[Michael Gastpar, Martin Vetterli, and Pier Luigi Dragotti]

Distributed Signal Processing in Sensor Networks

Sensing Reality and Communicating Bits: A Dangerous Liaison

[Is digital communication sufficient for sensor networks?]

The successful design of sensor network architectures depends crucially on the structure of the sampling, observation, and communication processes. One of the most fundamental questions concerns the sufficiency of *discrete* approximations in time, space, and amplitude. More explicitly, to capture the spatiotemporal variations of the underlying signals, when is it sufficient to build sensor network systems that work with discrete-*time* and -*space* representations? And can the underlying amplitude variations of interest be observed at the highest possible fidelity if the sensors *quantize* their observations, assuming that quantization is done in the most sophisticated fashion, exploiting the principles of (ideal) distributed source coding? The former can be rephrased as the question of whether there is a spatiotemporal sampling theorem for typical data sets in sensor networks. This question has a positive answer in many cases of interest, based on the physics of the processes to be observed. The latter can be expressed as the question of whether there is a

1053-5888/06/$20.00©2006IEEE

(source/channel) separation theorem for typical sensor network scenarios. We show that this question has in many cases a negative answer, and we show that the price of separation can be large. To illustrate the conceptual issues related to sampling, source representation/coding and communication in sensor networks, we review the underlying theory and discuss specific examples.

IN SENSOR NETWORKS, A BODY OF RESEARCH HAS DEVELOPED THAT ADDRESSES SIGNAL PROCESSING AND COMMUNICATION JOINTLY.

INTRODUCTION

To paraphrase Shannon, the goal of signal acquisition by means of sensor networks is to reproduce at a read-out station a distributed signal (or some of its key characteristics) under a fidelity constraint, using limited communication resources. In the nondistributed setting, this problem elegantly decomposes into a signal compression (or representation) problem and a communication problem [1]. Consequently, signal processing and communications have become separate topics over the past several decades, developing in quite different directions. In signal processing research, under the conventional paradigm, the data is first brought into a central location, where it is then processed jointly. Paradigmatic instances of this are signal transform techniques, where typically a large portion of the data (if not all of it) is processed simultaneously. The task of bringing the data to the processor is analyzed separately in the framework of communications research.

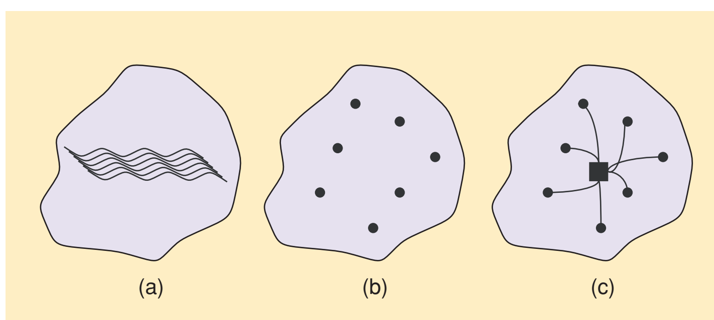
With the advent of sensor networks, a body of research has begun to develop that addresses signal processing and communication jointly. This originates from the insight that the new data sets look fundamentally different: sensors are capable of acquiring vast amounts of data, and there is little hope of first shipping all the data to a central location. Such undertaking would immediately drain the power supplies of all sensors and in a wireless setting would create major interference problems. More specifically, it can be shown that sensor network algorithms designed under such a paradigm may not scale. To address this problem, it has become imperative to process the data (at least partly) in a distributed fashion at the sensors. Such an approach may drastically reduce the communication needs.

In this article, we take a structure-driven, end-to-end approach to the sensor network problem, illustrated in Figure 1. Underlying the whole problem is the physics of the process of interest. This structures the data sets, points to sampling schemes, and indicates what types of correlation will be present in the sensor data. After sampling using the sensors, we are faced with the classic dilemma of the communication engineer: "to separate or not to separate." That is, we either go to the digital domain and apply discretization of the data through quantization and source compression, or we keep data in analog form. The former implicitly assumes a separation into source and channel coding and can be optimal in certain scenarios, while the latter permits any form of joint source-channel coding. Thus, the main focus and goal is to show how the structure of the distributed sensing and communication problem dictates new processing architectures. The key challenge lies in the discretization of space, time, and amplitude, since most of the advanced signal processing systems operate in discrete domain. In the sequel, we investigate and illustrate the sufficiency of such discretization, but also the lack thereof.

We will discuss the general framework, outlining that while the temporal and spatial discretization can be understood from (essentially) the same arguments as in the traditional signal processing applications, the situation is different for amplitude. We formalize this question in an information-theoretic way as one of *source-channel separation:* Can an optimal coding strategy be implemented by first compressing the source(s) into bit streams and communicate those via error-correcting codes? This question has a positive answer for the point-to-point link but not for general networks, and we outline some of the well-known key arguments.

We will also discuss the spatial structure of sensor data. The main insight is that this structure is governed by the *physics* of the underlying process. As we illustrate, in some cases, this leads to spatial sampling theorems, showing that a discrete-space consideration is sufficient.

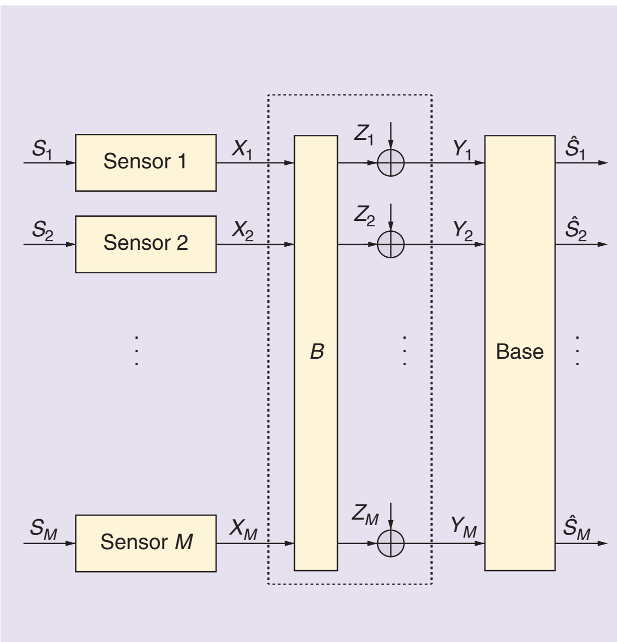
We will discuss and illustrate the sufficiency of discrete-amplitude (digital) processing, i.e., the question of where and when there is a *source-channel separation theorem* in sensor network situations. We show that the answer crucially depends on the interplay between the source structure, the source observation process, and the communication infrastructure. Via the following two paradigmatic examples, we illustrate this issue.



[FIG1] The end-to-end sensor network problem. (a) A physical environment, possibly driven by sources, generates a distributed data set, like a temperature distribution or a sound field. (b) A number of sensors acquire, through spatial sampling, a discrete space version of the physical data set (temporal sampling is typically also done). This leads to a spatiotemporal discrete signal. (c) The data set needs to be conveyed to a central location for reconstruction of the original field. This can be accomplished with standard (albeit distributed) source compression and appropriate communication or a joint source channel scheme.

EXPANDING SENSOR NETWORK

Consider the sensor network scenario illustrated in Figure 2, and suppose that the sources S_m, $m = 1, 2, \ldots, M$, are independent and identically distributed Gaussian with mean zero and unit variance, and that the matrix B is the identity matrix. Then, it is immediate to see that $D = 1/[P_{\text{tot}}/(M + \sigma_Z^2)]$, hence the distortion scales like $D(M) \sim M/P_{tot}(M)$. (For scaling law relationships, we use the notation $f(x) \sim g(x)$ if $\lim_{x\to\infty} f(x)/g(x) = c$ for a finite nonzero constant c.) Since in this case, we simply have M parallel channels, it is equally immediate to show that this distortion can be achieved via separately designed source and channel codes, and hence, a separation theorem applies. This insight can be extended to more general cases: Whenever the covariance matrix of the source vector $(S_1, S_2, \ldots, S_M)$ has full rank (and bounded condition number), and when the MIMO communication channel matrix B has full rank (and bounded condition number), it can again be shown that if $P_{\text{tot}} \sim M$, then $D(M) \sim$ const. To show that this is achievable, it suffices to combine standard distributed source coding (see, e.g., [47]) with standard channel coding. A lower bound follows from a consideration of the idealized scenario where all sensors in Figure 2 are merged into one "super sensor." For the resulting scenario, optimum performance is well known. Hence, this establishes a (scaling-law) separation theorem for a certain class of expanding sensor networks.



[FIG2] The expanding sensor network. Each new sensor also adds new data of interest. In this example, the communication infrastructure (the dashed box) is "rich," meaning that the rank of the channel matrix B is of the order of M. An example is for B to be the M-dimensional identity matrix, which represents the case of a *wired* sensor network. A scaling-law separation theorem applies.

