

# Perspectives on data compression for estimations from sensors

Mark L. Fowler\* and Mo Chen

Dept. of Electrical and Computer Eng., Binghamton University, Binghamton, NY, USA 13902;

## ABSTRACT

Data compression methods have mostly focused on achieving a desired perception quality for multi-media data for a given number of bits. However, there has been interest over the last several decades on compression for communicating data to a remote location where the data is used to compute estimates. This paper traces the perspectives in the research literature for compression-for-estimation. We discuss how these perspectives can all be cast in the following form: the source emits a signal – possibly dependent on some unknown parameter(s), the  $i^{\text{th}}$  sensor receives the signal and compresses it for transmission to a central processing center where it is used to make the estimate(s). The previous perspectives can be grouped as optimizing compression for the purpose of either (i) estimation of the source *signal* or (ii) the source *parameter*. Early results focused on restricting the encoder to being a scalar quantizer that is designed according to some optimization criteria. Later results focused on more general compression structures, although, most of those focus on establishing information theoretic results and bounds. Recent results by the authors use operational rate-distortion methods to develop task-driven compression algorithms that allow trade-offs between the multiple estimation tasks for a given rate.

**Keywords:** Data Compression, Remote Sensing, Sensor Networks, Estimation

## 1. INTRODUCTION

Advances in sensor and communication technology have focused interest on using networks of sensors. Sensor node tasks include collecting data, sharing the data, collectively making inferences (estimations and decisions) from the data, and communicating the collected data and/or the inference results to one or more information sinks where further inferences may be made; we'll focus here on making estimates. Sensor networks must be designed to satisfy constraints on metrics assessing energy efficiency, accuracy of the conveyed information, and communication latency, and there is a fundamental tradeoff among these metrics<sup>17</sup>.

Energy efficiency in sensor networks can be addressed through routing protocols, sleeping strategies, low-power architectures, and energy-efficient modulation schemes. Accuracy is generally controlled through optimal processing strategies as well as using accurate sensors deployed in optimal ways. Latency and channel capacity issues in sensor networks can be addressed through routing strategies and data compression<sup>15,16</sup>. However, it has become clear that data compression – normally viewed primarily as a tool for the capacity/latency vs. accuracy trade-off – is a natural tool for achieving the desired trade-off between energy, accuracy, and latency in sensor networks. This expanded role of compression arises as follows. Transmission of data is one of the most energy-expensive tasks a node undertakes – thus, using data compression to reduce the number of bits sent reduces energy expended for transmission. However, compression requires computation, which also expends energy. Fortunately, trading computation for transmission has potential to save energy: it has been asserted<sup>14</sup> that typically on the order of 3000 instructions can be executed for the energy cost required to transmit one bit over a distance of 100 m by radio – thus, the trade seems to favor spending computation to reduce the amount of transmission needed. Furthermore, since a primary task of sensor networks is to make estimates based on the data collected throughout the network, it is important to design compression methods that cause minimal degradation of the accuracy of these estimates.

One aspect of networked sensors that is not commonly discussed is the fact that there are many types of sensor networks (see Section 2) and many types of problems that such networks are envisioned to solve. Thus, it may not be possible to

---

\* mfowler@binghamton.edu; 607-777-6973; <http://www.ws.binghamton.edu/fowler/>

This work was supported in part by the Air Force Office of Scientific Research through Grant number FA9550-06-1-0249 and through the Air Force Summer Faculty Fellowship Program.

develop a universal approach to data compression that is applicable to each and every network of sensors. This paper will discuss the many different aspects of data compression that arise within sensor applications and outline the main perspectives that have been brought to bear on these problems.

## 2. IMPACT OF TYPES OF SENSOR NETWORK ON COMPRESSION

The area of sensor networks has recently attracted much interest from researchers after early papers (such as Pottie and Kaiser<sup>14</sup>) discussed the potential of using large numbers of densely-distributed miniature low-power sensors that are wirelessly networked together to share their collected data and make estimates from that data. Today, the term “sensor network” is almost universally taken to indicate this scenario of many small low-power sensors.

The benefit of data compression within this context is two-fold: shorten the time needed to share the data and reduce the energy expended to share the data. However, even within this fairly well-defined context of “many small, low-power sensors” there are many variations on the theme of “sensor network”, and those variations can impact the development of data compression methods suited to the application. Furthermore, there are many applications that exhibit aspects of networked sensors that can benefit from data compression but don’t require the “many sensors” and “low-power” aspects typically associated with “sensor networks”; yet we can still think of these applications as “sensor networks”. In this paper we use “sensor network” in a broader context than is typical in most recent literature. This broader context of sensor networks provides significant variety and therefore is likely to require a variety of different data compression approaches. Table 1 illustrates some of the significant variations in sensor networks characteristics. While the first two characteristics do impact the requirements placed on the compression algorithm, it is the last four that have the largest impact on compression. The sensor size and sensor power characteristics pertain to the physical characteristics of the sensor nodes that make up the network. The characteristic called hierarchy here pertains to where the processing of the data is performed: in a centralized hierarchy the raw sensor data is moved to a fusion center where it is processed en masse to obtain the required estimate, whereas in a non-centralized hierarchy the data is shared between sensors and the estimation processing is performed in a distributed fashion throughout the network. Routing characteristics pertain to how the data is passed between sensor nodes to progress from where it is collected to where it is used. The density and number of the sensor nodes characterizes the deployment of the sensor nodes. Finally, aggregativity describes from how many different sensors data is needed to support making the desired estimate; for example, if the estimate can be made based on each sensor’s data individually then we state the aggregativity as “Single-Sensor-Capable” but if two sensors’ data must be combined to make the estimate then we state the aggregativity as “Two-Sensor-Critical”<sup>6</sup>.

