Networked Slepian–Wolf: Theory, Algorithms, and Scaling Laws

Răzvan Cristescu, *Member, IEEE*, Baltasar Beferull-Lozano, *Member, IEEE*, and Martin Vetterli, *Fellow, IEEE*

***Abstract*—Consider a set of correlated sources located at the nodes of a network, and a set of sinks that are the destinations for some of the sources. The minimization of cost functions which are the product of a function of the rate and a function of the path weight is considered, for both the data-gathering scenario, which is relevant in sensor networks, and general traffic matrices, relevant for general networks. The minimization is achieved by jointly optimizing a) the transmission structure, which is shown to consist in general of a superposition of trees, and b) the rate allocation across the source nodes, which is done by Slepian–Wolf coding. The overall minimization can be achieved in two concatenated steps. First, the optimal transmission structure is found, which in general amounts to finding a Steiner tree, and second, the optimal rate allocation is obtained by solving an optimization problem with cost weights determined by the given optimal transmission structure, and with linear constraints given by the Slepian–Wolf rate region. For the case of data gathering, the optimal transmission structure is fully characterized and a closed-form solution for the optimal rate allocation is provided. For the general case of an arbitrary traffic matrix, the problem of finding the optimal transmission structure is NP-complete. For large networks, in some simplified scenarios, the total costs associated with Slepian–Wolf coding and explicit communication (conditional encoding based on explicitly communicated side information) are compared. Finally, the design of decentralized algorithms for the optimal rate allocation is analyzed.**

***Index Terms*—Energy efficiency, linear programming, sensor networks, shortest path tree, Slepian–Wolf coding.**

I. Introduction

A. Problem Motivation

CONSIDER networks that transport supplies among nodes. This is for instance the case of sensor networks that measure environmental data [2], [21], [24]. Nodes are supplied

Manuscript received December 27, 2003; revised May 25, 2005. This work was supported in part by the National Competence Center in Research on Mobile Information and Communications Systems (http://www.mics.org), a center supported by the Swiss National Science Foundation under Grant 5005-67322. The material in this paper was presented in part at INFOCOM 2004, Hong Kong, China, March 2004; the European Workshop on Sensor Networks, 2004, Berlin, Germany, January 2004; and the IEEE International Symposium on Information Theory, Chicago, IL, June/July 2004.

R. Cristescu was with the Laboratory for Audio-Visual Communications (LCAV), EPFL, Lausanne CH-1015, Switzerland. He is now with the Center for the Mathematics of Information, California Institute of Technology, Pasadena, CA 91125 USA (e-mail: razvanc@caltech.edu).

B. Beferull-Lozano is with the Laboratory for Audio-Visual Communications (LCAV), Swiss Federal Institute of Technology, Lausanne (EPFL), CH-1015 Lausanne, Switzerland (e-mail: Baltasar.Beferull@epfl.ch).

M. Vetterli is with the Laboratory for Audio-Visual Communications (LCAV), Swiss Federal Institute of Technology, Lausanne (EPFL), CH-1015 Lausanne, Switzerland, and with the Department of Electrical Engineering and Computer Science, University of California, Berkeley, Berkeley, CA 94720 USA (e-mail: martin.vetterli@epfl.ch).

Communicated by G. Sasaki, Associate Editor for Communication Networks.
Digital Object Identifier 10.1109/TIT.2005.858980

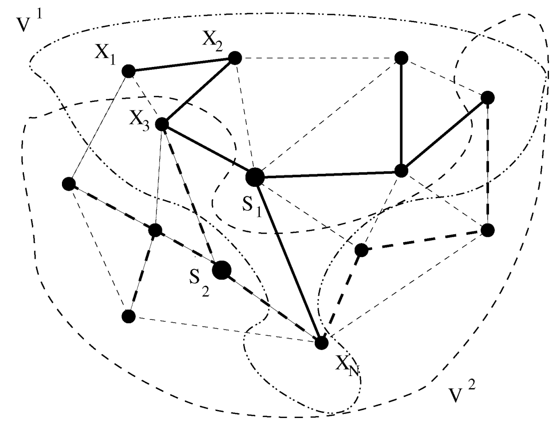


Fig. 1. An example of a network. Sources transmit their data to the sinks. Data from the sets V^1 and V^2 of sources need to arrive at sinks S_1 and S_2, respectively. A rate supply R_i is allocated to each node X_i. In thick solid and dashed lines, a chosen transmission structure is shown. In thin dashed lines, the other possible links of the network are shown.

amounts of measured data which need to be transmitted to end sites, called sinks, for control or storage purposes. The transmission topology in our model is assumed to be an undirected fully connected graph with point-to-point links. An example is shown in Fig. 1, where there are N nodes with sources $X_1, \ldots, X_N$, two of them being the sinks denoted by S_1 and S_2, and a graph of connectivity with edges connecting certain nodes. In this paper, we use interchangeably the notions of network and its graph representation. Sources corresponding to nodes in the sets V^1 and V^2 need to transmit their data, possibly using other nodes as relays, to sinks S_1 and S_2, respectively. A very important task in this scenario is to find a rate allocation at nodes and a transmission structure on the network graph that minimizes a cost function of interest (e.g., flow cost: [function (rate) $\times$ [path weight], total distance, etc.). This implies a joint treatment of source coding and optimization of the transmission structure.

The problem is trivial if the data measured at nodes are statistically independent: each node codes its data independently, and well-developed algorithms can be used to solve the minimum cost flow problem [5].

However, in many situations, data at nodes are *not* independent, such as in typical sensor networks. Thus, it can be expected that approaches that take advantage of the correlation present in the data can improve over existing algorithms, with regard to optimizing many cost functions of interest.

B. Correlated Network Data Gathering

The distributed source coding approach that exploits optimally the correlation of discrete sources is based on Slepian–

Wolf coding [29], where it is shown that when nodes generate correlated data, these data can be coded with a total rate not exceeding the joint entropy, even *without* the nodes explicitly communicating with each other (under some constraints on the rates, given by the so-called Slepian–Wolf region). This result provides the whole region of achievable rates for the rate allocation, that is, *all* the rates in that region are achievable.

In addition to encoding the data, these data need to be transmitted over the network from the sources to the sinks. In such situations, it is crucial to study the interplay between the rate allocation at the nodes and the transmission structure used to transport the data. In this work, we consider a joint treatment of the rate allocation and the chosen transmission structure, by means of cost functions that are functions of both. The cost functions usually found in practice separate the rate term from the path weight term. For instance, the [rate] $\times$ [path weight] cost function measures the transmission price in wired networks, or the power consumption in radio powered sensor networks that function at low signal-to-noise ratio [25], and the [$\exp(\text{rate})$] $\times$ [path weight] measures the power consumption in noisy point-to-point links, where the (path weight) term is a function of the total internode distance. In this work, we consider jointly the optimization of both source coding and transmission structure in the context of networks with correlated data at nodes. To the best of our knowledge, this is the first research work that addresses jointly Slepian–Wolf lossless source coding and network flow cost optimization.

Consider a network of N nodes. Let $\boldsymbol{X} = (X_1, \ldots, X_N)$ be the vector formed by the random variables representing the sources measured at the nodes $1, \ldots, N$. The samples taken at nodes are spatially correlated. We assume that each random variable X_i is taken from a discrete-time random process which is independent and identically distributed (i.i.d.) over time, and has a countable discrete alphabet (e.g., through quantization of a continuous valued random variable).[1]

For the sake of clarity, assume first that there is only a single sink, where the data from all the nodes has to arrive. A rate allocation $(R_1, \ldots, R_N)$ bits has to be assigned at the nodes so that the discretized samples are described losslessly.

In the case of sensor networks, where measurements are taken from the environment [2], [21], [24], a practical approach is to use nearest neighbor connectivity as a way to avoid the complexity of the wireless setting. Moreover, we assume that interference is addressed separately by the use of an adequate multiple-access protocol and/or the use of anisotropical antennas. In such a physical setup, the spatial correlation depends only on the distance distribution across space among nodes. Since battery power is a scarce resource for autonomous sensors, a meaningful cost function to minimize in the case of sensor networks is the total energy consumption (e.g., [rate] $\times$ [path weight]). The weight of the link between two nodes is a function of the distance d that separates the two nodes (e.g., kd^α [25] or $k\exp(\alpha d)$ [13], with k, α being constants that depend on the medium).

[1]Denote the continuous-space random process by $X_c(\boldsymbol{s})$, where $\boldsymbol{s}$ is the spatial coordinate. Then, the sources are obtained by the sampling and quantization of $X_c(\boldsymbol{s})$, namely, $X_i = X_c^q(\boldsymbol{s}_i)$, where $\{\boldsymbol{s}_i\}_{i=1}^N$ are the spatial positions of the nodes that generate the sources.

The model for a single sink can be extended to the case where there is a certain number of sinks to which data from different subsets of nodes have to be sent. Notice that in this case, it is also possible to allocate different rates at each node, depending on which sink the data is sent to, but this involves additional coding overhead, which might not be always feasible. We consider both cases in this paper, but for the sake of simplicity we concentrate our discussion on the case where there is a unique rate allocation. The sink-dependent rate allocation is a straightforward generalization of the unique rate allocation.

C. Related Work

Bounds on the performance of networks measuring correlated data have been derived in [19], [28]. On the other hand, progress toward practical implementation of Slepian–Wolf coding has been achieved in [1], [22], [23]. 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 [17], [26], 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.

A joint treatment of data aggregation and transmission structure is considered in [14]. The model in [14] does not take into account possible exploitation of common knowledge of the correlation structure, for joint coding among nodes. The novelty of our work stems from the fact that we consider the case of collaboration between nodes because we allow nodes to perform (jointly) Slepian–Wolf coding, and this is combined with the optimization of the transmission structure.

D. Main Contributions

We first show that if Slepian–Wolf coding is used in network-correlated data gathering scenarios, then the optimization separates: first an optimal transmission structure needs to be determined, and second the optimal rate allocation has to be found for this transmission structure. The optimal rate allocation is in general unique, except in some degenerate cases. We start our analysis with the important case of single-sink data gathering. We show that in this case the optimal transmission structure is the shortest path tree rooted at the sink. We fully solve the case of linear cost function by providing a closed-form solution for the rate allocation, and we analyze the complexity of the problem when the cost function has an exponential dependence on the rate.

Next, we consider the arbitrary traffic matrix case, and we prove that in this case finding the optimal transmission structure is NP-complete. Moreover, we show that, if the optimal transmission structure is approximated, then finding the optimal rate allocation is simple, by using centralized algorithms. However, in sensor network settings, the goal is to find distributed algorithms, and we show that in order to have a decentralized algorithm, we need a substantially large communication overhead in the network. We study further some particular cases of interest where the problem is more tractable. For example, in the single-sink data gathering scenario, we design a fully distributed algorithm for the rate allocation.

For some simplified scenarios, we compare the performance of the Slepian–Wolf coding, in terms of total flow cost, with another possible coding approach, explicit communication. We provide asymptotic behaviors and scaling laws for the total cost flows for various correlation structures, including band-limited processes and Gaussian random processes. We also provide the conditions on the correlation structure under which Slepian–Wolf coding in very large networks performs arbitrarily better than explicit communication.