PARADIGMATIC SENSOR NETWORK EXAMPLE 1

The Expanding Sensor Network with Rich (e.g., Wired) Communication Infrastructure

Consider the sensor network scenario of Figure 2, where we assume that the (continuous-time) source signals $S_1, S_2, \ldots, S_M$ are sufficiently independent of each other. Hence, this models the case where each sensor explores new territory, and thus the sensor network is expanding. The base station wishes to recover all of the (continuous-time) sensor readings. For this to be reasonably possible, it is necessary that the communication infrastructure be rich. In Figure 2, this means that the matrix B characterizing the communication channel is essentially of full rank. A special case is when B is a diagonal matrix. Then, the sensors are individually connected to the base station via wired links. To make matters somewhat more specific, for the purpose of this article, we will measure the quality at which the base station can recover the sensor readings by considering the mean-squared error, even though other distortion measures can be studied in an analogous fashion. For a compact parameterization, we will focus on the normalized sum of the M distortion terms, i.e.,

$$D = \frac{1}{M} \sum_{m=1}^{M} E\left[\left\|S_m - \hat{S}_m\right\|^2\right], \qquad (1)$$

where, in slight abuse of notation, we have used S_m to denote the entire source signal (either discrete time or continuous time), $\hat{S}_m$ to denote its estimate constructed at the base station, and $\|S_m - \hat{S}_m\|$ to denote the standard 2-norm between the two signals. The goal of our considerations is then to understand the relationship between the achievable distortion D, the source characteristics, and the communication infrastructure (the total transmitted power P_{tot} and the required bandwidth). We briefly consider a sensor network of this kind in "Expanding Sensor Network," where we show that in a certain sense, there is a source-channel separation theorem for such sensor network situations.

PARADIGMATIC SENSOR NETWORK EXAMPLE 2

The Refining Sensor Network with Poor Communication Infrastructure

By contrast to the previous example, consider now the sensor network scenario of Figure 3. There is a (relatively small) number of underlying sources (or degrees of freedom) that need to be observed, and each sensor picks up a merged and possibly noisy version of all of these underlying sources. As more sensors are added, a better and better reconstruction can be provided at the base station. Therefore, this can be thought of as a "refining" sensor network. Here, the interesting case is when the communication infrastructure is relatively poor, which, as expressed in Figure 3, we model by considering a matrix B with low rank. An example is the standard (scalar) multiple access channel where the rank of B is one. By analogy, we again consider the mean squared error

$$D = \frac{1}{L}\sum_{\ell=1}^{L} E\left[\left\|S_\ell - \hat{S}_\ell\right\|^2\right] \tag{2}$$

via the same slight abuse of notation as in (1). The goal of our considerations is again to understand the relationship between the achievable distortion D, the source and observation characteristics, and the communication infrastructure (the total transmitted power P_{tot} and the required bandwidth). In this article, we consider an example where there is no observation noise (see "Camera Sensor Network") and one where the source observation process is linear and the observations are noisy ("Digital Architectures for Sensor Network Example" and "Analog Architectures for Sensor Network Example"). We show that in such sensor network situations, the price of separately designed source and channel codes can be arbitrarily large.

[THE QUESTION IS WHETHER THERE IS A SOURCE/CHANNEL SEPARATION THEOREM FOR TYPICAL SENSOR NETWORKS.]

In summary, in the sensor network models of interest to our study, the observed source is (typically) continuous in time, space, and amplitude. The data collection point is required to reconstruct the *entire* source sequences, for all time (and space), with respect to an *average distortion* criterion, as expressed in (1) and (2) and subject to power constraints at the sensors.

It is important to point out that this is not the only interesting way of modeling sensor network situations. For example, one can remove the requirement that the source sequences must be estimated for all time. Instead, one can consider a (distributed) parameter estimation problem, such as in [2]–[4], or a (distributed) hypothesis testing or detection problem, such as in [5] and [6]. If the source sequence does not need to be estimated for all time, then it becomes interesting and meaningful to replace the power constraints with energy constraints, and to analyze the lifetime of the network. Such a perspective is taken in [6] and [7].

THE WAY OF THE BIT

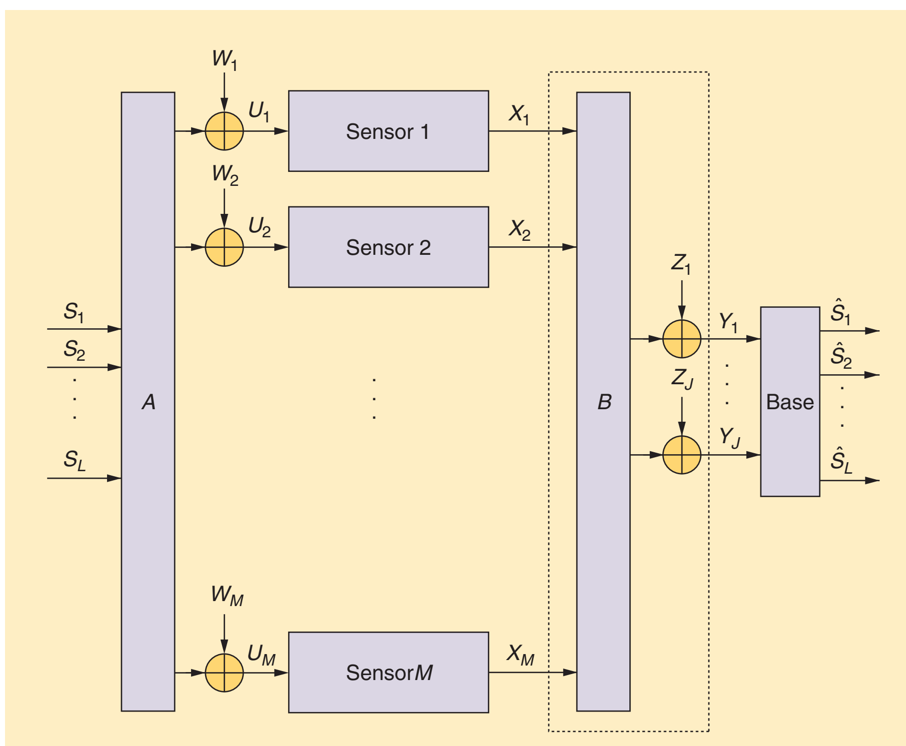
The physical reality observed by a sensor network typically lives in time and space, both of which are best thought of as continuous. The measurements taken by the sensor network are also often continuous in amplitude (and potentially phase). This is not different from the well-studied common communication scenarios, such as a telephone conversation across a wired or wireless connection.

In either case, due to the nature of the most interesting processing devices available today, one of the key questions for the engineer is whether discrete approximations in time, space, and amplitude are sufficient and, if not (or not quite), what kind of a loss they imply.

For the traditional point-to-point communication problem, this set of questions has been well studied and has led to a set of intuitively pleasing (if initially somewhat surprising) answers.

To understand the sufficiency of discrete-*time* approximations, the central result is the well-known sampling theorem for band-limited functions and extensions to other linear subspace cases [8]. When the functions of interest do not fit the model (for example, they are not bandlimited) then preprocessing (like low-pass filtering) has to be applied. This may or may not always be possible.

The problem of the sufficiency of discrete-*space* approximations bears some formal resemblance to the case of discrete-time approximations. From a practical point of view, however, the two problems are quite distinct. While discretization in



[FIG3] The refining sensor network. A vector source (with arbitrary distribution) is observed *M*-fold through a vector-valued function *A* (for example, *A* could be a matrix) and in additive noise, independently by *M* sensors. The *M* sensors communicate over an additive noise MIMO channel, characterized by the matrix *B*, to a base station that houses the central estimation officer. The sensors may have (generally limited) cooperation capabilities. No scaling-law separation theorem applies.

time can be seen as an engineering choice, discretization in space really is a physical necessity in most cases: sensors are spatially localized objects, and this leads necessarily to spatial sampling. However, one can again ask the question under what conditions discrete-space considerations are sufficient and what loss they imply otherwise. The fact that no spatial low-pass filtering is possible in general shows how critical spatial sampling and aliasing can be.

The remaining issue, then, is the question whether *discrete in amplitude* (often referred to as digital) is sufficient if the original data is *analog* in amplitude (such as a temperature, a sound pressure waveform) and if the communication medium is analog in nature (such as a voltage or an electromagnetic field). By analogy to spatiotemporal sampling, the question is again whether it is without loss of optimality to pass from an infinite set (continuous data) to a countable set (a set of

MULTIUSER INFORMATION THEORY AND SOURCE-CHANNEL SEPARATION

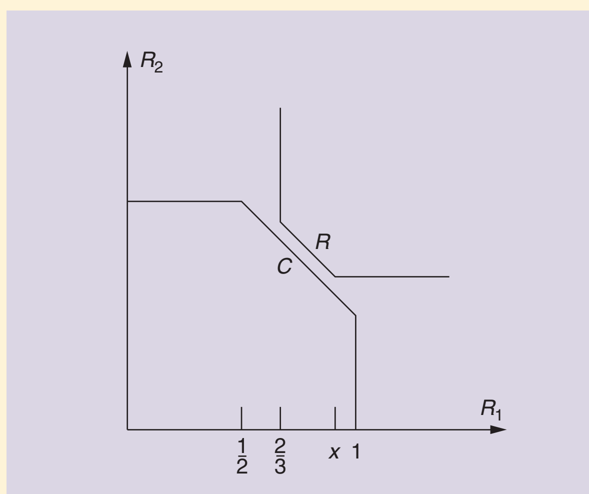
One of the most studied networks in multiuser information theory is the multiple access channel (MAC). A simple MAC is the scenario where two terminals transmit with power P each on the same frequency band to a single receiver (base station), subject to additive white Gaussian noise of variance σ^2. The capacity region $\mathcal{C}$ is the set of rate pairs (R_1, R_2) satisfying $R_1 \leq (1/2)\log_2(1 + (P/\sigma^2))$, $R_2 \leq (1/2) \log_2(1 + (P/\sigma^2))$, and $R_1 + R_2 \leq (1/2)\log_2(1 + (2P/\sigma^2))$. This leads to a pentagonal shape [9, Fig. 14.17].

