Scalability of Sensor Networks

Ameesh Pandya, Huiyu Luo, and Gregory Pottie
Department of Electrical Engineering
University of California, Los Angeles
{ameesh, huiyu, pottie}@ee.ucla.edu

Abstract

We consider the circumstances under which sensor networks can be scalable. The key attribute is a localization of the source-destination distribution, achievable through localized processing. We show that very simple strategies, viz., selection of the closest sensor or group of sensors, are sufficient to achieve scalability when distortion is permitted for point sources. When spatial distortion is permitted, networks are also scalable when sensing distributed phenomena.

I. Introduction

Large scale sensor networks consisting of hundreds or thousands of nodes will link the physical world to global communication networks for a broad set of applications [1], [2]. Individual nodes will have some combination of sensing, signal processing and communications capability and may self-organize for a variety of cooperative sensing and communication tasks, subject to resource constraints such as energy and bandwidth.

For purposes of this paper, the sensor network problem is for some end user to extract information concerning some source or set of sources to some desired level of fidelity, subject to resource constraints. Fidelity encompasses such concepts as spatial or temporal resolution, misidentification probability or other accuracy measures, and network quality of service related measures such as latency from initial observation. Resource constraints can include signal processing cycles, energy consumption, and information rate. Specific information theoretic problems for such networks include:

1) Scalability: whether decisions can be conveyed to users with the desired fidelity as the number of nodes grows without bound. Depending on the fidelity measure, this may be structured as a network capacity, reliability or congestion control problem.
2) Rate distortion limits: what is the best achievable tradeoff between fidelity and resource usage on both local and global scales? This topic includes formulation of optimal data fusion rules.

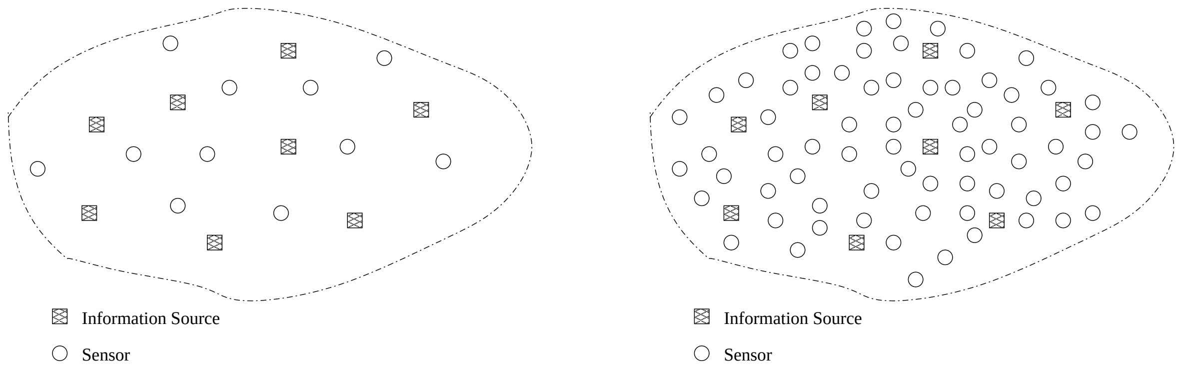
In this paper, we present conditions under which scalability is possible when observing both point and distributed sources, and discuss the implications of the solution to the q-helper problem [3] for local fusion of information from a Gaussian source.

The results in [4] suggest that static ad hoc networks are inherently non-scalable, i. e. per-node capacity tends to 0 with the increase in network size. In [5] , on the other hand, it was argued that the situation in sensor networks is not at all as pessimistic since the problem is often that of coding a correlated source. In this paper, we first study the scalability issue and show that the appropriate choice of source-destination pair distributions can lead to finite per node capacity in ad hoc networks. Applications of this result include sensor networks and hierarchical communication networks.

We then consider the rate-distortion region characterization for a sensor network with finite point sources and under a fidelity criterion . By contrast to [5], [6], in this paper, the sensor network problem consists of extracting information about sources in some region to some desired level of fidelity, and transmitting this information to some gateway(s). We assume point sources to be Gaussian for analytical simplicity. We show that the question of feasible rates for such a network can be made into a relatively easy problem to deal with by allowing the network to be dense, i. e. letting the number of sensors and communication relays be much larger than the number of sources. A sub-optimal decoupling of source and channel coding can be shown to be sufficient for achieving scalability in this context. We also explore how the issues of scalability, source separation, and information extraction can be dealt with by altering the relative densities of sources, communication relays and the sensors.

