Joint Source-Channel Coding on Graphs and Applications in Sensor Networks

Răzvan Cristescu* and Martin Vetterli†,‡

* Center for Mathematics of Information, Caltech, Pasadena CA 91125, USA

† School of IC, EPFL, Lausanne CH-1015, Switzerland

‡ Dept. of EECS, UC Berkeley, CA 94720, USA

razvanc@caltech.edu, martin.vetterli@epfl.ch

Abstract

We consider the interaction between the physical representation of information in a correlated data field, and the way that information is transmitted to a central processor. Namely, we study data gathering to a sink node, by considering a set of relevant metrics: first we consider energy efficiency, where the goal is to minimize the total transmission cost of transporting the information collected by the nodes, to the sink node, and second we consider mean-square-error (MSE) distortion, where the goal is to represent the information field with a maximal accuracy. The first problem requires a joint optimization of the data representation at the nodes and of the transmission structure; we study both the lossless scenario with Slepian-Wolf coding, and the high-resolution lossy scenario with optimal rate-distortion allocation. We show that the optimal transmission structure is the shortest path tree, and we find in closed-form the rate and distortion allocation. The second problem is relevant for data gathering of spatio-temporal correlated processes under delay constraints, where we show that both the temporal and the spatial distortion can be combined in a single measure of distortion, which depends on the density of the network. We prove that, for various standard correlation structures, there is an optimal *finite* density of the sensor network for which the total distortion is minimized.

The results in this chapter have been presented in part at the 23rd Conference of the IEEE Communications Society (INFOCOM 2004), at the 4th ACM International Symposium on Mobile Ad Hoc Networking and Computing (MobiHoc 2003), at the 3rd and 4th International Symposiums on Information Processing in Sensor Networks (IPSN'04, '05), and at the 2004 IEEE International Symposium on Information Theory (ISIT 2004).

CONTENTS

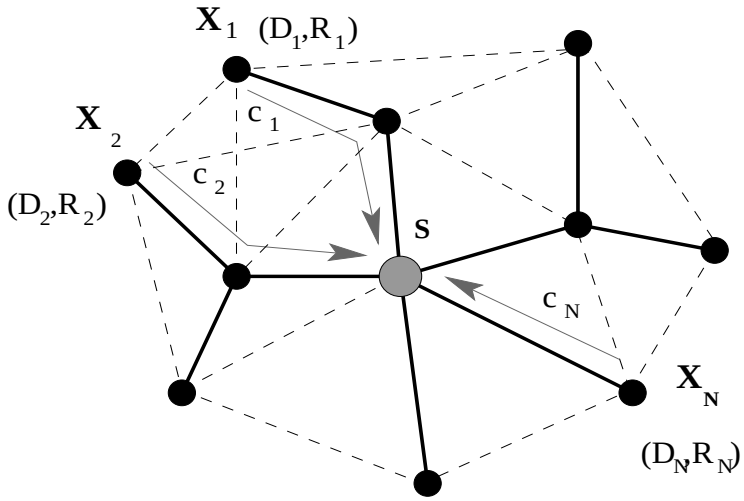


Fig. 1. In this example, data from nodes $X_1, X_2, \ldots, X_N$ need to arrive at sink S. A rate supply R_i is allocated to each node X_i, and, in the case of lossy coding, the distortion at that node is D_i. In thick solid lines, a chosen tree transmission structure is shown. In thin dashed lines, the other possible links are shown. The path from node i to the sink is shown as gray line, and its weight is c_i.

I. Introduction

A. Motivation

Consider a typical sensor network scenario [30], where sensors measure a data field (e.g. temperature) and the results of their measurements have to be transported across the network, to a certain designated node called the *sink* (see Figure 1). This is referred to as *data gathering*, and it is a relevant problem in various sensor network settings, where information from the network, in its coded form, is needed at a central base station node, for storage, monitoring or control purposes.

We address this problem from a joint source-channel coding perspective, corresponding to the network layer. Namely, given the statistical structure of the correlated data field, we study the interactions between the representation of the measured information, and the network over which this information needs to be transmitted. More specifically, the information is represented by means of the rate allocation employed for data coding at measuring nodes, and the actual placement in the field of those nodes. The transmission network is represented by its graph structure, formed by the nodes inter-connections, on which the information is transmitted to

the central processor that needs to reconstruct the data field. For quantifying these interactions, we consider a set of relevant metrics: energy used for the transmission and mean-square-error (MSE) distortion of the reconstruction. On one hand, the *source* to be coded is the correlated data field, and the task of the coder is twofold: first, to properly sample the correlated data field so that the result is a representation with a good accuracy, and second, to allocate coding rates at the nodes such that data can be reconstructed at the sink. Thus, in our setting, the *source coder* aims at a proper representation of information at the network nodes, as a function of the characteristics of the measured field, and of the nodes positions with respect to each other. On the other hand, the *channel* is the connectivity network formed by the nodes placed in the field. The task of the *channel coder* is to find a proper transmission structure (i.e. a subset of the edges of the connectivity graph) and/or a node placement that optimizes the metric of interest. Both the source and channel coding tasks have to be done under various restrictions, including energy and communication constraints[1].

B. Measured Data and Network Characteristics

There are several important issues specific to sensor networks measuring and transporting data [30]. First, the measured data have certain redundancy characteristics. For instance, if the measured data are random variables (e.g. temperature), the values at nodes are correlated and the data structure is given by the spatio-temporal correlation.

Second, the limited coverage and transmission capabilities of the sensor nodes induce limited connectivity and communication patterns in the network graph. Nodes usually have knowledge only about other sensors situated in a limited neighborhood, so efficient joint representation of data by groups of nodes has to be done in a decentralized manner. Also, due to the battery energy limitations, most nodes cannot send their data directly to the sink, and therefore data has to be relayed by other nodes. This implies that efficient routing is necessary, and moreover it has to be decentralized as well. Also, depending on the coding strategy that determines the amount of internode communication, the task of data representation at nodes may or may not separate from the task of routing that data across the network.

[1]Implicitly, we are using the assumption that the network is a bit-pipe, and that sources will be represented in a coded digital form. This level of abstraction makes the problem tractable, but can be suboptimal in the wireless case, as demonstrated in [19].

Third, the actual node positions influence both the accuracy of the measured data and the energy efficiency of the network. For instance, placing most nodes close to the sink will improve their lifetime since they only have to transmit data over small distances; however, this will leave areas that are far from the sink uncovered, which means high inaccuracy in the overall data measurement. On the contrary, an even distribution of nodes over the measured field will provide a good data accuracy, but at the same time, the energy consumption is large. Moreover, finding an optimized tradeoff has to take into account both data representation and routing.

Finally, in the case of data gathering of spatio-temporal processes under delay constraints, the network density influences the total distortion of reconstruction. Namely, a network with a small number of nodes results in a high spatial distortion of approximation, but the temporal distortion is small since data reaches the sink in a small number of hops. On the contrary, a network with a large density has a low spatial distortion, but the temporal distortion is high due to the large number of hops to the sink.

C. Metrics

There are certain specific metrics of interest for this type of applications, namely energy efficiency and accuracy of the data reconstruction at the sink. In sensor networks, the energy efficiency of the network depends on both the rate allocation at nodes and on the routing strategy (the paths chosen to transmit the data). Namely, the energy consumed by a node is usually proportional to the product [rate] $\times$ [path weight], where the [rate] term represents the data amount (in bits) sent by a node, and the [path weight] is an increasing function of the euclidean distance between nodes (for instance, [euclidean distance]$^{\kappa}$, where $\kappa \in [2, 4]$ is the path-loss exponent).

The accuracy of data reconstruction depends on the distortion allowed at measuring nodes and on the node placement, and influences the data representation as well. Namely, the desired accuracy of data representation at nodes determines the rate necessary to accommodate the corresponding distortion, and thus how much energy is needed to transmit that rate. Due to spatial representation reasons, the accuracy is influenced by the node placement as well. Moreover, for spatio-temporal processes, delay represents an additional issue that affects the accuracy of the reconstruction, since timely arrival of data reduces the total distortion.

To summarize, there is a strong interaction between data representation and routing, and the

metrics relevant for sensor network scenarios. The goal of this work is to study the interaction among these important parameters, for designing practical, efficient and accurate joint measurement (source coding) and transmission (channel coding) strategies. We will show that a joint consideration of the source and channel coding can result in significant improvements in terms of the metrics of interest.

D. Related Work

Bounds on the performance of networks measuring correlated data have been derived in [26], [33]. Progress towards practical implementation of Slepian-Wolf coding [35] has been achieved in [1], [31], [32]. However, none of these works takes into consideration the cost of transmitting the data over the links and the additional constraints that are imposed on the rate allocation by the joint treatment of source coding and transmission.

The problem of optimizing the transmission structure in the context of sensor networks has been considered in [24], [21], where the [energy], and the [energy] $\times$ [delay] metric are studied, and practical algorithms are proposed. But in these studies, the correlation present in the data is not exploited for the minimization of the metric.

Recent work that exploit correlation in the data in the context of sensor networks include [2], [15], [28].

In [28] an empirical data correlation model is used for a set of experimentally obtained data, and the authors propose cluster-based tree structures shown to have a good performance depending on the correlation level. The correlation function is derived as an approximation of the conditional entropy, and the cost function is the sum of bits transmitted by the network.

In [15], a circular-coverage correlation model on a grid is used, where correlation is modelled as a parameter proportional to the area covered by a sensor. The authors provide randomized shortest-path aggregation trees with constant-ratio approximations.

Some examples of network flow with joint coding of correlated sources under capacity constraints on the transmission links and Slepian-Wolf constraints on the rates are studied in [2], where trees are shown to perform suboptimally if splittable flows are allowed.

We note that in some scenarios, uncoded transmission is optimal [18].

A joint treatment of data aggregation and transmission structure is considered in [20], but the model does not take into account possible collaborations for joint coding among nodes. We

consider scenarios where joint source coding among nodes exploits inter-node correlations by means of Slepian-Wolf coding, and compare this approach with coding by explicit communication, where such collaborations are not allowed, and coding is rather done by using only available side information.

The rate-distortion region of coding with high-resolution for arbitrarily correlated sources has been found in [37]. We focus on finding the optimal rate-distortion operation point when additionally, energy constraints are imposed.

An analysis of the impact of data irregularity on the spatio-temporal sampling is done in [17]. Our novel take on the problem of data gathering of spatio-temporal processes is that we are able to formulate the problem in terms of a unique performance measure, namely the total distortion.

E. Main Contributions and Organization of the Chapter

The main contribution of this work is an unified treatment of data representation, routing and node placement in sensor networks, for the optimization of various metrics of interest. First, we consider energy efficiency, and we show that for data gathering of spatially correlated processes, the task of data representation at nodes separates from the transmission structure optimization. By using this separation result, we are able to find the optimal transmission structure, and the corresponding rate-distortion allocation. Second, we consider the distortion given by the total MSE of representation of the field at the sink, for energy efficient data gathering, and which includes spatial and temporal distortions. We show that for spatio-temporally processes, in general there is an optimal network density that minimizes this distortion.

In Section II we review some of the main concepts and methods used in this chapter. In Section III we present the network, transmission and signal models analyzed in this work. Section IV studies data gathering of spatially correlated but temporally i.i.d. random processes, namely the cases of lossless and high-resolution lossy coding, and addresses the node placement problem. Section V studies data gathering of spatio-temporally correlated processes: a performance measure of interest is defined, namely the total distortion, an analysis of this measure for the one-dimensional grid is performed, and the model is generalized for a two-dimensional grid. We conclude with Section VI.

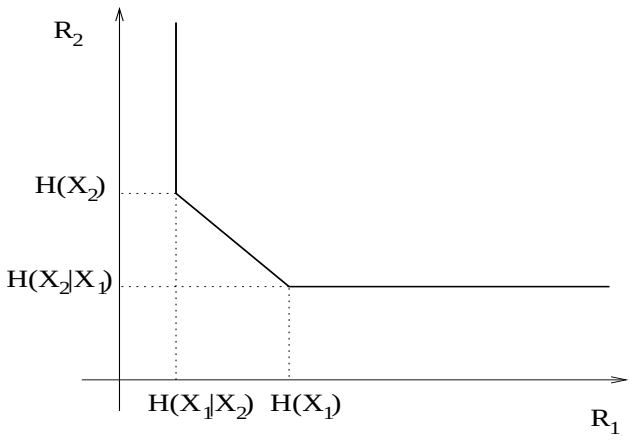


Fig. 2. The Slepian-Wolf region for two correlated sources X_1 and X_2.

II. BACKGROUND

In this section we review some of the main concepts used in this chapter.

Denote by $H(X)$ the entropy of a discrete random variable X. The entropy is a measure of uncertainty of a random variable [6]: $H(X) = -\sum_{x\in\mathcal{X}} p(x)\log p(x)$, where $\mathcal{X}$ is the discrete alphabet of X. Further, denote by $H(X|Y)$ the conditional entropy of a random variable X given that the random variable Y is known.

Consider the problem of lossless coding of two random sources X_1 and X_2 that are correlated. Intuitively each of the sources can code their data at a rate equal to at least their corresponding entropies, $R_1 = H(X_1)$, $R_2 = H(X_2)$, respectively. If they are able to communicate, then they could coordinate their coding and use together a total rate of $R_1+R_2 = H(X_1, X_2)$. Slepian and Wolf [35] showed that two correlated sources can be coded with a total rate equal to the joint entropy $H(X_1, X_2)$ even if they are *not* able to communicate with each other, as long as their individual rates are at least equal to the conditional entropies, $H(X_1|X_2)$ and $H(X_2|X_1)$ respectively (see Figure 2); this easily generalizes to the N-dimensional case.