The sensor size impacts the data compression algorithm mainly through the fact that micro sensors will typically have limited processing capabilities, thus requiring the development of compression algorithms that are computationally efficient. Likewise, low-power sensors will require computationally efficient compression algorithms that can run on low-power processors; however, as mentioned above, compression can be a tool to achieve the low-power requirement by reducing the amount of data needed to be sent by power-hungry transmission while incurring a small increase in power consumed for computation. Although there is much interest in applications of low-power micro-sensors, there are also many outstanding problems involved in the networking of medium-power macro-sensors (e.g., the networking of sensors on aircraft, unmanned aerial vehicles (UAVs), and satellites).

Table 1: Sensor networks have various characteristics, each of which can take on different values over a range.

<b>Characteristic</b>	<b>Range</b>	
<b>Sensor Size</b>	Micro	Macro
<b>Sensor Power</b>	Limited	Abundant
<b>Hierarchy</b>	Non-Centralized	Centralized
<b>Routing</b>	Single-Hop	Multiple-Hop
<b>Density &amp; Number</b>	Low	High
<b>Aggregativity</b>	Single-Sensor-Capable	Multi-Sensor-Critical

## 2.1 Hierarchy of the sensor network

There can be many aspects that could be considered part of the hierarchy of a sensor network but here the focus is on where the estimation processing is done. Figure 1 shows three typical types. In a non-centralized setting a sensor shares its data with its neighbors (typically using compression) and then processing is done in a distributed manner throughout the sensors. In a completely centralized hierarchy each sensor compresses its data and transmits directly to a fusion center.

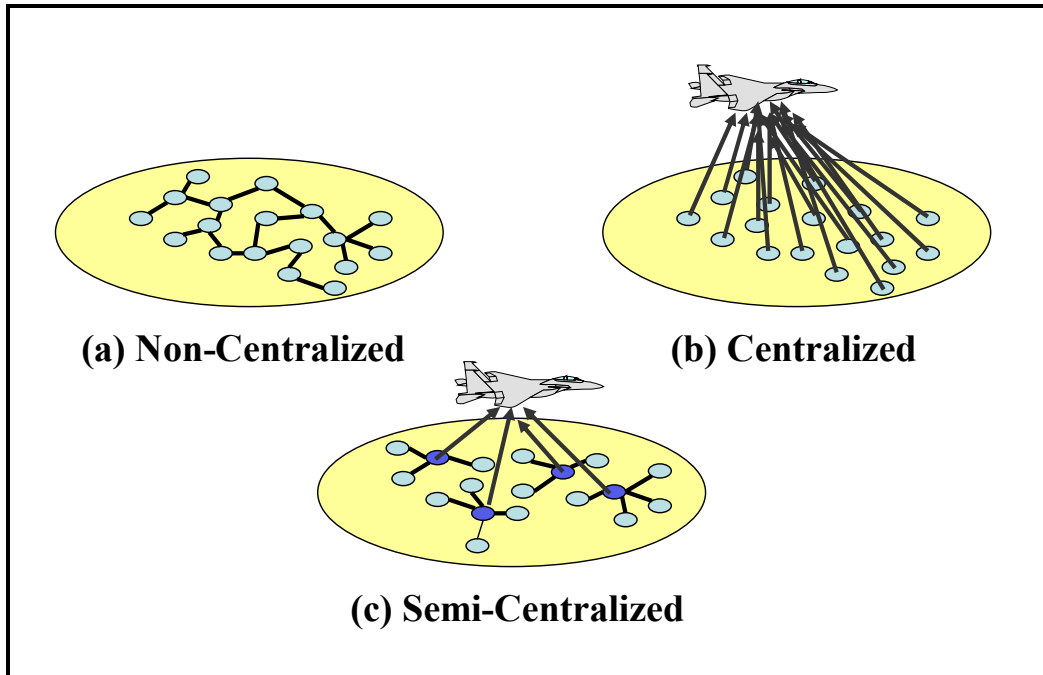


Figure 1: Three types of hierarchy. (a) Non-centralized sensor networks share the data between sensors and the estimation processing is done within the sensors in the networks. (b) Centralized sensor networks move the data back to a “fusion center” that may or may not be a sensor in the network (here the fusion center is shown as an aircraft flying over the sensor field). (c) Semi-centralized is a hybrid between the two where data is shared between sensors and then only certain sensors (“cluster heads”) communicate directly with the fusion center; various degrees of processing and/or compression could be performed at the cluster heads.

The trade-off between these two extremes is the following:

- Centralized Hierarchy
  - Typically results in simpler processing because processing is non-distributed and has access to all of the data
  - Typically results in large transmission costs because all sensor data must get moved to the fusion center; some data may have to transmit over significantly large distances.
- Non-Centralized Hierarchy
  - Typically results in more challenging processing because processing must be distributed between the sensors and there is no point where the processing has access to all the raw data simultaneously
  - Typically results in lower transmission costs because sensors only transmit data to nearby sensors
- Semi-Centralized Hierarchy

- Eases the challenge of decentralized processing
- Eases the need to transmit all the raw data back to the fusion center

Although the semi-centralized hierarchy appears to be a worthwhile approach there may be applications where either of the other two approaches may be preferred.