Finally, we consider the scalability problem in the context of distributed sources (e.g. reconstruction of a temperature profile over some volume to a desired level of fidelity). The perspective is that there is one distributed process, modelled using correlated point sources. As the network density increases without bound, we argue that as in the case of the point source the number of sensors that must report measurements for a given level of fidelity eventually becomes finite while the relay capacity of the network continues to grow, so that scalability is assured. We further consider a particular suboptimal but scalable interpolation strategy in order to explore network density tradeoffs in the presence of measurement error.

This material is based upon work supported by the National Science Foundation (NSF) under Cooperative Agreement #CCR-0121778.



(a) Finite Gaussian point sources and sensor nodes (b) Finite Gaussian point sources and sensor nodes $\rightarrow \infty$

Fig. 1. A region S with Gaussian point sources and sensor nodes

II. Networks and the Source-Destination Pair Distribution

The concept of source-destination pair distribution for achieving the scalability is motivated by the work in [7]. Successful communication is still possible with almost constant node bandwidth, if the distribution of the source-destination (S-D) pairs is such that the average hops per communication is small enough.

We consider the same 2-D framework as in [4]. We note that it is not the fact of correlated sources which is most fundamental to the result in [5], but rather the S-D pair distribution.

Observation 2.1: For the 2-D geometric model defined in [4], the average distance between source and destination should be $O(\frac{1}{\sqrt{n \log n}})$.

Clearly, in order for the average throughput per node to be constant, the average number of hops between source and destination should grow as $O\left(1\right)$. This follows immediately by observing that this criterion is necessary for transport capacity meeting the upper bound of $\sqrt{n \log n}$. The *zero-mean Truncated Gaussian* is one of the many distributions that achieve the finite per node capacity for the geometric model of [4]. The analytical details are skipped for the sake of brevity. This approach is based on the results in [5], but one can find the similar results with the alternative approach in [7]. J. Li, et. al. in [7] show that the traffic pattern determines whether an ad hoc network's per-node capacity will scale to large networks.

In practice, scalability can be achieved in two basic ways:

1) Local cooperative processing to produce decisions (e. g. in sensor networks).
2) Adding communications hierarchy so that communications in each level is local (e. g. telecommunication network).

The latter of course requires additional resources, but typically also provides latency benefits.

III. Consequences Of Decoupling Source, Sensor And Relay Densities

In sensor networks, the basic problem is to extract measurements of some physical phenomenon, to some desired level of fidelity, subject to constraints on energy consumption and bandwidth (resources). Nodes may also have explicit limits on signal processing and storage, which we will neglect here. By considering source, sensor and communications relay densities separately, we show that extraction of such information can be easily achieved without the requirement of complicated joint source-channel coding schemes, in the limit of high sensor and relay densities. Further, this formulation admits simple classification of a broad set of network information theory problems.

A. Spatial Source Separation

Consider any two dimensional region, S, with m ($m < \infty$) randomly distributed Gaussian point sources. The point sources are useful abstractions since many phenomena can be reasonably modelled as either a single point source, or constructed via interpolation from a set of point sources. Consider $p < \infty$ randomly distributed sensors in the region S to fulfill the functions of gathering information. Also, the sensors are assumed to be independent and identically distributed (i.i.d.) in location. Furthermore, the distribution by which the sensors are deployed is assumed to have positive density at every point over S (Fig. 1(a)). We also assume that the signal power decays with the distance. We need to know the achievable distortion $D\left(R\right)$, for the above described sensor network model. The distortion measure is assumed to be squared error.

When the number of sensor nodes in the network is of $O\left(m\right)$, not all D are achievable regardless of the rate constraints, since sensors may not be close enough to sources. Also, not all the rates are achievable because capacity may not be sufficient. Here, there is an interference among information streams and in the signals received by each sensor and so a joint source-channel coding approach would be needed to achieve a large fraction of the rate region. There is little prospect of actually implementing such a system for large m, although clearly it is a rich regime for future research.

