Routing and Data Compression in Sensor Networks: Stochastic Models for Sensor Data that guarantee Scalability

Anna Scaglione
School of Electrical & Computer Eng.
Cornell University, USA
{anna, servetto}@ece.cornell.edu

I. INTRODUCTION

In our problem setting (see Fig. I) N nodes v_i are placed on a square grid of unit area, at random locations (x_i, y_i). Each v_i observes a sample S_i of a spatial jointly normal stochastic process which is such that the correlation between samples increases as the distance between them in the grid decreases[1]. Each v_{ij} wants to communicate an approximation of its S_{ij} to every other node in the network. Each v_i can only send messages to and receive messages from nodes within distance C_N, where this connectivity radius is chosen so as to ensure that the network will be connected with high probability [8]. S is the vector that contains all the samples S_i collected by the sensors, $i = 1, \ldots, N$. Given these assumptions, we are interested in determining under what conditions it is possible for each node v_i to obtain an estimate $\hat{S}$ of S whose total distortion $\mathtt{E}(d(S, \hat{S})) < D$, for any prescribed value $D \geq 0$, and for an appropriate distortion measure $d(\cdot, \cdot)$ (Note: if $d(S, \hat{S}) = \sum_{i=1}^{N} d(S_i, \hat{S}_i)$ and if the distribution of the samples is identical, each sample will have average distortion $\mathtt{E}(d(S_i, \hat{S}_i)) < D/N$). The distortion measure we consider is the MSE.

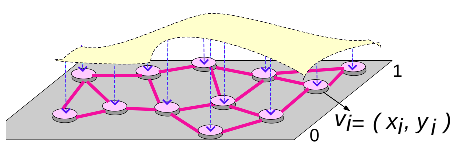

Fig. 1: The sensor network setting.

Recent work on the transport capacity of a large-scale multi-hop wireless network has established that the per-node transport capacity per square meter (i.e. the number of bits that is sent and successfully received over a square meter range) vanishes as an $O(N^{-\frac{1}{2}})$ as the number of nodes N becomes large [9]. The scenario analyzed in [9] is radically different from conventional cellular networks: in [9] all the nodes have identical functionalities, equal transmission bandwidth and are uniformly distributed over a region of fixed area. The result in [9] poses a serious challenge for the design of such networks—*some have even argued that large distributed networks are not feasible* for this reason.

However, in [9] there is one fundamental aspect of the sensor network problem which is not captured: as the density of the nodes increases the most likely scenario is that the nodes samples will be increasingly dependent. This important observation was made in [11, 16, 12] where the authors proposed to compress separately the correlated samples at each node by mean of distributed source coding techniques. The idea of distributed source coding was first introduced by Slepian and Wolf [] who quantified the number of bits that are necessary to encode a source X when the receiver has side Y information on the source. In [12], for the scenario that is described in Fig. I, it was shown that one can reduce the amount of bits per node per square meter to an $O(N^{-1})$.

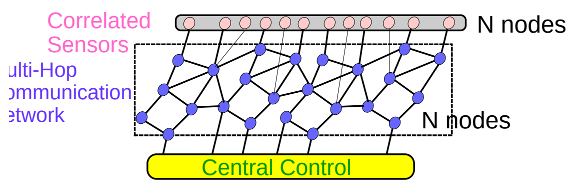

Fig. 2: The sensor network setting in [12].

The result in [12] provides the first theoretical evidence that coding techniques that exploit the dependence among the sensors' samples are key to counteract the vanishing throughput of multi-hop networks. In fact, even if the transport Capacity per node per square meter is vanishing so is the number of non-redundant bits that each node generates and, furthermore, the latter is vanishing at a faster rate than the throughput is.

There are two reasons why the results [11, 16, 12] can be improved: 1) the setting in which the bounds in [12] where derived (Fig. I) is restrictive for the sensors are in a one dimensional space, and the relays are in a two dimensional area which suggests that the nodes are physically separated, even if they are exactly in the same ratio; 2) the approach of distributed source coding requires complex encoders to achieve significant compression gains *without sharing one single bit of data among the nodes*. For example, the proof developed in [12] involves the use of codes for the problem of rate/distortion with side information [5, Ch. 14.9] which are efficient when all codewords are nearly uniformly distributed and this is true for highly correlated data only when the vector have large sizes. High-dimensional vector quantizers are not practical and in short-block settings the gains obtained are in general less significant.