Recently, a *scaling-law* perspective has been developed in multiuser information theory: How does capacity behave as the number of users in the network increases? For the Gaussian MAC, it is easy to see that the *sum* of the rates grows *logarithmically* in the number of nodes [9, p.407]. Recent work has shown that for the Gaussian (multiple) relay channel, it also grows logarithmically [10], [11]. Finally, for an ad hoc network scenario, the sum of all the rates grows like the square-root of the number of nodes [12]. These results are sometimes interpreted in a pessimistic fashion as saying that in all these networks, the rate *per user* vanishes as the number of nodes tends to infinity.

On the rate-distortion side, an interesting scenario for which a solution has been found is the so-called CEO problem [13]. Here, M agents all observe independently noisy versions of one and the same source and have to produce separate descriptions. If a total rate R bits per source sample is available, it has been shown that as the number of agents becomes large, the attainable distortion behaves inversely proportional to the rate. One interesting way of understanding this result is by noting that in the standard single-source rate-distortion problem, the distortion usually decreases exponentially in the rate. Some other interesting cases are discussed in "A Glimpse at Distributed Source Coding."

Unfortunately, even if the rate-distortion region $\mathcal{R}$ for *any* source coding problem and the capacity region $\mathcal{C}$ for *any* channel network were known, this would not solve the overall joint source-channel communication problem. A classical example illustrating the fact that source-channel separation does not hold for networks is usually given as follows [14]: The channel is the binary adder multiple access channel, taking two binary $\{0, 1\}$ inputs and outputting their sum $\{0, 1, 2\}$. The capacity region $\mathcal{C}$ of this channel has the pentagonal shape given in Figure 4; see [9, Fig.14.13] for more details. Now suppose that the two transmitting terminals each observe a binary sequence, call them S_1^n and S_2^n. The two sequences are correlated with each other such that for each time instant, the events $(S_1, S_2) = (0, 0), (0, 1)$, and $(1, 1)$ are all equally likely, and (1, 0) does not occur. Clearly, at least $H(S_1, S_2) = \log_2 3 \approx 1.58$ bits per source sample are required. The full Slepian-Wolf rate region $\mathcal{R}$ is also given in Figure 4; the point labeled x is $\log_2 3 - (2/3)$. The two regions do not intersect, and hence, one is tempted to guess that these two sources *cannot* be transmitted across this MAC. However, this conclusion is wrong: While there is no "digital" architecture that achieves this, there is a simple "analog" strategy: pure uncoded transmission will always permit to recover both source sequences without error, due to the fact that the dependence structure of the sources is perfectly matched to the channel. This illustrates that no separation theorem applies to general networks. An example where the gap between the best digital strategy (along the lines of the separation theorem) and the optimum scheme increases as the number of nodes in the network becomes large is analyzed in detail in "Analog Architectures."

While the general answer is unknown, there are also network cases known where a separation theorem of the shape of (5) can be given, including the transmission of independent sources with respect to independent fidelity criteria across any multiple access channel, see e.g., [15, Thm.1.9], and the error-free transmission of discrete correlated sources across separate (parallel) channels, see [16].



[FIG4] Capacity region $\mathcal{C}$ and rate-distortion region $\mathcal{R}$ do not intersect in this example.

messages). The simplest example is scalar quantization, a more sophisticated version being vector quantization. For the point-to-point communication problem if we allow *any* vector quantizer and *any* error-correcting code, even the abstract, information-theoretic ones, then it is well known that the answer to the above question is positive; it is given by Shannon's celebrated joint source-channel coding theorem, often referred to as (source-channel) separation theorem.

More precisely, the source coding problem can be characterized in terms of a rate-distortion function, often denoted as $R(D)$, and the channel coding problem in terms of a capacity-cost function, denoted as $C(P)$. The separation theorem then states that a distortion D is attainable if

$$R(D) < C \tag{3}$$

and *cannot* be attained if

$$R(D) > C. \tag{4}$$

The case $R(D) = C$ is attainable in some cases but not in all. (This issue shall not be discussed in any detail in this article.)

Owing to the stunning success of the digital communication paradigm in practical systems, it is clear that the same approach has been taken to the design of communication networks. Along the lines of the development for the point-to-point case, one can now consider the rate vectors $(R_1, \ldots, R_L)$, in bits per source symbol, required to maintain prescribed distortion levels on all sources. Generally, many different rate vectors will be permitted to achieve this, and one usually thinks of the corresponding *rate-distortion region*, denoted by $\mathcal{R}$. Similarly, for the communication network, one can determine the number of bits per channel used that can be simultaneously pushed through the inputs of the network. The vectors of reliably achievable rates can be captured in terms of a capacity region $\mathcal{C}$. It is then easy to see that a set of prescribed distortion levels is attainable if

$$\mathcal{R} \cap \mathcal{C} \neq \emptyset, \tag{5}$$

but this is not a necessary condition. In other words, even if the intersection of the rate-distortion region and the capacity region is empty, there may exist a code that achieves the prescribed distortion levels. However, that code is *not* a digital code; that is, it cannot be understood in terms of source compression followed by reliable communication across noisy channels. Rather, it requires joint source-channel coding.

DISCRETE SPACE: SAMPLING OF DISTRIBUTED SIGNALS

Sampling is so common that we sometimes forget it is a little miracle and that it comes with a few strings attached. In the case of sensor networks, the critical issue is certainly the sampling in space, inherent in the discrete nature of the sensors. Also, distributed signals exist in time and space and are thus inherently multidimensional. Distributed signal acquisition is thus the spatiotemporal sampling of such signals. Of course, the field of array signal processing has dealt with such problems in the past (see, e.g., [17]) but with a perspective that is different from the one used in sensor networks. In typical array signal processing, the array is one dimensional, regular, and the signals are often narrow band. In sensor networks, the array is irregular and two dimensional (random sensor placement on a plane), and the signals can be wide band (e.g., sound, images). The obvious question is one of spatial sampling, with the twist that there cannot be any spatial low-pass filtering before sampling. Thus, most sensor network data is aliased with respect to spatial frequency.

In "The Plenacoustic Function," we discuss the interaction between the physics of the process and spatial sampling, in particular in the case of distributed audio signals and the plenacoustic function [18]. Other examples of interest where such an analysis can be applied include distributed camera systems, where the plenoptic function [19] plays a key role. This function can be used as an underlying model for distributed image or video acquisition. Interestingly, a sampling theorem for spatial sampling can also be derived in this case [20]. Finally, the distribution of temperature, where the heat equation is central, has been considered in [21].

From the above, we can summarize the methodology. First, consider the physical process producing the quantity of interest for the sensor network. This leads to a specific spatiotemporal behavior. From there, analyze the possible sampling and aliasing, especially in the spatial dimension.

A final question of interest is if sampling can be considered in isolation, without referring to compression and communication issues. This is certainly of great interest but does not have a simple answer. Clearly, if we have the freedom to place sensors at will, we can optimize sensor placement so that energy usage is reduced, for example. At the same time, putting all the sensors close to the base station will lead to a very ill-conditioned interpolation of the original data, something to be avoided in the presence of noise. From results on irregular sampling, it is intuitively clear that to first order, it will not be possible to substantially deviate from uniform placement, and therefore, only limited gains are possible.

DIGITAL ARCHITECTURES

The term *digital* has become so pervasive that it is sometimes assumed to be given. That is, we go from whatever analog values to some discrete representation. This is what we will assume in this section. But then, in the next section, we will show that things are not always so simple in general.

To discuss digital architectures, we need to define them somewhat more formally. A digital architecture is a two-stage procedure, where each stage is designed independently, the only link between the two stages being digital rate constraints. Intuitively, any scheme whose performance can be expressed in terms of a rate-distortion behavior combined with a capacity region will be considered a "digital" architecture. In more detail, this can be described as follows:

1) The *source code* is designed with only the capacity (region) of the channel network available. No further information about the finer structure of the channel can be used.

2) The *channel code* is designed without any knowledge about the source at all. Its goal is to communicate messages in such a way as to avoid errors.

It is perhaps worth illustrating what constitutes digital architectures according to this definition and what does not. Clearly, any traditional digital communication strategy falls under this category, such as a system where the source is first passed through a vector quantizer, followed by, for example, an entropy coder, and where the resulting bit stream is communicated via an error-correcting code that avoids (block) errors on the noisy channel. On the other hand, a strategy where the channel provides *soft* information, and the source code is designed to work with such soft

[THE SENSOR NETWORK PERFORMS SAMPLING BUT WITHOUT SPATIAL FILTERING.]