When coding is done in a lossy manner, that is the sources are coded under distortion constraints, then the problem of finding the rate-distortion region becomes difficult [3], and for most scenarios it remains open. However, a recent result [37] finds the rate-distortion region for high resolution coding. The main idea of the proof in [37] is that at high rates, quantization followed by Slepian-Wolf coding is optimal. In the case of two sources coded with

high resolution, the rate-distortion region is similar to the Slepian-Wolf region, with the addition of terms related to the distortions D_1, D_2:

$$\begin{aligned} R_1 &\geq h(X_1|X_2) - \log 2\pi e D_1, \\ R_2 &\geq h(X_2|X_1) - \log 2\pi e D_2, \\ R_1 + R_2 &\geq h(X_1, X_2) - \log(2\pi e)^2 D_1 D_2, \end{aligned}$$

where $h(X)$ is the differential entropy of a continuous amplitude random variable [6].

Consider now the case where sources have to be sent over a transmission channel, to a certain destination, or receiver. One important problem in this case is *if* and *how* to adapt the characteristics of the source to those of the channel. This is called joint source-channel coding, and it is a subject that has been extensively studied in the literature (e.g. [4], [27]). Namely, an important result of information theory states that in the point-to-point case (a single source transmitting to a single receiver), and for infinitely long block-lengths, the separation of the source coder from the channel coder is optimal [34], [36]. However, in the multiuser case, which includes the scenario discussed in this chapter, separation of the source and channel coding may no longer be optimal. The results in this chapter reiterate this paradigm: a joint consideration of the source coder (rate and distortion allocation at the measuring nodes), and of the channel coder (the actual transmission structure used to transport the measured information to the destination) does improve the performance of the system, in terms of both energy used for communication and accuracy of the reconstruction.

III. Problem Setting

A. Network Model

Consider a network of N nodes. Let $\mathbf{X} = (X_1, \ldots, X_N)$ be the vector formed by the values representing the sources measured at nodes $1, \ldots, N$. The information measured at nodes has to be transmitted through the links of the network to the designated base station (see Figure 1). We will assume that the interference among nodes is negligible, and there are no capacity constraints on the links (the case of omnidirectional interfering wireless channels is beyond the scope of this study). Such assumptions are realistic in the case of wired networks or if the antennas are unidirectional. For such scenarios, the optimal gathering structure is a tree.

Fig. 3. One-dimensional grid network.

In some parts of this work, for the sake of simplicity, we use the one-dimensional network model in Figure 3 rather than the two-dimensional model in Figure 1.

B. Signal Model

1) Spatially Correlated Gaussian Random Field: We will consider first the case where $\mathbf{X} = (X_1, \ldots, X_N)$ is a vector formed by random variables representing the sources measured at nodes $1, \ldots, N$. The samples taken at nodes are spatially correlated and independent in time. We assume that the random variables are continuous and that there is a high-resolution quantizer in each sensor. In the lossless source coding case, a rate allocation $\{R_i\}_{i=1}^N$ (bits) has to be assigned at the nodes. In addition, if the data can be transmitted in a lossy manner, a distortion allocation $\{D_i\}_{i=1}^N$ has to be assigned as well, so that the quantized measured information samples are described with certain total distortion D and individual $\{D_i^{\max}\}_{i=1}^N$ distortion constraints.

For the sake of clarity, we use as an example a zero-mean *jointly Gaussian model* $\mathbf{X} \sim \mathcal{N}^N(0, \mathbf{K})$, with unit variances $\sigma_{ii} = 1$:

$$f(\mathbf{X}) = \frac{1}{\sqrt{2\pi}\det(\mathbf{K})^{1/2}} e^{-\left(\frac{1}{2}(\mathbf{X})^T \mathbf{K}^{-1}(\mathbf{X})\right)},$$

where $\mathbf{K}$ is the covariance matrix of $\mathbf{X}$, with elements depending on the distance between the corresponding nodes (e.g. $K_{ij} = \exp(-c d_{i,j}^{\beta}), c > 0, \beta \in \{1, 2\}$, where $d_{i,j}$ is the distance between nodes i and j [7], [9]). Although we will show numerical evaluations performed using the Gaussian random field model, our results are valid for any spatially correlated random processes, whose correlation decreases with distance.

2) Spatio-Temporally Correlated Gaussian Random Field: Further, we will consider spatio-temporal correlated processes $X(x,t)$, where x denotes the space dimension, and t denotes the time dimension. Our model for spatio-temporal processes is a generalization of the spatial model introduced in the previous Section III-B.1, by considering time as an additional dimension. Namely, we further assume that the process measured by the field is Gaussian distributed: each node measures a zero-mean and unit variance normal random variable $X(x,t) \approx \mathcal{N}(0,1)$, which is correlated both in space and in time with the rest of the network nodes. We consider correlation structures of the form

$$\begin{aligned} E\left[X(x_1,t_1)X(x_2,t_2)\right] &= \sigma_{X(x_1,t_1),X(x_2,t_2)} \\ &= \sigma(|x_1-x_2|,|t_1-t_2|) \\ &= e^{-c((x_1-x_2)^2+\gamma^2(t_1-t_2)^2)^{\frac{\beta}{2}}}, \end{aligned} \tag{1}$$

with γ the scaling constant for the time axis, and $\beta = 1$ corresponding to a Gauss-Markov field or $\beta = 2$ corresponding to a squared distance correlation model, and c a constant that measures the intensity of correlation [7].

IV. Data Gathering of Spatially Correlated Random Processes

A. Lossless Data Gathering

Consider data gathering of random processes. Denote by $\{R_i\}_{i=1}^N$ the rate allocation at nodes, by ST an arbitrary spanning tree of G, and by c_i the total weight of the path connecting node i to S on the spanning tree ST. For a given network with connectivity graph $G = (V,E)$, we formulate our problem as follows (see Figure 1):

$$\{\{R_i^*, c_i^*\}_{i=1}^N, ST^*\} = \arg\min_{R_i,c_i,ST} \sum_{i\in V} R_i c_i \tag{2}$$

under constraints

$$\sum_{i\in\mathcal{X}} R_i \geq H(\mathcal{X}|\mathcal{X}^C), \forall \mathcal{X} \subset V, \tag{3}$$

where (3) are the Slepian-Wolf constraints on rates, for joint data representation at nodes.

*1) Optimal Transmission Structure is the Shortest Path Tree (*SPT*):* Constraints (3) imply that nodes can code with any rate that obeys the constraint region *without explicitly exchanging data*. As a consequence, we can state the following theorem [9]:

Theorem 1: (Separation of the joint optimization of source coding and transmission structure:)

The overall joint optimization (2) can be achieved by first optimizing the transmission structure with respect only *to the link weights c_i, and then optimizing the rate allocation for the given transmission structure under the constraints (3).*

Proof: By definition, for any given node, the cost function in (2) is separable as the product of a function that depends only on the rate allocated at that node, and another function that depends only on the link weights. Once the rate allocation is *fixed*, the best way (least cost) to transport any amount of data from a given node i to the sink S does not depend on the value of the rate R_i. Since this holds for any rate allocation, it is also true for the minimizing rate allocation and the result follows. ■

Theorem 1 implies that the optimal transmission structure that optimizes (2) is the shortest path tree, given by the superposition of the shortest paths from all nodes to the sink (ST^*=*SPT*):

Corollary 1: (Optimality of the shortest path tree (SPT) for the single-sink data gathering problem)

When there is a single sink S in the data gathering problem and Slepian-Wolf coding is used, the shortest path tree (SPT) rooted in S is optimal, in terms of minimizing (2), for any rate allocation.

Proof: The best way to transport the data from any node to the sink is to use the shortest path. Minimizing the sum of costs under constraints in (2) becomes equivalent to minimizing the cost corresponding to each node independently. Since the shortest path tree is a superposition of all the individual shortest paths corresponding to the different nodes, it is optimal for any rate allocation that does not depend on the transmission structure, which is the case here. ■

2) Rate Allocation: Denote by $c_i^* = d_{SPT}(i, S)$ the total weight of the path from node i to the sink S. For the rest of this section, suppose without loss of generality that nodes are ordered in a list with increasing values of the weights corresponding to the shortest paths from each node to the sink, that is, $c_1^* \leq c_2^* \leq \cdots \leq c_N^*$.

From Corollary 1, it follows that the minimization of (2) becomes now a linear programming (LP) problem:

$$\{R_i^*\}_{i=1}^N = \arg \min_{\{R_i\}_{i=1}^N} \sum_{i \in V} R_i c_i^*, \tag{4}$$

under constraints (3).

Theorem 2: (LP solution)

The solution of the optimization problem given by (4) under constraints (3) is [9]:

$$\begin{aligned} R_1^* &= H(X_1), \\ R_2^* &= H(X_2|X_1), \\ \dots &\quad \dots \quad \dots \\ R_N^* &= H(X_N|X_{N-1}, X_{N-2}, \dots, X_1). \end{aligned} \tag{5}$$

Theorem 2 is proven in [10].

Thus, for optimal rate allocation, nodes code by conditioning on all the other nodes that are closer to the sink on the SPT. In words, the solution of this problem is given by the corner of the Slepian-Wolf region that intersects the cost function in exactly one point. The node with the smallest total weight on the *SPT* to the sink is coded with a rate equal to its unconditional entropy. Each of the other nodes is coded with a rate equal to its respective entropy conditioned on all other nodes which have a total smaller weight to the sink than itself.

Figure 4 gives an example involving only two nodes, and it is shown how the cost function is indeed minimized with such a rate allocation. The assignment (5) corresponds in this particular case to the point $(R_1, R_2) = (H(X_1|X_2), H(X_2))$.

Note that if two or more nodes are equally distanced from the sink on the SPT (e.g. $d_{SPT}(X_1, S) = d_{SPT}(X_2, S)$, in Figure 4) then the solution of (5) is not unique, since the cost function is parallel to one of the faces of the Slepian-Wolf region, so any point on the face is an optimal solution.

For an illustration of Theorem 2, in Figure 5 we plot the typical rate allocation as a function of the node index, for a one-dimensional network as in Figure 3 and a Gaussian process as introduced in Section III-B.1.

Even if the solution can be provided in the closed form (5), a distributed implementation of the optimal algorithm at each node implies knowledge of the overall structure of the network (total weights between nodes and total weights from the nodes to the sink). This knowledge is needed for:

1) Ordering the total weights on the *SPT* from the nodes to the sink: each node needs its index in the ordered sequence of nodes in order to determine on which other nodes to condition

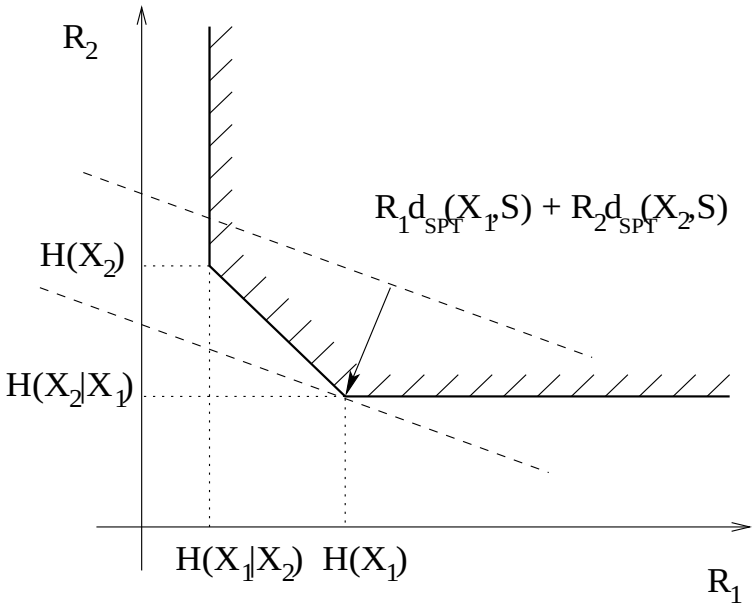


Fig. 4. A simple example with two nodes. The total weights from sources X_1, X_2 to the sinks, are respectively $d_{SPT}(X_1, S), d_{SPT}(X_2, S)$, $d_{SPT}(X_1, S) < d_{SPT}(X_2, S)$, in this particular case. In order to achieve the minimization, the cost line $R_1 d_{SPT}(X_1, S) + R_2 d_{SPT}(X_2, S)$ has to be tangent to the most interior point of the Slepian-Wolf rate region, given by $(R_1, R_2) = (H(X_1), H(X_2|X_1))$.

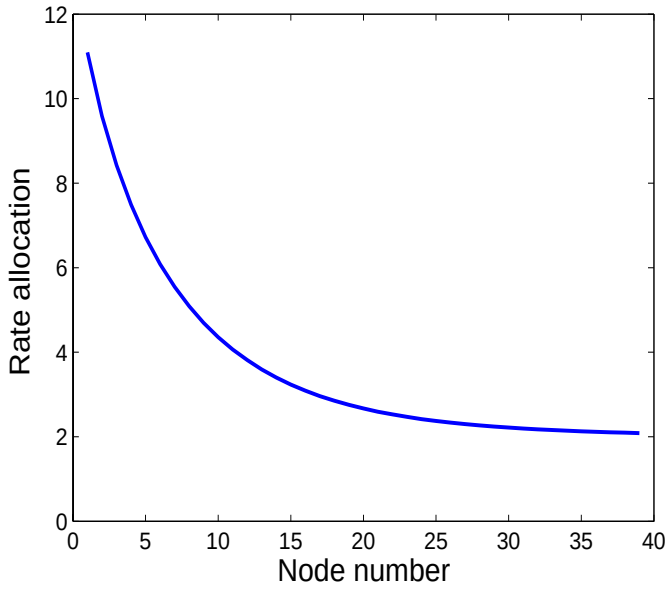


Fig. 5. Typical rate allocation as a function of the node index, for a one-dimensional network. The node index increases with the distance from the node to the sink.

when computing its rate assignment. For instance, it may happen that the distance on the graph between nodes X_i and X_{i-1} is large. Thus, closeness in the ordering on the *SPT* does not mean necessarily proximity in distance on the graph.

2) Computation of the rate assignment:

$$R_i = H(X_i|X_{i-1}, \ldots, X_1) = H(X_1, \ldots, X_i) - H(X_1, \ldots, X_{i-1}).$$

Note that for each node i we need to know locally *all* distances among the nodes $X_1, \ldots, X_i$, $i > 1$, in order to be able to compute this rate assignment, because the rate assignment involves a conditional entropy including all these nodes. However, for a static network, these distances can be calculated off-line at the deployment of the network, and as a result the optimal rate allocation can be computed at the beginning of the operation as well.