Finally, we present an approximation algorithm for the rate allocation in the single-sink data gathering case. If the correlation structure is distance dependent, namely, the correlation between nodes decreases with the distance, then our algorithm provides solutions very close to the optimal rate allocation, while using only local neighborhood information at each node. We illustrate with numerical simulations the performance of our algorithm in the case of Gaussian random processes.

E. Outline of the Paper

The rest of this paper is organized as follows. In Section II, we state the optimization problem and describe the Slepian–Wolf source coding approach. In Section III, we consider an important particular case, namely, the single-sink correlated data gathering problem with Slepian–Wolf source coding. In Section IV, we study the complexity of the problem for the case of a general traffic matrix problem. In Section V, we apply the results obtained in Section III to other particular cases of interests. In Section VI, we briefly introduce the Gaussian random processes as spatial correlation structure. In Section VII, we study the performance of Slepian–Wolf coding in large networks, in comparison with coding by explicit communication. In Section VIII, we present an efficient decentralized algorithm for approximating the optimal rate allocation in the single-sink data gathering case, and discuss how this algorithm can be used in the scenarios presented in Section V. We present some numerical simulations for Gaussian random processes in Section IX. We conclude and present directions of further work in Section X.

II. Problem Formulation

A. Optimization Problem

Consider a graph $G=(V,E)$, $|V|=N$. Each edge $e\in E$ is assigned a weight w_e. Each node in the graph generates a source of data. Some of the nodes are also sinks. Data has to be transported over the network from sources to sinks. Denote[2] by $S_1, S_2, \ldots, S_M$ the set of sinks and by $V^1, V^2, \ldots, V^M$ the set of subsets $V^j \subseteq V$ of sources; data measured at sources V^j have to be sent to sink S_j. Denote by S^i the set of sinks to which data from node i has to be sent. Denote by $E^i \subseteq E$ the subset of edges used to transmit data from node i to sinks S^i, which determines the transmission structure corresponding to node i.

[2]In this paper, we use subindices to denote nodes and upper indices to denote sets.

Definition 1 (Traffic Matrix): We call the traffic matrix T of a graph G, the $N\times N$ square matrix T that has elements given by

$$
\begin{aligned}
T_{ij} &= 1, && \text{if source } i \text{ is to be transmitted to sink } j\\
T_{ij} &= 0, && \text{else.}
\end{aligned}
$$

With this notation, $V^j=\{i: T_{ij}=1\}$ and $S^i=\{j: T_{ij}=1\}$.

We consider cost functions which consist of a product of a function that depends on the rate and another function that depends on the link weight. The overall task is to assign an optimal rate allocation $\{R_i^*\}_{i=1}^N$ and to find the optimal transmission structure on the graph G that minimizes the total flow cost $[F(\text{rate})]\times[\text{path weight}]$, where $F(\cdot)$ is a function of the rate allocated at a node. In all practical settings, the function $F(\cdot)$ is monotonically increasing. Thus, the optimization problem is

$$\{R_i^*, d_i^*\}_{i=1}^N = \arg\min_{\{R_i, d_i\}_{i=1}^N} \sum_{i=1}^N F(R_i) d_i \tag{1}$$

where d_i is the total weight of the transmission structure given by E^i that is chosen to transmit data from source i to the set of sinks S^i

$$d_i = \sum_{e\in E^i} w_e$$

where w_e denotes the weight (cost) associated to edge e. This weight w_e is in practice a function of the distance between the nodes connected by edge e. Thus, finding the optimal $\{d_i^*\}_{i=1}^N$ amounts to finding the optimal transmission structure. Note that the rate terms in the cost function (1) depend on the individual rates allocated at nodes, rather than on the total incoming flow at nodes.

In the next section, we show that when Slepian–Wolf coding is used, the tasks of finding the optimal $\{d_i^*\}_{i=1}^N$ and $\{R_i^*\}_{i=1}^N$, respectively, are separated, that is, one can first find the optimal transmission structure, which can be shown to always be a tree, and then find the optimal rate allocation.

B. Slepian–Wolf Coding

Consider the case of two random sources X_1 and X_2 that are correlated (see Fig. 2(b)). Intuitively, each of the sources can code their data at a rate greater or equal to their respective 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 equal to the joint entropy $R_1+R_2=H(X_1,X_2)$. This can be done, for instance, by using conditional entropy, that is, $R_1=H(X_1)$ and $R_2=H(X_2\mid X_1)$, since X_1 can be made available at node 2 through explicit communication. Slepian and Wolf [29] showed that two correlated sources can be coded with a total rate $H(X_1,X_2)$ even if they are *not* able to communicate with each other. Fig. 2(b) shows the Slepian–Wolf rate region for the case of two sources. This result can be generalized to the N-dimensional case.

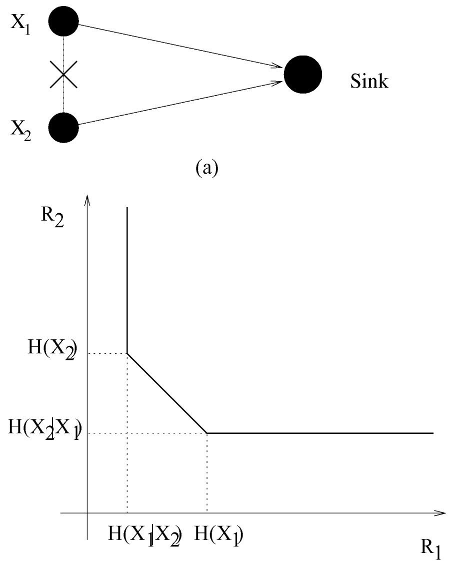


Fig. 2. Two correlated sources, and the Slepian–Wolf region for their rate allocation. (a) Two correlated sources X_1, X_2 send their data to one sink. (b) The Slepian–Wolf region shows the achievable pairs of rates that can be allocated to sources X_1 and X_2 for lossless data coding.

Consider again the example shown in Fig. 1. Assume that the set of sources that send their data to sink j, that is, the set of sources denoted

$$\{X_{j1}, \dots, X_{j|V^j|}\} \in V^j, \qquad j = 1, 2$$

know in advance the correlation structure in that set V^j. Then, nodes with sources in V^j can code their data jointly, without communicating with each other, with a total rate of $H(X_{j1}, X_{j2}, \dots, X_{j|V^j|})$ bits, as long as their individual rates obey the Slepian–Wolf constraints, which are related to the different conditional entropies [7], [29]. As a consequence of the possibility of joint source coding without sources communicating among them, we can state the following.

Proposition 1: Separation of source coding and transmission structure optimization.

If the joint cost function is separable as the product of a function that depends only on the rate and another function that depends only on the link weights of the transmission structure, and Slepian–Wolf coding is used, then, for any traffic matrix T, the overall joint optimization can be achieved by first optimizing the transmission structure with respect to only the link weights, and then optimizing the rate allocation for the given transmission structure.

Proof: Once the rate allocation is *fixed*, the best way to transport any amount of data from a given node i to the set of sinks S^i does not depend on the value of the rates. This is true because we consider separable flow cost functions, and the rate supplied at each node does not depend on the incoming flow at that node. Since this holds for any rate allocation, it is true for the minimizing rate allocation and the result follows □

For each node i, the optimal transmission structure is in fact a tree with root at node i and spanning the sinks S^i to which its data are sent [5]. Thus, the entire optimization problem can be separated into a spanning tree optimization (which is done only with respect to the link weights) for each node, and the rate allocation optimization. Then, after the optimal tree structure is formed, (1) becomes a problem of rate allocation that can be posed as an optimization problem under the usual Slepian–Wolf linear constraints, namely

$$\min_{\{R_i\}_{i=1}^N} \sum_{i=1}^{N} F(R_i) d_i^*$$

under constraints:

$$\sum_{l \in Y^j} R_l \geq H(Y^j | V^j - Y^j), \ \forall \quad V^j, Y^j \subseteq V^j \qquad (2)$$

that is, first the optimal weights $\{d_i^*\}_{i=1}^N$ are found (which determine uniquely the optimal transmission structure), and then the optimal rate allocation is found using the fixed values $\{d_i^*\}_{i=1}^N$ in (2). Note that there is one set of constraints for each set V^j.

The problem in (2) is an optimization problem under linear constraints. If the weights $\{d_i^*\}_{i=1}^N$ can be determined, the optimal allocation $\{R_i^*\}_{i=1}^N$ can be found easily with a *centralized* algorithm, by using Lagrange multipliers [18] in general for any monotonically increasing function $F(\cdot)$, or using linear programming in the case where $F(\cdot)$ is a linear function.

In Section III ,we show that for the case of single-sink data gathering, the weights $\{d_i^*\}_{i=1}^N$ correspond to the shortest paths from the nodes to the single sink, and thus they can be determined, with standard algorithms, in polynomial time. However, in Section IV, we show that for a general traffic matrix T, finding the optimal coefficients $\{d_i^*\}_{i=1}^N$ is NP-complete. Moreover, in general, even if the optimal structure is found, it is hard to *decentralize* the algorithm that finds the optimal solution $\{R_i^*\}_{i=1}^N$ of (2), as this implies a substantial amount of global knowledge of the network. As shown in Section III-B, a decentralized solution requires each node to know the total weight $\{d_i^*\}_{i=1}^N$ corresponding to all the other nodes.

Note that if the rate allocation at a source X_i is allowed to take different values R_{ij} depending on which sink j source X_i sends its data to, then it is straightforward to find this multiple rate assignment. In this case, the rate allocation for each cluster is independent from the rate allocations in the other clusters (see Sections III and VIII). However, this involves even more additional complexity in coding, so in many situations it might be desirable to assign a unique rate to each node, regardless of which sink the data is sent to. Thus, without loss of generality, in this work we concentrate on the study of the case where a single rate allocation is done at each node, regardless of the number of sinks to where data measured at that node has to arrive.

In the following sections, for various problem settings, we first show how the transmission structure can be found (i.e., the values of $\{d_i^*\}_{i=1}^N$), and then we discuss the complexity of finding the rate allocation (2) in a decentralized manner. We begin our analysis with a particular case widely encountered in practice, namely, the single-sink data gathering problem. In this case, there is only one sink to which data from all the other nodes have to arrive. We fully solve this case in Section III, and this result also provides useful insight into the structure of the general problem.

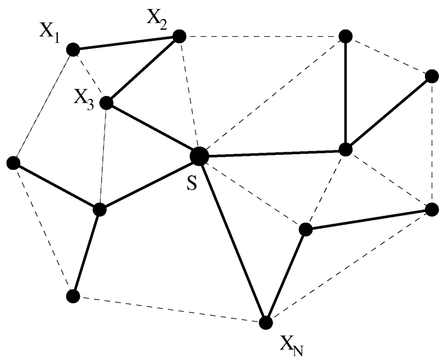


Fig. 3. In this example, data from all the nodes $X_1, X_2, \ldots, X_N$ need to arrive at sink S.

III. Single-Sink Data Gathering

In single-sink data gathering, the entire set of N sources ($V^j = V$) are sent to a single sink $S = j$. An example is shown in Fig. 3. First, we state the following corollary of Proposition 1, valid for the case of the single-sink data gathering scenario.

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 SPT rooted in S is optimal, in terms of minimizing (2), for any rate allocation.