TABLE I
SUMMARY OF SCALABILITY RESULTS

Question Asked	**Source Density**	**Relay Density**	**Sensor Density**	**Solution**
Scalability?	Same	Same	Same	only if we make S-D pair distribution local
Information Extraction?	Fixed	Same	Same	only if correlation increases fast enough [5]
Scalability and Information Extraction?	Fixed	Traffic Dependent	Distortion Dependent	Yes by Source Separation

As $p \to \infty$, we will have at least one sensor node in very close vicinity to the point source (Fig. 1(b)). In this scenario, the rate distortion bound for the network reduces to the individual rate distortion bounds, for each point source. Hence, we have the separation of point sources in the network. The clear distinction of our approach from [5], [6] is that we do not need every node to gather information, only the closest. We could, for lower distortion, use partial side information [8] at the decoder from a correlated sensor node for the rate-distortion coding locally to each point source. This is sufficient to achieve desired distortion level D.

Theorem 3.1 ((Source Separation)): A network with finite Gaussian point sources, say m, and number of sensor nodes, p, going to infinity can be considered as a network with m separate coding systems with partial side information at the decoder. We assume that power decays at least as the square of the distance.

Proof: Since $p \to \infty$, the frequency re-use distance goes to zero and thus there is no interference between information pathways through the network. Hence, a simple relay is adequate for carrying the traffic even if it is not capacity achieving. For the nodes that are in very close vicinity to the Gaussian point sources, SNR for those nodes goes to infinity. So, any value of distortion, D, will be achievable and there is no substantial interference among sources. Hence, local processing will be sufficient. In the scenario considered here, each Gaussian point source can be separated from the rest and source coding can be posed as a partial side information problem. This is because Gaussian point sources are independent of each other. Since we have m point sources in the network of Fig. 1(b), it can be seen as m separate coding system with partial side information at the decoder for the correlated Gaussian sources. The assumption that the signal decays with distance faster than some rate is required to avoid interference growing without bound for large fields of sources. This assumption is quite reasonable for typical deployments of sensor networks for many physical phenomena of interest. ■

From Theorem 3.1 and limiting cooperation to only two nodes per source, it is evident that the data rate, R_{X_i}, $i = 1, 2, , \dots, m$, associated with each point source is [8]

$$\mathcal{R}_{X_i}\left(D_{X_i}\right) \geq \frac{1}{2}\log^{+}\left[\frac{\sigma^2_{X_i}}{D_{X_i}}\left(1-\rho^2+\rho^2 2^{-2R_{Y_i}}\right)\right] \tag{1}$$

where X_i is the main source and Y_i is a helper such that $(i = \{1, 2, \dots, m\})$. The rate distortion bound for the network will be the ensemble over the position of all the point sources. Clearly, extending to the q-helpers [3] per source $(q > 1)$, lower distortions would be achievable for a given density of sensors.

In terms of achievable distortion for a point source we have,

$$D_{X_i}\left(R_{X_i}\right) \geq \frac{\sigma^2_{X_i}}{2^{2R_{X_i}}}\left[\frac{\sigma^2_{X_i}}{D_{X_i}}\left(1-\rho^2+\rho^2 2^{-2R_{Y_i}}\right)\right] \tag{2}$$

for $i = \{1, 2, \dots, m\}$.

The achievable distortion for the network will be $D = \sum_{i=1}^{m} D_{X_i}$. For a point source, if we fix the node density and allow an algorithm to select a node closest to the source while making the rest of the nodes inactive, it is less likely to achieve the desired distortion. But this approach is a practical way to deal with achieving the desired distortion, at the cost of increased density. Note that we consider only those sensor nodes that are near to the source and it is sufficient to have only partial side information at the decoder. Now it is also possible that for given values of m, p and D, the capacity of the network may be inadequate. However, by allowing the number of communication relays $n \gg p$ then the information can be extracted. Large over-provisioning will enable decoupling of source and channel coding. The scalability results are summarized in Table I.