This implies that, for a distributed algorithm, global knowledge should be available at nodes, which might not be the case in a practical situation.

However, note that if the correlation decreases with distance, as it is usual in sensor networks, it is intuitive that each node i could condition only on a small neighborhood, incurring only a small penalty. In the next section, we propose a fully distributed heuristic approximation algorithm, which avoids the need for each node to have global knowledge of the network, and which provides solutions for the rate allocation which are very close to the optimum.

3) Heuristic Approximation Algorithm: So far, we found the optimal solution of the LP problem for the rate assignment under the Slepian-Wolf constraints. In this section, we consider the design of a distributed heuristic approximation algorithm for the case of single-sink data gathering.

Suppose each node i has complete information (distances between nodes and total weights to the sink) only about a local vicinity $\mathcal{N}_1(i)$ formed by its immediate neighbors on the connectivity graph G. All this information can be computed in a distributed manner by running for example a distributed algorithm for finding the *SPT* (e.g. Bellman-Ford [5]). By allowing a higher degree of (local) overhead communication, it is also possible for each node i to learn this information for a neighborhood $\mathcal{N}_k(i)$ of k-hop neighbors. The approximation algorithm we propose is based on the observation that nodes that are outside this neighborhood count very little, in terms of rate, in the local entropy conditioning.

This means that data are coded locally at the node with a rate equal to the conditional entropy,

Algorithm 1 Approximated Slepian-Wolf coding

for each node i **do**

Set the neighborhood range k (only k-hop neighbors).

Find the *SPT* using a distributed Bellman-Ford algorithm.

end for

for each node i **do**

Using local communication, obtain all the information from the neighborhood $\mathcal{N}_k(i)$ of node i.

Find the set $\mathcal{C}_i$ of nodes in the neighborhood $\mathcal{N}_k(i)$ that are closer to the sink, on the *SPT*, than the node i.

Transmit at rate $R_i^\dagger = H(X_i|X_j, j \in \mathcal{C}_i)$.

end for

where the conditioning is performed *only* on the subset $\mathcal{C}_i$ formed by the neighbor nodes which are closer to the sink than the respective node.

The proposed algorithm needs only local information, so it is completely distributed. For a given correlation model, depending on the reachable neighborhood range, this algorithm gives a solution close to the optimum since the neglected conditioning is small in terms of rate for a correlation function that decays sufficiently fast with the distance.

We present numerical simulations that show the performance of this approximation algorithm, for the case of single-sink data gathering. We consider the stochastic data model introduced in Section III-B.1, given by a multi-variate Gaussian random field, and a correlation model where the inter-node correlation decays exponentially with the distance between the nodes. More specifically, we use an exponential model of the covariance $K_{ij} = \exp(-cd_{i,j}^2)$. The weight of an edge (i, j) is $w_{i,j} = d_{i,j}^2$ and the total cost is given by expression (4). Figure 6 presents the average ratio of total costs between the Slepian-Wolf approximated solution using a neighborhood of $\mathcal{N}_1(i)$ for each node, and the optimal one. In Figure 7, we show a comparison of our different

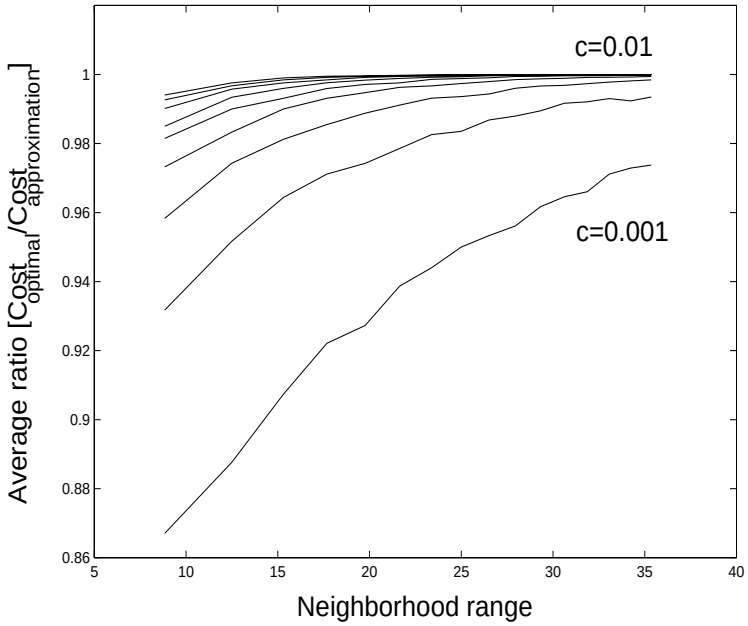


Fig. 6. Slepian-Wolf coding: average value of the ratio between the optimal and the approximated solution, in terms of total cost, vs. the neighborhood range. Every network instance has 50 nodes uniformly distributed on a square area of size 100×100, and the correlation exponent varies from $c = 0.001$ (high correlation) to $c = 0.01$ (low correlation). The average has been computed over 20 instances for each (c, radius) value pair.

approaches for the rate allocation, as a function of the distances from the nodes to the sink[2].

4) Scaling Laws: A Comparison between Slepian-Wolf Coding and Explicit Communication Based Coding : The alternative to Slepian-Wolf coding is coding by explicit communication, which is considered in [12], [14]. In this case, compression at nodes is done only using *explicit communication* among nodes, namely, a node can reduce its rate only when data from other nodes that use it as relay is available (as opposed to Slepian-Wolf coding where no communication among nodes is required for joint optimal rate allocation). The study of the complexity of joint rate allocation and transmission structure optimization with explicit communication can be found in [12], [14].

In this section, we compare the asymptotic behavior (large networks) of the total cost using Slepian-Wolf coding and the total cost with coding by explicit communication. The advantages

[2]Note that the slight increase in rate allocation with Slepian-Wolf coding for the furthest nodes from the sink is a boundary effect, namely nodes that are at the extremity of the square simulation area that we use have a smaller number of close neighbors on which to condition, as compared to nodes which are located at an intermediate distance from the sink.

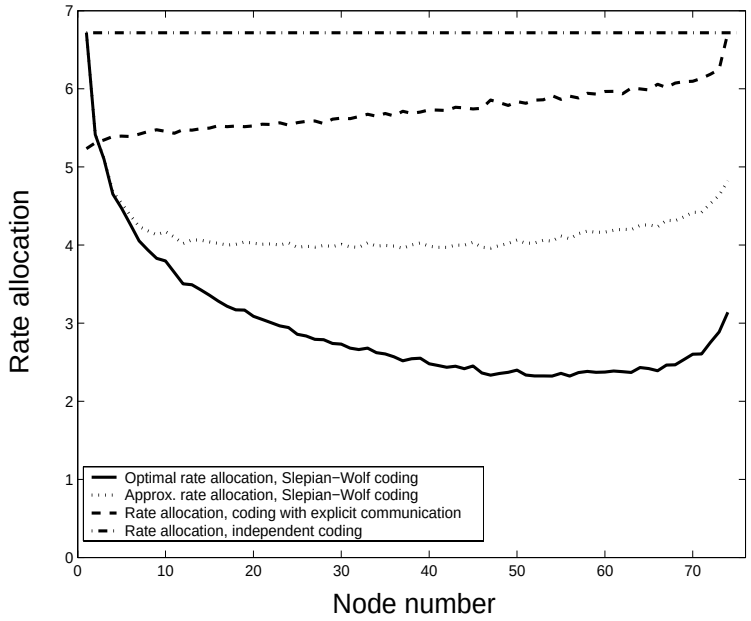


Fig. 7. Average rate allocation for 1000 network instances with 75 nodes, and a correlation exponent $c = 0.0008$ (strong correlation). On the x-axis, nodes are numbered in increasing order as the total weight from the sink increases, on the corresponding *SPT*.

that coding by explicit communication has over Slepian-Wolf coding are (i) no a-priori knowledge of the correlation structure is needed, and (ii) the compression, which is done by conditional encoding, is easily performed at the nodes relaying data. However, even for a simple one dimensional setting presented in this section, our analysis shows that in large networks, for some cases of correlation models and network scalability, Slepian-Wolf coding can provide very important gains over coding by explicit communication, in terms of total flow cost.

For the sake of simplicity in the analysis, we consider a one-dimensional network model where there are N nodes placed uniformly on a line (see Figure 8). The distance between two consecutive nodes is d. The nodes need to send their correlated data to the sink S.

For this scenario, the *SPT* is clearly the optimal data gathering structure for both coding approaches. Thus, the overall optimization problem (2) simplifies, and we can compare the two different rate allocation strategies in terms of how they influence the total cost.

Within the one-dimensional model, we consider two important cases of network scalability, namely, the *expanding network*, where the inter-node distance is kept constant and equal to $d = 1$ (that is, by increasing N we increase the distance between the node N and the sink S), and the

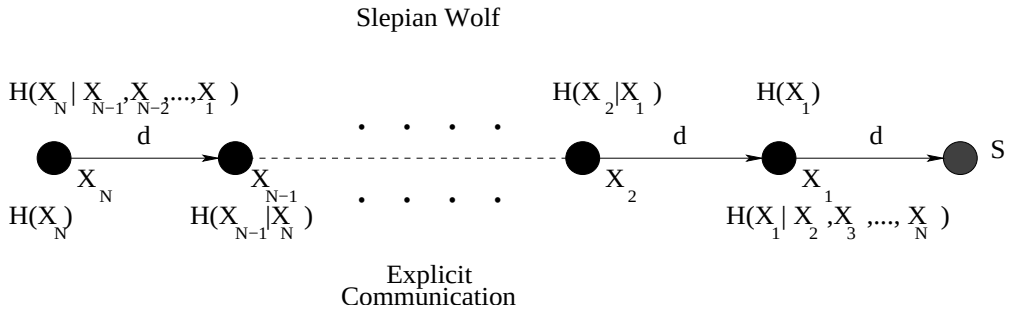


Fig. 8. A one dimensional example: The rate allocations for Slepian-Wolf (above the line) and explicit communication (below the line).

refinement network, where the total distance from node N to the sink S is kept constant, namely $Nd = 1$ (that is, nodes are uniformly placed on a line of length 1, and hence, by adding nodes, the inter-node distance goes to zero).

As mentioned in Section III-B.1, we consider that the nodes of the network are sampling a Gaussian continuous-space wide-sense-stationary (WSS) random process $X_c(s)$, where s denotes the position. Thus, we have a vector of correlated sources $\mathbf{X} = (X_1, \ldots, X_N)$ where $X_i = X_c(id)$ and where the correlation structure for the vector $\mathbf{X}$ is inherited from the correlation present in the original process $X_c(s)$. As N goes to infinity, the set of correlated sources represents a discrete-space random process denoted by $X_d(i)$, with the index set given by the node positions. Thus the spatial data vector $\mathbf{X}$ measured at the nodes has an N-dimensional multivariate normal distribution $G_N(\mu, \mathbf{K})$. In particular, we consider two classes of random processes:

(a) Non-bandlimited processes, namely (a.1): $K_{ij} = \sigma_{ij}^2 \exp(-c|d_{i,j}|)$, which corresponds to a regular continuous-space process [23], and (a.2): $K_{ij} = \sigma_{ij}^2 \exp(-c|d_{i,j}|^2)$, which corresponds to a singular continuous-space process [23], where $c > 0$.

(b) Bandlimited process with bandwidth B, that is, there exists a continuous angular frequency such that $S_{X_c}(\Omega) = 0$, for $|\Omega| \geq \Omega_0$, where $S_{X_c}(\Omega)$ is the spectral density and $B = 2\Omega_0$. This process can also be shown to be a singular continuous-space process[3] [23].

Let us denote the conditional entropies by $a_i = H(X_i|X_{i-1}, \ldots, X_1)$. Note that for any correlation structure, the sequence a_i is monotonically decreasing (because conditioning cannot increase

[3]Actually, it can be shown that the same singularity property holds as long as $S_{X_c}(\Omega) = 0$ on some frequency interval of non-zero measure [23].

entropy), and is bounded from below by zero (because the entropy cannot be negative). Since the nodes are equally spaced, and the correlation function of a WSS process is symmetric, it is clear that $H(X_I|X_{I-1}, X_{I-2}, \ldots, X_{I-i}) = H(X_I|X_{I+1}, X_{I+2}, \ldots, X_{I+i})$, for any $I, 0 \leq i \leq I-1$.

Let us denote by $\psi(N)$ the ratio between the total cost associated to Slepian-Wolf coding ($\text{cost}_{SW}(N)$) and the total cost corresponding to coding by explicit communication ($\text{cost}_{EC}(N)$), that is:

$$\psi(N) = \frac{\text{cost}_{SW}(N)}{\text{cost}_{EC}(N)} = \frac{\sum_{i=1}^{N} i a_i}{\sum_{i=1}^{N} (N-i+1) a_i}. \tag{6}$$

Then, the following theorem holds [11]:

Theorem 3: (Scaling Laws)

Asymptotically, we have the following results:

(i) *If* $\lim_{i\to\infty} a_i = C > 0$,

- $\lim_{N\to\infty} \psi(N) = 1$,
- $\textit{cost}_{SW}(N) = \Theta(\textit{cost}_{EC}(N))$.

(ii) *If* $\lim_{i\to\infty} a_i = 0$,

(ii)-1 *If* $a_i = \Theta(1/i^p), p \in (0,1)$,

- $\lim_{N\to\infty} \psi(N) = 1 - p$,
- $\textit{cost}_{SW}(N) = \Theta(\textit{cost}_{EC}(N))$.

(ii)-2 *If* $a_i = \Theta(1/i^p), p \geq 1$,

- $\lim_{N\to\infty} \psi(N) = 0$,
- $\textit{cost}_{SW}(N) = o(\textit{cost}_{EC}(N))$,
- *If* $p = 1$, $\psi(N) = \Theta(1/\log N)$,
- *If* $p \in (1,2)$, $\psi(N) = \Theta(1/N^{p-1})$,
- *If* $p = 2$, $\psi(N) = \Theta(\log N/N)$,
- *If* $p > 2$, $\psi(N) = \Theta(1/N)$.