Proof: Once the rate allocation is *fixed*, the best way to transport the data from any node to the sink is to use a shortest path. Minimizing the sum of costs under constraints in (2) becomes equivalent to minimizing the cost corresponding to each node independently. Since the SPT is a superposition of 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 □

Thus, once the SPT is found, which provides $\{d_i^*\}_{i=1}^N$, the overall optimization reduces to solving (2). We consider two important types of functions of the rate in the cost function, which are widely used in practice. Namely, in Section III-B, we study the case where $F(R) = R$, which corresponds to wired networks and power constrained sensor networks, and in Section III-C, the case where $F(R) = e^R$, which corresponds to point-to-point links in noisy wireless networks.

A. Linear Cost Function: Optimal Solution

For the single-sink data gathering, the cost function in (2) that has to be minimized can be rewritten as

$$\sum_{i \in V} R_i d_{ST}(i, S) \tag{3}$$

where $d_{ST}(i, S)$ is the total weight of the path connecting node i to S on the spanning tree ST, and the constraints are given by

$$\sum_{i \in Y} R_i \geq H(Y|Y^C) \tag{4}$$

for any of the $2^N - 1$ subsets $Y \subseteq V$ (see Fig. 2(b) for the achievable rate region corresponding to two sources).

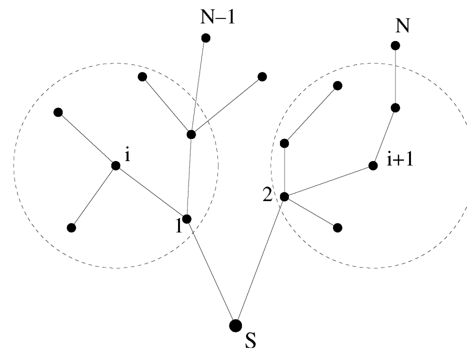


Fig. 4. Nodes are ordered $1, \ldots, N$ with increasing total weight of their path in the SPT to the sink S. For an optimal rate allocation, node $i+1$ is assigned an entropy rate obtained by conditioning its data on data measured at nodes i, $i-1, \ldots, 1$. Even if nodes $i+1$ and i are consecutive in the ordering, they are not necessarily in each other's neighborhood.

From Corollary 1, it follows that the minimization of (3) becomes now a linear programming (LP) problem

$$\{R_i^*\}_{i=1}^N = \arg \min_{\{R_i\}_{i=1}^N} \sum_{i \in V} R_i d_{\mathrm{SPT}}(i, S) \tag{5}$$

under constraints (4).

Suppose, without loss of generality, that nodes are numbered in increasing order of the total weight of their path to the sink on the SPT, that is,

$$(X_1, X_2, \ldots, X_N)$$

with

$$d_{\mathrm{SPT}}(X_1, S) \leq d_{\mathrm{SPT}}(X_2, S) \leq \cdots \leq d_{\mathrm{SPT}}(X_N, S).$$

Thus, nodes X_1 and X_N are, respectively, the nodes corresponding to the smallest and the largest total weight in the SPT to the sink. A network example with nodes numbered as above is shown in Fig. 4.

Theorem 1 (LP Solution): The solution of the optimization problem given by (5) under constraints (4) is [10]

$$\begin{aligned} R_1^* &= H(X_1) \\ R_2^* &= H(X_2|X_1) \\ \cdots &\quad \cdots \quad \cdots \\ R_N^* &= H(X_N|X_{N-1}, X_{N-2}, \ldots, X_1). \end{aligned} \tag{6}$$

We prove Theorem 1 in Appendix I.

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.

Fig. 5 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 (6) 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., $d_{\mathrm{SPT}}(X_1, S) = d_{\mathrm{SPT}}(X_2, S)$, in Fig. 5)

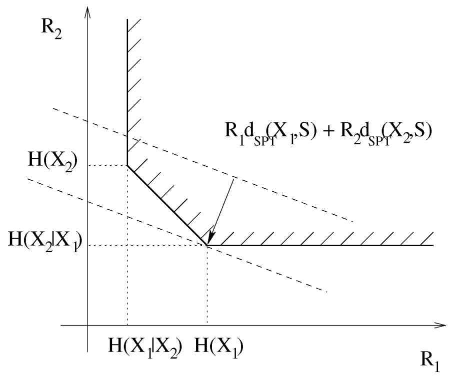


Fig. 5. A simple example with two nodes. The total weights from sources X_1, X_2 to the sinks, are, respectively, $d_{\rm SPT}(X_1, S)$, $d_{\rm SPT}(X_2, S)$, $d_{\rm SPT}(X_1, S) < d_{\rm SPT}(X_2, S)$, in this particular case. In order to achieve the minimization, the cost line $R_1 d_{\rm SPT}(X_1, S) + R_2 d_{\rm SPT}(X_2, S)$ has to be tangent to the most interior point of the Slepian–Wolf rate region, given by $(R_1, R_2) = (H(X_1), H(X_2|X_1))$.

then the solution of (6) is not unique, since the cost function is parallel to one of the faces of the Slepian–Wolf region.

Even if the solution can be provided in the closed form (6), 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 (see Fig. 4) the following.

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 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

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

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.

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, notice 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 Section VIII, 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.

B. Cost Function With an Exponential Dependence on the Rate

We consider now the cost function $[\exp(\text{rate})] \times [\text{path weight}]$. This cost function is typical in the case of noisy wireless networks with point-to-point links.[3] Note that the (path weight) in this case is the sum of link weights, each of them depending on a power α of the corresponding Euclidean distance (usually $\alpha \in [2, 4]$).

We consider the case when relay nodes transmit data to be forwarded sequentially as it arrives, rather than waiting to aggregate the whole incoming data into a single packet. In other words, each rate R_i (coming from source X_i) is associated to a single data packet. Namely, we consider the minimization of (1), that is a cost function in which the term corresponding to each node i depends on the individual rate R_i allocated at that node, rather than on the total sum of rates forwarded by the node. Note that this is a reasonable assumption for the type of scenarios considered in this paper, when some scheduling policy is employed to avoid wireless interference [15]. In such scenarios, a node need not wait for all incoming data before making a transmission, but rather data should be sent a packet at a time by forwarding nodes.[4] Moreover, in the case of a cost function exponential in the rate term, this strategy is usually more power efficient (e.g., for R_i, $R_j > 1$, $\exp(R_i + R_j) > \exp(R_i) + \exp(R_j)$). Optimization with nonlinear cost functions of the link flow is significantly more difficult and is not considered further in this paper.

Thus, assuming a cost function which is exponential in the rate, the optimization problem (3) becomes now

$$\{R_i^*\}_{i=1}^N = \arg \min_{\{R_i\}_{i=1}^N} \sum_{i \in V} e^{R_i} d_{\rm SPT}(i, S) \tag{7}$$

under the Slepian–Wolf constraints (4).

By making the change of variables $P_i = e^{R_i}$, for all $i = 1, \ldots, N$, (7) becomes equivalent to

$$\{P_i^*\}_{i=1}^N = \arg \min_{\{P_i\}_{i=1}^N} \sum_{i \in V} P_i d_{\rm SPT}(i, S)$$

under the constraints

$$\sum_{i \in Y} \log P_i \geq H(Y|Y^C)$$

for all $Y \subseteq X$.

For the sake of clarity, consider the simple case of two sources. The problem in this case is

$$\{P_1^*, P_2^*\} = \arg \min_{P_1, P_2} \left(P_1 d_{\rm SPT}(1, S) + P_2 d_{\rm SPT}(2, S)\right) \tag{8}$$

under the constraints

$$\begin{aligned} P_1 &\geq e^{H(X_1|X_2)} \\ P_2 &\geq e^{H(X_2|X_1)} \\ P_1 P_2 &\geq e^{H(X_1, X_2)}. \end{aligned} \tag{9}$$

[3]We can write the maximum rate that can be transmitted over a wireless channel as [7] $R = \frac{1}{2} \log(1 + \frac{\mathcal{P}_s}{\mathcal{P}_n})$ where $\mathcal{P}_s$ is the power of the transmitted signal, and $\mathcal{P}_n$ is the power of the noise at the receiver; in practice, $\mathcal{P}_n$ depends on the transmission distance. Then $\mathcal{P}_s \approx e^R \times \mathcal{P}_n$.

[4]The related latency problem where additional delay constraints are considered is outside the scope of this paper.

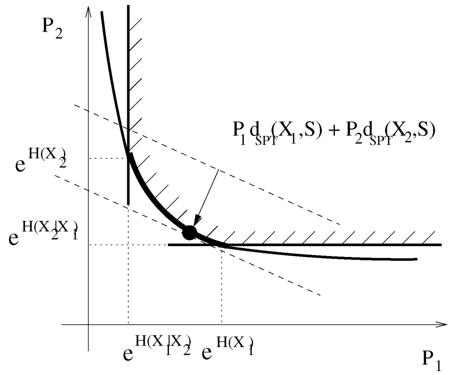


Fig. 6. The same example with two nodes. The minimization is achieved at the point where the cost line $P_1 d_{\rm SPT}(X_1, S) + P_2 d_{\rm SPT}(X_2, S)$ is tangent to the transformed Slepian–Wolf rate region.

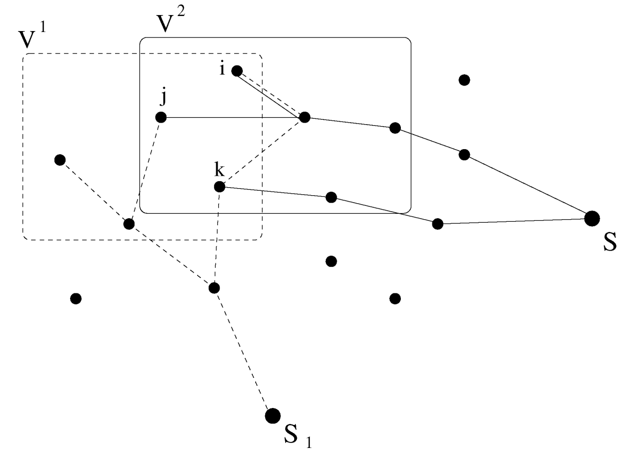


Fig. 7. A network example with two sinks. Data from the subsets V^1 (dotted) and V^2 (solid) have to be transmitted to sinks S_1 and S_2, respectively. For each of the nodes i, j and k, the transmission structure is the Steiner tree covering the respective node and both sinks S_1 and S_2. For the rest of the nodes, the transmission structure is the shortest path from that node to the corresponding sink. The overall optimal transmission structure is the superposition of all these individual Steiner trees.

The solution of this minimization problem is illustrated in Fig. 6. Notice that, with the change of variables, the cost function remains linear, while the constraints become multiplicative instead of additive.

We observe that the rate allocation depends on the slope of the cost function, that is, on the total weights $d_{\rm SPT}(i, S)$. Depending on these weights, the tangent point of the cost function to the rate region curve can be either on the joint constraint curve given by the third constraint in (9) (thicker curve in Fig. 6), which corresponds to having active only the third constraint in (9), or on one of the corner points, which corresponds to one of the two first constraints in (9) being active only. As the number of sources N increases, it becomes cumbersome to provide a direct closed form for the optimal rate allocation, although it can be obtained systematically using Lagrangian optimization. For the simple case of two sources, depending on the ratio $\eta = d_{\rm SPT}(X_1, S)/d_{\rm SPT}(X_2, S)$, the optimal rate allocation is given in the equation at the bottom of the page. We show in Appendix II how this rate allocation is obtained.