IV. SENSING DISTRIBUTED SOURCES

The sensed data are transmitted to the fusion center. Due to the stringent constraint on available data rate, the correlation among sensors should be exploited to bring down the communication cost. At the fusion center, an interpolation or approximation algorithm is used to reconstruct the source. During this process, we single out three types of errors. Sensing error is due to the ambient and circuit noise of sensor nodes, and can be reduced by increase the number of independent observations. Quantization error is bitterly constrained by the limited data rate of the network. The interpolation error is affected by both

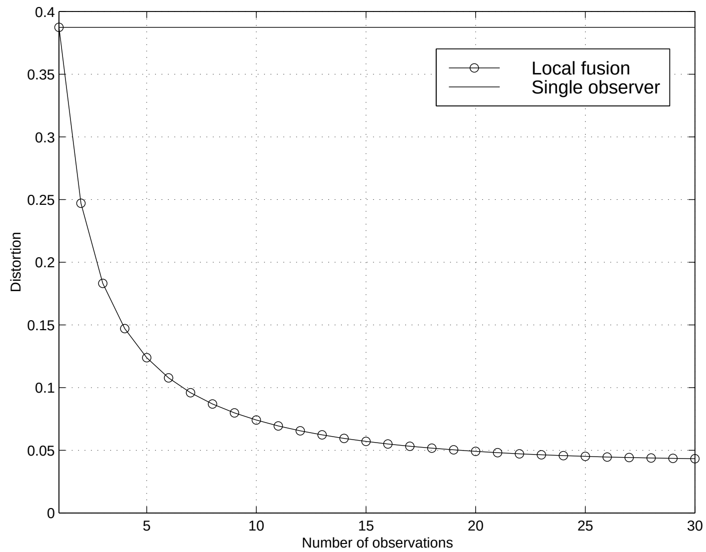


Fig. 2. Comparing distortions using single and multiple observations. $a_i = [0.1 - 0.001(i-1)], (i = 1, 2, \ldots, N_m)$, $\sigma_X^2 = 1$, $\sigma_Z^2 = 0.003162$, $D_l = D_g = \sigma_Z^2$.

the spatial sampling rate (or mesh size) and the former two types of errors. A hierarchical structure can be formed to further distribute the transmission and processing cost.

A. Sensing Point Sources

A simple sensing model for the point source is as follows:

$$Y_i = a_i X + Z_i$$

with

$$X \sim \mathcal{N}(0, \sigma_X^2), \quad Z_i \sim \mathcal{N}(0, \sigma_Z^2), \quad a_i = \frac{1}{1 + \kappa r_i^2}.$$

Signal attenuation is determined by source-sensor separation. Let us suppose that the sensors are deployed in a region according to a uniform distribution. It can be proved that if the region is a disc, the average distance between the source and the closest sensor is proportional to $1/\sqrt{n}$ i. e. $E(R_{min}) \propto \frac{1}{\sqrt{n}}$. Similar behavior can be expected for different geometry. Consider the distortion measure, $d(X, \hat{X}) = \frac{1}{n}\sum_{i=1}^{n}(X_i - \hat{X}_i)^2$ and optimal coding and estimation. The estimation and error at final reconstruction are as follows:

$$\hat{X} = \frac{a_i \sigma_X^2 Y}{a_i^2 \sigma_X^2 + \sigma_Z^2 + D_q} \tag{3}$$

$$D(n) = E_{R_{min}}\left[\frac{\sigma_X^2(\sigma_Z^2 + D_q)}{a_i^2 \sigma_X^2 + \sigma_Z^2 + D_q}\right]. \tag{4}$$

If more observations are used to estimate the source, the distortion can be effectively reduced.

$$\hat{X} = \frac{\sigma_X^2 \sum_{i=1}^{N_m} a_i (Y_i + Z_{li})}{D_l + \sigma_Z^2 + \sigma_X^2 \sum_{i=1}^{N_m} a_i^2} \tag{5}$$

$$D = \frac{(D_l + \sigma_Z^2)\sigma_X^2}{D_l + \sigma_Z^2 + \sigma_X^2 \sum_{i=1}^{N_m} a_i^2} + D_g \tag{6}$$

Fig. 2 compares the distortion using single and multiple observations. As the number of observer surpass a certain number, the return diminishes.

B. Sensing and Transmission of Distributed Phenomena

The observations at different sensors of distributed phenomenon can be considered as data generated from correlated point sources. Due to increasing number of sensors and data, the data rate and energy is sorely constrained. Hence, the correlation among sensors ought to be exploited to cut the rate needed to transmit the data samples. Although a great deal of local interaction and fusion may be required, this is still beneficial because the cost of local transmission (whose range is bounded) is far less expensive than global transmission (whose range is unbounded). The data rate limits on coding correlated sources have long been studied by the image processing community.

