IEEE Trans. Inf. Theory*, vol. 52, pp. 3369–3395, August 2006.
- [19] S. Tatikonda, S. Yang, and A. Kavčić, "Feedback capacity of finite-state machine channels," *IEEE Trans. Inf. Theory*, vol. 51, pp. 799–810, March 2005.
- [20] Y.-H. Kim, "Feedback capacity of the first-order moving average Gaussian channel," *IEEE Trans. Inf. Theory*, vol. 52, pp. 3063–3079, July 2006.
- [21] Y.-H. Kim, "Feedback capacity of stationary Gaussian channels," *IEEE Trans. Inf. Theory*, vol. 56, pp. 57–85, January 2010.
- [22] H. Permuter, P. Cuff, B. V. Roy, and T. Weissman, "Capacity of the trapdoor channel with feedback," *IEEE Trans. Inf. Theory*, vol. 54, pp. 3150–3165, July 2008.
- [23] H. Permuter, T. Weissman, and A. Goldsmith, "Finite state channels with time-invariant deterministic feedback," *IEEE Trans. Inf. Theory*, vol. 55, pp. 644–662, February 2009.
- [24] N. Gaarder and J. Wolf, "The capacity region of a multiple-access discrete memoryless channel can increase with feedback," *IEEE Trans. Inf. Theory*, vol. 21, no. 1, pp. 100–102, 1975.
- [25] T. Cover and C. Leung, "An achievable rate region for the multiple-access channel with feedback," *IEEE Trans. Inf. Theory*, vol. 27, no. 3, pp. 292–298, 1981.
- [26] 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, 1984.
- [27] G. Kramer, I. Marić, and R. Yates, *Cooperative communications*, vol. 1 of *Foundations and Trends in Networking*. Hanover, MA, USA: Now Publishers Inc., August 2008.
- [28] O. Simeone, D. Gündüz, H. Poor, A. Goldsmith, and S. Shamai (Shitz), "Compound multiple access channels with partial cooperation," *IEEE Trans. Inf. Theory*, vol. 55, pp. 2425–2441, June 2009.
- [29] A. Tchamkerten and I. E. Telatar, "Variable length coding over an unknown channel," *IEEE Trans. Inf. Theory*, vol. 52, pp. 2126–2145, May 2006.
- [30] O. Shayevitz and M. Feder, "Achieving the empirical capacity using feedback: Memoryless additive models," *IEEE Trans. Inf. Theory*, vol. 55, pp. 1269–1295, March 2009.
- [31] K. Eswaran, A. D. Sarwate, A. Sahai, and M. Gastpar, "Zero-rate feedback can achieve the empirical capacity," *IEEE Trans. Inf. Theory*, vol. 56, pp. 25–39, January 2010.
- [32] A. D. Sarwate, "Observation uncertainty in Gaussian sensor networks," Master's thesis, University of California, Berkeley, 2005.
- [33] A. Dembo and T. Weissman, "The minimax distortion redundancy in noisy source coding," *IEEE Trans. Inf. Theory*, vol. 49, no. 11, pp. 3020–3030, 2003.
- [34] A. Dimakis, S. Kar, J. Moura, M. Rabbat, and A. Scaglione, "In-network signal processing with gossip algorithms." to appear in Proc. IEEE.
- [35] C. Shannon, "A mathematical theory of communication," *Bell System Technical Journal*, vol. 27, pp. 379–423, 623–656, 1948.
- [36] M. Gastpar, "Separation theorems and partial orderings for sensor networks," in *Networked Sensing Information and Control* (V. Saligrama, ed.), New York, NY: Springer, 2008.
- [37] H. Viswanathan and T. Berger, "The quadratic Gaussian CEO problem," *IEEE Trans. Inf. Theory*, vol. 43, no. 5, pp. 1549–1559, 1997.
- [38] Y. Oohama, "The Rate-Distortion Function for the Quadratic Gaussian CEO Problem," *IEEE Trans. Inf. Theory*, vol. 44, no. 3, pp. 1057–1070, 1998.
- [39] V. Prabhakaran, D. Tse, and K. Ramchandran, "Rate region of the quadratic Gaussian CEO problem," in *IEEE International Symposium on Information Theory Proceedings (ISIT 2004)*, p. 117, 2004.
- [40] Y. Oohama, "Rate-distortion theory for Gaussian multiterminal source coding systems with several side informations at the decoder," *IEEE Trans. Inf. Theory*, vol. 51, no. 7, pp. 2577–2593, 2005.
- [41] M. Schwartz and J. Hayes, "A history of transatlantic cables," *IEEE Commun. Mag.*, vol. 46, pp. 42–48, September 2008.
- [42] M. Gastpar and M. Vetterli, "Power, spatio-temporal bandwidth, and distortion in large sensor networks," *IEEE J. Sel. Areas Commun. (Special Issue on Self-Organizing Distributive Collaborative Sensor Networks)*, vol. 23, pp. 745–754, April 2005.
- [43] H. S. Witsenhausen, "On sequences of pairs of dependent random variables," *SIAM J. Appl. Math.*, vol. 28, pp. 100–113, January 1975.
- [44] R. Durrett, *Probability: Theory and Examples*. Belmont, CA: Duxbury, 2nd ed., 1995.