Fortunately, there is no need to impose the constraint of encoding without exchanging information among the nodes. The separation of source-coding an routing in different layers of the communication system architecture does not reflect a physical separation of functionalities in the multi-hop sensor network setting of Fig. I. After all, the trademark of multi-hop networks is that power efficient transmission is achieved when the *data travel through several intermediate close-by nodes* before reaching their final destination. Clearly, if the neighboring nodes are jointly compressing the data in their queues before forwarding them remotely, when the network is dense and the field is smooth, they can drastically reduce the number of bits per sample while transmitting with the same or even greater precision. To accomplish this gain the nodes can use a variety of techniques that are used to compress sequences with no need of resorting to highly complex

[1]The samples are assumed to be temporally independent, hence we will focus on one vector of samples only S, and drop the time index. In any case, it is not a restrictive assumption and further gains in terms of compression could be obtained exploiting the temporal dependence of the samples.

distributed source coding techniques. This is the truly novel and interesting aspect of our paper: the combination of classical source coding methods and routing algorithms which, to the best of our knowledge, has not been explored by other authors.

The goal of this paper is to determine, under plausible models for the sensor samples, what is the order of the number of bits generated by the entire network as a function of the number of nodes N. By looking at the entire network as a unique source we derive the asymptotic rate distortion function of the network: in this way our analysis shows that the minimum total number of bits that the network can produce is a $O(\log(N/D))$.

Our work shows that high density sensor networks are not only possible but increased density can potentially be used to increase the precision of the measurements or decrease the transmission error.

II. Independent encoders lead to network congestion

If the nodes encode independently the data a measure of the amount of traffic the network generates depends on the marginal statistics of the samples and on the quantizers. Let us consider the case where S_i are uniformly distributed in the range $[0,1]$ and each node uses a scalar quantizer with B bits of resolution (i.e., the quantization step is 2^{-B}), and the distortion is the mean-square error. In this basic example of quantization theory the average sample distortion of the uniform quantizer is $\frac{1}{12}2^{-2B}$ [6] and, thus, the operational distortion-rate function is $\delta(B) = \sum_{i=1}^{N} E(d(S_i, \hat{S}_i)) = \frac{1}{12}N2^{-2B}$. Solving for B in $D = \frac{1}{12}N2^{-2B}$ we can observe that each sample requires $B = \lceil \frac{1}{2}\log_2(\frac{N}{12D})\rceil$ bits and, correspondingly the total amount of traffic generated by the whole network scales like $O(N\log N)$ in network size. In fact, even if each node utilizes a k dimesional vector quantizer, high resolution methods show that the operational distortion rate function would be

$$\delta_k(B) \approx c_k 2^{2h_k(S)}2^{-2B}. \tag{1}$$

where $h_k(S) = 1/kh(S(t_1), ...S(t_k))$, $h(S(t_1), ...S(t_k))$ is the differential entropy of k successive sensor's samples and c_k a source specific constat. Hence, the $O(N\log N)$ scaling behavior of the network traffic is encountered under wide assumptions on the source statistics and even when optimum vector quantizers are used as encoding technique, so long as the encoding is done locally at the node, without assuming any correlation with the other sensor mearurements (i.e. node by node) [3] (other examples can be found in [12, Sec. 3] and [14, Sec. 7]).

If we compare this result with the findings of [9] we have to conclude that the problem we described in the introduction has no solution in the sense that the network inevitably will enter in a state of congestion, as the number of nodes will increases. The question is: are there *other* coding strategies that can compress sensor data enough? As discussed in the introduction, intuitively when the network density increases the nodes their samples will be increasingly similar and, thus, representing them independently will be increasingly redundant. Distributed source coding is the alternative that has been explored recently with some success [11, 16, 12] which interestingly does not require basic modifications in the layers of the transmission architecture. The solution we propose is to abandon the traditional layered structure, where source coding and routing functionalities are separated by two layers, and combine the two operations. Instead of complicated vector quantizers required by distributed source coding, the processing at each node that receives a vector of samples $(S_1, \ldots, S_k)$ can be done using any of the standard compression technique which are used normally to compress sequences from analog sources (DPCM, Σ–Δ, subband-coding such as JPEG, simplified Vector Coding techniques, etc.). Of course, the data exchanged are also quantized, therefore our quantization scheme has to be viewed more precisely as a form of compression rather than quantization. Also, note that high-resolution analysis shows that if S_i are quantized with a uniform quantizer with cell size Δ individually, their joint entropy is [6]:

$$H(q(S_1), \ldots, q(S_k)) \approx h(S_1, \ldots, S_k) - \log\Delta, \tag{2}$$

and the vector $(q(S_1), \ldots, q(S_k))$ can potentially be compressed using variable length entropy coding or universal lossless compression algorithms, such as Lempel-Ziv. To better address practical scenarios we are currently investigating simpler and/or more effective forms of compression techniques.