The proof of Theorem 3 can be found in [10]. In Figure 9, we show typical behaviors of the ratio of total flow costs for the two coding approaches.

We apply now Theorem 3 to the correlation models we consider in this work:

- For an expanding network: In cases (a.1) and (a.2), the result of sampling is a discrete-space regular process [29], thus $\lim_{i\to\infty} a_i = C > 0$, and it follows that $\lim_{N\to\infty} \psi(N) = 1$. In case (b), if the spatial sampling period d is smaller than the Nyquist sampling rate $1/B$ of the

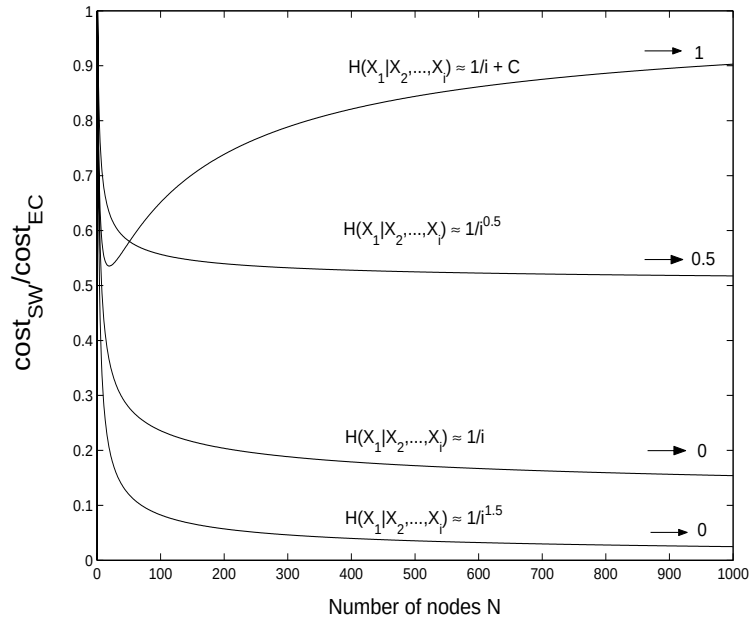
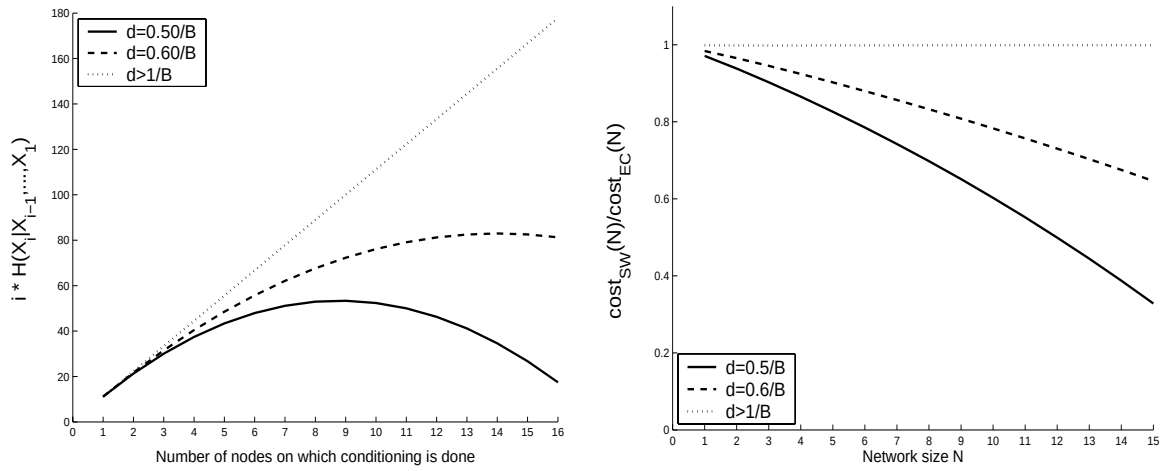


Fig. 9. Typical behavior of the ratio of the total costs $\text{cost}_{SW}(N)/\text{cost}_{EC}(N)$.

corresponding original continuous-space process, then $\lim_{i\to\infty} a_i = 0$. The specific speed of convergence of a_i depends on the spatial sampling period (that is, how small it is with respect to $1/B$) and the specific (bandlimited) power-spectrum density function of the process. In Figure 10, we show the bandlimited example with correlation $K(\tau) = B\text{sinc}(B\tau)$. It can be seen that when $d < 1/B$, $a_i = o(1/i)$ and thus, the ratio of total costs goes to zero. Also, the smaller d is, the faster the convergence is.

- For a refinement network: In case (a.1), we show in [10] that $H(X_i|X_{i-1},\ldots,X_1) = H(X_i|X_{i-1})$, thus $a_i = H(X_i|X_{i-1})$ for any $i \geq 2$. Then, for any finite N, $a_N > 0$. Since ia_i does not converge to zero (see Figure 11), then it follows from Theorem 3 that in the limit, the ratio of total costs is $\lim_{N\to\infty} \psi(N) = 1$. In case (a.2), a closed form expression for the conditional entropy is difficult to derive. However, we show numerically in Figure 11(a) that in this case a_i decreases faster[4] than $1/i$. For comparison purposes, we show in Figure 11(a) also the behavior for case (a.1). Thus, from Theorem 3, $\lim_{N\to\infty} \psi(N) = 0$. In Figure 11(b), we plot also the ratio of total costs for both correlation models. Finally, in

[4]Since these two processes are both non-bandlimited, sampling them results in discrete-space regular processes [29]. However, the sampled model (a.2) inherits a 'superior predictability' than (a.1), which makes a_i decrease faster than $1/i$.

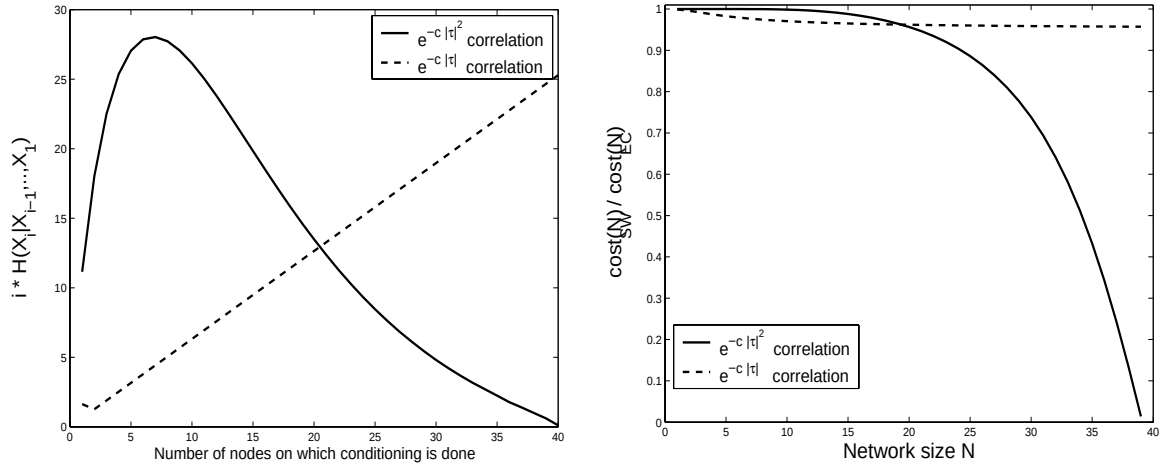


(a) (b)

Fig. 10. Expanding network sampling a bandlimited process with correlation model given by $K(\tau) = B\text{sinc}(B\tau)$: (a) The conditional entropy $H(X_i|X_{i-1}, \ldots, X_1)$ decreases faster than $1/i$ if $d < 1/B$; (b) The behavior of the ratio of total $\text{cost}_{SW}(N)/\text{cost}_{EC}(N)$ as a function of the size of the network.

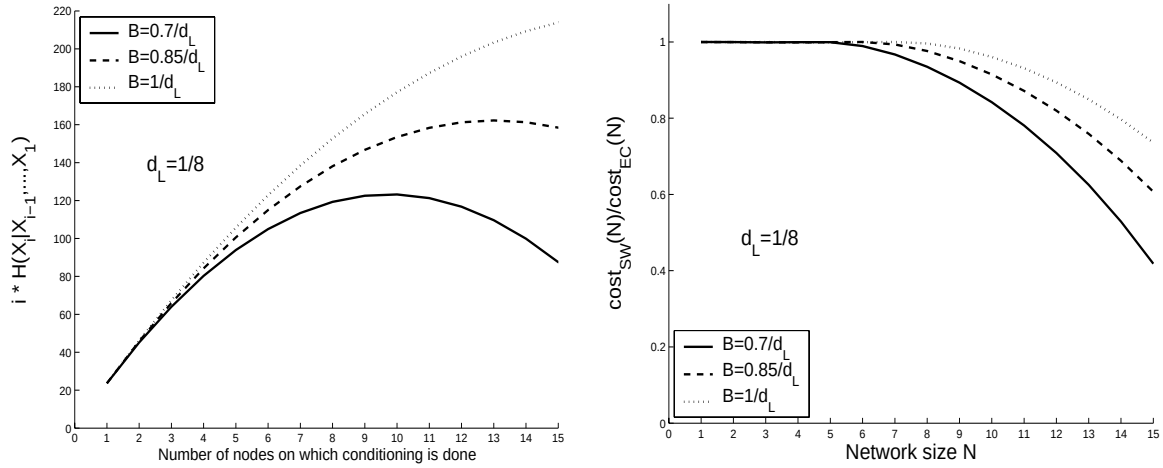
case (b), a_i goes to zero very fast, as for the case (a.2), because of the singularity of the original bandlimited process. It can be seen in Figure 12 how the ratio of costs starts to decrease as soon as $d < 1/B$, thus $\lim_{N\to\infty} \psi(N) = 0$.

Intuitively, similar results to the ones presented in this section hold also for higher dimensions, when the transmission structure that is used is the same (e.g. *SPT*) for both types of coding. The ideas leading to the results for the one-dimensional network can be generalized to two-dimensional networks. For instance, one can consider a two-dimensional wheel structure with the sink in the center of the wheel, where entropy conditioning at the nodes on any spoke is done as in the one-dimensional case (see Figure 13). The same analysis as in the one-dimensional case holds, with the additional twist that, according to Theorem 2, Slepian-Wolf coding at node i is done by conditioning not only on the nodes closer to the sink on its spoke, but also on the nodes on the other spokes closer to the sink on the SPT than node i (the dashed circle in Figure 13). However, the explicit communication coding is still done only on the nodes on the spoke that forward their data to node i (the solid circle in Figure 13). Thus, the ratio of costs $\psi(N)$ in the two-dimensional case is upper bounded by its counter-part in the one-dimensional case, which means that the results of Theorem 3 apply for the two-dimensional case as well.



(a) (b)

Fig. 11. We consider a correlation dependence on the inter-node distance d given by $\exp(-c|\tau|^{\beta}), \beta \in \{1, 2\}$: (a) The conditional entropy $H(X_i|X_{i-1}, \ldots, X_1)$ decreases faster than $1/i$ for $\beta = 2$, but is constant for $\beta = 1$ (after $i \geq 2$); (b) The behavior of the ratio of total $\text{cost}_{SW}(N)/\text{cost}_{EC}(N)$ with as a function of the size of the network.



(a) (b)

Fig. 12. Refinement network sampling a bandlimited process; we denote the reference bandwidth with $B_L = 1/d_L$: (a) The conditional entropy $H(X_i|X_{i-1}, \ldots, X_1)$ decreases faster than $1/i$ as soon as $B < 1/d$, that is, $N > 1/d_L$; (b) The behavior of the ratio of total $\text{cost}_{SW}(N)/\text{cost}_{EC}(N)$ as a function of the size of the network.

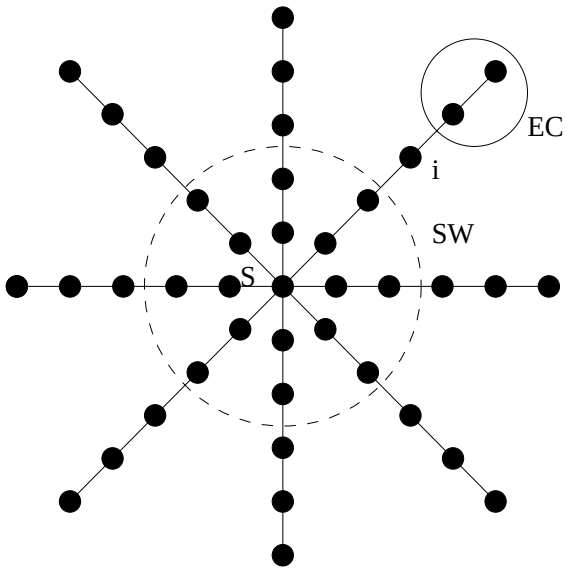


Fig. 13. A two-dimensional network with a wheel structure, with the sink S in the center. Slepian-Wolf coding for node i is done by conditioning on the nodes in the dashed region (denoted by SW). Explicit communication coding for node i is done by conditioning on nodes in the solid region (denoted by EC).

B. Node Placement

Further, we consider the related problem where a given number of nodes N is placed in a field such that the sensed data can be reconstructed at the sink within specified distortion bounds while minimizing the energy consumed for communication. For the sake of simplicity, consider the one-dimensional network in Figure 3, where the nodes can now be placed in arbitrary position on the line. Denote by w_i the distance between node i and node $i-1$. The optimization problem considered in this section aims to minimize the total transmission cost under distortion constraints. Namely, the distortion constraints impose a prescribed accuracy of representation at the sink for all the points in the network, when only the measurements at the sensor nodes are available.