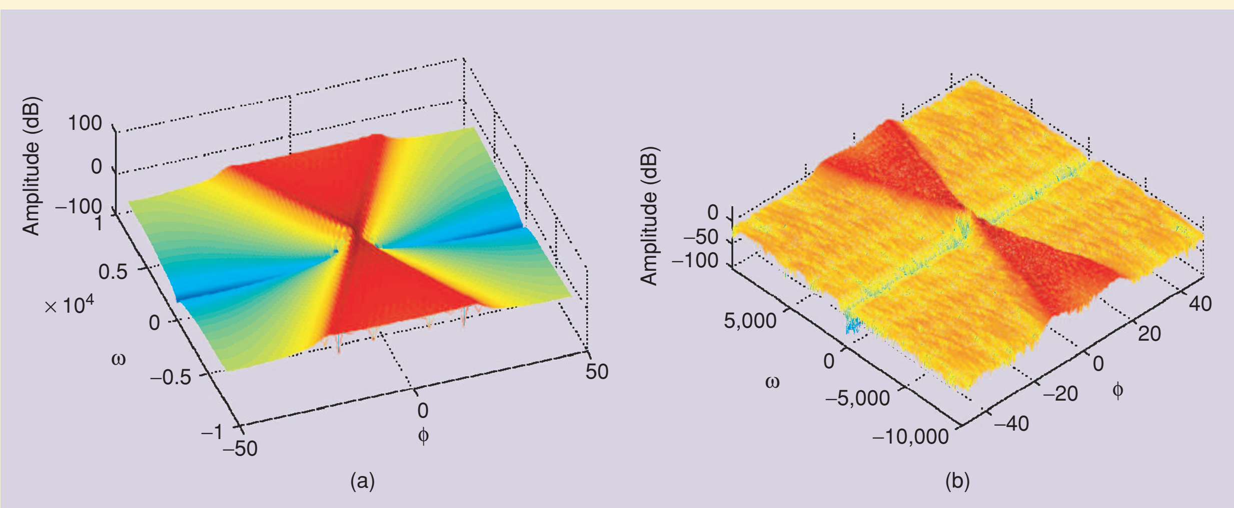
THE PLENACOUSTIC FUNCTION

To make matters specific, we first consider the concrete case of acoustics signals and microphone arrays. The sound field, be it in open space or inside a room, is the solution of a second-order partial differential equation called the wave equation. The driving term in the differential equation is given by the various sound sources. The key is thus the kernel of the wave equation, since the source distribution is convolved with the kernel to produce the actual acoustic field. This kernel, also known as the Green function, has a particular form. Its Fourier transform for a particular temporal frequency is essentially band-limited in spatial frequency. For a concrete example, consider a line in a room, and the spatiotemporal room impulse response $h(x, t)$ with respect to a source. The Fourier transform $H(\omega)$ is essentially supported on a triangle with

$$\phi \leq \frac{\omega}{c}, \qquad (6)$$

where c is the speed of sound and ϕ and ω are the spatial and temporal frequencies, respectively. Figure 5 shows a simulated and a measured spectrum of the Green function or plenacoustic function of a room, indicating clearly the bow-tie shape of the spectrum that can be used in sampling. For details, we refer to [18].

Now we are in a position to address the sampling question. First, it is worth remembering that while the temporal frequency can be limited using low-pass filtering, there is no such possibility over space. That is, spatial sampling cannot be preceded by any spatial filtering. Nonetheless, thanks to the shape of the spectrum, if the maximal temporal frequency is ω_0, then the spatial spectrum is limited to ω_0/c. That is, spatial sampling with a distance between microphones of the order of $d = c/\omega_0$ is adequate to obtain a good representation of the acoustic field. Such a rule of thumb is well known in array signal processing [17]. A precise analysis is given in [18], where the decay of the spectrum and the analysis of the resulting signal-noise ration is given. It is to be noted that the discrete spectrum over time and space is not white, and thus residual correlation is present and can be used in distributed compression. One such scheme is analyzed in [22], where it is shown that distributed compression using quincunx sampling achieves the same $D(R)$ as centralized compression. This points to the close interaction of signal structure, spatiotemporal sampling, and distributed compression.



[FIG5] The Fourier transform of the plenacoustic function, with spatial and temporal frequencies. (a) Simulated and (b) measured plenacoustic function of a room. The triangular shape of the Fourier transform is clearly visible, which leads to a sampling theorem over space when the temporal frequency is limited.

information, is not considered a digital architecture since the two stages are not truly designed independently of each other. It is clear that such a strategy really constitutes a joint source-channel code.

DISTRIBUTED SOURCE CODING IMPLICITLY ASSUMES A DIGITAL ARCHITECTURE.

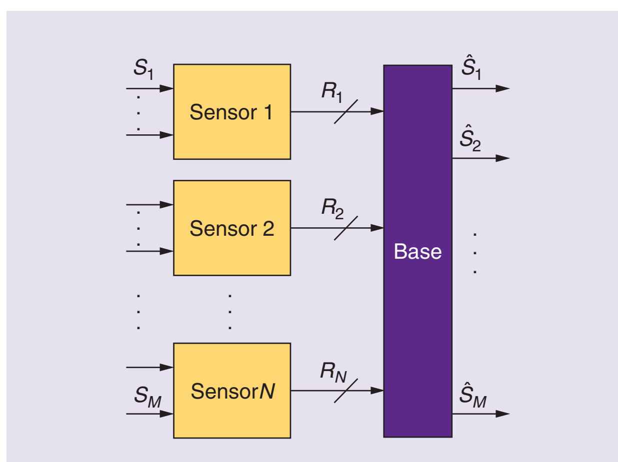
In this section, we discuss some of the key aspects of digital architectures, focusing primarily on the source coding side. To compress a single source, one can think of applying a suitable vector quantizer to an entire vector of source symbols. Unless the vector quantizer is an information-theoretically optimal construction, the resulting quantization indices still have redundancy in them, and it is customary to pass them through an entropy coder. An alternative and very popular approach known as transform coding consists in applying a suitable (linear) transform to the vector before quantizing each transformed component independently and, in the case of jointly Gaussian vector, it is well known that the optimal transform is the Karhunen-Loève transform (KLT). To be more precise, if we denote the covariance matrix of the input vector S by Σ_S and assume that the vector is jointly Gaussian, then the optimal transform coder operates as follows: The input vector S is first transformed with a KLT, then the transformed components are quantized independently, and more rate is allocated to the components related to the largest eigenvalues of Σ_S.

Let us now consider the sensor network problem where we have multiple correlated sources that need to be compressed in a distributed fashion. Specifically, consider the source coding problem illustrated in Figure 6. If sensors could collaborate among themselves (at no cost), then the distributed source coding problem would be mute: we could merely apply the same algorithms as in the single-source case. However, such sensor collaboration is usually not feasible since it would require an elaborate intersensor communication and would consume most of the power of the sensors. In other words, it is no longer possible to apply a vector quantizer or a transform coder to the entire source vector. Instead, these algorithms have to be approximated in a distributed fashion. Suppose that each sensor has applied a suitable vector quantizer to its observed data and is now left with quantization indices. There are two different kinds of redundancies that can still be exploited. On the one hand, each sensor's indices may be dependent; on the other hand, and more interestingly, the quantization indices of sensor 1 may be correlated with the indices of sensor 2, and so on. This type of redundancy can be removed in a very elegant fashion, pioneered by Slepian and Wolf [23], and further developed (and extended to the case of lossy compression) by Wyner and Ziv in [24]. An overview of these fundamental results on distributed source coding is given in "A Glimpse at Distributed Source Coding." Constructive distributed encoders have been developed more recently in [25]–[33].

This leaves us with the question of how the transform coding paradigm changes in this new distributed context. Namely, if each sensor were to apply a transform to the observed subvector, should this transform be the same as in the single-source case or should it be modified, and should the quantization and bit allocation strategies be modified as well? The interesting answer is that, in the new distributed scenario, the optimal solution usually requires not only a modification of the structure of the KLT (leading, for instance to the conditional KLT) but also to a modification of the bit allocation strategy and of the quantization process [34], [35]. Extensions to the centralized transform coding paradigm have also been investigated in [36].

Let us now return to the overall design of the digital architecture, specifically to the interactions between the observation process, the source coding, and the channel coding. While the general problem is hard and comes in many flavors, we want to consider three special cases. The first special case is discussed in "Expanding Sensor Network" and is related to the expanding sensor network of Figure 2. We show that, in this case, the architecture scales properly with the number of sensors and that separation holds in a scaling-law sense.

The two examples found in "Camera Sensor Network" and "Digital Architectures for Sensor Network Example" show, however, that there exist other instances where separation does not hold. In particular, in "Camera Sensor Network," we assume sensors are digital cameras, and we show that we incur a small penalty by doing separation. "Digital Architectures for Sensor Network Example," finally, concerns the wireless sensor network (WSN) with a structure as given in Figure 3. For this case, we explicitly evaluate the performance to compare it to analog architectures. As it will become evident in the next section, the digital architecture of this second example does not scale properly with the number of sensors and this leads to vastly suboptimal performance.



[FIG6] Distributed transform coding: The full data vector $(S_1, \ldots, S_M)$ is observed in multiple partitions. Each terminal provides a compressed version of its observation.

The case of transmission of correlated sources through networked independent channels has been investigated in [16], [37], and [38]. Other digital approaches have been studied for example in [39]–[42].

[WHILE DIGITAL IS CONVENIENT, ANALOG MIGHT BE OPTIMAL.]