To see weather the correlation among data can be sufficient to ensure favorable scaling laws for the problem examined, it is necessary to define rigorously the concept of *dense* network and to tie it to the statistical properties of the measurements. Our goal is to prove that for reasonable models of the sensor data, if the intermediate nodes compress jointly the data before forwarding them the broadcast problem requires transferring a minimum number of bits that grows slower than an $O(N^{-1/2})$.

We provide the rate of growth of the number of bits produced by network we derive next the rate-distortion function of the data generated by all the sensors through the following fairly general model (see Section III): (a) the data are Gaussian random variables (b) the correlation among samples is an arbitrary spatially homogeneous function; and (c), as we let the number of nodes in the network grow, the correlation matrix converges to a *smooth* two-dimensional function. The smoothness, which will be defined more rigorously in Section VII, guarantees that the sensors' measurements become increasingly correlated as the density of the nodes becomes large. Assumption (a) is a worse case scenario as far as compression is concerned, as a consequence of Berger's result [2], stating that the rate-distortion function of a memoryless, continuous-amplitude source with zero mean and finite variance σ_S^2 with respect to the MSE distortion measure is upper-bounded as:

$$R(D) \le \frac{1}{2}\log(\sigma_S^2/D) \qquad \text{if } \ 0 \le D \le \sigma_S^2. \tag{3}$$

Spatial stationarity, even though not totally general is a technical assumption common to many statistical analysis and captures well local properties of random processes.

III. Source Model for the combined Routing and Data Compression

This section establishes the basic model upon which we will base our asymptotic analysis. Let $S(t)$ denote the random vector $\{S(t)\}_i = S_i(t)$ of the samples collected by the sensors at time t. Our first assumption is:

(a) $S(t)$ is a spatially correlated random Gaussian vector $\sim \mathcal{N}(0, \boldsymbol{R})$. The samples are temporally uncorrelated.

The samples are temporally independent if we assume that the power spectrum of $S(t)$ is band-limited and the data are sampled at the Nyquist rate. In any case, it is not a restrictive assumption and further gains in terms of compression could be obtained exploiting the temporal dependence of the samples. Because of the temporal independence, we will focus on one vector of samples only S, and from now on we will drop the time index.

With the mean square error (MSE) as distortion measure $d(S, \hat{S}) = \|S - \hat{S}\|^2$ and with the constraint

$$E(\|S - \hat{S}\|^2) < D,$$

we can calculate under assumption (a1) the rate/distortion function of the network using the *reverse water-filling* result [5]. Indicating by $\lambda_1 \geq \lambda_2 \ldots \geq \lambda_N$ the ordered eigenvalues of $\boldsymbol{R}$, the rate/distortion function is

$$R(D) = \sum_{n=1}^{N} \frac{1}{2} \log \quad \frac{\lambda_n}{D_n} \tag{4}$$

where

$$D_n = \begin{cases} K & \text{if } K < \lambda_n, \\ \lambda_n & \text{otherwise.} \end{cases} \tag{5}$$

and K is such that

$$\sum_{n=1}^{N} D_n = D. \tag{6}$$

For $\sum_{n=1}^{N} \lambda_n \geq D$, there exists an $\bar{n} \leq N$ such that $K < \lambda_{\bar{n}}$ and $K \geq \lambda_{\bar{n}+1}$, therefore:

$$K = \frac{D - \sum_{n=\bar{n}+1}^{N} \lambda_n}{\bar{n}} \tag{7}$$

and

$$R(D) = \sum_{n=1}^{\bar{n}} \frac{1}{2} \log \quad \frac{\lambda_n}{K} \quad . \tag{8}$$

The rate-distortion function is a function of the eigenvalues of $\boldsymbol{R}$ only and $\boldsymbol{R}$ is formed with the samples of the continuous multivariate function that represents the correlation between the samples taken two arbitrary points in the network:

$$\{\boldsymbol{R}\}_{i,k} \triangleq E\{S_i S_k\} = R(x_i, x_k, y_i, y_k). \tag{9}$$

The eigenvectors $\boldsymbol{u}$, with entries $\{\boldsymbol{u}\}_i = u(x_i, y_i)$ satisfy the following equation:

$$\lambda^{(N)} u^{(N)}(x_i, y_i) = \sum_{k=1}^{N} R(x_i, x_k, y_i, y_k) u^{(N)}(x_k, y_k). \tag{10}$$