Note that for the one-dimensional network in Figure 3 and for a space-dependent correlation model as considered in this work, the distortion constraints translate into spatial constraints. Assume the value at intermediate nodes is approximated by the sensing node corresponding to the Voronoi cell to which that intermediate node belongs. Then, our optimization becomes:

$$\{R_i^*, w_i^*\}_{i=1}^N = \arg\min_{R_i, w_i} \sum_{i=1}^{N} \sum_{j=i}^{N} R_j \times w_i^{\kappa} \quad (7)$$

under constraints

$$\sum_{i=1}^{N-1} w_i + \frac{w_N}{2} = L, \quad (8)$$

$$w_i \leq W^{\max}, \quad (9)$$

where (8) is a constraint on the average distortion, and (9) is a constraint on the individual constraints per Voronoi cell.

For the sake of simplicity, we only consider in this chapter the case where data at nodes is independent, namely $\{R_i\}_{i=1}^N = R$. A complete analysis of the case when Slepian-Wolf coding and explicit communication based coding are exploited is provided in [16].

Then the optimal placement $\{w_i^*\}_{i=1}^N$ is obtained by solving:

$$\{w_i^*\}_{i=1}^N = \arg\min_{w_i} \sum_{i=1}^{N} (N-i+1) w_i^{\kappa}, \quad (10)$$

under the distortion constraints (8), (9).

By using Lagrangian optimization with a multiplier λ, we obtain the optimal solution:

$$w_i = \left(\frac{\lambda}{\kappa(N-i+1)}\right)^{\frac{1}{\kappa-1}}, i = 1 \ldots N-1,$$

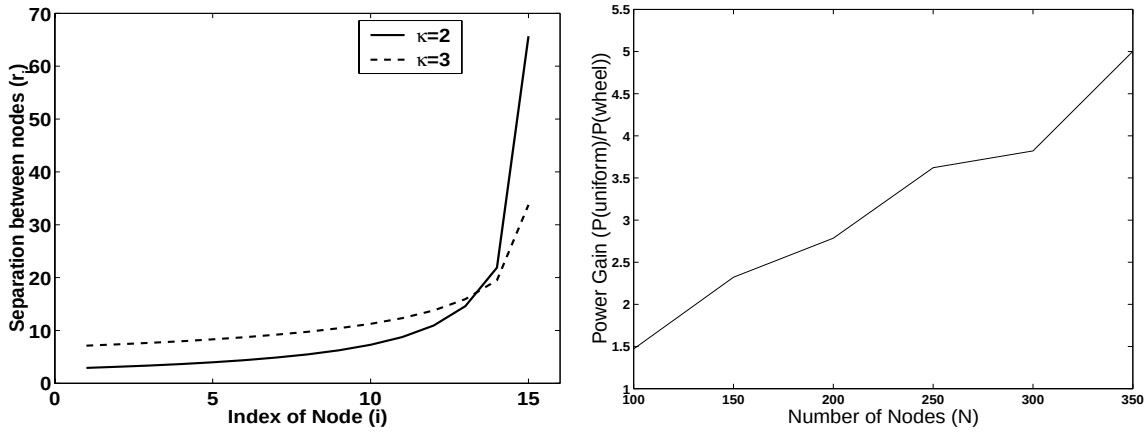
$$w_N = \left(\frac{3\lambda}{2\kappa}\right)^{\frac{1}{\kappa-1}},$$

$$\lambda = \left(\frac{L}{\sum_{i=1}^{N-1} (\frac{1}{\kappa(N-i+1)})^{\frac{1}{\kappa-1}} + \frac{3}{2}(\frac{3}{2\kappa})^{\frac{1}{\kappa-1}}}\right)^{\kappa-1}.$$

Such a placement provides important energy performance improvements as compared to uniform placement (see Figure 14); exploiting data correlation by using Slepian-Wolf coding further improves the results [16].

C. Lossy Data Gathering

1) Rate-Distortion Allocation: Let us further consider the case when lossy coding is used at nodes, but with high-resolution [37]. The information measured by the nodes should be available



(a) (b)

Fig. 14. Optimal placement for one-dimensional network, for different path-loss exponents (left); energy improvement over uniform placement for the two-dimensional network (right), as a function of the number of network nodes.

at the sink within certain total and individual distortion bounds. A rate-distortion allocation $\{(R_i, D_i)\}_{i=1}^N$ (bits) has to be assigned at the nodes so that the quantized measured information samples are described with certain total D and individual $D_i^{\max}, i = 1, \ldots, N$ distortions. Denote by ST the spanning tree to be found, which defines the transmission structure; $c_i, i = 1 \ldots N$ are the total weights of the path from node i to the sink on the spanning tree ST, thus $c_i = \sum_{e \in \mathcal{E}_i} w_e^\kappa$, where $e \in \mathcal{E}_i$, and $\mathcal{E}_i$ is the set of edges linking node i to the sink S on ST, w_e is the Euclidean distance of edge e, and κ is the path-loss exponent; $h(\cdot)$ denotes the differential entropy. Then, the most general form of our optimization problem is given as follows:

$$\{\{R_i^*, D_i^*, c_i^*\}_{i=1}^N, ST^*\} = \arg \min_{R_i, D_i, c_i, ST} \sum_{i=1}^N c_i R_i \tag{11}$$

under constraints

$$\sum_{i \in \mathcal{X}} R_i \geq h(\mathcal{X}|V \backslash \mathcal{X}) - \log \prod_{i \in \mathcal{X}} 2\pi e D_i, \forall \mathcal{X} \subset V, \tag{12}$$

$$\sum_{i=1}^N D_i \leq D, \qquad D_i \leq D_i^{\max}, i = 1 \ldots N, \tag{13}$$

where (12) express the rate-distortion region constraints given in [37], namely that the sum of rates for any given subset of nodes has to be larger than the entropy of the random variables measured at those nodes, conditioned on the random variables measured at all other nodes. In constraints (13), D is the maximum allowed total distortion and $D_i^{\max}, i = 1 \ldots N$ are the

maximum individual constraints.

In the high rate regime, uniform quantization and Slepian-Wolf coding is optimal [37]. Thus, a similar result as the one stated in Theorem 1 holds in the lossy case about the separation between transmission optimization and rate-distortion allocation, since nodes do not need to communicate explicitly to code data with a given rate-distortion allocation. As a result, the SPT is the optimal transmission structure in this case as well. Thus, we can prove in the lossy case a result similar to Theorem 2 which corresponds to the lossless case; namely, for any set of distortion values $\{D_i\}_{i=1}^N$, the rate allocation is given by:

Theorem 4: (Optimal rate allocation):

$$\begin{aligned} R_1^* &= h(X_1) - \log 2\pi e D_1, \\ R_2^* &= h(X_2|X_1) - \log 2\pi e D_2, \\ &\dots \\ R_N^* &= h(X_N|X_{N-1}, \dots, X_1) - \log 2\pi e D_N. \end{aligned} \tag{14}$$

Theorem is proven in 4 in [8].

Next, we consider optimization of (11) for the case where we assume that the constraints given by (13) are not active. By Theorem 4, $\{R_i^*\}_{i=1}^N$ *only* depends on $\{D_i\}_{i=1}^N$. Therefore, at this point, we can insert in (11) the values for $\{R_i^*\}_{i=1}^N$ given by Theorem 4, thus obtaining an optimization problem having as argument the set of distortions $\{D_i\}_{i=1}^N$ only, that is:

$$\begin{aligned} \{D_i^*\}_{i=1}^N &= \arg \min_{\{D_i\}_{i=1}^N} \sum_{i=1}^N c_i^* \cdot (h(X_i|X_{i-1}, \dots, X_1) - \log 2\pi e D_i) \\ &\text{under the constraint} \\ \sum_{i=1}^N D_i &\leq D. \end{aligned} \tag{15}$$

Note that the differential entropy terms in (15) do not depend on the distortions D_i. Thus, (15) can be equivalently written as:

$$\begin{aligned} \{D_i^*\}_{i=1}^N &= \arg \max_{\{D_i\}_{i=1}^N} \sum_{i=1}^N c_i^* \log 2\pi e D_i \\ &\text{under the constraint} \\ \sum_{i=1}^N D_i &\leq D. \end{aligned} \tag{16}$$

Denote $\sum_{i=1}^{N} c_i^* = C$. The solution of the optimization problem (16) is easily obtained, using Lagrange multipliers, giving a linear distribution of distortions:

$$D_i^* = D \cdot \frac{c_i^*}{C}, \quad i = 1 \dots N. \tag{17}$$

By combining (14) and (17), we obtain that the rate-distortion allocation at nodes is given by:

$$R_i^* = h(X_i | X_{i-1}, \dots, X_1) - \log\left(\frac{2\pi e D c_i^*}{C}\right), \quad i = 1 \dots N. \tag{18}$$

Note that for correlation functions that depend only on the distance among nodes, as we consider in this work, and for an uniform placement of nodes, the differential entropy monotonically decreases as the number of nodes on which conditioning is done increases. Also, by definition, the sequence $\{c_i^*\}_{i=1}^N$ is monotonically increasing with i. As a result, the rate allocation R_i in (18) is a function that monotonically decreases with the node index i.

Further, (18) essentially depends only on the path weights ordering of the nodes on the SPT, given by $\{c_i^*\}_{i=1}^N$. For example, in the case of correlated Gaussian random fields, (18) can be written as:

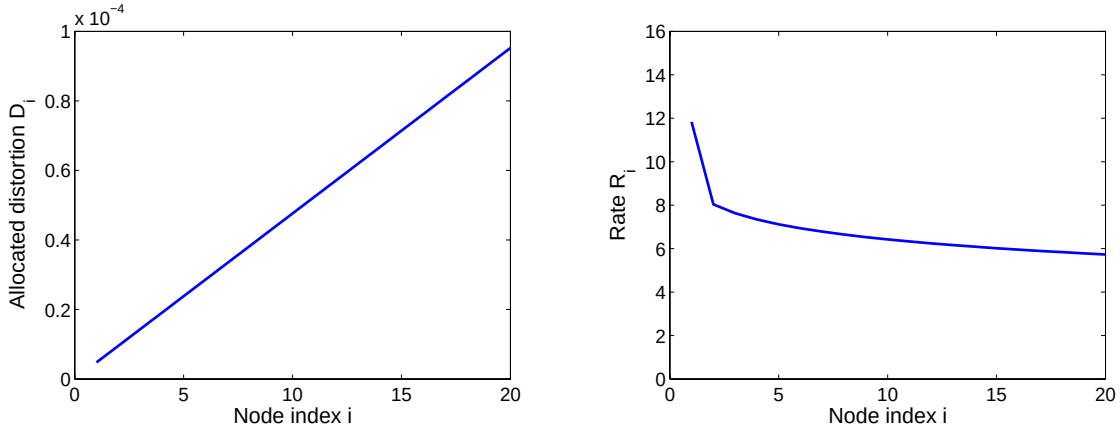
$$R_i^* = \log \frac{\det(\mathbf{K}(1, \dots i))}{\det(\mathbf{K}(1, \dots, i-1))} \frac{C}{c_i^* \cdot D}, \quad i = 1 \dots N, \tag{19}$$

where $\mathbf{K}(1, \dots, i)$ is the correlation matrix corresponding to nodes $1, \dots, i$.

In Figure 15, we illustrate the distortion and rate allocations provided by (19) for the one-dimensional grid with uniform inter-node distances, measuring a correlated Gaussian random field with $\beta = 1, c = 10^{-3}$, $N = 20$, $D = 10^{-3}$, $D_i^{\max} = 0.7 \cdot 10^{-3}$, and $\kappa = 2$ (which corresponds to sampling a continuous Gauss-Markov process). The same analysis holds for arbitrary two-dimensional networks. Similar results are obtained for many other network parameter settings, and in the case of individual distortion constraints [8].

2) Node Placement: In some scenarios, the positions of the nodes are not fixed in advance, but it is possible to place the nodes optimally so as to minimize various resources [16]. Since the study in this work is concerned with energy efficient scenarios, one possible task to be achieved when the node placement can be chosen is the total energy efficiency. Namely, we consider the optimization (11) where the positions of the nodes are additional variables, and the optimization is done additionally over the weights of the paths from the nodes to the sink:

$$\begin{aligned} \{R_i^*, D_i^*, c_i^*\}_{i=1}^N &= \arg \min_{\{R_i, D_i, c_i\}_{i=1}^N} \sum_{i=1}^{N} c_i R_i \\ \text{under the constraints} \quad & (12), (13), \end{aligned} \tag{20}$$



(a) (b)

Fig. 15. One-dimensional network, average distortion constraint: (a) distortion and (b) rate allocations as a function of the node index.

with additional constraints on $\max\{c_i\}_{i=1}^{N}$ (coverage constraint) and on $\max\{c_i - c_{i-1}\}_{i=2}^{N}$ (inter-node space constraint). The coverage constraint imposes that the entire area is covered. The inter-node space constraint ensures that any given point in the measured area is close enough to a sensor node, such that the data corresponding to that point can be reconstructed with a certain minimal accuracy at the sink, by approximating it with the value measured by the closest sensor.