We have studied so far the case when there is a single sink. For simplicity, we limit the discussion of the next sections to the case of a *linear* separable cost function, namely, $F(R) = R$.

IV. Arbitrary Traffic Matrix

We begin the analysis with the most general case, that is, when the traffic matrix T is arbitrary, by showing the following proposition [9].

Proposition 2: The optimal transmission structure for a general traffic matrix T is a superposition of Steiner trees: Given an arbitrary traffic matrix T, if $F(R) = R$ then, a) for any node i, the optimal value d_i^* in (2) is given by the minimum-weight tree rooted in node i and which spans the nodes in S^i; this is by definition the minimum Steiner tree that has node i as root and which spans S^i, and b) the problem is NP-complete.

Proof: The proof is straightforward: first, notice that we can optimize the transmission structure regardless of the rate allocation. Data from node i has to be sent over the transmission structure with minimum total weight to the nodes in S^i, possibly via nodes with sources in $V - \{i, S^i\}$ (see Fig. 7). This is a minimum Steiner tree problem for the graph G, thus, it is NP-complete □

The approximation ratio of an algorithm that finds an approximate solution for an optimization problem is defined as the guaranteed ratio between the cost of the approximate solution and the optimal one [6]. If the weights of the graph are the Euclidean distances ($w_e = l_e$ for all $e \in E$), then the problem becomes the Euclidean Steiner tree problem, and it admits a polynomial time approximation scheme (PTAS) [3] (that is, for any $\epsilon > 0$, there is a polynomial time approximation algorithm with an approximation ratio of $1 + \epsilon$). However, in general, the link weights are not the Euclidean distances (e.g., if $w_e = l_e^2$ etc.). Then finding the optimal Steiner tree is APX-complete (that is, there is a hard lower bound on the approximation ratio that can be achieved by any polynomial time approximation algorithm), and is only approximable (with polynomial time in the input instance size, namely, the number of nodes) within a constant factor of $(1 + \ln 3)/2$ [4], [27].

Once the optimal weights $\{d_i^*\}_{i=1}^N$ are found (i.e., approximated by some approximation algorithm for solving the Steiner tree), then, as mentioned in Proposition 2, (2) becomes an LP problem and the solution of this problem is given by a corner of the Slepian–Wolf region (see Fig. 5). Consequently, it can be readily solved in a simple way with a centralized algorithm, where global knowledge of the network is allowed. However, it

$$(R_1, R_2) = \begin{cases} (H(X_1), H(X_2|X_1)), & \text{if } \eta \le e^{H(X_2|X_1) - H(X_1)} \\ (\frac{1}{2}(H(X_1, X_2) - \log \eta, H(X_1, X_2) + \log \eta)), & \text{if } e^{H(X_2|X_1) - H(X_1)} < \eta < e^{H(X_2) - H(X_1|X_2)} \\ (H(X_1|X_2), H(X_2)), & \text{if } \eta \ge e^{H(X_2) - H(X_1|X_2)}. \end{cases}$$

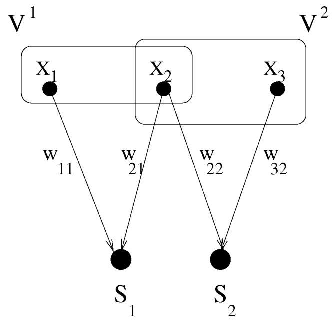


Fig. 8. Two sets of sources transmit their correlated data to two sinks.

is not possible in general to find a closed-form expression for the optimal rate allocation solution. Notice also that the derivation of a decentralized algorithm for the rate allocation optimization is difficult, as this involves exchange of network knowledge among the subsets of nodes.

Example: Fig. 8 shows a simple example, but sufficiently complete to illustrate the impossibility of solving this problem without having global knowledge of the network structure. Suppose that the optimal total weights $\{d_i^*\}_{i=1}^3$ in (2) have been approximated by some algorithm. Then, the cost function to be minimized, by finding the optimal rates $\{R_i\}_{i=1}^3$, is

$$R_1 w_{11} + R_2(w_{21} + w_{22}) + R_3 w_{32}$$

with $d_1^* = w_{11}$, $d_2^* = w_{21} + w_{22}$, $d_3^* = w_{32}$, and the Slepian–Wolf constraints are given by

$$\begin{aligned} R_1 + R_2 &\geq H(X_1, X_2) \\ R_1 &\geq H(X_1|X_2) \\ R_2 &\geq H(X_2|X_1) \end{aligned}$$

for set $V^1 = \{X_1, X_2\}$, and, respectively,

$$\begin{aligned} R_2 + R_3 &\geq H(X_2, X_3), \\ R_2 &\geq H(X_2|X_3), \\ R_3 &\geq H(X_3|X_2) \end{aligned}$$

for set $V^2 = \{X_2, X_3\}$.

Suppose the weights are such that $w_{11} < w_{21} + w_{22} < w_{32}$. A decentralized algorithm has to use only local information, that is, information only available in a certain local range or neighborhood. In this case, this means that each node only knows a) the Euclidean distance to its neighbors in its corresponding cluster and b) the total weights d_i^* from its neighbors to the corresponding sinks. Thus, we assume that only local Slepian–Wolf constraints are considered in each set V^j for the rate allocation, and for the nodes in one subset no knowledge about the total weights d_i^* from nodes in other subsets is available. Then, as we have seen in Section III, by solving each of the two LP problems, it readily follows that the optimal rate allocations in each of the two subsets are, respectively, given by

$$\begin{aligned} R_1^1 &= H(X_1) \\ R_2^1 &= H(X_2|X_1), \qquad \text{for set } V^1 \end{aligned}$$

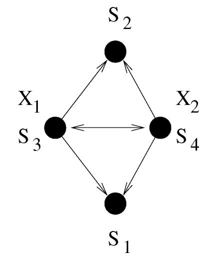


Fig. 9. Data from X_1, X_2 need to be transmitted to all nodes S_1, S_2, S_3, S_4.

$$\begin{aligned} R_2^2 &= H(X_2) \\ R_3^2 &= H(X_3|X_2), \qquad \text{for set } V^2 \end{aligned}$$

where the upper index indicates the corresponding set.

If each node can send a different rate depending on the sink it has to send a certain data to, then, there is no problem in decentralizing the algorithm for finding the optimal rate allocation. But this implies, on the other hand, more complexity for the coding at each node.

Thus, we can see from this simple example that we obtain different values for the rate associated to a node (in this particular case, node 2), depending on the set. Unless every node (not only node 2 in this particular case) has global knowledge of the total weights from nodes 1, 2, 3 to the sinks S_1 and S_2, which allows for getting the necessary global ordering in total weight, it is not possible to find the optimal rate allocation for this problem. As explained in Section III-B, the optimal rate allocation involves solving an LP problem which depends jointly on all the weights w_{11}, w_{21}, w_{22}, w_{32}. Therefore, from this example we can see that for this network topology, it is necessary to have global information at each node. The same problem will appear in any network topology where $\exists j, k$ with $V^j \cap V^k \neq \emptyset$. This makes it necessary to have an additional important communication overhead (which grows exponentially with the number of nodes) in order to transfer all this topology information before the nodes can start encoding.[5]

There are however some other important special cases of interest where the problem is more tractable, from a decentralized point of view, and we treat them in the following section.

V. Other Cases of Interest

A. Broadcast of Correlated Data

This case corresponds to the scenario where a set of L sources are sent to all nodes ($S^i = V$). A simple example is shown in Fig. 9. In this case, the traffic matrix is $T_{ij} = 1$, for $i \in \{i_1, \ldots, i_L\} \subset V$ and $1 \leq j \leq N$.

In this case, for any node i, the value d_i^* in (2) is given by the tree of minimum weight which spans V; this is the minimum spanning tree (MST), and thus, by definition, it does not depend on i. Thus, in this case all weights $\{d_i^*\}_{i=1}^N$ are equal. Notice that this case trivially includes the typical single-source broadcast scenario where one node transmits its source to all the nodes in the network.

[5]We do not consider the optimization of this overhead in this paper.

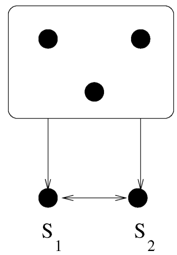


Fig. 10. Data from all nodes have to be transmitted to the set of sinks $S^a = \{S_1, S_2\}$. Each sink has to receive data from *all* the sources.

B. Multiple-Sink Data Gathering

This case corresponds to the scenario where all sources ($V^j = V$) are sent to some set S^a of sinks. In this case (see Fig. 10), finding the optimal weights $\{d_i^*\}_{i=1}^N$ is as difficult as in the arbitrary matrix case presented in Section IV, because for every i, the optimal weight d_i^* is equal to the weight of the minimum Steiner tree rooted at i and spanning the nodes in the set S^a.

However, given the optimal transmission structure and assuming global knowledge of the network at each node, the optimal rate allocation can be easily found in a similar manner as in Section III-B. First, we order the nodes by increasing distance $d_1^* < d_2^* < \cdots < d_N^*$, and then the optimal rate allocation is as given in (6).

C. Localized Data Gathering

This case corresponds to the scenario where data from *disjoint* sets $\{V^1, V^2, \ldots, V^L\}$ are sent to some sinks $\{S_1, S_2, \ldots, S_L\}$. In this case, for each node i, the solution for the optimal weights $\{d_i^*\}_{i=1}^N$ is again the corresponding Steiner tree rooted at i and that spans S^i. Assuming d_i^* (or an approximate value) is found, then the optimal rate allocation can be found for each set $\{V^j\}_{j=1}^L$, in the same way as in Section III-B, that is, we solve L LP programs independently. Notice that in this case, since $V^j \cap V^k = \emptyset, \forall j \neq k$, it is possible (as opposed to the example in Section IV) to solve this problem in a decentralized manner, that is, assuming that each node in a set V^j has only knowledge about the nodes in that set.

Algorithm 1: Optimization for localized data gathering.

- For each set V^j, order nodes $\{i, i \in V^j\}$ as a function of the total weight $\{d_i^*\}_{i=1}^N$.
- Assign rates in each V^j as in (6), taking into account this order.

Thus, we see that in this case, although finding the optimal transmission structure is NP-complete, if a good approximation for the transmission structure can be determined, then the optimal rate allocation is straightforward.

In our further discussion, we make use of Gaussian random processes as our source model. We describe it briefly for the sake of completeness.

VI. Source Model: Gaussian Random Processes

A model frequently encountered in practice is the Gaussian random field. This has also the nice property that the dependence in data at different nodes is fully expressed by the covariance matrix $\boldsymbol{K}$, which makes it more suitable for analysis. Thus, we assume a jointly Gaussian model for the spatial data $\boldsymbol{X}$ measured at nodes, with an N-dimensional multivariate normal distribution $G_N(\boldsymbol{\mu}, \boldsymbol{K})$

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

where $\boldsymbol{K}$ is the covariance matrix (positive definite) of $\boldsymbol{X}$, and $\boldsymbol{\mu}$ the mean vector. The diagonal elements of $\boldsymbol{K}$ are the individual variances $K_{ii} = \sigma_i^2$. The off-diagonal elements K_{ij}, $i \neq j$, depend on the distance between the corresponding nodes (e.g., $K_{ij} = \sigma^2 \exp(-c d_{i,j}^\beta)$, with $\beta \in \{1, 2\}$). Then, for any index combination $I = \{i_1, \ldots, i_k\} \subseteq \{1, \ldots, N\}$, $k \leq N$, $\boldsymbol{W}_I = (X_{i_1}, \ldots, X_{i_k})$ is k-dimensional normally distributed with covariance matrix being the submatrix $\boldsymbol{K}[I]$ selected from $\boldsymbol{K}$, with rows and columns corresponding to $\{i_1, \ldots, i_k\}$.