- [45] R. Mudumbai, G. Barriac, and U. Madhow, "On the feasibility of distributed beamforming in wireless networks," *IEEE Trans. Wireless Commun.*, vol. 6, pp. 1754–1763, May 2007.
- [46] R. Mudumbai, D. R. Brown III, U. Madhow, and H. V. Poor, "Distributed transmit beamforming: Challenges and recent progress," *IEEE Commun. Mag.*, vol. 47, pp. 102–110, February 2009.
- [47] M. Johnson, K. Ramchandran, and M. Mitzenmacher, "Distributed beamforming with binary signaling," in *Proceedings of the 2008 IEEE International Symposium on Information Theory*, (Toronto, Canada), 2008.
- [48] J. Bucklew and W. Sethares, "Convergence of a class of decentralized beamforming algorithms," in *Proc. IEEE Workshop of Statistical Signal Processing*, (Madison, WI), August 2007.
- [49] J. Thukral and H. Bölcskei, "Distributed spatial multiplexing with 1-bit feedback," in *Proceedings of the Allerton Conference on Communication Control, and Computing*, (Monticello, IL), September 2007.



Anand D. Sarwate (S'99–M'09) received B.S. degrees in electrical engineering and computer science and mathematics from the Massachusetts Institute of Technology (MIT), Cambridge, in 2002 and the M.S. and Ph.D. degrees in electrical engineering in 2005 and 2008, respectively, from the University of California, Berkeley. He is currently a postdoctoral researcher at the Information Theory and Applications Center at the University of California, San Diego. His research interests include information theory, distributed signal processing, machine learn-

ing, communications, and randomized algorithms for communications and signal processing in sensor networks.

Dr. Sarwate received the Laya and Jerome B. Wiesner Student Art Award from MIT, and the Samuel Silver Memorial Scholarship Award and Demetri Angelakos Memorial Achievement Award from the EECS Department at University of California at Berkeley. He was awarded an NDSEG Fellowship from 2002 to 2005. He is a member of Phi Beta Kappa and Eta Kappa Nu.



Michael C. Gastpar (M'04) received the Dipl. El.-Ing. degree from the Swiss Federal Institute of Technology (ETH), Zurich, Switzerland, in 1997, the M.S. degree from the University of Illinois at Urbana-Champaign, Urbana, IL, in 1999, and the Doctorat es Science degree from the Swiss Federal Institute of Technology (EPFL), Lausanne, Switzerland, in 2002, all in electrical engineering. He was also a student in engineering and philosophy at the University of Edinburgh, Edinburgh, U.K., and the University of Lausanne.

He is currently an Associate Professor in the Department of Electrical Engineering and Computer Sciences at the University of California, Berkeley, and a Professor in the Department of Electrical Engineering, Mathematics, and Computer Science at Delft University of Technology, The Netherlands. He was a summer researcher in the Mathematics of Communications Department at Bell Labs, Lucent Technologies, Murray Hill, NJ. His research interests are in network information theory and related coding and signal processing techniques, with applications to sensor networks and neuroscience.

Prof. Gastpar won the 2002 EPFL Best Thesis Award, an NSF CAREER award in 2004, and an Okawa Foundation Research Grant in 2008. He is an Information Theory Society Distinguished Lecturer (2009–2010). He is currently an Associate Editor for Shannon Theory for the IEEE Transactions on Information Theory, and he will serve as Technical Program Committee Co-Chair for the 2010 International Symposium on Information Theory, Austin, TX.