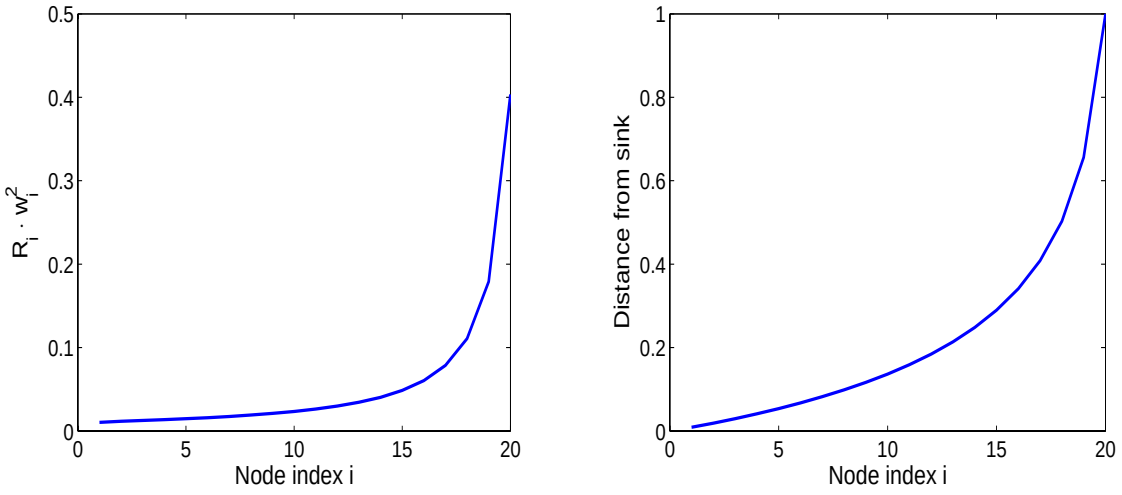
We study the problem of optimal placement for two energy efficiency targets of interests, namely total energy and network lifetime[5], and compare the tradeoffs involved. For the one-dimensional example in Figure 3, the optimal placement is [8]:

$$w_i^* = \frac{L}{(\sum_{j=i}^{N} R_j^*)^{\delta} \left(\sum_{l=1}^{N} \frac{1}{(\sum_{j=l}^{N} R_j^*)^{\delta}} \right)}, i = 1 \dots N,$$

where $\{w_i^*\}_{i=1}^{N}$ is the distance between nodes $i-1$ and i; $\delta = 1$ for total energy minimization, and $\delta = 1/2$ for lifetime maximization. The optimal joint solution for the placement and rate allocation is obtained by using an iterative algorithm:

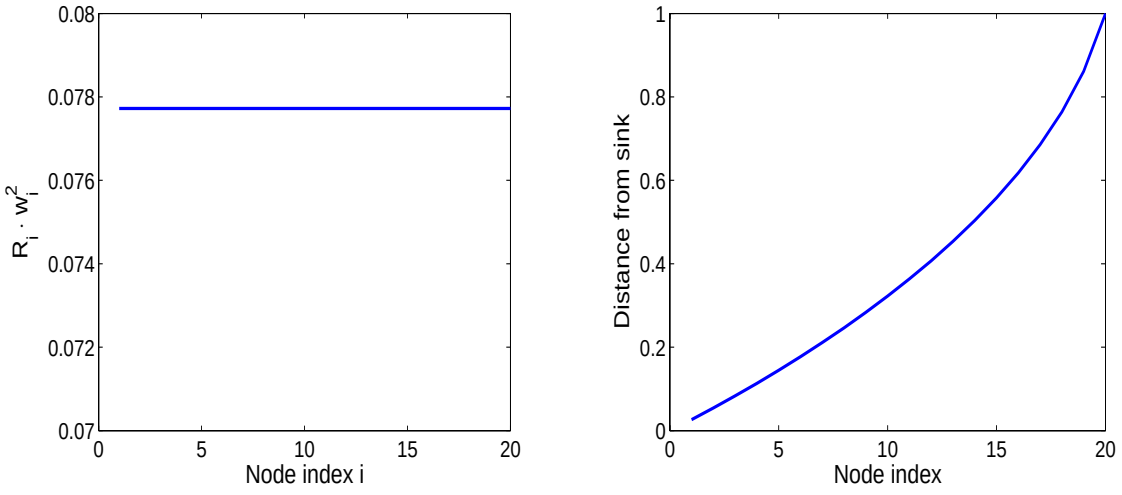
The energy consumption and node positioning resulting from running Algorithm 1 are presented in Figure 16 (for the case of total energy minimization) and Figure 17 (for the case of lifetime maximization). A complete study of the node placement problem for efficient lossy network data gathering is provided in [8].

[5]We address network lifetime optimization by considering the constraint that all nodes use equal energy.



(a) (b)

Fig. 16. Placement optimization for minimizing the total energy: (a) energy consumption as a function of node index; (b) distance from sink as a function of node index.



(a) (b)

Fig. 17. Placement for lifetime optimization: (a) energy consumption as a function of node index; (b) distance from sink as a function of node index.

Algorithm 2 Placement and rate allocation

Initialize uniformly the node placement $\{w_i\}_{i=1}^N = L/N$.

repeat

Given $\{w_i\}_{i=1}^N$, solve the energy minimization problem for $\{R_j\}_{j=1}^N$.

Re-write $\{w_i\}_{i=1}^N$ as a function of $\{R_j\}_{j=1}^N$.

until convergence.

We studied the problem of data gathering from a network of nodes measuring correlated data to a sink node. We considered an energy related metric, and showed that, in both the cases of lossless and of lossy coding of the measured information, the optimal transmission structure is the shortest path tree of the network graph. Further, we determined the optimal rate and distortion allocations for coding at nodes. Moreover, for the case when the positions of the nodes can be chosen, we provided strategies for energy-efficient node placement.

Our discussion so far considered the case of spatially correlated but temporally i.i.d. processes. In the following section we will discuss the case of processes that are correlated both in time and in space.

V. Data Gathering of Spatio-Temporal Processes

In this section, we consider the usual scenario of a sensor network with a sink. The goal of the sink is now to reconstruct the *entire* field over space and time, with a certain minimum distortion, and only based on measurements at the sensors [13]. For this, a fixed number N of nodes are placed in the field (see Figure 18). Nodes transmit their measurements to the sink, at given time instants, by using a subset of the links of the graph. The sink needs to reconstruct the whole field with a minimal total distortion. We consider settings where data is time critical, and thus delay results in distortion. Such settings include scenarios for fire prevention, or seismic awareness; extremal settings include sensor networks measuring phenomena where abrupt transitions are critical (e.g. cracks in a massive structure, or mudslides over a large terrain area). Another class of relevant scenarios is when sink feedback or control is needed at nodes, and where the effect of delay in reporting the data induces suboptimal estimation and control.

If the network is dense (large N), the data has a good spatial approximation. However, for energy efficiency reasons, nodes in sensor networks cannot transmit their data directly to the

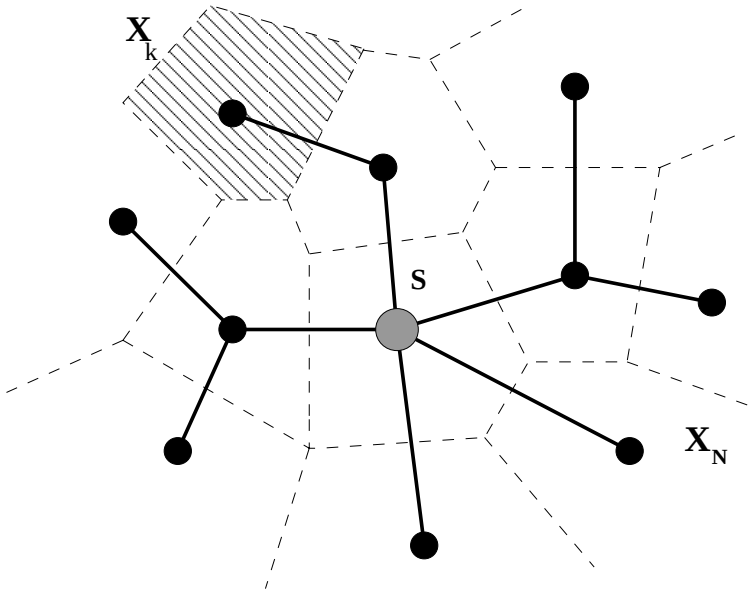


Fig. 18. In this example, spatio-temporally correlated data from nodes $1, \ldots, N$ need to arrive at sink S. The sink needs to reconstruct the *whole* data field using only the measured values $X_1, \ldots, X_N$. Arbitrary points in the two-dimensional space are approximated by the measurements of the sensor node corresponding to the Voronoi cell to which they belong. The dashed zone corresponding to node k represents the area of the field approximated by the value X_k in node k. In thick solid lines, a chosen transmission structure is shown (here, the shortest path tree *SPT*). Data from node k reaches the sink after being relayed by one other sensor node.

sink, but rather communication is usually done via energy efficient transmitting structures. This implies long delays until the data sent from nodes far from the sink reach the sink, which thus results in weak temporal approximation.

On the contrary, the opposite effects take place when N is small. Namely, the spatial approximation of the data is poor, but on the other hand data has to travel over only a limited number of hops, which results in a good temporal approximation. Thus, as we will show in this section, for a given spatio-temporal correlation structure, there usually exists a *finite optimal* N that minimizes the overall distortion of the field reconstruction at the sink.

A. Problem Setup

We study the influence of node density on the total distortion of estimation, when several aspects specific to sensor nodes are considered, namely delay and energy efficiency. Our setting takes into account two important issues typical in sensor networks scenarios: the precision of

estimation [22], and the energy efficiency [30].

First, since the measured data is correlated and the number of available nodes is limited, the sink can reconstruct the values of the field at each point by approximating them with the values at the points where the actual sensor nodes are placed. In those points, full measurements are available. Also, no other information except the values measured at sensor nodes is available at the sink about that region of the field. The precision of the approximation depends both on the level of spatial correlation in the data and on the number of sensors available. This approximation introduces a first factor of distortion, which we call 'spatial distortion'.

Second, since the nodes have limited battery energy, a good strategy is to send data via relaying nodes rather than directly to the sink (multi-hopping). However, multi-hopping results in data delay, since data from the extremities of the network need to be transmitted via multiple relays until they reach the sink. In various practical sensor network scenarios, data is needed at the sink in real-time. For instance, for the tasks of control or active monitoring, data may become useless if it arrives at the sink with too large a delay. For a spatio-temporal correlated process, the data that arrive at the sink are distorted from the original measured values, however they can be reconstructed with a certain precision given by the intensity of temporal correlation of the process. Thus, delay introduces a second factor of distortion, which we call 'time distortion'.

In our study, the two types of distortion are modelled as a single *distortion per field point* quantity. Their combined effect results in the total distortion of the field at the sink, and the goal of this work is to study how this quantity is influenced by the density of nodes of the sensor network. Namely, we argue that for various typical spatio-temporal correlation models of the data field, there is a unique optimal value for the number of placed nodes N that minimizes the total spatio-temporal distortion.

B. One-Dimensional Network

1) Transmission Model: For simplicity of the analysis, in this work we consider networks with nodes uniformly placed on one- and two-dimensional grids, for which the distortion optimization is done only with respect to the size of the network. In the case of arbitrary networks with position-dependent correlation structure, the optimization becomes a function of the node placement as well. In our study, we consider Gaussian random processes which exhibit spatio-temporal correlation (introduced in Section III-B.2).

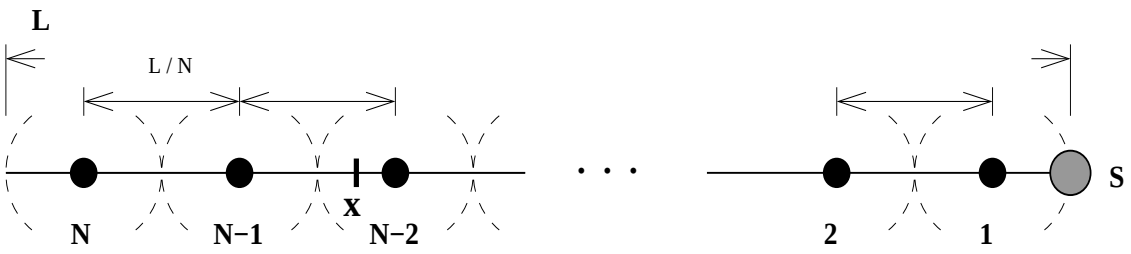


Fig. 19. A one-dimensional example of a data gathering network, where each sensor node covers an area of the whole network. An *instantaneous point of reconstruction* is denoted by x. In this example, x belongs to the Voronoi cell of node $N-2$, thus its value at time t is approximated, with a certain distance dependent distortion, by $X(x_{N-2}, t)$.

Consider N nodes placed on a line of fixed length L (see Figure 19). The inter-node distance is $d = L/N$. An additional node S at the extreme right of the line represents the sink to which all data should arrive. The task of the sink is to reconstruct the *whole* field on the line.

We assume that the quantization done at nodes is very fine, namely we assume the reconstruction error at the sink is only an estimation error. For that, points on the line that belong to the space intervals among the N nodes are assigned to Voronoi cells of the sensor nodes; these cells are delimited by mid-interval points. Therefore, each sensor node covers an interval of length d around its position. The values of the intermediate points are estimated at the sink by the value of the corresponding sensor node in the middle of the Voronoi cell[6] (see Figure 19 and Figure 20). In this work, we assume that the measurements at nodes have the same variance, and that the coding rates allocated at nodes are equal. This implies that nodes use equal transmission energy, and that the spatial distortion per cell does not depend on the node identity. The study of networks with different rates at nodes is a subject of our current research.

Multi-hopping, as mentioned in Section V-A, inherently introduces delay. The delay has two causes. On the one hand, if relaying nodes have finite buffers then packet forwarding results in a certain processing time due to buffering, and thus the delay is proportional to the number of hops that data needs to travel. On the other hand, if communication edges are lossy channels then this requires retransmissions of data, and thus the delay is proportional to the number of edges that

[6]More complex estimation strategies can be designed, in which the field value at an arbitrary position is based on the measurements of more than a single sensor node [25], [26].

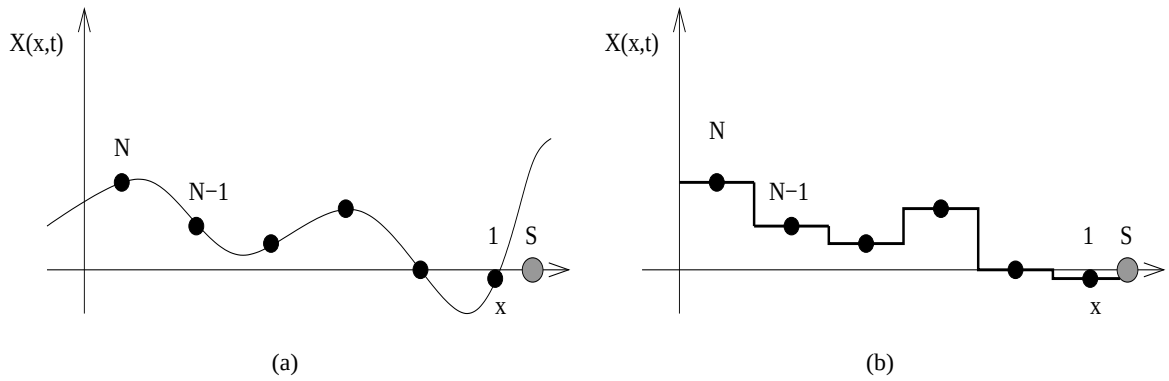


Fig. 20. The reconstruction at the sink is based on the values measured by the sensing nodes: (a) the original signal ; (b) the approximation by interpolation at the sink.

data needs to travel. For the sake of simplicity, we assume that the delay is proportional to the number of hops. Therefore, since the measured process is correlated both in time and in space, the reconstruction is further distorted from the real-time value due to delay caused by relaying.