Without loss of generality, we use here differential entropy instead of the usual entropy, since we assume that the data at all nodes is quantized independently with the same quantization step, and the differential entropy differs from the usual entropy by a constant for uniformly quantized variables [7], assuming a sufficiently small stepsize. Let $Y = (X_{i_1}, \ldots, X_{i_k})$ be a set of k discrete random variables, obtained by performing independent quantization with stepsize Δ. Then, it follows that [7]

$$H(Y) \approx h(G_k(\boldsymbol{\mu}, \boldsymbol{K})) - k \log \Delta, \qquad \text{as } \Delta \to 0$$

where the differential entropy of a k-dimensional multivariate normal distribution $G_k(\boldsymbol{\mu}, \boldsymbol{K})$ is

$$h(G_k(\boldsymbol{\mu}, \boldsymbol{K})) = \frac{1}{2} \log(2\pi e)^k \det(\boldsymbol{K}).$$

In this paper, we use this approximation by considering a sufficiently small stepsize Δ.

The Slepian–Wolf constraints (4) can readily be expressed as

$$\begin{aligned} h(Y|Y^C) &= h(Y, Y^C) - h(Y^C) \\ &= \frac{1}{2} \log \left((2\pi e)^{N-|Y^C|} \frac{\det(\boldsymbol{K})}{\det(\boldsymbol{K}[Y^C])} \right) \end{aligned}$$

where $\boldsymbol{K}[Y^C]$ is the selected matrix out of $\boldsymbol{K}$, with indices corresponding to the elements of Y^C, respectively. Notice that the matrix notation $\boldsymbol{K}[Y]$ implies that a certain ordering is done; this is valid for *any* ordering for the nodes.

This natural correlation model is useful for us because our approximation algorithm can be easily tested. Consider, for example the case where the correlation decays exponentially with the distance. Then, the performance of our approximation algorithm is close to optimal even for small neighborhoods $\mathcal{N}(i)$.

VII. 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], [8]. 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 (as opposed to

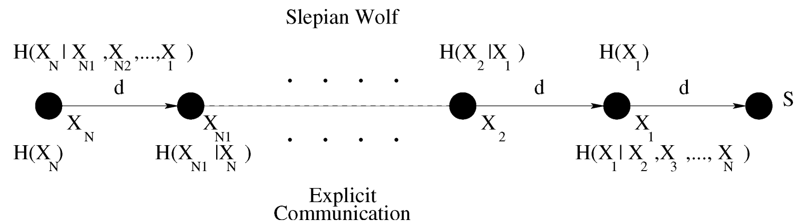


Fig. 11. A one-dimensional example: The rate allocations for Slepian–Wolf (above the line) and explicit communication (below the line).

Slepian–Wolf coding where no communication among nodes is required for joint optimal rate allocation) is available. We study the complexity of joint rate allocation and transmission structure optimization with explicit communication in a separate work [12], [8].

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 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 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 Fig. 11). 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 (3) 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 internode 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 *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 internode distance goes to zero).

As mentioned in Section VI, 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 $\boldsymbol{X} = (X_1, \ldots, X_N)$ where $X_i = X_c(id)$ and where the correlation structure for the vector $\boldsymbol{X}$ 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 $\boldsymbol{X}$ measured at the nodes has an N-dimensional multivariate normal distribution $G_N(\boldsymbol{\mu}, \boldsymbol{K})$. In particular, we consider two classes of random processes.

(a) Non-band-limited processes, namely (a.1): $K_{ij} = \sigma_{ij}^2 \exp(-c|d_{i,j}|)$, which corresponds to a regular continuous-space process [16], and (a.2): $K_{ij} = \sigma_{ij}^2 \exp(-c|d_{i,j}|^2)$, which corresponds to a singular continuous-space process [16], where $c > 0$.

(b) Band-limited 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[6] [16].

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 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 $\gamma(N)$ 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,

$$\gamma(N) = \frac{\text{cost}_{SW}(N)}{\text{cost}_{EC}(N)} = \frac{\sum_{i=1}^{N} i a_i}{\sum_{i=1}^{N} (N-i+1) a_i}. \tag{10}$$

Then, the following theorem holds [11].

Theorem 2 (Scaling Laws): Asymptotically, we have the following results.

(i) If $\lim_{i \to \infty} a_i = C > 0$:
- $\lim_{N \to \infty} \gamma(N) = 1$,
- $\text{cost}_{SW}(N) = \Theta(\text{cost}_{EC}(N))$.

[6]Actually, it can be shown that the same singularity property holds as long as $S_{X_c}(\Omega) = 0$ on some frequency interval of nonzero measure [16].

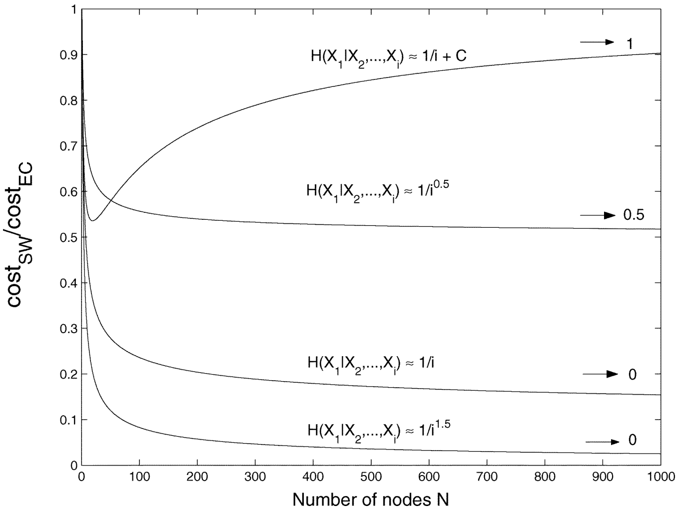


Fig. 12. Typical behavior of the ratio of the total costs $\mathrm{cost}_{SW}(N)/\mathrm{cost}_{EC}(N)$.

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

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

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

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

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

We prove Theorem 2 in Appendix III. In Fig. 12, we show typical behaviors of the ratio of total flow costs for the two coding approaches.

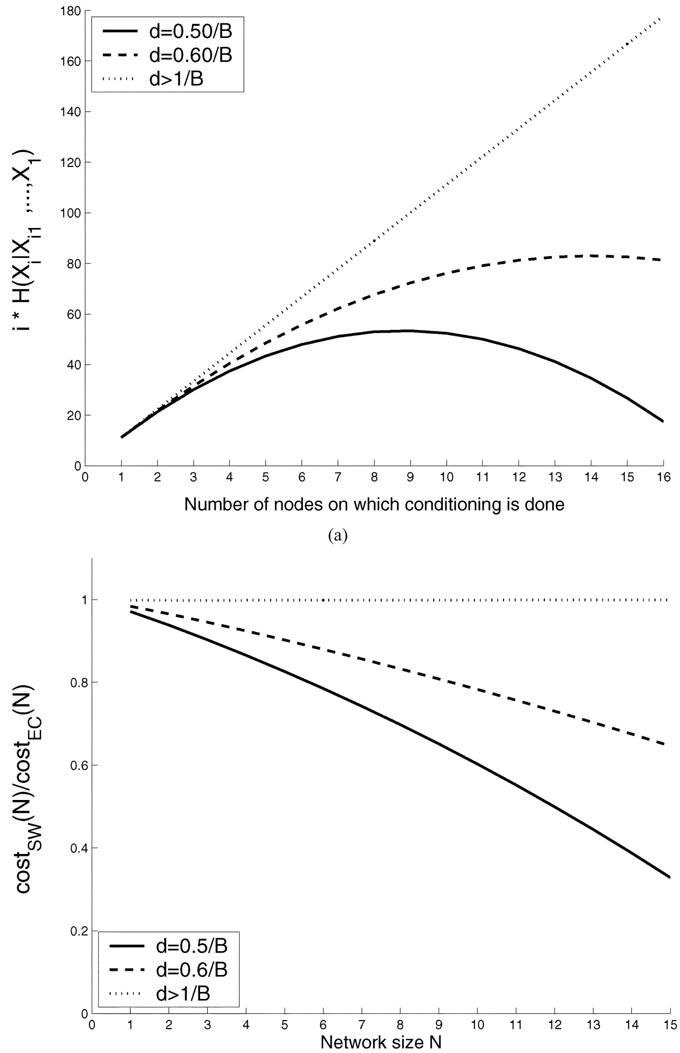
We apply now Theorem 2 to the correlation models we consider in this paper.

- For an expanding network: In cases (a.1) and (a.2), the result of sampling is a discrete-space regular process [20], thus, $\lim_{i\to\infty} a_i = C > 0$, and it follows that $\lim_{N\to\infty} \gamma(N) = 1$. In case (b), if the spatial sampling period d is smaller than the Nyquist sampling rate $1/B$ of the 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 Fig. 13, we show the bandlimited example with correlation $K(\tau) = B\mathrm{sinc}(B\tau)$. 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 Appendix IV 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 Fig. 14), then it follows from Theorem 2 that in the limit, the ratio of total costs is $\lim_{N\to\infty} \gamma(N) = 1$. In case (a.2), a closed-form expression for the conditional entropy is difficult to derive. However, we show numerically in Fig. 14(a) that in this case a_i decreases faster[7] than $1/i$. For comparison purposes, we show in Fig. 14(a) also the behavior for case (a.1). Thus, from Theorem 2, $\lim_{N\to\infty} \gamma(N) = 0$. In Fig. 14(b), we also plot the ratio of total costs for both correlation models. Finally, in Fig. 14(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 Fig. 15 how the ratio of costs starts to decrease as soon as $d < 1/B$, thus, $\lim_{N\to\infty} \gamma(N) = 0$.



(a)

(b)

Fig. 13. Expanding network sampling a band-limited process with correlation model given by $K(\tau) = B\mathrm{sinc}(B\tau)$. (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 $\mathrm{cost}_{SW}(N)/\mathrm{cost}_{EC}(N)$ as a function of the size of the network.

[7]Since these two processes are both non-band-limited, sampling them results in discrete-space regular processes [20]. However, the sampled model (a.2) inherits a "superior predictability" than (a.1), which makes a_i decrease faster than $1/i$.