## 2.2 Routing within the sensor network

Many researchers have studied the need for and the advantages of routing within a sensor network.<sup>10,16,17</sup> One goal of routing is to allow efficient communication of the data either (i) back to a fusion center or (ii) throughout all the nodes in the sensor network.

One of the first published results on the interconnection between data compression and routing actually didn't use the term "data compression" but none-the-less demonstrated the usefulness of coupling routing with data compression.<sup>10</sup> That paper considered a very specific application of sensor networks: nodes in the network are organized into clusters, data from nodes within a cluster is sent to the cluster head where it is combined through beamforming into a single data stream. Thus, the beamforming processing performed at each cluster head is a form of data compression: if there are  $M$  nodes in the cluster then the compression ratio achieved is  $M:1$ . The beamformed data from each cluster head is then sent directly back to a fusion center; this approach was dubbed LEACH. A short step from this beamforming-specific approach to a more general approach is to replace the beamforming with a general data compression performed at each node and then send directly to the fusion center. Results showing the capabilities of this compression-with-direct-transmission approach have been previously presented<sup>2</sup> and are shown in Figure 2.

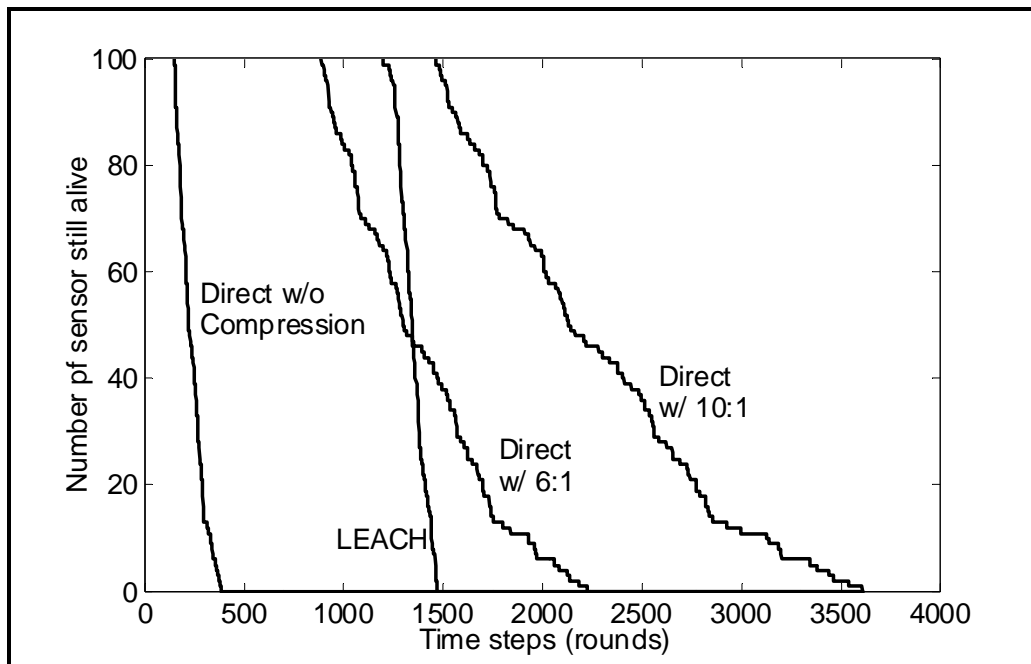


Figure 2: Non-distributed compression improves network lifespan when using direct transmission in a way that is similar to the improvement obtained using the LEACH method of routing and beamforming.

These results show that the gain in lifetime obtained by LEACH is not due to its *routing* but rather to the beamforming's effective data compression. However, the next subsection will discuss that routing is inherently linked with data compression.

### 2.3 Density and number of sensors

A common current viewpoint of researchers in the sensor network area is toward sensor networks that have a large number of sensors that are densely distributed over a physical region to be monitored. However, it should be mentioned that many applications exist that would not benefit from (or can't afford) a dense deployment of a large number of sensors.

When a sensor network is densely deployed it has a major impact on the data compression algorithm. For example, a dense deployment will likely result in a large correlation in the sensed data observed at adjacent sensors. Thus, a data compression algorithm for such an application should take advantage of this correlation. Many examples of this are available in the literature and a few<sup>2,15,16</sup> are given in the references section.

A demonstration of the potential gain from exploiting inter-sensor correlation has been previously presented by the current authors.<sup>2</sup> As a *simple* demonstration we ran simulations to characterize the effect of exploiting this spatial redundancy. The results are shown in Figure 3 where we have simulated the effect of signals within a radius of 10m of a randomly selected set of primary nodes as having virtually the same information content. The spatially correlated data is compressed at each cluster head in a semi-centralized hierarchy and then the remaining data sets are compressed using non-spatial methods having a compression ratio of 6:1, after which all compressed data is sent to the fusion center using direct transmission. This simulation considers repeatedly collecting, compressing and transmitting a set of data at a sequence of time steps (or rounds). The results in Figure 3 show that by exploiting this spatial redundancy through compression it is possible to extend the life of the sensor network (as measured through the number of sensors still having power): one curve shows the results when a 6:1 compression ratio is applied to the data at each sensor; a second curve shows the results when spatial correlation is first exploited at a cluster head and then the remaining data is compressed 6:1 and sent from the cluster head to the fusion center.