2) Point-wise Distortion: For instance, consider the value $X(x,t_0)$ of an arbitrary point x on the line at an arbitrary time t_0 (see Figure 21). Assume that the sink approximates the value at point x by considering the value $X(x_0,t_0)$, at point x_0, placed k hops away from the sink. For any data packet, we assume that the relation between the time delay t_k of that packet and the number of hops k it has to travel is $k=\gamma t_k$. Assume the delay per hop is a constant T, thus $t_k = kT$. Then, the mean-square error (MSE) of $X(x,t_0)$ at the sink, when $X(x_0,t_0)$ is known, is expressed by:

$$\begin{aligned} D_{x,t_0,x_0,k} &= E\left[(X(x,t_0)-X(x_0,t_0+kT))^2\right] \\ &= E\left[X(x,t_0)^2\right]+E\left[X(x_0,t_0+kT)^2\right]-2E\left[X(x,t_0)X(x_0,t_0+kT)\right] \\ &= 2-2\sigma_{X(x,t_0)X(x_0,t_0+kT)} \\ &= 2-2\sigma(|x-x_0|,kT) \\ &= 2-2e^{-c((x-x_0)^2+(\gamma kT)^2)^{\frac{\beta}{2}}}. \end{aligned} \tag{21}$$

In other words, for any point in time and space, the generalized distance between the approximated and the real value as seen by the sink is $\sqrt{(x-x_0)^2+(\gamma kT)^2}$, and the corresponding *distortion per field point* of point x as seen by the sink is given by $2(1-\sigma(|x-x_0|,kT))$.

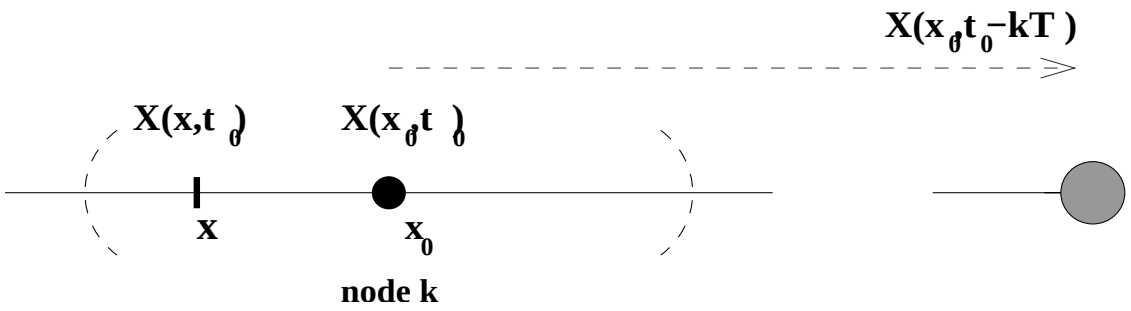


Fig. 21. The value at point x is approximated by the value of the kth sensor node placed at x_0, resulting in *spatial distortion*. Moreover, due to transmission over k hops, the version that reaches the sink is delayed with kT, which results in *time distortion*. The combined result of the two distortion effects is the *total distortion* $D_x(N)$.

In general, the statistics of the correlated data field might not be known, but they can be measured on-line, during the network deployment period (for instance, if the correlation is distance dependent, then nodes can make use of the distance information acquired from the neighbors for constructing routing tables, for the additional tasks of estimating the correlation structure and fitting the measurement parameters to a valid correlation model [7]).

3) Total Distortion: We compute now the total distortion of the data estimated by the sink at a snapshot in time, in MSE sense. Consider node k, which is placed k hops away from the sink. Denote the position of node k as x_k, and the data that node k measures at time t as $X(x_k, t)$. The data sent at time t to the sink about the region $[x_k - d/2, x_k + d/2]$ is $X(x_k, t)$, but since it is *delayed* with k clock ticks, this packet actually reaches the sink at time $t + kT$ (see Figure 21). In fact, at time t, the actual available data at the sink about node k is $X(x_k, t - kT)$. Thus, the corresponding distortion of reconstruction of the region covered by node k is:

$$D_k(N) = 4 \int_{x=x_k}^{x_k + \frac{L}{2N}} (1 - \sigma(x - x_k, kT))dx. \tag{22}$$

For simplification, we can consider $x_k = 0$ as axis origin for each node k, and then (22) becomes:

$$D_k(N) = 4 \int_{x=0}^{\frac{L}{2N}} (1 - \sigma(x, kT))dx. \tag{23}$$

The total distortion $D(N)$ is simply obtained by summing (23) over all nodes $k = 0 \ldots N-1$:

$$D(N) = \sum_{k=0}^{N-1} 4 \int_{x=0}^{\frac{L}{2N}} (1 - \sigma(x, kT))dx. \tag{24}$$

Further, if we insert the correlation model for a Gaussian spatio-temporal process (1), we can finally write the total distortion of reconstruction of the whole field by the sink, as a function of N:

$$D(N) = \sum_{k=0}^{N-1} 4 \int_{x=0}^{\frac{L}{2N}} (1 - \exp(-c(x^2 + \gamma^2(kT)^2)^{\frac{\beta}{2}}))dx, \tag{25}$$

where k counts the number of hops from a node to the sink, T is the time delay per hop, γ is the time scaling constant, and c is a constant quantifying the intensity of correlation of the field. The term which is integrated is the distortion incurred by approximating the field between $[-\frac{L}{2N}, \frac{L}{2N}]$, around the node which is at k hops away from the sink, with the value of that node delayed k time steps.

The expression in (25) cannot be expressed in a closed form. However, an experimental analysis shows that (25) has always a minimum as a function of N. Moreover, in Section V-B.5, we will use a strong correlation approximation to derive in a closed-form the optimal value N for which (25) is minimized. In general, the optimal value of N is obtained by setting $\frac{\delta D(N)}{N} = 0$ and numerically solving for N, by rounding the solution to the closest integer.

Since spatially correlated Gaussian processes can only have certain structures for the correlation dependence on the distance [7], we will restrict our analysis to the models introduced in Section V-B.2.

4) Optimum N is Finite: In this section we show that for the Gaussian correlation models introduced in Section V-B.2, there is indeed a ***finite*** optimum N_0 that minimizes (25). Denote:

$$a_n = \sum_{k=0}^{n-1} \int_{x=0}^{\frac{L}{2n}} \sigma(x, kT)dx. \tag{26}$$

Note that by definition a_n is lower bounded by 0. Thus, from (24) we can see that a sufficient condition for the existence of a finite optimum N is that there exists N_0 such that for all $n > N_0$, a_n is a decreasing sequence.

(i) *Correlation model*: $\exp(-c(x^2 + \gamma^2(kT)^2))$

In this case, we can rewrite a_n as

$$a_n = \int_{x=0}^{\frac{L}{2N}} e^{-cx^2} dx \cdot \sum_{k=0}^{n-1} \frac{1}{e^{\gamma^2(kT)^2}}. \tag{27}$$

But $\lim_{n\to\infty} a_n \searrow 0$, since the first term in the product converges to zero (the error function) and the second one can be easily shown to be upper bounded by a finite positive constant. Thus, for $\beta = 2$, the optimum N that minimizes (25) is finite.

(ii) *Correlation model*: $\exp(-c\sqrt{x^2+\gamma^2(kT)^2})$

This case is difficult to analyze analytically, due to the function that is integrated. However, our simulations in Section V-D show that in this case too there is a finite optimal N_0.

5) Strong Correlation Approximation: We study the case when both L/N and the time scale γ are small. In other words, data is strongly correlated both spatially and temporally. In this case, we can make the approximation:

$$\begin{aligned} 1 - e^{-c(x^2+\gamma^2(kT)^2)^{\frac{\beta}{2}}} &\approx 1 - (1 - c(x^2+\gamma^2(kT)^2)^{\frac{\beta}{2}}) \\ &= c(x^2+\gamma^2(kT)^2)^{\frac{\beta}{2}}, \end{aligned}$$

which simplifies our analysis further.

(i) *Correlation model*: $\exp(-c(x^2+\gamma^2(kT)^2))$

First, we can write:

$$\int_{x=0}^{\frac{L}{2N}} c(x^2+(\gamma kT)^2)dx = c(\frac{L^3}{24N^3} + \frac{L(\gamma kT)^2}{2N}). \tag{28}$$

Further, from rewriting (25), it results:

$$D(N) = c(\frac{L^3}{2}\cdot\frac{1}{N^2} + \frac{2L(\gamma T)^2}{3}\cdot N^2 - L(\gamma T)^2\cdot N + \frac{1}{3}L(\gamma T)^2).$$

Now we take the partial derivative of $D(N)$ with respect to N and make it equal to zero. We obtain that N_0 is a solution of the equation:

$$4N^4 - 3N^3 - \alpha = 0, \tag{29}$$

where $\alpha = \frac{L^2}{\gamma^2T^2}$. For $\alpha > 0$, (29) has a single positive solution N_0.

This gives a good indication of the optimal value of N_0 as a function of $\frac{L^2}{\gamma^2T^2}$; intuitively, N_0 increases with the decrease of the importance of delay in the distortion function, given by the time scale γ.

Note that, as expected, in the approximation of very strong correlation, the optimal N does not depend on the value of c (which models the strength of correlation). The optimal value of N only depends on the ratio between the length of the field L and the time scaling parameter γT,

that models the relative importance of delay in the distortion function as compared to spatial distortion.

(ii) *Correlation model*: $\exp(-c\sqrt{x^2+\gamma^2(kT)^2})$

In this case, we compute:

$$\int_{x=0}^{\frac{L}{2N}} c(x^2+(\gamma kT)^2)^{\frac{1}{2}}dx = -\frac{c}{8N}(-L\sqrt{\frac{L^2+4(\gamma Tk)^2N^2}{N^2}}+2(\gamma Tk)^2\ln(\gamma Tk)^2N+$$

$$+4(\gamma Tk)^2N\ln 2-4(\gamma Tk)^2N\ln\frac{L+\sqrt{\frac{L^2+4(\gamma Tk)^2N^2}{N^2}}N}{N}. \quad (30)$$

When N is large, the second, third and fourth terms in the summation in the paranthesis cancel each other out, and it can be easily shown that (30) simplifies to

$$\int_{x=0}^{\frac{L}{2N}} c(x^2+(\gamma kT)^2)^{\frac{1}{2}}dx \approx \frac{cL\gamma Tk}{4N}. \quad (31)$$

By summing (31) over k, we can see that the resulting sum is a strictly increasing sequence in N. This only happens for N large enough to guarantee the strong correlation approximation, however this is enough to show that the optimum N_0 has to be finite.

C. Two-Dimensional Model

1) Total Distortion: The case of a two-dimensional grid network (see Figure 22) is studied similarly to the one-dimensional model. Consider a square area $L \times L$, on which N^2 nodes are uniformly placed on a square grid. The network is divided into Voronoi cells centered in the sensor nodes. We count the number of hops from each node to the sink on the most energy-efficient transmission structure for gathering uncorrelated data, which is the shortest path tree (*SPT*). Note that, in general, since data at nodes are spatially correlated, the shortest path tree is *not* the most energy-efficient transmission structure if in-network fusion by coding with side information is performed at nodes; moreover, finding the optimal transmission structure for such scenarios is NP-hard [12], [9]. Thus, in our analysis, we make the assumption that, due to limited resources, relay sensor nodes do not perform in-network fusion, namely they do not use as side information data from other nodes that use them as relay, to reduce the amount of data themselves need to transmit about their own measurements. In short, data is relayed without being processed.

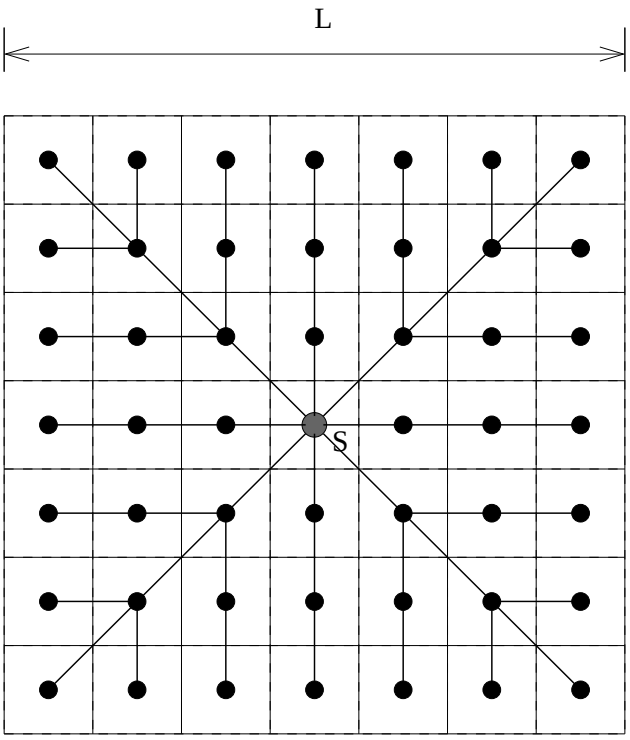


Fig. 22. A two-dimensional square grid. The Voronoi cell partition is drawn is dashed lines, and the shortest path tree (*SPT*) is drawn in bold solid lines.

In order to simplify the analysis, we consider a slightly modified setting for the two-dimensional square grid as compared to the one-dimensional model (see Figure 22). Namely, the modification from the one-dimensional study is that for the two-dimensional model we assume that the sink gathers with no delay data in its corresponding Voronoi cell (in other words, the sink itself is considered as a regular sensor).