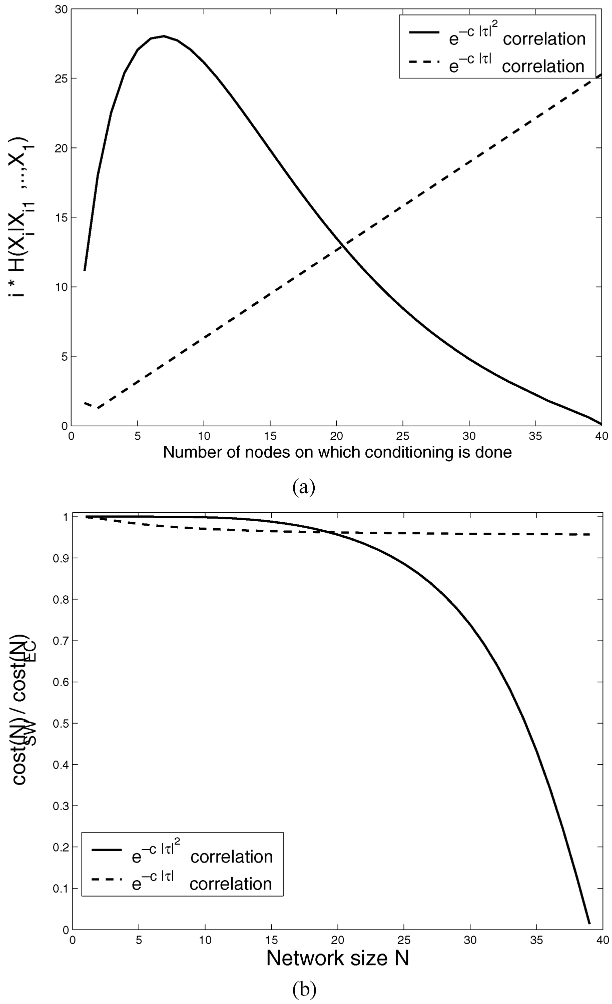


Fig. 14. We consider a correlation dependence on the internode distance d given by $\exp(-c|\tau|^{\beta})$, $\beta \in \{1, 2\}$. (a) The conditional entropy $H(X_i|X_{i-1}, \ldots, X_1)$ decreases faster than $1/i$ for $\beta = 2$, but is constant for $\beta = 1$ (after $i \geq 2$). (b) The behavior of the ratio of total $\text{cost}_{SW}(N)/\text{cost}_{EC}(N)$ with as a function of the size of the network.

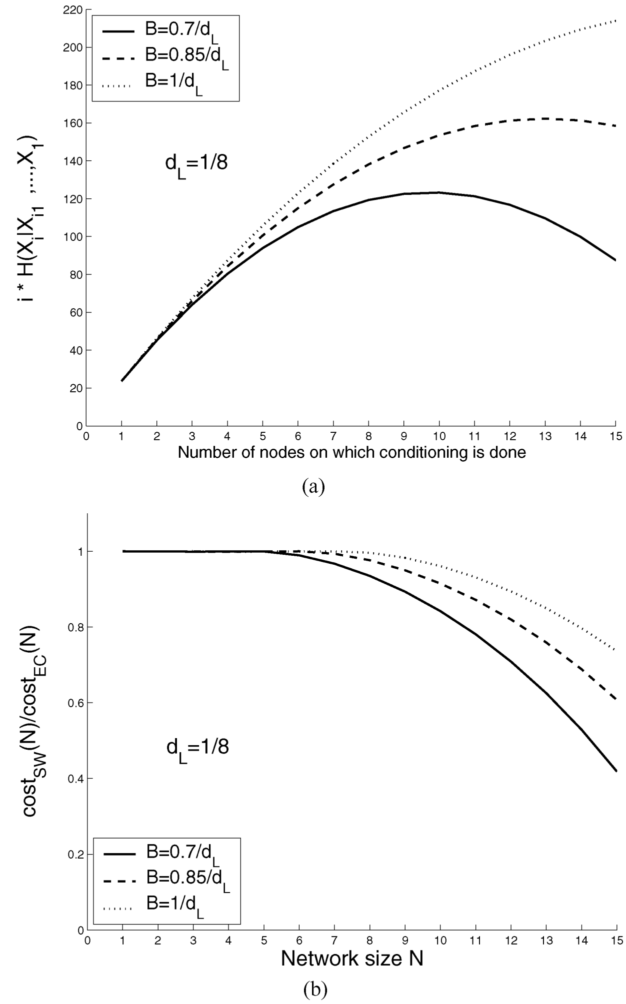


Fig. 15. Refinement network sampling a band-limited process; we denote the reference bandwidth with $B_L = 1/d_L$. (a) The conditional entropy $H(X_i|X_{i-1}, \ldots, X_1)$ decreases faster than $1/i$ as soon as $B < 1/d$, that is, $N > 1/d_L$. (b) The behavior of the ratio of total $\text{cost}_{SW}(N)/\text{cost}_{EC}(N)$ as a function of the size of the network.

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 Fig. 16). The same analysis as in the one-dimensional case holds, with the additional twist that, according to Theorem 1, 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 Fig. 16). 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 Fig. 16). Thus, the ratio of costs $\gamma(N)$ in the two-dimensional case is upper-bounded by its counterpart in the one-dimensional case, which means that the results of Theorem 2 apply for the two-dimensional case as well.

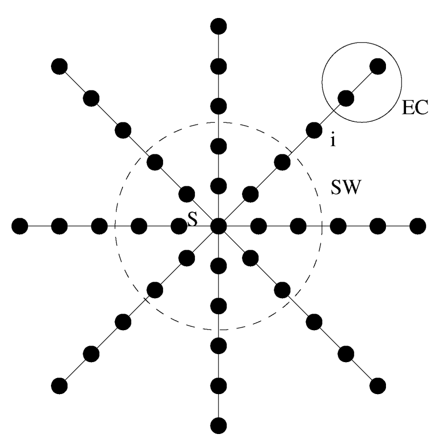


Fig. 16. A two-dimensional network with a wheel structure, with the sink S in the center. Slepian-Wolf coding for node i is done by conditioning on the nodes in the dashed region (denoted by *SW*). Explicit communication coding for node i is done by conditioning on nodes in the solid region (denoted by *EC*).

VIII. Heuristic Approximation Algorithms

In Section III-B, we found the optimal solution of the LP problem for the rate assignment under the Slepian-Wolf constraints. In this section, for the sake of clarity, 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). 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, under the assumption that local correlation is dominant. For instance, in sensor networks, this is a natural assumption, since usually the correlation decreases with the increase of the distance between nodes.

Algorithm 2: Approximated Slepian-Wolf coding:

- For each node i, set the neighborhood range k (only k-hop neighbors).
- Find the SPT using a distributed Bellman–Ford algorithm.
- For each node i, using local communication, obtain all the information from the neighborhood $\mathcal{N}_k(i)$:
  - — find in the neighborhood $\mathcal{N}_k(i)$ the set $\mathcal{C}_i$ of nodes that are closer to the sink, on the SPT, than the node i itself;
  - — transmit at rate $R_i^\dagger = H(X_i | X_j, j \in \mathcal{C}_i)$.

This means that data are coded locally at the node with a rate equal to the conditional entropy, 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 (see Section IX for numerical experiments).

Although we only consider in this section the single-sink data gathering case, similar techniques can be used to derive decentralized heuristic approximation algorithms for all network scenarios where $V^j \cap V^k = \emptyset, \forall j \neq k$, and each node from a set can only send to one sink.

IX. Numerical Simulations

We present numerical simulations that show the performance of the approximation algorithm introduced in Section VIII, for the case of single-sink data gathering. We consider the stochastic data model introduced in Section VI, given by a multivariate Gaussian random field, and a correlation model where the internode correlation decays exponentially with the distance between the nodes.

More specifically, we use an exponential model of the covariance $K_{ij} = \exp(-c d_{i,j}^2)$, where $d_{i,j}$ denotes the distance between nodes i and j, and several values for the correlation exponent c. The weight of an edge (i,j) is $w_{i,j} = d_{i,j}^2$ and the total cost is given by expression (5). Fig. 17 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 Fig. 18, we show a comparison of our different approaches for the rate allocation, as a function of the distances from the nodes to the sink. 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 grid 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.

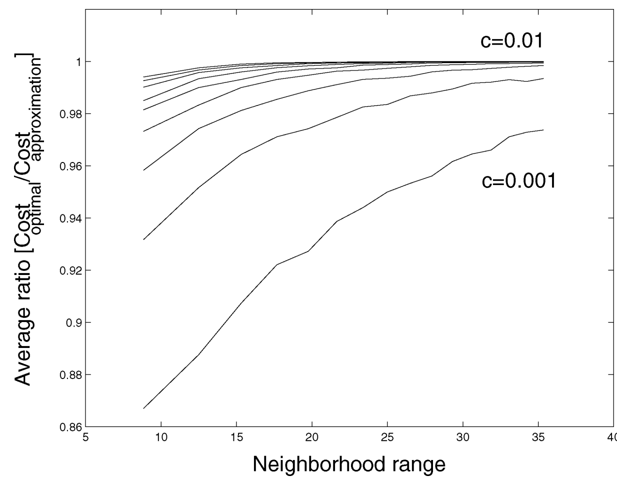


Fig. 17. Slepian–Wolf coding: average value of the ratio between the optimal and the approximated solution, in terms of total cost versus the neighborhood range. Every network instance has 50 nodes uniformly distributed on a square area of size 100×100, and the correlation exponent varies from $c = 0.001$ (high correlation) to $c = 0.01$ (low correlation). The average has been computed over 20 instances for each (c, radius) value pair.

X. Conclusion and Future Work

We addressed the problem of joint rate allocation and transmission structure optimization for sensor networks, when the flow cost cost function [function (rate)] × [path weight] is considered. We showed that if the cost function is separable, then the tasks of optimal rate allocation and transmission structure optimization separate. We assess the difficulty of the problem, namely, we showed that for an arbitrary transfer matrix the problem of finding the optimal transmission structure is NP-complete. For linear cost functions, the problem of optimal rate allocation can be posed as an LP problem, while in the general case, it can be posed as a Lagrangian optimization problem. It is difficult in general to find decentralized algorithms that use only local information for this task. We also studied some cases of interest where the problem becomes easier, leading to a closed-form solution and where efficient approximation algorithms can be derived. For a one-dimensional network, we provide scaling laws for asymptotic behavior and limits of the ratio of total costs associated to the two coding approaches. In particular, we show

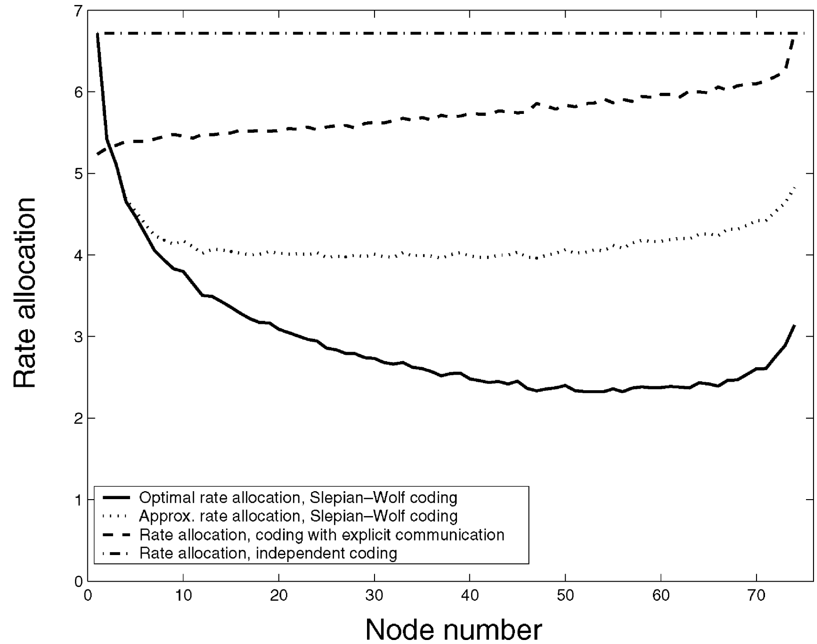


Fig. 18. Average rate allocation for 1000 network instances with 75 nodes, and a correlation exponent $c = 0.0008$ (strong correlation). On the x-axis, nodes are numbered in increasing order as the total weight from the sink increases, on the corresponding SPT.

that for some conditions on the correlation structure, the use of Slepian–Wolf coding techniques can result in unbounded gains in terms of total flow cost over simple coding by explicit communication. Joint rate allocation and transmission structure optimization with explicit communication is studied in detail in [12], [8].