Consider N point sources $\mathbf{X} = \{X_i, i = 1, 2, \ldots, N\}$ in the space. The sensor measurements $\mathbf{Y} = \{Y_i : a_i X_i + Z_i, i = 1, 2, \ldots, N\}$ are attenuated signals corrupted by noise. At discrete times, the measurements $Y_i^1, Y_i^2, Y_i^3 \ldots$ at each sensor are zero-mean i.i.d. random variables. The samples $Y_1^n, Y_2^n, \ldots, Y_N^n$ measured at the same time are correlated with covariance matrix $\mathbf{Q_N}$, whose positive eigenvalues are given by $\lambda_1, \lambda_2, \ldots, \lambda_N$. No further information about the probability distribution of $\mathbf{Y}$ is assumed. Therefore, we are facing the problem of jointly coding the i.i.d. blocks of N samples generated from a class of random variables with the distortion requirement D_q such that $E[d(\mathbf{Y}, \hat{\mathbf{Y}})] \leq D_q$. For the distortion measure defined as $d(\mathbf{Y}, \hat{\mathbf{Y}}) = \frac{1}{N}\sum_{i=1}^{N}(Y_i - \hat{Y}_i)^2$, it can be shown that the superior of the minimum rate of this class of random variables is attained when $\mathbf{Y}$ is Gaussian [9], [10]. Thus, to be able to code this whole class of random variables and satisfy the given constraint, the minimum rate required is [9], [11]

$$R_g = \sum_{i=1}^{N} \frac{1}{2} \log\left(\frac{\lambda_i}{\min[\lambda_i, D^*]}\right) \tag{7}$$

and,

$$D_q = \frac{1}{N} \sum_{i=k}^{N} \min(\lambda_i, D^*) \tag{8}$$

C. Source Reconstruction

At the final reconstruction, two types of errors are to be distinguished. One is the interpolation error due to the discretizing process, and error propagated from the sensing and quantization process. The former is determined by the sampling mesh sizes. We shall consider the latter type of error.

Many of the interpolation algorithms such as cubic Hermite, cubic Bessel, cubic B-splines, etc. are local. Due to the nature of these local method, the error does not propagate to distant nodes. Therefore, it is expected that the error will be bounded. For example, consider the uniform cubic B-spline approximation. The approximate curve is constructed from bases by a set of control vertices V_i.

$$Q(\bar{u}) = \sum_i V_i B_i(\bar{u}) = \sum_i (x_i B_i(\bar{u}), y_i B_i(\bar{u}))$$

The resulting curve is in C^2 and satisfies the convex hull property ($Q = \sum_i w_i V_i$ where $w_i \geq 0$ and $\sum_i w_i = 1$). In other words, the i^{th} segment of a uniform cubic B-spline curve lies within the convex hull of the vertices $V_{i-3}, V_{i-2}, V_{i-1}$, and V_i. The error due to noisy measurement is to be characterized since the reconstruction is done by linear combining of the proximate vertices (using basis functions as the weights). For simplicity consider only the x- axis.

$$|\Delta Q_x| = \left|\frac{\partial Q_x(\bar{u})}{\partial x_i}\Delta x_i\right| = |B_i \Delta x_i| \leq |\Delta x_i|$$

Notice that $B_i \leq 1$ due to the normalization assumption.

Finally, consider a global algorithm: the cubic spline. Given the locations of $(N+1)$ points and a set of associated ordinates:

$$\Delta : \quad a = x_0 < x_1 < \cdots < x_N = b.$$

$$Y : \quad y_0, y_1, \cdots, y_N.$$

the spline function on $[x_{j-1}, x_j]$, $(j = 1, 2, \ldots, N)$ is defined as follows

$$S_\Delta = M_{j-1}\frac{(x_j - x)^3}{6h_j} + M_j \frac{(x - x_j)^3}{6h_j} + \left(y_{j-1} - \frac{M_{j-1}h_j^2}{6}\right)\frac{x_j - x}{h_j} + \left(y_j - \frac{M_j h_j^2}{6}\right)\frac{x - x_{j-1}}{h_j}$$

in which $h_j = x_j - x_{j-1}$. In general, the interpolation error uniformly converges with respect to x in $[a, b]$ as follows (Theorem 2.3.1, 2, 3 and 4 [12]):

$$e_1 = |f(x) - S_\Delta(x)| \leq K_1 \parallel \Delta_k \parallel^{n+\alpha}, \quad \text{for some constant } K_1$$