ANALOG ARCHITECTURES

By contrast to the digital architectures discussed earlier, there are ways of "coding" that are not based on the representation of all information in terms of discrete messages (such as bit streams). For the purpose of this exposition, we will refer to any such approach as *analog* architecture. Specifically, it should be noted that analog is not taken to imply *linear* processing nor any other constraint of this form. Rather, *analog* should be defined negatively as *nondigital*, and the point of the article is to show that some sensor network scenarios strictly require nondigital architectures. Such nondigital architectures are, in certain contexts, also referred to as joint source-channel coding.

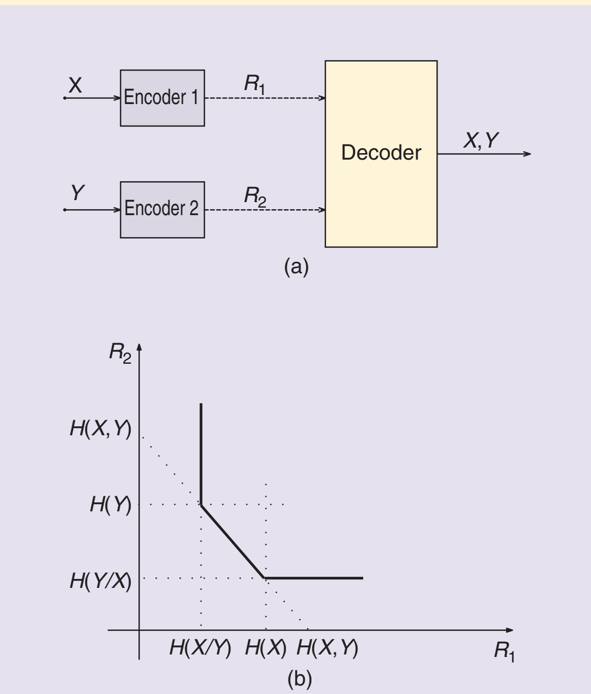
As we have argued, a set of powerful tools has been developed over the past five decades that facilitates the design of algorithms for handling discrete information,

A GLIMPSE AT DISTRIBUTED SOURCE CODING

Consider two discrete memoryless sources X and Y that have to be encoded at rates R_1 and R_2, respectively. Clearly, this can be achieved with no loss of information using $R_1 \geq H(X)$ and $R_2 \geq H(Y)$ bits where $H(\cdot)$ denotes the source entropy. If X and Y are correlated and a single encoder has access to both sources, lossless compression is achieved when $R_1 + R_2 \geq H(X, Y)$.

Now assume that these two sources are separated, and two separate encoders need to be used as illustrated in Figure 7(a). Slepian and Wolf [23] showed that lossless compression can still be achieved with R_1 and R_2 satisfying:

$$R_1 \geq H(X|Y), \quad R_2 \geq H(Y|X), \quad R_1 + R_2 \geq H(X, Y).$$



[FIG7] (a) The Slepian-Wolf problem: distributed encoding of *X* and *Y*. (b) The Slepian-Wolf rate region.

This means, surprisingly, that there is no loss in terms of the overall rates even though the encoders are separated. The Slepian-Wolf rate region is sketched in Figure 7(b).

Of particular interest is the asymmetric case $(R_1, R_2) = (H(X|Y), H(Y))$. Since $R_2 = H(Y)$, Y can be assumed available at the decoder and the only challenge is to find an efficient way to encode X. This is normally known as the source coding problem with side information at the decoder. This case is important because, if we can show that the rate pair $(R_1, R_2) = (H(X|Y), H(Y))$ is achievable, then by exchanging the role of X and Y we can achieve $(H(X), H(Y|X)$ as well and, finally, all the points on the line connecting $(H(X|Y), H(Y))$ with $(H(X), H(Y|X))$ can be achieved using time-sharing arguments.

The proof of Slepian and Wolf of the achievability of $(H(X|Y), H(Y))$ is based on classical information theoretic arguments. However, it contains already all the main ingredients and intuitions that have been used more recently to design practical distributed source codes. The Slepian and Wolf main intuition goes along the following lines: Since X and Y are correlated and Y is available at the decoder, one can view X as the input and Y as the output of a noisy communication channel. To quote [23, p. 474]: "from the fact that $p_{X,Y}(x, y) = p_{Y|X}(y|x)p_X(x)$, we can think of the Y-sequences of the correlated source as being generated by applying successive characters of the X-sequence as inputs to a discrete memoryless channel with transition probabilities $p_{Y|X}(y|x)$."

This means that, if we design a capacity-achieving channel code for that channel, we can ensure reliable transmission of a sub-set of X. More precisely, a capacity-achieving code contains on average $2^{I(X,Y)} = 2^{H(X)-H(Y|X)}$ elements that can be used as inputs to this channel and decoded with vanishing error probability when Y is observed. Now, the good news is that we can design many such codes and, since the source X produces on average $2^{H(X)}$ different symbols, by designing $2^{H(Y|X)}$ disjoint capacity-achieving codes, we can including source codes as well as channel codes. No similarly general tools are known for the design of analog architectures. Rather, these techniques are usually designed on a case-by-case basis, and it is often hard to analyze their performance in a precise fashion.

NO GENERAL TOOLS ARE KNOWN FOR THE DESIGN OF ANALOG ARCHITECTURES.

We focus on paradigmatic exemplary cases that illustrate the need for the development of a more systematic framework for the design of nondigital communication system architectures.

The key case is the wireless refining sensor network example that was introduced earlier. In "Digital Architectures Sensor Network Example," we used known techniques to bound the best possible performance of any digital architecture. Specifically, from (10), we concluded that as the number of sensors becomes large, the distortion decays at best like $1/\log(MP_{\text{tot}})$. The question, then, is whether there is any nondigital approach that can outperform this or whether this is a fundamental bound for the considered communication problem. While the optimal strategy for this case to date is unknown, we consider a very simple analog architecture: Each sensor basically scales its noisy observation by an

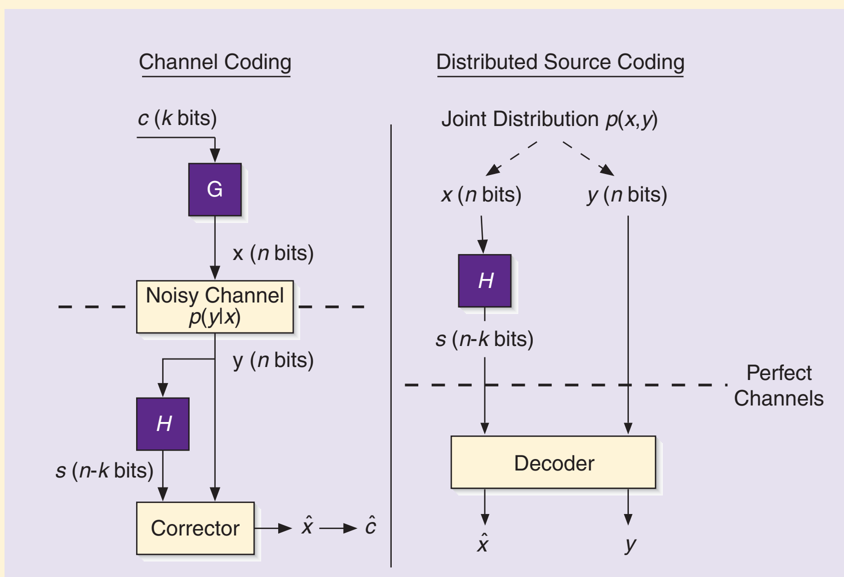
associate any symbol produced by X to one capacity-achieving channel code. Then the coding strategy is as follows: Encoder 2 transmits Y using $R_2 = H(Y)$ bits. Encoder 1 does not need to send the code word representing X, instead the encoder simply indicates which channel code X belongs to. This requires on average $H(X|Y)$ bits. The receiver can then use the decoder appropriate for the channel code specified by Encoder 1 to retrieve X from Y with no error. The rate pair $(H(X|Y), H(Y))$ is thus achievable.

This connection between source coding with side information and channel coding principles, which is highlighted in Figure 8, was made more evident by Wyner [43] and Berger [44], and has been used recently to design constructive distributed codes, see [25], [28], and [29] for early examples. Extensions to an arbitrary number of correlated sources and ergodic processes were presented by Cover in [45], [46].

The case of lossy coding of correlated sources, in particular, of continuous-valued sources is more involved and much less is known. An important case studied by Wyner and Ziv [24] is the one where Y is available at the decoder and X has to be reconstructed within a certain distortion D. Even though the minimum rate R_1 necessary to achieve this distortion is usually greater than the rate used in the case where Y is available at both the encoder and the decoder, Wyner and Ziv showed that there is no rate loss in the particular case of MSE distortion and jointly Gaussian sources. In particular, if $X \sim \mathcal{N}(0, \sigma_X^2)$ and $Y = X + U$ with U independent of X and $U \sim \mathcal{N}(0, \sigma_U^2)$, we have that

$$R_{WZ}(D) = R_{X|Y}(D) = \begin{cases} \dfrac{1}{2}\log_2 \dfrac{\sigma_X^2\sigma_U^2}{(\sigma_X^2+\sigma_U^2)\,D} & \text{if } 0 \le D \le \dfrac{\sigma_X^2\sigma_U^2}{(\sigma_X^2+\sigma_U^2)} \\ 0 & \text{if } D > \dfrac{\sigma_X^2\sigma_U^2}{(\sigma_X^2+\sigma_U^2)}. \end{cases}$$

The exact solution to the fully distributed compression problem (i.e., when both X and Y are compressed) is unknown to date. The best known bounds where provided by Berger in [44].