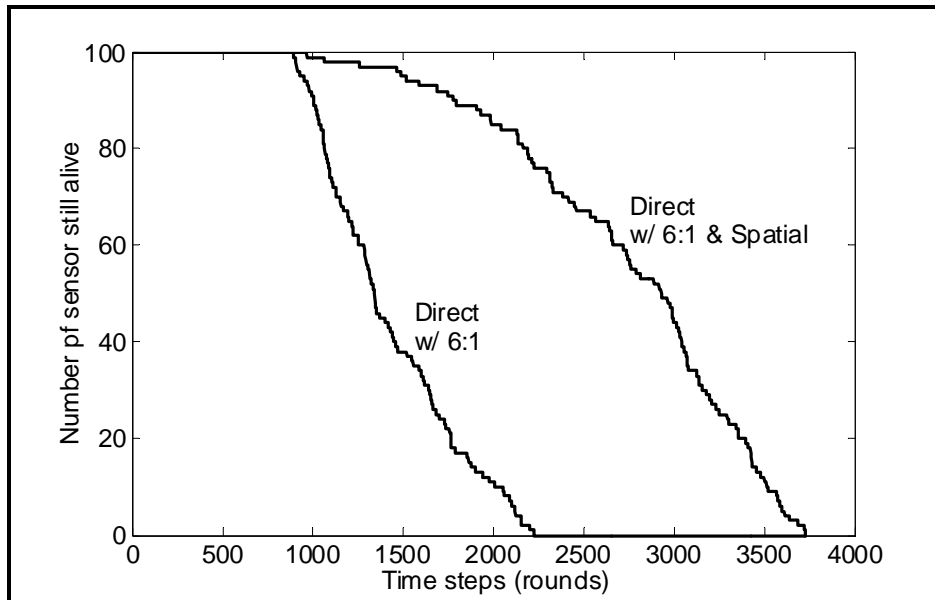


Figure 3: Results showing improvement using compression that exploits spatial correlation within a semi-centralized hierarchy.

A compression algorithm for exploiting spatial correlation in a distributed compression setting has been given by Ramchandran and his students.<sup>15</sup> Unlike that illustrated in Figure 3, the approach here requires no sharing between the two correlated sensors – although it does rely on some level of knowledge of an inter-sensor correlation model. The importance of this paper lies in the fact that it was the first demonstration of a practical way to perform true distributed compression of correlated sensor data by exploiting the ideas put forth in the Slepian-Wolf theorem. Unfortunately, this distributed compression approach is not suited to some sensor network applications due to the lack of a reliable model

for the spatial correlation. However, Scaglione and Servetto<sup>16</sup> showed that it is possible to achieve high-capacity data transport within a large and dense sensor network without resorting to such distributed compression approaches. Their results were the first to theoretically show a tight interdependence between data compression and routing within a sensor network. It should be noted that neither of these results<sup>15,16</sup> explicitly considers the fact that the sensor-collected data is to be used for estimation processing – the specifics of compression within the estimation-task context is discussed in Section 3.

## 2.4 Aggregativity

It is important to point out that in the vast number of research papers published on sensor networks it is possible to find a great many different signal models and application scenarios being considered. For example, a very common one is that the signal collected by each sensor *could* be used directly to compute the estimate (and then the individual estimates could be fused) but this is not desired because better performance can be achieved by combining data from multiple sensors. This is what we call a Single-Sensor-Capable problem. In such problems there is an underlying assumption that there is an advantage to be gained by having the data from all the sensors (or at least a large number of them) be processed jointly (either at a fusion center or using some distributed scheme).

However, another scenario that arises in some applications is that the signal collected by each sensor *can't* be used directly to compute the estimate but rather you must have data from two (or more) sensors to estimate the parameter of interest. For example, if using simple sensors that each have a single omni-directional antenna it is not possible to estimate a signal source's location from a single sensor's data. Instead you must jointly process data from two or more sensors. This is what we call a Multiple-Sensor-Critical problem. Another such scenario might be estimating the spatial gradient field of a chemical in the air as a function of time. In such cases there seems to be motivation for doing some degree of processing within the sensors rather than sending all data to a fusion center. In fact, this is exactly the scenario that is well-suited to a semi-centralized hierarchy. As an example, consider a network having  $N$  sensors that are grouped into  $N/2$  pairs of sensors; one sensor in each pair compresses its raw signal data and shares it to its paired sensor where it is used to estimate a TDOA/FDOA\* value for the pair, then these  $N$  numbers ( $N/2$  TDOA values and  $N/2$  FDOA values) are then sent to a fusion center (most likely one of the original sensor nodes). Note that in this setting there would be little advantage to having all the sensors' raw data available at one location but the communication cost would be great. Notice also that the estimation of the TDOA/FDOA within a pair is essentially compression at a large compression ratio!

## 3. HISTORICAL PERSPECTIVES OF DATA COMPRESSION FOR ESTIMATION

In the section above we have described many different scenarios that impact the requirements of the data compression approach for a sensor network. In this section we take a look at some of published results for developing data compression algorithms that are tailored for the task of estimation processing.