IV. Cases where Routing and Data Compression does not prevent congestion

Just observing (8) it is not difficult to conclude that if the model of $\boldsymbol{R}$ is such that $\boldsymbol{R}$ is asymptotically full rank as $N \to \infty$, then $R(D) \geq O(N)$ and the broadcast problem would not have a solution. In fact, if $\boldsymbol{R}$ is asymptotically full rank with an asymptotically continuous distribution of the positive eigenvalues, then $\bar{n} = O(N)$ and the rate-distortion function will be at least growing in the order of N. For example, if $\boldsymbol{R}$ is a random symmetric matrix (could be a sample of a Wishart distribution) with independent i.i.d. entries having unit variance, it is known that the empirical distribution of the normalized eigenvalues (i.e. the eigenvalues of $\boldsymbol{R}/N$) follows the well know quarter of circle law [15]:

$$\mu_\lambda(x) = 2/\pi \sqrt{1 - x^2/4} \qquad 0 \leq x \leq 2. \tag{11}$$

Therefore the eigenvalues in (8), which are not normalized, grow as an $O(N)$ leading to the negative conclusion that $R(D)$ in (8) grows overall like an $O(N \log N)$, despite the fact that many of the data are possibly highly correlated. In these cases the scaling law of the independent encoder is no worse than the one of dependent encoders!

However, considering that the network area is limited, modelling $\boldsymbol{R}$ as a matrix with random entries and assuming that its range space grows as an $O(N)$ is equivalent to assume that the random process sampled by the sensors has *infinite many spatial degrees of freedom over a finite area*, a situation that will rarely apply in practice. If instead $R(x_i, x_k, y_i, y_k)$ is a smooth function of the four variables (x_i, x_k, y_i, y_k) what will happen is that the matrix $\boldsymbol{R}$ will have an increasingly large null space, while the size of the range space will tend to saturate to a finite number and so will $\bar{n}$ for $N \gg 1$. Correspondingly the entries of the vector S will be an highly over-sampled representation of the spatial process: most of the entries could in fact be interpolated with small error from a subset of them, whose cardinality will tend to be independent of N as N grows. These are the key ideas behind the derivations in the following sections.

V. Asymptotic Distribution of the Eigenvalues

Our derivations in the following try to capture the fact that the process cannot have infinite spatial degrees of freedom. The asymptotic rate-distortion function is obtained using the following two basic steps: 1) we prove that the eigenvalues of the correlation matrix $\boldsymbol{R}$ tend to the eigenvalues of the continuous integral equation:

$$\lambda^{(\infty)} u^{(\infty)}(x, y) = \iint R(x, \xi, y, \nu) u^{(\infty)}(\xi, \nu) d\xi d\nu; \tag{12}$$

2) we provide a model for the kernel of the continuous integral equation (12) which is bandlimited in the spatial frequencies, and this allows us to obtain the asymptotic rate distortion bound.

As we said, the first step is to rewrite (10) as a quadrature formula which approximate the integral equation (12). For a general integral, there will be quadrature coefficients α_i such that the approximation of (12) holds:

$$\begin{aligned} &\sum_{k=1}^{N} R(x_i, x_k, y_i, y_k) u^{(N)}(x_k, y_k) \alpha_k \\ &\approx \iint R(x, \xi, y, \nu) u^{(\infty)}(\xi, \nu) d\xi d\nu. \end{aligned} \tag{13}$$

Since we want to explore the convergence of the eigenvalues of (10) we can set $\alpha_i = 1/N$ in which case the first side (13) is equivalent to the right side of (10) normalized by N. Therefore, when (13) is valid, the left sides of (10) normalized by N is approximately equal to the left side of (12) which leads to the following approximation:

$$\lambda^{(N)}/N \approx \lambda^{(\infty)}. \tag{14}$$

The error in the approximation (13) determines the error in (14). The two errors are related by the following theorem, derived from [10, Sec. 5.4]:

Lemma 1 *Denoting by $\lambda^{(\infty)}$ an arbitrary eigenvalue of (12) and by $u^{(\infty)}(\xi, \nu)$ the corresponding normalized eigenvector, for N sufficiently large there exist an eigenvalue of $\boldsymbol{R}$ such that:*

$$|\lambda^{(N)} - N\lambda^{(\infty)}|^2 \leq \frac{\frac{1}{N}\sum_{i=1}^{N} \mathcal{E}_N\{R(x_i, y_i, \xi, \nu) u(\xi, \nu)\}}{1 + \mathcal{E}_N\{u^2(\xi, \nu)\}}, \tag{15}$$

where $\mathcal{E}_N\{f(x, y)\}$ denotes the quadrature error, i.e.:

$$\iint f(x, y) dx dy = \sum_{i=1}^{N} f(x_i, y_i) \alpha_i + \mathcal{E}_N\{f(x, y)\}.$$