Further research directions include the related problems of joint efficient placement and transmission, the case when the links are capacity constraint, and the derivation of efficient distributed approximation algorithms for both finding the optimal transmission structure and the optimal distribution of rates among the various subsets of sources for various more general cases of transmission matrices. Moreover, an interesting research issue is to find tight bounds for the approximation ratios, in terms of total costs, for these distributed algorithms.

Appendix I
Proof of Theorem 1

First, we prove that (6) is indeed a feasible solution for (5), that is, it satisfies all the constraints given by (4). Consider any constraint from (4), for some subset $Y \subset V$. Denote by $M = |Y|$ the number of elements in Y. Order the indices of $X_i \in Y$ as $i_1, i_2, i_3, \dots i_M$, with i_1 closest and i_M furthest from the sink on the SPT.

If we rewrite the left-hand side in terms of the solutions that we provide in the theorem statement, we have

$$\begin{aligned}\sum_{i \in Y} R_i = &H(X_{i_M}|X_{i_M-1}, \dots, X_1)\\ &+ H(X_{i_{M-1}}|X_{i_{M-1}-1}, \dots, X_1)\\ &+ \dots + H(X_{i_1}|X_{i_1-1}, \dots, X_1). \end{aligned} \tag{11}$$

Expanding the right-hand side terms with the chain law for conditional entropies, we obtain

$$\begin{aligned}H(Y|Y^C) = &H(X_{i_M}|Y^C \cup \{Y - \{X_{i_M}\}\})\\ &H(X_{i_{M-1}}|Y^C \cup \{Y - \{X_{i_M}, X_{i_{M-1}}\}\})\\ &+ \dots + H(X_{i_1}|Y^C \cup \{Y - \{X_{i_M}, \dots, X_{i_1}\}\})\\ = &H(X_{i_M}|V - \{X_{i_M}\})\\ &+ H(X_{i_{M-1}}|V - \{X_{i_M}, X_{i_{M-1}}\})\\ &+ \dots + H(X_{i_1}|V - \{X_{i_M}, X_{i_{M-1}}, \dots, X_{i_1}\}). \end{aligned} \tag{12}$$

Consider the terms on the right-hand side in expressions (11) and (12). It is clear that for any $i_k \in Y$, the term corresponding to X_{i_k} in (12) is at most equal to its counterpart in (11). This is because the set of nodes on which the entropy conditioning is done for each term in (11) is a subset of its counterpart in (12). Since the choice of Y was arbitrary, then any constraint in (4) is satisfied by the assignment (6).

On the other hand, note also that the rate allocation in (6) satisfies with equality the constraint on the total sum of rates

$$\sum_{i \in V} R_i \geq H(X_1, \dots, X_N). \tag{13}$$

This proves that (6) is a valid rate allocation. We have to prove now that the assignment in (6) makes the expression to be minimized in (5) smaller than any other valid assignment.

We prove this by recursion. Note first that the rate allocation to node N is minimal. That is, we cannot allocate to X_N less than $H(X_N|X_{N-1}, X_{N-2}, \dots, X_1)$ bits, due to the Slepian–Wolf constraint corresponding to $Y = \{X_N\}$. Assume now that a solution that assigns $H(X_N|X_{N-1}, X_{N-2}, \dots, X_1)$ bits to X_N is not optimal, and X_N is assigned $H(X_N|X_{N-1}, \dots, X_1) + b$ bits. Due to (13), at most b bits in total can be extracted from the rates assigned to some of the other nodes. But since $d_{\mathrm{SPT}}(X_N, S)$

is the largest coefficient in the optimization problem (5), it is straightforward to see that any such change in rate allocation increases the cost function in (5). Thus, assigning

$$R_N = H(X_N|X_{N-1},\dots,X_1)$$

bits to X_N is indeed optimal.

Consider now the rate assigned to X_{N-1}. From the rate constraint corresponding to $Y=\{X_{N-1},X_N\}$, it follows that

$$\begin{aligned} R_N + R_{N-1} \geq\, & H(X_N, X_{N-1}|X_{N-2},\dots,X_1) \\ =\, & H(X_N|X_{N-1},X_{N-2},\dots,X_1) \\ & + H(X_{N-1}|X_{N-2},\dots,X_1). \end{aligned}$$

Since for optimality R_N must be given by

$$R_N = H(X_N|X_{N-1},X_{N-2},\dots,X_1)$$

it follows that

$$R_{N-1} \geq H(X_{N-1}|X_{N-2},\dots,X_1).$$

Following a similar argument as for X_N, we can show in the same way that the optimal solution allocates

$$R_{N-1} = H(X_{N-1}|X_{N-2},\dots,X_1).$$

The rest of the proof follows similarly by considering successively the constraints corresponding to the subsets $Y=\{X_i,X_{i+1},\dots,X_N\}$, with $i=N-2,N-3,\dots 1$. □

Appendix II
Exponential Dependence in Rate: Optimal Rate Allocation for the Case of Two Nodes

As seen in Fig. 6, the solution of (8) can be either on the thicker curve in Fig. 6, when the first two (linear) constraints in (9) are inactive, or on one of the two corners, when either of the first two constraints become active. First, we put to inactive the first two (linear) constraints in (9). The Lagrangian corresponding to the minimization of (8) under constraints (9) is

$$\begin{aligned} L(P_1,P_2) = d_{\mathrm{SPT}}(X_1,S)P_1 + d_{\mathrm{SPT}}(X_2,S)P_2 + \\ + \lambda(P_1P_2 - e^{H(X_1,X_2)}). \end{aligned}$$

Then, by taking partial derivatives of $L(P_1,P_2)$ with respect to P_1 and P_2 and after some computations, we obtain

$$P_1 = \sqrt{\frac{d_{\mathrm{SPT}}(X_2,S)}{d_{\mathrm{SPT}}(X_1,S)} e^{H(X_1,X_2)}}$$

$$P_2 = \sqrt{\frac{d_{\mathrm{SPT}}(X_1,S)}{d_{\mathrm{SPT}}(X_2,S)} e^{H(X_1,X_2)}}.$$

By enforcing the first two constraints active, it follows that the optimal solution is on either corner on the transformed Slepian–Wolf region if

$$\frac{d_{\mathrm{SPT}}(X_2,S)}{d_{\mathrm{SPT}}(X_1,S)} \leq e^{H(X_1|X_2)-H(X_2)}$$

corresponding to the corner $(e^{H(X_2)},\ e^{H(X_1|X_2)})$

$$\frac{d_{\mathrm{SPT}}(X_2,S)}{d_{\mathrm{SPT}}(X_1,S)} \geq e^{H(X_1)-H(X_2|X_1)}$$

corresponding to the corner $(e^{H(X_1)},\ e^{H(X_2|X_1)})$

and otherwise it lies on the thicker curve shown in Fig. 6.

Letting $\nu = \frac{d_{\mathrm{SPT}}(X_2,S)}{d_{\mathrm{SPT}}(X_1,S)}$ and undoing the change of variable the result follows. □

Appendix III
Proof of Theorem 2

Denote $\gamma = \lim_{N\to\infty}\gamma(N)$.

A. Case (a)

$a_i \to C, C > 0$

Lemma 1: If $g_i \to 0$, and g_i is monotonically decreasing, then

$$\lim_{N\to\infty}\frac{1}{N}\sum_{i=1}^{N} g_i = 0.$$

Proof: Since $g_i \to 0$ and g_i is monotonically decreasing, it results that for any $\epsilon > 0$, there exists an integer N_ϵ such that for any $i \geq N_\epsilon$, $g_i < \epsilon$. Then, for any $\epsilon > 0$

$$\begin{aligned} \lim_{N\to\infty}\frac{1}{N}\sum_{i=1}^{N} g_i \leq & \lim_{N\to\infty}\frac{1}{N}\left[N_\epsilon d_0 + (N-N_\epsilon)\epsilon\right] \\ =& 0 + \lim_{N\to\infty}\frac{(N-N_\epsilon)\epsilon}{N} \\ =& \epsilon. \end{aligned}$$

Since this happens for any $\epsilon > 0$, the result follows □

Lemma 2: Let $a_i \to C, C>0$. Then $\gamma = 1$

$$\lim_{N\to\infty}\frac{\sum_{i=1}^N ia_i}{\sum_{i=1}^N (N-i+1)a_i} = 1. \tag{14}$$

Proof: We can write

$$\lim_{N\to\infty}\frac{\sum_{i=1}^N ia_i}{\sum_{i=1}^N (N-i+1)a_i} = \lim_{N\to\infty}\frac{1}{(N+1)\frac{\sum_{i=1}^N a_i}{\sum_{i=1}^N ia_i} - 1}. \tag{15}$$

Now, since $a_i \to C$, then we can write

$$a_i = g_i + C$$

where $g_i \to 0$ and g_i is monotonically decreasing.

Denote

$$l = \lim_{N\to\infty}(N+1)\frac{\sum_{i=1}^N a_i}{\sum_{i=1}^N ia_i}.$$

Then

$$\begin{aligned} l =& \lim_{N\to\infty}\frac{(N+1)\sum_{i=1}^N (g_i + C)}{\sum_{i=1}^N i(g_i+C)} \\ =& \lim_{N\to\infty}\frac{(N+1)\sum_{i=1}^N g_i + N(N+1)C}{\sum_{i=1}^N ig_i + C\frac{N(N+1)}{2}} \\ =& \lim_{N\to\infty}\frac{\frac{1}{N}\sum_{i=1}^N g_i + C}{\frac{1}{N(N+1)}\sum_{i=1}^N ig_i + C/2}. \end{aligned}$$

We can easily prove that

$$0 \leq \lim_{N\to\infty}\frac{1}{N(N+1)}\sum_{i=1}^N ig_i \leq \frac{1}{N}\lim_{N\to\infty}\sum_{i=1}^N g_i$$

and thus, it is enough to apply Lemma 1 and obtain that

$$\gamma = \frac{1}{2-1} = 1. \qquad \square$$

B. Case (b)

$a_i \to 0$

Lemma 3: Let $a_i \to 0$. If a_i decreases faster than $1/i$, then $\gamma = 0$. If a_i decreases as fast as $1/i^p, 0 < p < 1$, then $\gamma = 1-p$.

Proof: From (15), we see that the limit $\gamma = 0$ iff the ratio

$$\frac{\sum_{i=1}^{N} i a_i}{(N+1)\sum_{i=1}^{N} a_i} \tag{16}$$

goes to zero as $N \to \infty$.