Over the past several decades many sensor-relevant compression results have been published. Generally, these results are cast in the form shown in Figure 4. The source emits a source signal  $S(\theta)$  depending on an unknown source parameter  $\theta$ , the  $i^{\text{th}}$  sensor receives signal  $X_i(\theta)$  that is related to  $S(\theta)$  and encodes it as  $\hat{X}_i(\theta)$ , which is transmitted to a central processing center where the encoded signals are decoded and then used to make an estimate; Figure 4 shows the estimate being either estimating the parameter  $\theta$  or estimating the signal  $S(\theta)$ .

These compression-estimation results can be grouped in several ways. The focus is either on (i) optimizing compression for the purpose of estimation of the source *signal*<sup>1,5,15,16</sup> or (ii) optimizing compression for the purpose of estimating the source *parameter*<sup>8,9,11,19</sup>. Many focus on restricting the encoder to being a scalar quantizer that is then designed according to some optimization criteria<sup>1,5,8,11</sup>. Others focus on more general compression structures<sup>9,15,16,19</sup>, although, most of these focus on establishing information theoretic results and bounds<sup>9,16,19</sup>. Only a few results specifically address issues in sensor networks<sup>15,16</sup>.

---

\* TDOA = "Time-Difference-of-Arrival" and FDOA = "Frequency-Difference-of-Arrival"

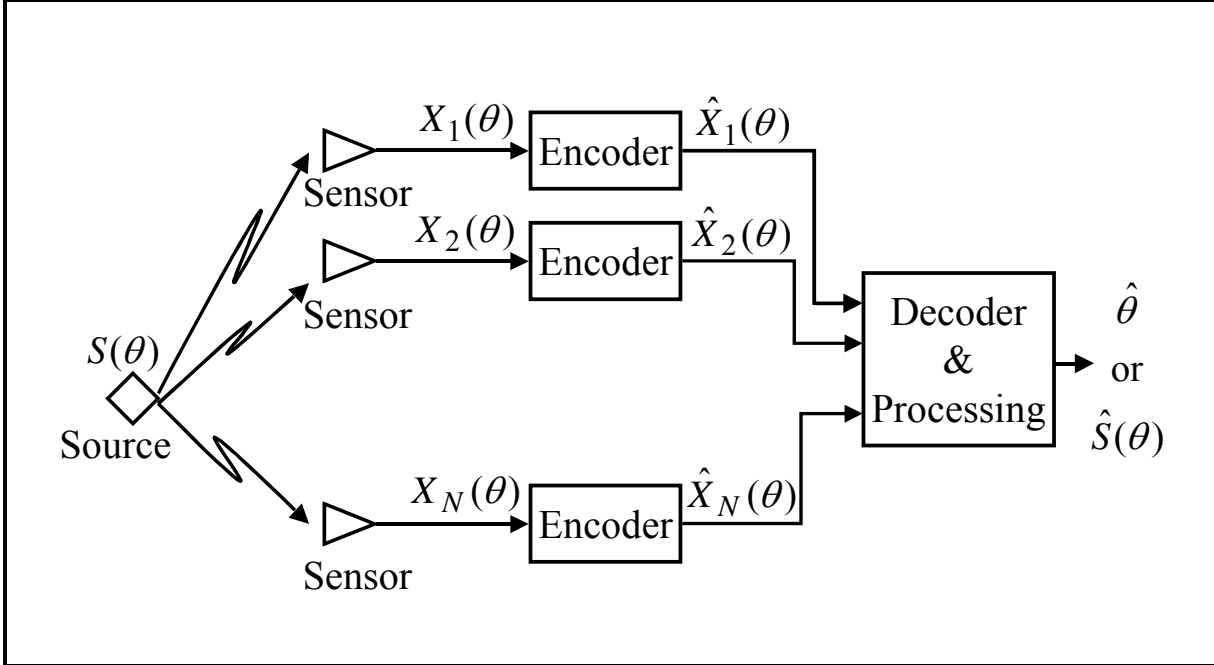


Figure 4: The classical set-up for compression in a distributed sensor system consists of  $N$  distributed sensors that compress and send their data to a central fusion center that decodes the data and processes it.

In the remainder of this section we give an historical overview (roughly in chronological order) of some significant results in the area of compression for the purpose of estimation. We have by no means included all significant results (due to space considerations) but rather have selected a set of representative papers that illustrates the progression of ideas within the field. We focus on papers that address the scenario of estimating a *deterministic parameter* from sensor-collected signal data (rather than on those that focus on estimation of the underlying *signal* itself).

### 3.1 Early results on compression for estimation

Perhaps one of the earliest papers in this area was by Flynn and Gray<sup>5</sup> in 1987. This paper first establishes rate-distortion bounds for the problem then constrains compression at sensors to being quantization. Then the focus is turned to finding optimal quantizers and fusion center processing, where distortion function is the MSE of the parameter estimate. This is done under the assumption that the designer knows the probability structure of the data to be compressed. An alternate approach is developed that uses Ali-Silvey distances to measure the impact of the quantization on the data's ability to allow estimation of the parameter.