However, the spline reconstructed at the fusion center is not $S_\Delta(x)$ but a shifted spline $S^e_\Delta(x)$ due to the sensing and quantization error at measured points. It is given that the noise at prescribed points is bounded by: $E|\delta y_i| \leq e_{sq}$ and $E(\delta y_i)^2 \leq D_{sq}$. Firstly, consider the absolute error.

$$\begin{aligned} e &= E|f(x) - S^e_\Delta(x)| \leq |f(x) - S_\Delta(x)| + E|S_\Delta(x) - S^e_\Delta(x)| \\ &= e_1 + e_2 \end{aligned}$$

Note that since both $f(x)$ and $S_\Delta(x)$ are considered deterministic, the mean operation disappears for e_1. As for e_2, we have the following:

$$e_2 = E\left|\sum_{i=0}^{N} \frac{\partial S_\Delta}{\partial y_i}\,\delta y_i\right| \leq e_{sq}\sum_{i=0}^{N}\left|\frac{\partial S_\Delta}{\partial y_i}\right| \tag{9}$$

It can be shown that the following holds.

$$\beta = \sum_{i=0}^{N}\left|\frac{\partial S_\Delta}{\partial y_i}\right| \leq K_2, \qquad \text{for some finite number } K_2$$

Hence the total absolute error is bounded by

$$e \leq e_1 + \beta e_{sq} \tag{10}$$

The result can be easily extended to two-dimensional cubic splines. Thus, in either case with finite spatial sampling the distributed source can be reconstructed with bounded error, even with relatively simple techniques. If the density of communication relays increases, sufficient capacity will be available for extracting the desired information. Thus, with the addition of spatial fidelity constraints, sensing of distributed phenomena is also scalable.

V. Conclusion

The scalability results presented here reflect local processing to change the source-destination distribution, namely, to suppress message-generation activity in most of the nodes so that they serve only as communication relays. Once sufficient sensing density is achieved to meet the distortion requirements, adding more nodes only increases the communication resources. Clearly however more sophisticated strategies can lead to meeting the information extraction requirements with lower densities. These are interesting objects of research.

References

[1] D. Estrin, L. Girod, G. Pottie, and M. Srivastava, “Instrumenting the world with wireless sensor networks”, *In Proceedings of the International Conference on Acoustics, Speech, and Signal Processing*, vol. 4, pp. 2033, Salt Lake City, UT.

[2] J. M. Kahn, R. H. Katz and K. S. J. Pister, “Emerging challenges: Mobile Networking for Smart Dust”, *Journal of Communications And Networks*, vol. 2(3), pp. 188-196, September 2000.

[3] A. Pandya, A. Kansal, G. Pottie, and M. Srivastava, “Bounds on the Rate-Distortion of Multiple Cooperative Gaussian Sources,” *submitted to IEEE Tran. Info. Theory* for review and possible publication, *CENS Tech. Report*, Sep. 2003

[4] P. Gupta and P. R. Kumar, “The capacity of wireless networks,” *IEEE Tran. Inform. Theory*, vol. 46, Mar. 2000.

[5] S. D. Servetto, “On the feasibility of large scale wireless sensor networks,” in *Proc. 40th annual Allerton conf. on commun., contr. and comput.*, 2002.

[6] 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. Int. Workshop Inform. Processing in Sensor Networks (IPSN)*, April 2003.

[7] Jinyang Li, Charles Blake, Douglas S. J. De Couto, Hu Imm Lee, and Robert Morris, “Capacity of ad hoc wireless networks”, *In Proceedings of the 7th ACM International Conference on Mobile Computing and Networking*, pp. 61-69, Rome, Italy, July 2001.

[8] Y. Oohama, “Gaussian multiterminal source coding,” *IEEE Trans. Inform. Theory*, vol. 43, pp. 1912-1923, Nov. 1997.

[9] D. J. Sakrison, “The rate of a class of random processes,” *IEEE Trans. Inform. Theory*, vol. IT-16, pp. 10-16, Jan. 1970.

[10] I. E. Telatar, “Capacity of multiple-antenna Gaussian channel,” *Technical Report*, AT&T-Bell labs, June 1995.

[11] L. D. Davisson, “Rate-distortion theory and application,” *Proc. IEEE*, vol. 60, no. 7, July 1972.

[12] J. H. Ahlberg et al. *The Theory of Splines and Their Applications*, Academic Press, 1967.