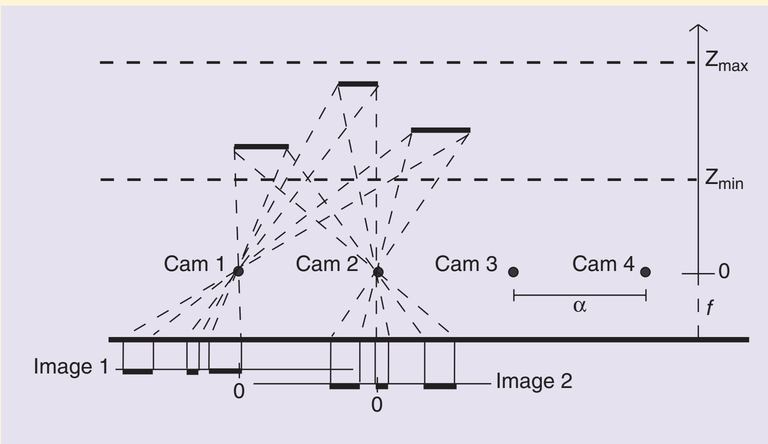
[FIG8] Channel coding and distributed source coding. In channel coding, the syndrome of *Y* is used to retrieve the original symbol *X*. In distributed source coding, the syndrome of *X* is transmitted to the decoder. By observing *Y* and the syndrome, the decoder can reconstruct *X*.

appropriately chosen factor and transmits this on the channel. This generates very strong interference between all the sensors, but this interference is designed so that a cooperation gain results. To make this explicit, we reconsider the simple version of the WSN example that we studied in detail in the context of digital architectures in "Digital Architectures for Sensor Network Example." Specifically, in "Analog Architectures for Sensor Network Example," we present a detailed argument that shows that there is an analog architecture that incurs a distortion that scales like

[FOR THIS EXAMPLE, THE DIGITAL ARCHITECTURE REQUIRES EXPONENTIALLY MORE SENSORS THAN THE ANALOG ONE.]

CAMERA SENSOR NETWORK

Consider the simplified camera sensor network setup shown in Figure 9. There are M digital pinhole cameras that are located along a line. We assume that camera locations are known and denote with α the distance between two consecutive cameras. The visual scene that is perpendicular to this line



[FIG9] Our camera sensor network configuration.

is made of L Lambertian planes. Plane locations are unknown, but the minimum and maximum possible distances of the planes to the line are known and are denoted by z_{min} and $z_{\max}$ ($z_{\max}$ can be infinity and $z_{\min} > 0$). Cameras communicate to a single base station through a classical multiaccess Gaussian channel with capacity $C = 1/2\log_2(1 + [P_{\text{tot}}/\sigma^2])$ where σ^2 is the variance of the noise and P_{tot} is the total power used by the sensors (see "Multiuser Information Theory and Source-Channel Separation" for more details).

Because of the pinhole model, each camera observes a perspective projection of the visual scene. Since the scene is made of Lambertian planes, these projections are piecewise constant functions. The acquisition process at each camera can be modeled as a linear filtering followed by sampling (we assume noiseless measurements for the sake of simplicity). Thus, each camera observes a *blurred* and sampled version of the original piecewise constant projection, and it is possible to show that, in many cases, exact reconstruction of the original projection from the samples is possible [48]. The reconstruction of the original visual scene is then obtained by back-projecting the reconstructed perspective projections and is exact when there are at least $M \geq L + 1$ cameras observing the visual scene without suffering occlusion. This means that, in this particular context, there exists an exact answer to the sampling problem.

Now assume that no occlusion occurs at any of the M cameras. The perspective projections have been reconstructed and each projection is piecewise constant with L pieces and $2L$ discontinuities. Each projection is therefore specified by $3L$ parameters. The distributed compression is then performed as follows: each sensor quantizes the $3L$ parameters independently and then a Slepian and Wolf (S-W) encoder is used to remove the remaining redundancy. The interesting element here is that the design of the S-W encoders depends on the properties of the physical phenomenon and, since we are assuming that $z_{\min}$, $z_{\max}$ and α are known, the practical implementation of the S-W encoders is almost straightforward [49]. It is then possible to show that, if the total bit budget is R, the distortion-rate behavior at high rates is given by

$$D(R) \sim c_1 2^{-2R\tilde{\gamma}},$$

where $\tilde{\gamma} = 1/[9L + (36/5)L^2]$ [50]. Notice that $D(R)$ does not depend on the number of sensors involved, but only on the total bit budget. We thus have an exact bit conservation principle in this case.

Since we can transmit only $R = C = (1/2)\log_2[1 + (P_{\text{tot}}(M)/\sigma^2)]$ bits per channel use, we obtain that the distortion at the base station behaves like $D_{\text{digital}}(M) \sim c_1(1/(1 + P_{\text{tot}}(M)/\sigma^2))^{\tilde{\gamma}}$, which implies the following scaling behavior for a (distributed) digital code:

$$D_{\text{digital}}(M) \sim \frac{1}{(P_{\text{tot}}(M))^{\tilde{\gamma}}}. \quad (7)$$

A lower bound to the distortion can be obtained by considering the scenario where all the cameras are linked with ideal cables. For this idealized scenario, it can be shown that the optimal distortion behaves like

$$D_{\text{lowerbound}}(M) \sim \frac{1}{(MP_{tot}(M))^{\gamma}}, \quad (8)$$

where $\gamma = 1/(9L)$.

DIGITAL ARCHITECTURES FOR SENSOR NETWORK EXAMPLE

For the sensor network topology illustrated in Figure 3, suppose that we use a digital architecture. The corresponding (sum-)rate-distortion function for the case $L = 1$ and when S is distributed according to a Gaussian law is called the quadratic Gaussian CEO problem [51], [52] and can be expressed as

$$R^{\text{CEO}}(D) = \log_2\left(\frac{\sigma_S^2}{D}\right) + M\log_2\left(\frac{M\sigma_S^2}{M\sigma_S^2 - \sigma_W^2\left(\frac{\sigma_S^2}{D}-1\right)}\right). \quad (9)$$

The rate available is no larger than the capacity of the Gaussian MIMO channel with input vector $(X_1, \ldots, X_M)$ and output vector $(Y_1, \ldots, Y_J)$, with reference to Figure 3. Assuming that J is held fixed, this rate increases logarithmically with M. Then, it can be shown easily that the distortion, as a function of the number of sensors M and the total sensor power P_{tot} behaves as

$$D_{\text{digital}}(M) \sim \frac{1}{\log(MP_{\text{tot}}(M))}. \quad (10)$$

The same scaling behavior can be established for the case where the distribution of S is more general [53]. This example is explained in more detail in [54].

$$D_{\text{analog}}(M) \sim \frac{1}{MP_{\text{tot}}(M)}. \quad (11)$$

To compare this to the digital architectures discussed earlier, suppose now that a minimum tolerable distortion D_0 and a power budget $P_{\text{tot}}(M) = P_0$ is fixed. How many sensors M_{analog} and M_{digital} do the analog and the digital architectures, respectively, require? By comparing (10) and (15), we find that

$$M_{\text{digital}} \approx e^{M_{\text{analog}}}. \quad (12)$$

That is, the digital architecture will require exponentially more sensors than the analog.

This shows that the question of how much information is acquired by a sensor network *cannot* be generally expressed in terms of bits, a somewhat counterintuitive insight. Assessing in an operationally meaningful way the "amount" of information depends on the overall structure of the network under consideration.

The fact that pure analog transmission outperforms the most sophisticated digital architecture may seem counterintuitive at first, but there is definitely no reason to believe that pure analog transmission should be the best possible strategy. The latter is unknown at this point, and one has to resort to lower bounds to the distortion instead. The currently available tools to develop such lower bounds are rather limited in their generality. Specifically, the only general techniques are of the *cut-set* type, i.e., they essentially partition the network into two sides and analyze the performance of the resulting point-to-point system. That performance cannot be worse than the performance of the original network. In fact, it will generally be much better, leading to overly optimistic bounds. This is discussed in more detail in [54]. Somewhat surprisingly, in spite of the overly optimistic nature of the bounds, they are sufficient to confirm that for the sensor network of Figure 3, the "scaling behavior" of the simple analog architecture considered above, i.e., the dependence of its performance on the number M of sensors as given in (11), coincides with the optimum scheme insofar as the dominant term is concerned.