A year later Poor published a paper<sup>13</sup> that explored ideas along the same lines of using the Ali-Silvey distances used by Flynn and Gray. Poor proposed designing a quantizer to maximize  $f$ -divergences, which are measures of discrimination information between statistical hypotheses. He shows that when applied to parameter estimation problems this is essentially the same as maximizing the Fisher information (FI) after quantization. In that same year Zhang and Berger<sup>19</sup> established information theoretic results that illustrate how the variance of an unbiased estimator using compressed data depends on the communication rate. Although no algorithmic approach was given the results provided an early characterization of the problem.

In 1993 Lam and Reibman further explored the idea introduced by Poor of maximizing the FI after quantization<sup>11</sup>. They used a generalization of the Lloyd-Max algorithm to design vector quantizers that maximize the FI for a single scalar parameter to be estimated. This quantizer then must be designed in advance based on a known probability structure or a set of training data. If there is much variation in the type or structure of the signals to be compressed, then the design of such quantizers becomes difficult if not impossible.

In 1995 Han and Amari published an extensive and intricate paper<sup>9</sup> establishing a rich theoretical basis for the problem. In particular, they essentially extended the classical Cramer-Rao bound (the inverse of the FI) result to the case where the bound is now a function of the rate used to encode the sensors' signals. They also established the theoretical basis for the resulting maximum-likelihood estimators to use with rate-constrained sources.

One should notice that all of these results are quite theoretical in nature; they provide excellent insight into the underlying problem but provide results that are hard to apply in practice – especially in a sensor network setting where the signals to be encountered can be quite variable in nature making methods based Lloyd-Max-type quantizers difficult. Also, none of these results focused on the specific characteristics of compression within a sensor network; a few results specifically aimed at sensor networks will be discussed in subsection 3.2. One should also note that with the exception of Han and Amari, all of these results focused on deriving optimal quantizers; some results using transform coding will be presented in subsection 3.3.

### 3.2 Compression developed specifically for sensor networks

As we mentioned above, Ramchandran and his students<sup>15</sup> have developed a distributed compression method specifically aimed at the sensor network arena that attempts to exploit the correlation between the data at two sensors for transmission to a third location without having to share any side information between the two compressing sensors. To achieve this, however, it is necessary to have a model for the inter-sensor correlation. Many sensor network applications do not admit such a model that can be known for all cases of expected data collection. Nonetheless, the results here provide important theoretical insight into the problem.

Also mentioned above, Scaglione and Servetto<sup>16</sup> showed that by using routing together with non-distributed compression (i.e., standard compression approaches) it is possible to achieve flexible compression methods for sensor networks that can exploit the inter-sensor correlation without having an *a priori* correlation model.

With the rapidly growing recent interest in sensor networks there are many newer results that we are unable to discuss in detail here due to space limitations. As a good starting point see the very recent paper by Xiao et. al<sup>18</sup> and the references contained therein. The current viewpoint of these researchers is to seek optimal designs of quantizers for compressing measurements from sensors, where the estimate is made from one data point from each sensor.

### 3.3 Transform coding for parameter estimation

Unlike the results discussed above, we have explored the development of transform coding methods for sensor networks rather than the design of quantizers<sup>2,3,4,6,7</sup>. Like Poor<sup>13</sup> and Lam and Reibman<sup>11</sup> we use the Fisher information as a distortion metric<sup>3,4,6</sup>; however, our approach differs in that we attempt to measure the FI directly from the received data<sup>7</sup>.

We model the collected sensor data as a deterministic signal plus additive Gaussian noise having variance of  $\sigma^2$ . We seek to compress a block of data collected at a sensor  $S_1$  so that it can be transmitted to another sensor  $S_2$  using no more than a budgeted  $R$  bits while making the estimate at  $S_2$  (using the compressed data and  $S_2$ -local data) with the lowest possible RMS estimation error. Our approach is to develop a transform-based compression scheme that is operationally optimized with respect to a distortion measure that uses Fisher information to quantify the degrading effect of compression on estimation accuracy.

Let the real data vector  $\mathbf{x}$  be drawn from a probability density function (PDF)  $p(\mathbf{x}; \theta)$  that is parameterized by  $\theta$ , which is to be estimated. The Fisher information of this estimation problem is defined to be

$$J(\theta; \mathbf{x}) = E \left\{ \left[ \frac{\partial \ln p(\mathbf{x}; \theta)}{\partial \theta} \right]^2 \right\}, \quad (1)$$

where the expected value is taken with respect to  $p(\mathbf{x}; \theta)$  and therefore the Fisher information is not a function of the data vector  $\mathbf{x}$ . However, the *notational* dependence on  $\mathbf{x}$  shown on the left-hand side of (1) is included merely to keep track of the data set or data subset for which the Fisher information is computed. As indicated on the left-hand side of (1), the Fisher information can be a function of the parameter to be estimated, although in many cases it is not.



Clearly, compression of the data vector  $\mathbf{x}$  using a lossy algorithm changes the underlying PDF and therefore changes the Fisher information. Roughly, then, our goal here is to seek *operational* rate-distortion methods to maximize the amount of Fisher information remaining in the data set while satisfying a budget  $R$  on the rate used to represent the data set. Our approach (see Figure 5) is to transform the original data into some appropriate set of coefficients  $\boldsymbol{\chi} = \{\chi_n | n=1, 2, \dots, N\}$ , only some of which are then selected ( $\Omega$  is set of indices for the selected coefficients) and quantized to give the set  $\hat{\boldsymbol{\chi}} = \{\hat{\chi}_n | n \in \Omega\}$ . The resulting set of transform coefficients  $\hat{\boldsymbol{\chi}}$  then has Fisher information  $J(\hat{\boldsymbol{\chi}}) \leq J(\mathbf{x})$ , where the reduction in Fisher information is due to the compression processing of selection and quantization. For notational use let  $\chi_n = \xi_n + \omega_n$  where  $\chi_n$  is a coefficient of the “signal+noise” while  $\xi_n$  and  $\omega_n$  are the signal coefficient and noise coefficient, respectively.