Assuming that:

(c.I) *For any continuous $f(x, y)$ the grid (x_i, y_i) is such that with $\alpha_i = 1/N$ the quadrature error $\mathcal{E}_N\{f(x, y)\} \to 0$;*

then (15) implies that:

$$\lim_{N \to \infty} \lambda^{(N)}/N = \lambda^{(\infty)}.$$

Condition **(c.I)** in Lemma 1 is the operational condition on the distribution of the sensor nodes: the nodes should be distributed in such a way that the quadrature error $\mathcal{E}_N\{R(x_i, y_i, \xi, \nu)u(\xi, \nu)\}$ vanishes as $N \to \infty$.

There is another interesting and intuitively obvious consequence of Lemma 1 and **(c.I)**, which is summarized in the following corollary (it can be easily proved using the bounds in Lemma 1 and the triangular inequality):

Corollary 1 *The eigenvalues of* $\boldsymbol{R}$ *and* $\boldsymbol{R}'$ *corresponding to two different grids are such that*

$$|\lambda^{(N)} - \lambda'^{(N)}|^2 \leq \frac{\frac{1}{N}\sum_{i=1}^{N} \mathcal{E}_N\{R(x_i, y_i, \xi, \nu)u(\xi, \nu)\}}{1 + \mathcal{E}_N\{u^2(\xi, \nu)\}} + \frac{\frac{1}{N}\sum_{i=1}^{N} \mathcal{E}_N\{R(x_i', y_i', \xi, \nu)u(\xi, \nu)\}}{1 + \mathcal{E}_N\{u^2(\xi, \nu)\}}, \quad (16)$$

Hence, if **(c.I)** *is satisfied by both grids,* $\boldsymbol{R}$ *and* $\boldsymbol{R}'$ *have the same eigenvalues asymptotically.*

Corollary 1 implies that we can rely on any grid that has the same asymptotic behavior, such as for example a regular lattice, and extrapolate the asymptotic behavior of the eigenvalues from the latter.

VI. A Tractable Case

We assume that

(b) the correlation between points in (9) is spatially homogeneous:

$$R(x, \xi, y, \nu) = R(x - \xi, y - \nu), \quad (17)$$

The consequent structure of $\boldsymbol{R}$ on a regular grid is also known as *doubly Toeplitz*, i.e. $\boldsymbol{R}$ is a block Toeplitz matrix with blocks that are Toeplitz themselves.

Assumption (b) implies that

$$\lambda^{(N)}u^{(N)}(x_i, y_i) = \sum_{k=1}^{N} R(x_i - x_k, y_i - y_k)u^{(N)}(x_k, y_k) \quad (18)$$

$$\lambda^{(\infty)}u^{(\infty)}(x, y) = \iint R(x - \xi, y - \nu)u^{(\infty)}(\xi, \nu)d\xi d\nu. \quad (19)$$

The useful aspect of this model is that the empirical distribution of the eigenvalues of $\boldsymbol{R}$ converges under mild assumptions to the 2-D Fourier Spectrum of (17), as we see next.

Under the assumption **(b)** we can define:

$$\psi(\xi_{i,k}, \nu_{i,k}) \triangleq R(x_i - x_k, y_i - y_k) = \{\boldsymbol{R}\}_{i,k}, \quad (20)$$

Adopting a regular grid covering the square of area 1, the spacing between them is $x_i - x_k = \xi_{i,k} = (i - k)/\sqrt{N}$ and like-wise $\nu_{i,k}$. Szegö's theorem [7] establishes that asymptotically the eigenvalues of a Toeplitz matrix converge to the spectrum of the correlation function. Essentially, Szegö's theorem establishes that the eigenfunctions of an homogeneous kernel tend to be the Fourier basis of complex exponentials. The result can be generalized to the two dimensional case when the matrix is doubly Toeplitz, i.e.:

$$\lim_{N\to\infty} \lambda_n(\boldsymbol{R}) = \iint \psi(\xi/\sqrt{N}, \nu/\sqrt{N})e^{-j2\pi(f\xi + f'\nu)}d\xi d\nu \quad (21)$$

$$= N\Psi(\sqrt{N}f, \sqrt{N}f). \quad (22)$$

Before proceeding, the final modelling assumption is:

(c.II) $\Psi(f, f')$id bandlimited with respect to f, f' bandwidth $\leq W$, i.e. $\Psi(f, f') = 0$ for $|f, f'| > W$.