ANALOG ARCHITECTURES FOR SENSOR NETWORK EXAMPLE

Consider the sensor network topology illustrated in Figure 3 with $L = J = 1$ and $A = B^T = (1, 1, \ldots, 1)$, and let the underlying source sequence $\{S[i]\}_{i=1}^{\infty}$ be a sequence of independent Gaussian random variables of mean zero and variance σ_S^2. Consider the analog architecture where each sensor, at time n, transmits $X_m[n] = \sqrt{P_{\text{tot}}/M(\sigma_S^2 + \sigma_W^2)}U_m[n]$. Hence, the receiver observes

$$\begin{aligned} Y[n] &= Z[n] + \sum_{m=1}^{M} X_m[n] \\ &= Z[n] + \sqrt{\frac{MP_{tot}}{(\sigma_S^2+\sigma_W^2)}}S[n] + \sum_{m=1}^{M}\sqrt{\frac{P_{tot}}{M(\sigma_S^2+\sigma_W^2)}}W_m[n]. \end{aligned} \quad (13)$$

Clearly, for this scenario, the optimum estimator of $S[n]$ given the entire received sequence $\{Y[i]\}_{i=1}^{\infty}$ needs to only take into account $Y[n]$ and due to the fact that all random variables are jointly Gaussian, it is merely a linear operation, given by $\hat{S}[n] = (E[SY]/E[Y^2])Y[n]$. The resulting distortion is the well-known formula

$$D = \sigma_S^2 - \frac{(E[SY])^2}{E[Y^2]}, \quad (14)$$

where we can evaluate $(E[SY])^2 = MP_{\text{tot}}\sigma_S^4/(\sigma_S^2+\sigma_W^2)$ and $E[Y^2] = \sigma_Z^2 + MP_{tot}\sigma_S^2/(\sigma_S^2+\sigma_W^2) + P_{\text{tot}}\sigma_W^2/(\sigma_S^2+\sigma_W^2)$, leading to the following overall distortion:

$$D_{\text{analog}} = \frac{\sigma_S^2\sigma_W^2}{M\sigma_S^2+\sigma_W^2}\left(1 + \frac{M(\sigma_S^2\sigma_Z^2/\sigma_W^2)}{\frac{M\sigma_S^2+\sigma_W^2}{\sigma_S^2+\sigma_W^2}P_{tot}(M) + \sigma_Z^2}\right). \quad (15)$$

The information-theoretic optimality of this simple approach was first pointed out in [10].

CONCLUSIONS AND KEY CHALLENGES FOR FUTURE RESEARCH

We have considered sensor networks acquiring data from the physical world to reproduce a physical phenomenon at a central location. Since discrete representations of information are at the heart of current technology, a fundamental question concerns the problem of whether *any* signal can be sufficiently characterized in discrete form. For sensor networks, there are three fundamental dimensions: time, space, and amplitude. All three are typically best thought of as continuous initially, and we have illustrated that there are reasons to believe that in interesting cases, discrete-time and discrete-space representations are sufficient. The question of discrete-amplitude representations is a more subtle one, and, in the WSN case, a double answer must be accepted. In some cases, discrete-amplitude representations are sufficient, but in other cases, analog architectures using joint source-channel communication perform optimally in a scaling sense, while digital architectures would lead to a vastly suboptimal solution.

The challenge now is to understand precisely when a separation principle holds in a scaling sense. While the fully general solution to this problem is still open, our findings suggest the following overall picture:

- *Expanding sensor network (with rich communication infrastructure):* For the example presented in "Expanding Sensor Network," we were able to show that the distortion behaves at best like $D(M) \sim M/P_{\text{tot}}(M)$, which can be achieved via separately designed source and channel codes. Hence, it appears that for such scenarios, a (scaling-law) separation theorem holds.
- *Refining sensor network (with poor communication infrastructure):* If the number of degrees of freedom in the source network increases *slowly* with the number of sensors, then
  — If the sensor observations are noiseless, a conclusive answer appears more difficult to obtain in general. However, as the example presented in "Camera Sensor Network" suggests, separately designed source and channel codes incur a polynomial gap, and hence, a slightly weaker form of a scaling-law separation theorem may apply. In our example, we showed that a distortion that behaves as $D_{\text{digital}}(M) \sim 1/(P_{\text{tot}}(M))^{\tilde{\gamma}}$ is achievable, and that the distortion may scale no better than $D_{\text{lowerbound}}(M) \sim 1/(MP_{\text{tot}}(M))^{\gamma}$. If we assume that $P_{\text{tot}}(M) \sim M$ and that $\gamma \ll 1$ (as in the specific example in "Camera Sensor Network"), then these two bounds are not far apart from each other, suggesting that an approximate (scaling-law) separation theorem applies.
  — If the sensor observations are subject to noise, then no (scaling-law) separation theorem seems to apply. Indeed, the example discussed in "Digital Architectures for Sensor Network Example" and "Analog Architecture for Sensor Network Example" shows that a distortion that behaves like $D \sim 1/(MP_{\text{tot}}(M))$ is optimal, but separately designed source and channel codes only achieve $D_{\text{digital}}(M) \sim 1/\log(MP_{\text{tot}}(M))$.

When separation does not hold, the exponential gap between the two architectures points to a vast space for new, creative designs. Are there multiuser joint source-channel codes that could reap some of that exponential gain? These are certainly among the most intriguing research challenges for WSNs.

ACKNOWLEDGMENTS

The difficulty of network information theory and its relative virginity is nicely conveyed in the title of [55], and we acknowledge gratefully this influence on our title. P.L. Dragotti would like to thank N. Gehrig for helping with the preparation of Figure 5 and for providing Figure 8. M. Vetterli would like to thank T. Ajdler for interaction on the plenacoustic function and providing Figure 5. The material in this article was supported in part by the NSF Award CCF-0347298, the Swiss National Science Foundation in the framework of the National Competence Center in Research for Mobile Information and Communication Systems (http://mics.org), and the UK Defence Technology Centre DIF-DTC (project 12.6.2).

AUTHORS

Michael Gastpar (gastpar@eecs.berkeley.edu) received his engineering degree from ETH in Zurich in 1997, his M.S. from the University of Illinois at Urbana-Champaign in 1999, and his Ph.D. from EPFL in Lausanne in 2002. He was also student in engineering and philosophy at the universities of Edinburgh and Lausanne, and a summer researcher in the Mathematics of Communications Department at Bell Labs, Lucent Technologies, Murray Hill, New Jersey. He is now an assistant professor at the University of California, Berkeley. His research interests are in network information theory and related coding and signal processing techniques, with applications to sensor networks and neuroscience. He won the 2002 EPFL Best Thesis Award and an NSF CAREER award in 2004. He is a Member of the IEEE.

Martin Vetterli (martin.vetterli@epfl.ch) received his engineering degree from ETH in Zurich, his M.S. from Stanford University, and his Ph.D. from EPFL in Lausanne. In 1986, he joined Columbia University as an associate professor of electrical engineering. In 1993, he joined the University of California at Berkeley, were he was a professor. Since 1995, he has been a professor at EPFL, where he headed the Communication Systems Division and now heads the Audiovisual Communications Laboratory. He is also a vice president for international affairs. From 2001 to 2004 he directed the National Competence Center in Research on mobile information and communication systems. His research interests are in the areas of applied mathematics, signal processing, and communications. He is a Fellow of the IEEE.

Pier Luigi Dragotti (p.dragotti@imperial.ac.uk) is a lecturer in the Electrical and Electronic Engineering Department at Imperial College, London. He received the Laurea in electrical engineering from the University Federico II, Naples, in 1997, the Master degree in Communications Systems from EPFL, Lausanne in 1998, and the Ph.D. degree from EPFL in 2002. He was a visiting student at Stanford University and a summer researcher in the Mathematics of Communications Department

at Bell Labs, Lucent Technologies. His research interests include wavelet theory, image and video processing and compression, joint source-channel coding, and signal processing for sensor networks. He is a Member of the IEEE.

REFERENCES

[1] C.E. Shannon, "A mathematical theory of communication," *Bell Syst. Tech. J.*, vol. 27, pp. 379–423, 623–656, July and Oct. 1948.

[2] M. Çetin, L. Chen, J.W. Fisher III, A.T. Ihler, R.L. Moses, M.J. Wainwright, and A.S. Willsky, "Distributed fusion in sensor networks: A graphical models perspective," *IEEE Signal Processing Mag.*, vol. 23, no. 4, pp. 42–55, 2006.

[3] J.B. Predd, S.R. Kulkarni, and H.V. Poor, "Distributed learning in wireless sensor networks," *IEEE Signal Processing Mag.*, vol. 23, no. 4, pp. 56–69, 2006.

[4] J. Xiao, A. Ribeiro, Z. Luo, and G.B. Giannakis, "Distributed compression-estimation using wireless sensor networks," *IEEE Signal Processing Mag.*, vol. 23, no. 4, pp. 26–41, 2006.

[5] B. Chen, L. Tong, and P.K. Varshney, "Channel aware distributed detection in wireless sensor networks," *IEEE Signal Processing Mag.*, vol. 23, no. 4, pp. 16–25, 2006.

[6] E.B. Ermis, M. Alanyali, and V. Saligrama, "Search and discovery in an uncertain networked world," *IEEE Signal Processing Mag.*, vol. 23, no. 4, pp. 107–118, 2006.

[7] Q. Zhao, A. Swami, and L. Tong, "The interplay between signal processing and networking in sensor networks for military applications," *IEEE Signal Processing Mag.*, vol. 23, no. 4, pp. 84–93, 2006.

[8] M. Unser, "Sampling—50 years after Shannon," *Proc. IEEE*, vol. 88, pp. 569–587, Apr. 2000 [Online]. Available: http://bigwww.epfl.ch/publications/unser0001.html

[9] T.M. Cover and J.A. Thomas, *Elements of Information Theory*. New York: Wiley, 1991.

[10] M. Gastpar and M. Vetterli, "On the capacity of wireless networks: The relay case," in *Proc IEEE Infocom 2002*, New York, June 2002, pp. 1577–1586.