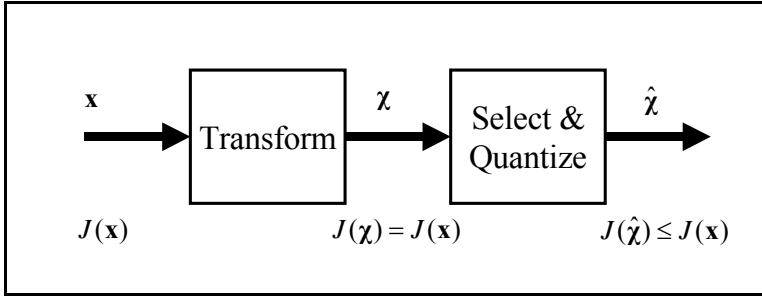


Figure 5: Single-estimation-task compression processing, the data vectors, and their corresponding Fisher informations.

We have previously shown<sup>6</sup> that the Fisher information after selection/quantization can be expressed as

$$\begin{aligned}
 J(\hat{\boldsymbol{\chi}}) &= \sum_{n \in \Omega} \left[ \frac{\partial \xi_n(\theta)}{\partial \theta} \right]^2 \frac{1}{\sigma^2 + q_n^2} \\
 &= \sum_{n \in \Omega} \frac{\Gamma^2(\xi_n)}{\sigma^2 + q_n^2},
 \end{aligned} \tag{2}$$

where  $\Gamma(\xi_n) = \frac{\partial \xi_n(\theta)}{\partial \theta}$  captures the signal’s sensitivity to a change in the parameter and  $q_n^2$  is the variance of the quantization noise in the  $n^{\text{th}}$  quantized coefficient. Note that the form of (2) has a nice interpretation: we are compressing to maintain a high level of what we call the “signal-sensitivity-to-noise-ratio” (SSNR). Our goal then is to select coefficients and allocate bits  $\{b_n\}$  to them so as to maximize the SSNR in (2) subject to a rate constraint  $R$ . To implement our method, one must derive the form of  $\Gamma(\xi_n)$  for the desired estimation task as a function of the *signal* coefficients  $\xi_n$ . However, because we must compute  $J(\hat{\boldsymbol{\chi}})$  from our *data* coefficients  $\chi_n$ , we must instead use a noisy version given by

$$\hat{J}(\hat{\boldsymbol{\chi}}) = \sum_{n \in \Omega} \frac{\Gamma^2(\chi_n)}{\sigma^2 + q_n^2}, \tag{3}$$

where the “hat” is used to indicate that the quantity uses the noisy quantities that are available from the data rather than the noise-free values really needed. Furthermore, we define

$$\hat{J}_n = \begin{cases} 0, & \text{if } b_n = 0 \\ \frac{\Gamma^2(\chi_n)}{\sigma^2 + q_n^2}, & \text{if } b_n > 0, \end{cases} \quad (4)$$

which combines the selection and quantization by explicitly saying that  $\hat{J}_n = 0$  when  $b_n = 0$ ; that is, a coefficient that is allocated 0 bits is not sent and can't possibly contribute any information to the estimation.

Thus our method can be stated as follows. Given an explicit form for the signal sensitivity function  $\Gamma(\cdot)$  derived for the desired estimation problem, find a bit allocation set  $B = \{b_n \geq 0 | n \in 1, 2, \dots, N\}$  that solves

$$\max_B \left\{ \sum_{n=1}^N \hat{J}_n \right\} \quad \text{subject to} \quad \sum_{n=1}^N b_n \leq R, \quad (5)$$

where  $\hat{J}_n$  is as given in (4).

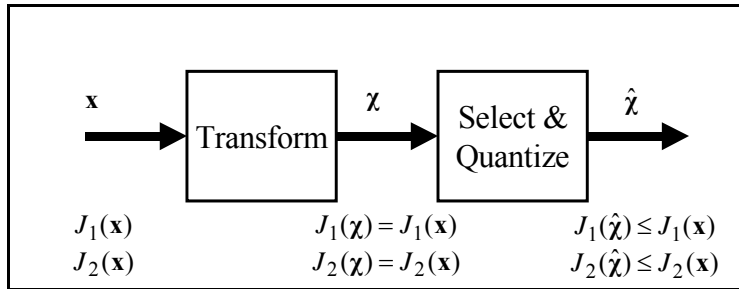
When there are multiple parameters to be estimated from the data set  $\mathbf{x}$  obviously it is desirable to maximize each Fisher information; Figure 6 shows the case for two estimates, where  $J_1$  and  $J_2$  are the two Fisher informations. However, it is likely that the optimal compression that maximizes  $J_1(\hat{\chi})$  will not also maximize  $J_2(\hat{\chi})$ . A further challenge lies in the fact that the importance of the multiple estimation tasks are not equal. This motivates the following: if we let  $J_k(\hat{\chi})$ ,  $k = 1, 2, \dots, K$  be the Fisher informations for the  $K$  desired estimation tasks and each has an importance weight  $\alpha_k$  with

$$\sum_{k=1}^K \alpha_k = 1; \quad (6)$$

then we strive to find the bit allocation that maximizes

$$\sum_{k=1}^K \alpha_k J_k(\hat{\chi}) \quad \text{subject to} \quad \sum_{n \in \Omega} b_n \leq R. \quad (7)$$

By using (7) we can easily apply our single-estimate results to the multi-estimate case.