- If a_i decreases faster than $1/i$, then $i a_i \to 0$. Then we can use directly Lemma 1 for $i a_i$

$$\lim_{N\to\infty} \sum_{i=1}^{N} \frac{i a_i}{N+1} = 0.$$

Since $\sum_{i=1}^{N} a_i$ is monotonically increasing, the result $\gamma = 0$ follows.

- If a_i decreases as fast as $1/i$, then, without loss of generality, take $a_i = 1/i$. Then

$$\gamma = \lim_{N\to\infty} \frac{1}{(N+1)\frac{\sum_{i=1}^{N} 1/i}{N} - 1}.$$

But, since $\sum_{i=1}^{N} 1/i$ is divergent, it follows again $\gamma = 0$.

- If a_i decreases slower than $1/i$, suppose without loss of generality that $a_i = 1/i^p$, with $p \in (0,1)$.

Then, by using the integral test, we obtain

$$\begin{aligned}\lim_{N\to\infty} \frac{\sum_{i=1}^{N} i^{1-p}}{(N+1)\sum_{i=1}^{N} i^{-p}} &= \lim_{T\to\infty} \frac{\int_0^T t^{1-p} dt}{(T+1)\int_0^T t^{-p} dt} \\ &= \lim_{T\to\infty} \frac{\frac{t^{2-p}}{2-p}\Big|_{t=0}^{t=T}}{(T+1)\frac{t^{1-p}}{1-p}\Big|_{t=0}^{t=T}} \\ &= \lim_{T\to\infty} \frac{\frac{T^{2-p}}{2-p}}{(T+1)\frac{T^{1-p}}{1-p}} \\ &= \frac{1-p}{2-p}.\end{aligned}$$

This means that the sought limit in the case $p \in (0,1)$ is

$$\gamma = \frac{1}{\frac{2-p}{1-p} - 1} = 1 - p.$$

So far, we have seen that

$$\gamma(N) = \frac{1}{(N+1)\frac{\sum_{i=1}^{N} a_i}{\sum_{i=1}^{N} i a_i} - 1}$$

converges to zero if $i a_i$ goes to zero in the limit $i \to \infty$. Suppose without loss of generality $a_i = 1/i^p$, so the condition for $\lim_{N\to\infty} \gamma(N) = 0$ is $p \geq 1$.

We note that the rate of decay of the ratio of costs is directly related to the rate of increase with N of the partial sum $s^N(q) = \sum_{i=1}^{N} \frac{1}{i^q}$ for $q > 0$, namely

- $s^N(q) = \Theta(N^{1-q})$, if $q \in (0,1)$;
- $s^N(q) = \Theta(\log N)$, if $q = 1$;
- $s^N(q) = \Theta(1)$, if $q > 1$.

We thus obtain the following.

Case $p = 1$: In this case

$$\gamma(N) = \Theta\left(\frac{1}{N\frac{\log N}{N} - 1}\right) = \Theta\left(\frac{1}{\log N}\right).$$

Case $p \in (1,2)$: In this case

$$\gamma(N) = \Theta\left(\frac{1}{N\frac{1}{N^{2-p}} - 1}\right) = \Theta\left(\frac{1}{N^{p-1}}\right).$$

Case $p = 2$: In this case

$$\gamma(N) = \Theta\left(\frac{1}{N\frac{1}{\log N} - 1}\right) = \Theta\left(\frac{\log N}{N}\right).$$

Case $p > 2$: In this case

$$\gamma(N) = \Theta\left(\frac{1}{N-1}\right) = \Theta\left(\frac{1}{N}\right). \qquad \square$$

Appendix IV
Conditional Entropy for Correlation Law $\exp(-c|\tau|)$

Consider the one-dimensional example in Fig. 11. For the sake of simplicity, assume that the variance $\sigma^2 = 1$. The correlation between nodes l and j is $K_{lj} = \exp(-c|l-j|d)$. Denote by $\rho = \exp(-cd)$. Then we can write the covariance matrix of any i consecutive nodes on the line as

$$K^i = \begin{pmatrix} 1 & \rho & \rho^2 & \dots & \rho^{i-1} \\ \rho & 1 & \rho & \dots & \rho^{i-2} \\ \dots & \dots & \dots & \dots & \dots \\ \rho^{i-2} & \rho^{i-1} & \dots & 1 & \rho \\ \rho^{i-1} & \dots & \rho^2 & \rho & 1 \end{pmatrix}$$

and thus, their joint differential entropy is

$$h(X_1,\dots,X_i) = \log(2\pi e)^i \det(K^i) = \log(2\pi e)^i (1-\rho^2)^{i-1}.$$

It follows that we can write the conditional differential entropy as

$$\begin{aligned} h(X_i|X_{i-1},\dots,X_1) &= h(X_i, X_{i-1},\dots,X_1) - \\ &\quad - h(X_{i-1},\dots,X_1) \\ &= \log(2\pi e)^i (1-\rho^2)^{i-1} - \\ &\quad - \log(2\pi e)^{i-1}(1-\rho^2)^{i-2} \\ &= \log \frac{(2\pi e)^i (1-\rho^2)^{i-1}}{(2\pi e)^{i-1}(1-\rho^2)^{i-2}} \\ &= \log(2\pi e)(1-\rho^2) \\ &= h(X_i|X_{i-1}) \end{aligned}$$

which depends on the internode distance d, but not on the number of nodes i.

Since the conditional entropy $a_i = H(X_i|X_{i-1}, \ldots, X_1)$ differs in approximation by only a constant from the conditional differential entropy, it follows that $a_i \approx H(X_i|X_{i-1})$, for any i.

REFERENCES

[1] A. Aaron and B. Girod, "Compression with side information using turbo codes," in *Proc. IEEE Data Compression Conf., DCC 2002*, Snowbird, UT, Apr. 2002, pp. 252–261.

[2] I. F. Akyildiz, W. Su, Y. Sankarasubramaniam, and E. Cayirci, "A survey on sensor networks," *IEEE Commun. Mag.*, vol. 40, no. 8, pp. 102–116, 2002.

[3] S. Arora, "Polynomial time approximation scheme for Euclidean TSP and other geometric problems," in *Proc. 37th Ann. IEEE Symp. Foundations of Computer Science*, Burlington, VT, Oct. 1996, pp. 2–11.

[4] M. Bern and P. Plassmann, "The Steiner problem with edge lengths 1 and 2," *Inf. Process. Lett.*, vol. 32, pp. 171–176, 1989.

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

[6] T. H. Cormen, C. E. Leiserson, and R. L. Rivest, *Introduction to Algorithms*. New York: MIT Press/McGraw-Hill, 1990.

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

[8] R. Cristescu, B. Beferull-Lozano, M. Vetterli, and R. Wattenhofer, "Network correlated data gathering with explicit communication: NP-completeness and algorithms," *IEEE/ACM Trans. Netw.*, to be published.

[9] R. Cristescu, B. Beferull-Lozano, and M. Vetterli, "Networked Slepian-Wolf: Theory and algorithms," in *Proc. 1st Europ. Workshop on Sensor Networks*, EWSN 2004, Berlin, Germany, Jan. 2004.

[10] ——, "On network correlated data gathering," in *Proc. INFOCOM 2004*, Hong Kong, Mar. 2004.

[11] ——, "Scaling laws for correlated data gathering," in *Proc. IEEE Int. Symp. Information Theory (ISIT 2004)*, Chicago, IL, Jun./Jul. 2004, p. 471.

[12] R. Cristescu and M. Vetterli, "Power efficient gathering of correlated data: Optimization, NP-completeness and heuristics," in *Proc. MobiHoc 2003 (poster)*, Annapolis, MD, Jun. 2003.

[13] M. Franceschetti, J. Bruck, and L. J. Schulman, "A random walk model of wave propagation," *IEEE Trans. Antennas Propag.*, vol. 52, no. 5, pp. 1304–1317, May 2004.

[14] A. Goel and D. Estrin, "Simultaneous optimization for concave costs: Single sink aggregation or single source buy-at-bulk," in *Proc. ACM-SIAM Symp. Discrete Algorithms*, Baltimore, MD, Jan. 2003.

[15] P. Gupta and P. R. Kumar, "The capacity of wireless networks," *IEEE Trans. Inf. Theory*, vol. 46, no. 2, pp. 388–404, Mar. 2000.

[16] H. Larson and B. Shubert, *Probabilistic Models in Engineering Sciences*. New York: Wiley, 1979.

[17] S. Lindsey, C. S. Raghavendra, and K. Sivalingam, "Data gathering in sensor networks using the $\mathrm{Energy} * \mathrm{Delay\ Metric}$," in *Proc. IPDPS Workshop on Issues in Wireless Networks and Mobile Computing*, San Francisco, CA, Apr. 2001.

[18] D. Luenberger, *Linear and Nonlinear Programming*. Reading, MA: Addison-Wesley, 1984.

[19] D. Marco, E. Duarte-Melo, M. Liu, and D. L. Neuhoff, "On the many-to-one transport capacity of a dense wireless sensor network and the compressibility of its data," in *Proc. 2nd Int. Workshop on Information Processing in Sensor Networks (IPSN'03) (Lecture Notes in Computer Science)*. New York: Springer-Verlag, Apr. 2003, vol. 2634, , pp. 1–16.

[20] NJB. Porat, *Digital Processing of Random Signals*. Englewood Cliffs: Prentice-Hall, 1994.

[21] G. J. Pottie and W. J. Kaiser, "Wireless integrated sensor networks," *Commun. ACM*, vol. 43, no. 5, pp. 51–58, 2000.

[22] S. Pradhan, "Distributed source coding using syndromes (DISCUS)," Ph.D. dissertation, Univ. California, Berkeley, CA, 2001.

[23] S. Pradhan and K. Ramchandran, "Distributed source coding using syndromes (DISCUS): design and construction," in *Proc. DCC'99 Data Compression Conf.*. Los Alamitos, CA: IEEE Comput. Soc., 1999, pp. 414–423.

[24] J. Rabaey, M. J. Ammer, J. L. da Silva, D. Patel, and S. Roundy, "Picoradio supports ad-hoc ultra-low power wireless networking," *IEEE Comput.*, vol. 33, no. 7, pp. 42–48, Jul. 2000.

[25] T. Rappaport, *Wireless Communication: Principles and Practices*. Englewood Cliffs, NJ: Prentice-Hall, 1996.

[26] W. Rabiner-Heinzelman, A. Chandrakasan, and H. Balakrishnan, "Energy-efficient communication protocol for wireless microsensor networks," in *Proc. 33rd Int. Conf. System Sciences (HICSS '00)*, Hawaii, HI, Jan. 2000.

[27] G. Robins and A. Zelikovsky, "Improved Steiner tree approximation in graphs," in *Proc. 10th Ann. ACM-SIAM Symp. Discrete Algorithms*, San Francisco, CA, Jan. 2000, pp. 770–779.

[28] A. Scaglione and S. D. Servetto, "On the interdependence of routing and data compression in multi-hop sensor networks," in *Proc. 8th ACM Int. Conf. Mobile Computing and Networking (MobiCom)*, Atlanta, GA, Sep. 2002, pp. 140–147.

[29] D. Slepian and J. K. Wolf, "Noiseless coding of correlated information sources," *IEEE Trans. Inf. Theory*, vol. IT-19, no. 4, pp. 471–480, Jul. 1973.