With **(c.II)**, we capture the notion that the limit covariance function varies smoothly in space.

VII. Asymptotic Rate Distortion Function of the Network $R(D) = O(\log(N/D))$

The asymptotic rate-distortion function is obtained replacing the summations in (7) and (8) with integrals over (f_x, f_y). Because the eigenvalues become asymptotically a continuous function in the evaluation of D_n in (5) there will be points (f_x', f_y') where $N\Psi(\sqrt{N}f_x, \sqrt{N}f_y)$ will cross the threshold K, i.e. $K = N\Psi(\sqrt{N}f_x', \sqrt{N}f_y')$.

Let us denote the sets of points f_x and f_y where $0 \leq N\Psi(\sqrt{N}f_x, \sqrt{N}f_y) \leq K$ as $\mathcal{I}$, i.e.:

$$\mathcal{I} \triangleq \{(f_x, f_y) : 0 \leq N\Psi(\sqrt{N}f_x, \sqrt{N}f_y) \leq K, |f_x, f_y| \leq W\}. \quad (23)$$

Let us also indicate as $\bar{\mathcal{I}}$ the set

$$\bar{\mathcal{I}} \triangleq \{(f_x, f_y) : N\Psi(\sqrt{N}f_x, \sqrt{N}f_y) > K\} \quad (24)$$

$$K \approx \frac{\frac{D}{N} - \iint_{\mathcal{I}} N\Psi(\sqrt{N}f_x, \sqrt{N}f_y)df_x df_y}{\iint_{\bar{\mathcal{I}}} df_x df_y}. \quad (25)$$

The rate/distortion function is:

$$R(D) = \frac{N}{2}\iint_{\bar{\mathcal{I}}} \log\frac{N\Psi(\sqrt{N}f_x, \sqrt{N}f_y)}{K}df_x df_y. \quad (26)$$

Because of (c.II) the areas of both $\mathcal{I}$ and $\bar{\mathcal{I}}$ are smaller than $4W^2/N$. Thus, we have the following lower bound on K (also illustrated in Fig. 3):

$$K \geq \frac{\frac{D}{N} - K\frac{4W^2}{N}}{\frac{4W^2}{N}} \quad \Rightarrow \quad K \geq \frac{D}{8W^2}, \quad (27)$$

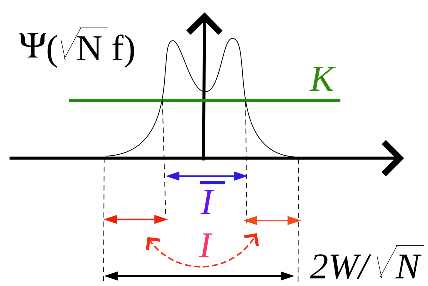

Figure 3: One dimensional illustration of how the lower bound on K from eqn. (27) is obtained by bounding the integrals that define K in eqn. (25).

Together with (c.II), this justifies the following upper-bound on the rate/distortion function:

$$R(D) \leq \frac{N}{2}\iint_{\bar{\mathcal{I}}} \log\frac{N\Psi(\sqrt{N}f_x, \sqrt{N}f_y)}{D/8W^2}df_x df_y.$$

$$\leq 2W^2 \log(ND^{-1}\sup(\Psi(f_x, f_y))8W^2). \quad (28)$$

So, we see that the total rate-distortion function over the entire network is $O(\log(N/D))$, and because D/N is the average distortion per sample if that is kept fixed, then the total amount of traffic generated by the network is upper bounded by a constant, irrespective of network size. Alternatively, if we keep the total distortion D fixed, by considering increasingly large N we let the average distortion $D/N \to 0$. Even though the rate distortion function is an asymptotic bound, the result is very significant because we can observe that the amount of traffic generated by the entire network grows only *logarithmically* in N, well below the capacity bound $O(\sqrt{N})$ proved in [9].