[11] M. Gastpar and M. Vetterli, "On the capacity of large Gaussian relay networks," *IEEE Trans. Inform. Theory*, vol. IT–51, pp. 765–779, Mar. 2005.

[12] P. Gupta and P.R. Kumar, "The capacity of wireless networks," *IEEE Trans. Inform. Theory*, vol. IT–46, pp. 388–404, Mar. 2000.

[13] Y. Oohama, "The rate-distortion function for the quadratic Gaussian CEO problem," *IEEE Trans. Inform. Theory*, vol. IT–44, pp. 1057–1070, May 1998.

[14] T.M. Cover, A.A.E. Gamal, and M. Salehi, "Multiple access channels with arbitrarily correlated sources," *IEEE Trans. Inform. Theory*, vol. 26, pp. 648–657, Nov. 1980.

[15] M. Gastpar, "To code or not to code," Ph.D. dissertation, Ecole Polytechnique Fédérale (EPFL), Lausanne, Switzerland, 2002.

[16] J. Barros and S.D. Servetto, "Network information flow with correlated sources," *IEEE Trans. Inform. Theory*, vol. 52, no. 1, pp. 155–170, Jan. 2006.

[17] D.H. Johnson and D.E. Dugeon, *Array Signal Processing: Concepts and Techniques*. Englewood Cliffs, NJ: Prentice Hall, 1993.

[18] T. Ajdler, L. Sbaiz, and M. Vetterli, "The plenacoustic function and its sampling," *IEEE Trans. Signal Processing*, to be published.

[19] E. Adelson and J. Bergen, "The plenoptic function and the elements of early vision," in *Computational Models of Visual Processing*. Cambridge, MA: MIT Press, 1991, pp. 3–20.

[20] J. Chai, X. Tong, S. Chan, and H. Shum, "Plenoptic sampling," in *Proc. Conf. Computer Graphics*, 2000, pp. 307–318.

[21] B. Beferull-Lozano, R.L. Konsbruck, and M. Vetterli, "Rate-distortion problem for physics based distributed sensing," in *Proc. 3rd Int. Symp. Information Processing Sensor Networks (IPSN'04)*, Berkeley, CA, Apr. 2004, pp. 330–339.

[22] R.L. Konsbruck, E. Telatar, and M. Vetterli, "On the multiterminal rate-distortion function for acoustic sensing," in *Proc. IEEE Int. Conf. Acoustic Speech, Signal Processing*, Toulouse, France, May 2006.

[23] D. Slepian and J.K. Wolf, "Noiseless coding of correlated information sources," *IEEE Trans. Inform. Theory*, vol. IT–19, pp. 471–480, 1973.

[24] A. Wyner and J. Ziv, "The rate-distortion function for source coding with side information at the decoder," *IEEE Trans. Inform. Theory*, vol. 22, pp. 1–10, Jan. 1976.

[25] S. Pradhan and K. Ramchandran, "Distributed source coding using syndromes (DISCUS): Design and construction," in *Proc. Data Compression Conf.*, Snowbird, Utah, Mar. 1999, pp. 158–167.

[26] S. Pradhan and K. Ramchandran, "Distributed source coding: Symmetric rates and applications to sensor network," in *Proc. Data Compression Conf.*, Snowbird, Utah, Mar. 2000, pp. 363–372.

[27] S. Pradhan and K. Ramchandran, "Distributed source coding using syndromes (DISCUS): Design and construction," *IEEE Trans. Inform. Theory*, vol. 48, no. 3, pp. 626–643, Mar. 2003.

[28] J. Garcia-Frias and Y. Zhao, "Compression of correlated binary sources using turbo codes," *IEEE Commun. Lett.*, vol. 5, pp. 417–419, Oct. 2001.

[29] A. Aaron and B. Girod, "Compression with side information using turbo codes," in *Proc. Data Compression Conf.*, Snowbird, Utah, Mar. 2002, pp. 252–261.

[30] A. Liveris, Z. Xiong, and C. Georghiades, "Compression of binary sources with side information at the decoder using ldpc codes," *IEEE Commun. Lett.*, vol. 6, pp. 440–442, Oct. 2002.

[31] V. Stanković, A. Liveris, Z. Xiong, and N. Georghiades, "Design of Slepian-Wolf codes by channel code partitioning," in *Proc. Data Compression Conf.*, Snowbird, Utah, Mar. 2004, pp. 302–311.

[32] N. Gehrig and P. Dragotti, "Symmetric and asymmetric Slepian-Wolf codes with systematic and non-systematic linear codes," *IEEE Commun. Lett.*, vol. 9, no. 1, pp. 61–63, Jan. 2005.

[33] R. Puri, A. Majumdar, P. Ishwar, and K. Ramchandran, "Distributed video coding in broadband wireless sensor networks," *IEEE Signal Processing Mag.*, vol. 23, no. 4, pp. 94–105, 2006.

[34] M. Gastpar, P. Dragotti, and M. Vetterli, "The distributed Karhunen-Loève transform," in *Proc. Multimedia Signal Processing Workshop*, Virgin Islands, Dec. 2002, pp. 57–60.

[35] M. Gastpar, P. Dragotti, and M. Vetterli, "The distributed Karhunen-Loève transform," submitted for publication.

[36] H. Nurdin, R.R. Mazumdar, and A. Bagchi, "On estimation and compression of distributed correlated signals with incomplete observations," in *Proc. Mathematical Theory Networks Systems (MTNS 2004)*, Leuven, Belgium, July 2004.

[37] T. Han, "Cover-Slepian-Wolf theorem for a network of channels," *Inf. Control*, vol. 47, no. 1, pp. 67–83, 1980.

[38] J.-J. Xiao and Z.-Q. Luo, "Multiterminal source-channel communication under orthogonal multiple access," in *Proc. Alerton Conf. Communication, Control, Computing*, Urbana, Illinois, Sept. 2005, pp. 1247–1256.

[39] D. Marco, E.J. Duarte-Melo, M. Liu, and D.L. Neuhoff, "On the many-to-one transport capacity of a dense wireless sensor network and compressibility of its data," in *Proc. 2nd Int. Workshop Information Processing Sensor Networks (IPSN'03)*, Apr. 2003, pp. 1–16.

[40] D. Marco and D. Neuhoff, "Reliability versus efficiency in distributed source coding for field-gathering sensor networks," in *Proc. 3rd Int. Symp. Information Processing Sensor Networks (IPSN'04)*, Berkeley, California, Apr. 2004, pp. 161–168.

[41] P. Ishwar, A. Kumar, and K. Ramchandran, "Distributed sampling for dense sensor networks: A bit-conservation principle," in *Proc. 2nd Int. Workshop Information Processing Sensor Networks (IPSN'03)*, Apr. 2003, pp. 17–31.

[42] A. Kashyap, L. Lastras-Montaño, C. Xia, and Z. Liu, "Distributed source coding in dense sensor networks," in *Proc. Data Compression Conf.*, Snowbird, Utah, Mar. 2005, pp. 13–22.

[43] A. Wyner, "Recent results in the Shannon theory," *IEEE Trans. Inform. Theory*, vol. 20, pp. 2–10, Jan. 1974.

[44] T. Berger, "Multiterminal source coding," in *Proc. Lectures CISM Summer School Information Theory Approach Communications*, July 1977, pp. 171–231.

[45] T. Cover, "A proof of the data compression theorem of Slepian and Wolf for ergodic sources," *IEEE Trans. Inform. Theory*, vol. 21, pp. 226–228, Mar. 1975.

[46] T. Cover and J. Thomas, *Elements of Information Theory*. New York: Wiley, 1991.

[47] K. Eswaran and M. Gastpar, "On the significance of binning in a scaling-law sense," in *Proc. IEEE Information Theory Workshop*, Punta del Este, Uruguay, Mar. 2006, pp. 258–262.

[48] M. Vetterli, P. Marziliano, and T. Blu, "Sampling signals with finite rate of innovation," *IEEE Trans. Signal Processing*, vol. 50, no. 6, pp. 1417–1428, June 2002.

[49] N. Gehrig and P. Dragotti, "Different-distributed and fully flexible image encoders for camera sensor networks," in *Proc. IEEE Int. Conf. Image Processing (ICIP)*, Genova, Italy, Sept. 2005, pp. 690–693.

[50] N. Gehrig and P. Dragotti, "Distributed sampling and compression of scenes with finite rate of innovation in camera sensor networks," in *Proc. Data Compression Conf.*, Snowbird, Utah, Mar. 2006, pp. 83–92.

[51] T. Berger, Z. Zhang, and H. Viswanathan, "The CEO problem," *IEEE Trans. Inform. Theory*, vol. IT–42, pp. 887–902, May 1996.

[52] Y. Oohama, "Gaussian multiterminal source coding," *IEEE Trans. Inform. Theory*, vol. IT–43, pp. 1912–1923, Nov. 1997.

[53] K. Eswaran, "Remote source coding and AWGN CEO problems," Master's thesis, UC, Berkeley, 2005.

[54] M. Gastpar and M. Vetterli, "Power, spatio-temporal bandwidth, and distortion in large sensor networks," *IEEE J. Select. Areas Commun.*, vol. 23, pp. 745–754, Apr. 2005.

[55] A. Ephremides and B. Hajek, "Information theory and communication networks: an unconsummated union," *IEEE Trans. Inform. Theory*, vol. IT–44, pp. 2416–2434, Oct. 1998.

SP