We plot in Figure 22 the energy efficient paths from the nodes to the sink. Note that for every $k = 0 \ldots N-1$ there are $8k$ cells situated at k hops away from the sink. Therefore, analogously to the one-dimensional case, we can write the total distortion for the two-dimensional case as:

$$D(N) \;=\; \sum_{k=0}^{N-1} 4 \cdot 8k \int_{x=0}^{\frac{L}{2N}} \int_{y=0}^{\frac{L}{2N}} (1 - \exp(-c(x^2 + y^2 + \gamma^2 (kT)^2)^{\frac{\beta}{2}}))dydx, \qquad (32)$$

where N is now the number of hops from the sink to the extremity of the square network.

2) Strong Correlation Approximation: In this section we use an approximation similar to the one in Section V-B.5, namely

$$1-\exp(-c(x^2+y^2+\gamma^2(kT)^2)^{\frac{\beta}{2}}) \approx c(x^2+y^2+\gamma^2(kT)^2)^{\frac{\beta}{2}}. \tag{33}$$

For the sake of simplicity, we analyze only the case $\beta = 2$, since the resulting optimization is easier. Namely, after some straightforward manipulations including taking the partial derivative of the resulting $D(N)$ with respect to N, we obtain that the optimal N_0 is a solution of the equation:

$$N^5 - N^4 - \frac{\alpha}{3}N + \frac{\alpha}{2} = 0. \tag{34}$$

where $\alpha = \frac{L^2}{\gamma^2 T^2}$. A full analysis of the behavior of this polynomial is outside the scope of this study. However, by numerical experiments, we are able to provide a set of insights:

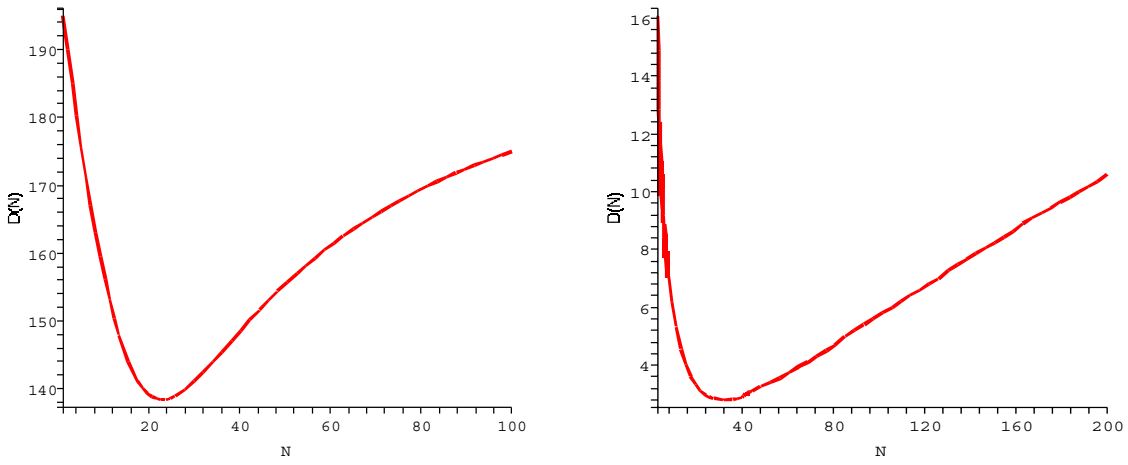
- For $0 < \alpha < 83.9$, this equation has no real positive solution (namely, it is strictly increasing and thus its optimum is attained at $N_0 = 1$, which means that in such a case the distortion caused by delay becomes so important that the optimal solution is to not place any sensor and let the sink estimate the whole field!)
- For $\alpha \geq 83.9$, the equation has two positive real solutions, one ($N_1 \in (1, 2)$) corresponding to a maximum of the function $D(N)$, and the other $N_2 > 2$ to a minimum. For $N > N_2$, $D(N)$ is strictly increasing. Thus, the optimum solution is either in $N_0 = 1$, or in N_2, both being finite integers.

D. Numerical Simulations

In this section we do not use the approximation of strong correlation, but rather use the rough total distortion formulae given by (25) and (32).

We use Maple to plot in Figure 23(a) the distortion $D(N)$ for the one-dimensional case, as expressed in equation (25), as a function of N, for typical values of the constants involved: $c = 0.5$ (reasonable correlation decay), $\gamma T = 0.1$ (the constant scaling the time axis), $L = 100$, and $\beta = 2$. In Figure 23(b) we illustrate with a similar plot the case when $\beta = 1$, with $c = 0.05$, $\gamma T = 0.05$, $L = 100$. We observe that, in general, there is an optimal N, that depends on the few constants involved in our model: c, γ, β, L.

Finally, in Figure 24, we plot the distortion $D(N)$ for the two-dimensional case, as expressed in equation (32), as a function of N, for typical values of the constants involved: $c = 0.05$,



(a) (b)

Fig. 23. Total estimation distortion of the field at the sink $D(N)$ as a function of the number of nodes N, for a one-dimensional network: (a) $\frac{\beta}{2} = 1$; (b) $\frac{\beta}{2} = 0.5$.

$\gamma T = 0.05$, $L = 10$, and $\beta = 2$. We observe that again there is an optimal N minimizing the total distortion. The ripples in the plots are due to Maple's graphical interpolation.

E. Remarks

In this section, we studied data gathering of spatio-correlated processes from a correlated data field to a sink. The task was to perform at the sink the reconstruction of the data measured at all the points in the field, with maximal accuracy, when the only available information is the data at the sensor nodes. We defined a single measure of accuracy, that combines the distortions due to the spatial approximation and to the delay in the network. We showed that, in general, there is a finite optimal density of sampling the field. Future work includes the analysis of more complex interpolation strategies and the study of a wider class of random processes.

VI. Conclusions

We studied the interaction between data representation at nodes, rate allocation, routing and node placement, for gathering of correlated data in sensor networks. The results of our work show that a joint consideration of these issues provides important improvements in the overall data gathering energy efficiency and accuracy of representation.

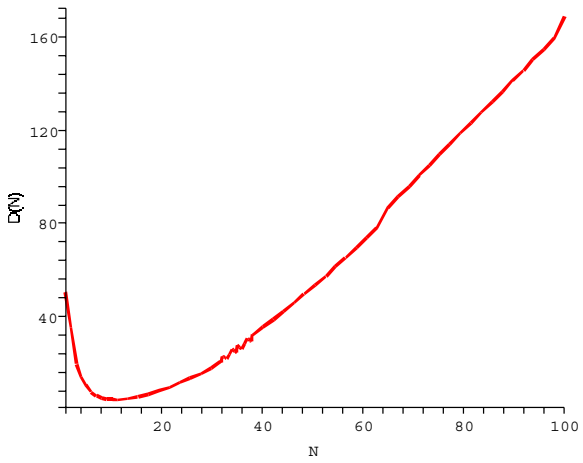


Fig. 24. Total estimation distortion of the field at the sink $D(N)$ as a function of the number of nodes N for a two-dimensional network.

We first analyzed energy efficient data gathering of random spatially correlated processes with lossless and lossy coding. We successively found the optimal transmission structure and rate-distortion allocations. Moreover, we considered the problem of energy-efficient optimal node placement. Further, we considered data gathering of spatio-temporally correlated processes under delay constraints. Namely, we defined a distortion measure that includes both the effects of spatial approximation and delay. We showed that, in general, there is an optimal finite density of nodes that should be placed in the field, for minimizing the total distortion of reconstruction at the sink.

In all the scenarios, the non-trivial interaction between the structure of the data (or underlying signal) and the transport mechanism to a central sink has been highlighted. In our view, this is a central challenge in the design and operation of sensor networks. No simple 'separation' can be used, and substantial gains are obtained by a joint analysis and design.

REFERENCES

[1] A. Aaron, B. Girod: *Compression with Side Information Using Turbo Codes*, in Proc. DCC 2002.

[2] J. Barros, C. Peraki, S. D. Servetto: *Efficient Network Architectures for Sensor Reachback*, in Proc. of the IEEE International Zurich Seminar on Communications, 2004.

[3] T. Berger: *Multiterminal Source Coding*, lecture notes presented at the CISM summer school, 1977.

[4] T. Berger, J.D. Gibson: *Lossy Source Coding*, IEEE Trans. on Inf. Theory, 1998.

[5] D. Bertsekas: *Network Optimization: Continuous and Discrete Models*, Athena Scientific, 1998.

[6] T.M. Cover, J.A. Thomas: *Elements of Information Theory*, John Wiley and Sons, Inc., 1991.

[7] N. Cressie: *Statistics for Spatial Data*, Wiley 1991.

[8] R. Cristescu, B. Beferull-Lozano: *Lossy Network Correlated Data Gathering with High-Resolution Coding*, journal paper submitted, 2005.

[9] R. Cristescu, B. Beferull-Lozano, M. Vetterli: *On Network Correlated Data Gathering*, in Proc. INFOCOM 2004.

[10] R. Cristescu, B. Beferull-Lozano, M. Vetterli.: *Networked Slepian-Wolf: Theory, Algorithms and Scaling Laws*, submitted to IEEE Trans. on Inf. Th., 2003.

[11] R. Cristescu, B. Beferull-Lozano, M. Vetterli: *Scaling Laws for Correlated Data Gathering*, in Proc. ISIT 2004, Chicago, IL, June 2004.

[12] R. Cristescu, B. Beferull-Lozano, M. Vetterli, R. Wattenhofer: *Network Correlated Data Gathering with Explicit Communication: NP-Completeness and Algorithms*, to appear in IEEE/ACM Trans. on Networking, 2006.

[13] R. Cristescu, M. Vetterli: *On the Optimal Density for Real-Time Data Gathering of Spatio-Temporal Processes in Sensor Networks*, in Proc. IPSN'05.

[14] R. Cristescu, M. Vetterli: *Power Efficient Gathering of Correlated Data: Optimization, NP-Completeness and Heuristics*, in Proc. MobiHoc 2003 (poster), Annapolis, MD, June 2003.

[15] M. Enachescu, A. Goel, R. Govindan, R. Motwani: *Scale Free Aggregation in Sensor Networks*, in *Algosensors*, 2004.

[16] D. Ganesan, R. Cristescu, B. Beferull-Lozano: *Power-Efficient Sensor Placement and Transmission Structure for Data Gathering under Distortion Constraints*, submitted to ACM Trans. on Sensor Networks, Oct. 2004.

[17] D. Ganesan, S. Ratnasamy, H. Wang, D. Estrin: *Coping with Irregular Spatio-Temporal Sampling in Sensor Networks*, in Proc. HotNets-II, 2003.

[18] M. Gastpar, B. Rimoldi, M. Vetterli: *To Code, or Not to Code: Lossy Source-Channel Communication Revisited*, IEEE Tr. on Inf. Th., 2003.

[19] M. Gastpar, M. Vetterli: *Power, Spatio-Temporal Bandwidth, and Distortion in Large Sensor Networks*, IEEE Journal on Selected Areas in Communications, Vol. 23, Nr. 4, April 2005.

[20] A. Goel, D. Estrin: *Simultaneous Optimization for Concave Costs: Single Sink Aggregation or Single Source Buy-at-Bulk*, ACM-SIAM Symp. on Discrete Alg., 2003.

[21] W. Rabiner Heinzelman, A. Chandrakasan, H. Balakrishnan: *Energy-Efficient Communication Protocol for Wireless Microsensor Networks*, in Proc. HICSS 2000.

[22] A. Kumar, P. Ishwar, K. Ramchandran: *On Distributed Sampling of Bandlimited and Non-Bandlimited Sensor Fields*, in Proc. ICASSP 2004.

[23] H. Larson, B. Shubert: *Probabilistic Models in Engineering Sciences*, John Wiley and Sons, Inc., 1979.

[24] S. Lindsey, C.S. Raghavendra, K. Sivalingam: *Data Gathering in Sensor Networks Using the Energy*Delay Metric*, in Proc. of IPDPS Workshop on Issues in Wireless Networks and Mobile Computing, April 2001.

[25] P. S. Maybeck, S. Peter: *Stochastic Models, Estimation, and Control*, Mathematics in Science and Engineering vol. 141, 1979.

[26] D. Marco, E. Duarte-Melo, M. Liu, D. L. Neuhoff: *On the Many-to-One Transport Capacity of a Dense Wireless Sensor Network and the Compressibility of its Data*, in Proc. IPSN 2003.

[27] J. Massey: *Joint-Source Channel Coding*, in Communication Systems and Random Process Theory, 1978.

[28] S. Pattem, B. Krishnamachari, R. Govindan: *The Impact of Spatial Correlation on Routing with Compression in Wireless Sensor Networks*, in Proc. IPSN 2004.

[29] B. Porat: *Digital Processing of Random Signals*, Prentice Hall, 1994.

[30] G. J. Pottie, W. J. Kaiser: *Wireless Integrated Sensor Networks*, Communications of ACM 2000.

[31] S. Pradhan: *Distributed Source Coding Using Syndromes (DISCUS)*, Ph.D. thesis, U.C. Berkeley, 2001.

[32] S. Pradhan, K. Ramchandran: *Distributed Source Coding Using Syndromes: Design and construction*, in Proc. DCC 1999.

[33] A. Scaglione, S.D. Servetto: *On the Interdependence of Routing and Data Compression in Multi-Hop Sensor Networks*, in MobiCom 2002.

[34] C.E. Shannon: *A Mathematical Theory of Communication*, Bell Sys. Tech. J., 1948.

[35] D. Slepian, J.K. Wolf: *Noiseless Coding of Correlated Information Sources*, IEEE Trans. Inf. Theory, 1973.

[36] S. Vembu, S. Verdu and Y. Steinberg: *The Source-Channel Separation Theorem Revisited*, IEEE Trans. Inf. Theory, 1995.

[37] R. Zamir, T. Berger, *Multiterminal Source Coding with High Resolution*, IEEE Trans. on Inf. Th., 1999.