VIII. A SIMPLE COMPRESSION STRATEGY: DOWN-SAMPLING WHILE ROUTING

A simple-minded compression strategy that the nodes can implement is to *down-sample* appropriately the sensor measurements as they are spread through the network. This simple strategy allows us the reach with some approximations the same conclusion of our asymptotic analysis. In fact, even though a spatially bandlimited process requires infinite samples to be correctly reconstructed through interpolation, Nyquist theorem indicates that if condition **c.II** is met, we can sample the field with frequency W in the x and y axis respectively. Because the network area is equal to one, even if we over-sample to reduce the interpolation error at the borders, we need $O(W^2)$ (10, 100 times W^2...) samples to represent it fairly accurately. However, the number of samples does not grow with N. On the other hand, if we constrain the total distortion error to be D, the average distortion per sample has to be $D_i = D/N$. Let us assume that the variance of each sample is $\sigma^2_{S_i} = 1$: the interpolated samples will have distortion which is greater or equal to the distortion of the non interpolated ones which implies that the non interpolated samples have to be quantized at a rate that is at least:

$$R(D_i) \geq 1/2 \log(\sigma^2_{S_i}/D_i) = 1/2 \log(N/D). \qquad (29)$$

Therefore, the total amount of traffic produced by the network will be in the order of:

$$R = O(W^2 \log(N/D)), \qquad (30)$$

q.e.d.

IX. NUMERICAL EXAMPLES

In this section we provide numerical evidence that validates our asymptotic claims.

The first numerical example is aimed at corroborating Corollary 1. We assumed the area of the network is normalized to one and that the function defined in (17) is:

$$R(\xi, \nu) = sinc(\pi W \xi) sinc(\pi W \nu), \qquad (31)$$

where $W = 0.5$ and it can be easily verified that condition **C.II** is met. In fig. 4 we show samples obtained over the regular grid that are $\mathcal{N}(\mathbf{0}, \boldsymbol{R})$ and in fig. 5 we show the eigenvalues of the matrix $\boldsymbol{R}$, whose entries are $R(x_i - x_j, y_i - y_j)$ where in one case (x_i, y_i), $i = 1, \ldots, N$ are on a random grid (red line) and in the other case they are on a regular lattice (blue line): we can observe that they are nearly identical in both cases and that the support of the non zero eigenvalues does not grow with N while their values increase proportionally to it.

Finally, in fig. 6 for the same covariance model in (31) and total distortion $D = 1$, we show the rate-distortion function calculated numerically using the inverse water-filling in (4). As expected, the growth is clearly logarithmic.

X. CONCLUSIONS

In this work we have shown that to work around the vanishing per-node throughput constraints of sensor networks [9] one can combine classical source codes combined with suitable routing algorithms and re-encoding of the sensor data at intermediate relay nodes. To the best of our knowledge, these are the first results in which interdependencies between routing and data compression problems are captured in a system model that is also analytically tractable. And a key (and enabling) step in our derivation was the construction of a family of spatial processes satisfying some fairly mild (and easily justifiable from a physical point of view) regularity conditions, for which we were able to show that the amount of data generated by the sensor network is well below its transport capacity. This provides further evidence that large-scale multi-hop sensor networks are perfectly feasible, even under the network model considered in [9]. We also have shown that there are cases where compression is not going to prevent congestion, specifically in the case where the correlation between the data is an arbitrary symmetric random matrix.

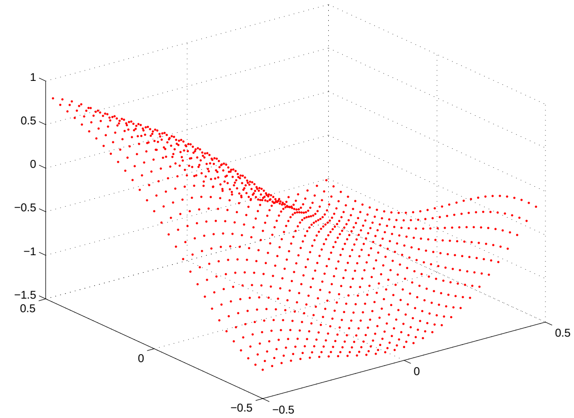

Figure 4: Samples $\sim \mathcal{N}(\mathbf{0}, \boldsymbol{R})$ where $\boldsymbol{R}$ is defined in (31) obtained over the regular grid of sensors.

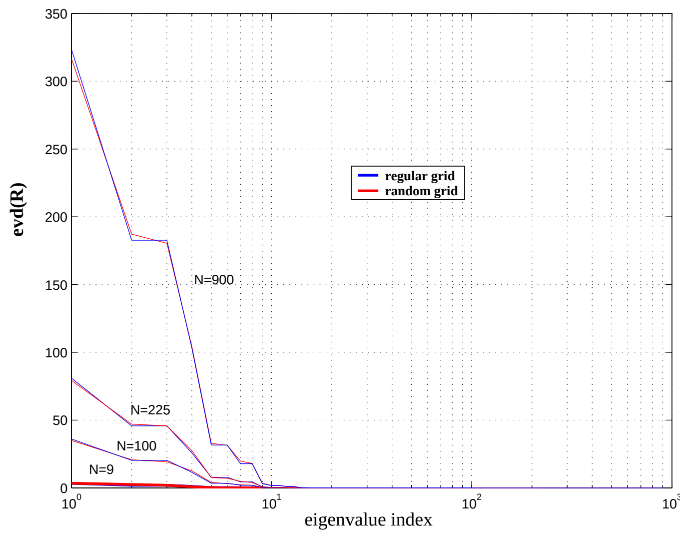

Figure 5: Eigenvalues of $\boldsymbol{R}$ for various values of N.