**Figure 6: Double-estimation-task compression processing, the data vectors, and their corresponding Fisher informations.**

To apply this approach to a particular estimation problem the parameter sensitivity function  $\Gamma(\xi_n)$  must be derived for each estimation task and appropriate weights  $\alpha_k$  are specified; then (7) is maximized using an efficient Lagrange multiplier approach.

We have applied this approach to the problem of locating an RF emitter using a network of aircraft-mounted sensors<sup>3,4</sup>, where there is a fundamental need to accomplish trade-offs between the impact of compression of TDOA accuracy and FDOA accuracy. For this application it is clear that the quantization approaches that were developed by early researchers<sup>11,13</sup> as well as current researchers in sensor networks<sup>18</sup> can not handle this trade-off issue. However, for other sensor network applications the transform coding approaches may not be applicable due to their complexity. Thus, there seems to be a need for a wide variety of different compression approaches within the field of sensor networks.

## REFERENCES

1. M. Di Biseegli and M. Longo, "Decentralized encoding of a remote source," *Signal Processing*, 55 (1996), pp. 15 – 29.
2. M. Chen and M. L. Fowler, "The Importance of Data Compression for Energy Efficiency in Sensor Networks," Conference on Information Sciences and Systems, Johns Hopkins University, March 12-14, 2003.
3. M. Chen and M. L. Fowler, "Geometry-Adaptive Data Compression For TDOA/FDOA Location," in Proceedings of IEEE ICASSP 2005, March 18 – 23, 2005, pp. 1069 – 1072, Vol. 4.
4. M. Chen, M. L. Fowler, A. Noga, "Data Compression for Simultaneous/Sequential Inference Tasks in Sensor Networks," Invited Paper, IEEE ICASSP 2006, Toulouse, France, May 14 – 19, 2006.
5. T. J. Flynn and R. M. Gray, "Encoding of correlated observations," *IEEE Transactions on Information Theory*, vol. IT-33, no. 6, Nov. 1987, pp. 773 – 211.
6. M. L. Fowler and M. Chen, "Fisher-Information-Based Data Compression for Estimation Using Two Sensors," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 41, no. 3, July 2005, pp. 1131 - 1137.
7. M. L. Fowler and M. Chen, "Evaluating Fisher Information From Data for Task-Driven Data Compression," Proceedings of Conference on Information Sciences and Systems, Princeton University, March 22-24, 2004, pp. 976 – 972.
8. J. A. Gubner, "Distributed estimation and quantization," *IEEE Transactions on Information Theory*, vol. 39, no. 4, July 1993, pp. 1456 – 1459.
9. T. S. Han and S. Amari, "Parameter estimation with multiterminal data compression," *IEEE Transactions on Information Theory*, vol. 41, no. 6, Nov. 1995, pp. 1802 – 1833.
10. W. Heinzelman, A. Chandrakasan, and H. Balakrishnan, "Energy-efficient communication protocol for wireless microsensor networks," Proc. of the Hawaii Conf. on System Science, Jan. 2000.
11. W. Lam and A. R. Reibman, "Design of quantizers for decentralized estimation systems," *IEEE Transactions on Communications*, vol. 41, no. 11, Nov. 1993, pp. 1602 – 1605.
12. Y. Liu and G. L. Wise, "A Note on Estimation with Quantized Data," *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. ASSP-33, no. 6, Dec. 1985, pp. 1619 – 1621.
13. H. V. Poor, "Fine Quantization in Signal Detection and Estimation," *IEEE Transactions on Information Theory*, vol. 34, no. 5, Sept. 1988, pp. 960 – 972.
14. J. Pottie and W. J. Kaiser, "Wireless integrated network sensors," *Communications of the ACM*, vol. 43, pp. 51 – 58, May 2000.
15. S. S. Pradhan, J. Kusuma, and K. Ramchandran, "Distributed compression in a dense microsensor network," *IEEE Signal Processing Magazine*, pp. 51– 60, March 2002.
16. A. Scaglione and S. Servetto, "On the interdependence of routing and data compression in multi-hop sensor networks," MOBICOM'02, Sept. 23 – 26, 2002, Atlanta, Ga.
17. S. Tilak, N. B. Abu-Ghazaleh, and W. Heinzelman, "A taxonomy of wireless micro-sensor network models," *ACM Mobile Computing and Communications Review (MC2R)*, vol. 6, pp. 1 – 8, April 2002.
18. J.-J. Xiao, A. Ribeiro, Z.-Q. Luo, and G. B. Giannakis, "Distributed compression-estimation using wireless sensor networks," *IEEE Signal Processing Magazine*, vol. 23, no. 4, July 2006, pp. 27 – 41.
19. Z. Zhang and T. Berger, "Estimation via compressed information," *IEEE Transactions on Information Theory*, vol. 34, no. 2, March 1988, pp. 198 – 211.