REFERENCES

[1] R. Ahlswede, N. Cai, S.-Y. R. Li, and R. W. Yeung. Network Information Flow. *IEEE Trans. Inform. Theory*, 46(4):1204–1216, 2000.

[2] T. Berger. *Rate distortion Theory*, Prentice Hall, Englewood Cliffs, NJ.

[3] J. H. Conway and N. J. A. Sloane. *Sphere Packings, Lattices and Groups*. Springer Verlag, 3rd edition, 1998.

[4] T. Cormen, C. Leiserson, and R. Rivest. *Introduction to Algorithms*. MIT Press, 1990.

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

[6] A. Gersho and R. M. Gray. *Vector Quantization and Signal Compression*. Kluwer Academic Publishers, 1992.

[7] U. Grenander and G. Szegő. *Toeplitz Forms and their Applications*. University of California Press, 1958.

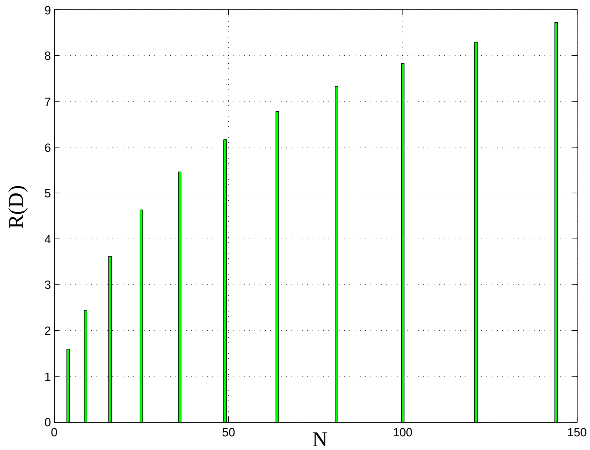

Figure 6: Rate distortion function $R(D)$ for $D = 1$ versus number of nodes N in the network.

[8] P. Gupta and P. R. Kumar. Critical Power for Asymptotic Connectivity in Wireless Networks. In W. M. McEneany, G. Yin, and Q. Zhang, editors, *Stochastic Analysis, Control, Optimization and Applications: A Volume in Honor of W. H. Fleming*. Birkhauser, 1998.

[9] P. Gupta and P. R. Kumar. The Capacity of Wireless Networks. *IEEE Trans. Inform. Theory*, 46(2):388–404, 2000.

[10] H. B. Keller Numerical Methods For Two-Point Boundary-Value Problems. *Dover Publications Inc.*, 1992.

[11] S. S. Pradhan and K. Ramchandran. Distributed Source Coding: Symmetric Rates and Applications to Sensor Networks. In *Proc. IEEE Data Compression Conf. (DCC)*, Snowbird, UT, 2000.

[12] S. D. Servetto. Lattice Quantization with Side Information. In *Proc. IEEE Data Compression Conf. (DCC)*, Snowbird, UT, 2000. An extended journal version, with title *Quantization with Side Information: Lattice Codes, Asymptotics, and Applications in Wireless Networks*, has been submitted for publication, and is available from `http://people.ece.cornell.edu/servetto/`.

[13] C. E. Shannon, Coding Theorems for a Discrete Source With a Fidelity Criterion *Institute of Radio Engineers, International Convention Record*, Vol. 7 (Part 4), pp. 142–163, 1959.

[14] V. A. Vaishampayan, N. J. A. Sloane, and S. D. Servetto. Multiple Description Vector Quantization with Lattice Codebooks: Design and Analysis. *IEEE Trans. Inform. Theory*, 47(5):1718–1734, 2001.

[15] E. Wigner, “On the distribution of the roots of certain symmetric matrices,” *Annales of Math.* Vol. 67, pp.325–327, 1958.

[16] Q. Zhao and M. Effros. Optimal Code Design for Lossless and Near Lossless Source Coding in Multiple Access Networks. In *Proc. Data Compression Conf. (DCC)*, Snowbird, UT, 2001.