Data compression trade-offs in sensor networks

Mo Chen and Mark L. Fowler†

Department of Electrical and Computer Engineering
State University of New York at Binghamton
Binghamton, NY 13902

ABSTRACT

This paper first discusses the need for data compression within sensor networks and argues that data compression is a fundamental tool for achieving trade-offs in sensor networks among three important sensor network parameters: energy-efficiency, accuracy, and latency. Next, it discusses how to use Fisher information to design data compression algorithms that address the trade-offs inherent in accomplishing multiple estimation tasks within sensor networks. Results for specific examples demonstrate that such trades can be made using optimization frameworks for the data compression algorithms.

Keywords: data compression, sensor networks, estimation, Fisher information, emitter location, time-difference-of-arrival, TDOA, frequency-difference-of-arrival, FDOA

1. INTRODUCTION

Advances in sensor technology have focused interest towards using networks of sensors to collect useful information from an environment[1]. Tasks given to sensor nodes include collecting signal data, sharing the data between themselves, making inferences (estimations and decisions) from the data, and communicating the collected data and/or the inference results to one or more information sinks. To be useful, sensor networks must be designed to satisfy constraints on metrics assessing energy efficiency, communication latency, and accuracy of the conveyed information and there is a fundamental tradeoff over these metrics[2]; there are also constraints on fault tolerance and scalability[2] but we don't address those here. The challenges facing engineers seeking to satisfy these requirements include:

- Data volumes within sensor networks can be extremely large
- Communication capacity within wireless sensor networks can be quite low
- Reducing the data volume (via compression or other means) often results in excessive degradation of accuracy
- Wireless communication consumes a significant portion of a sensor node's energy
- Energy resources on each sensor are severely limited

Historically, data compression has been used to address the first three of these problems: minimizing the impact on accuracy of reducing a data volume that is ill-supported by the communication capacity. However, as we discuss in Section 2, it is also possible to use data compression to attack the energy issues. A further feature in our approach is to design the compression algorithm to address the impact on the achievable accuracy of the multiple estimation tasks the sensor network is to perform.

Some past efforts of other researchers in the area of data compression within sensor networks have focused on the data volume vs. communication capacity. For example, distributed compression methods have been proposed[3] that remove statistically redundant information between two nodes without sharing any data between them, provided a statistical model is available for the cross-correlation between the two nodes' data. Other recent results[4] establish fundamental information-theoretic limits on the rate of information transferal through the network and show that data compression combined with routing can be used to achieve a latency constraint. However, they don't directly address impact on estimation accuracy as we do.

Although energy efficiency can obviously be addressed through development of more efficient hardware, it is also possible to address energy efficiency through processing and communication algorithms that reduce wasted energy. For example, one such approach[5] proposes combining routing with data aggregation to reduce energy usage. The data

† Correspondence: mfowler@binghamton.edu

aggregation in this case consists of beamforming together the signals received at a handful of sensors surrounding a designated cluster head sensor. The result of this beamforming is to reduce the amount of data that must be transmitted across the network (i.e., a form of data compression), thus providing a significant energy saving. However, in many applications beamforming the data from a handful of sensors is not appropriate, so we explore compression methods.

Communication of data is one of the most energy-expensive tasks a node undertakes – using data compression to reduce the number of bits sent reduces energy expended for communication. However, compression requires computation, which also expends energy. Fortunately, trading computation for communication can save energy since a recent paper[1] asserts that typically on the order of 3000 instructions can be executed for the energy cost required to communicate one bit over a distance of 100 m by radio. Using that idea, we have shown[6] that general data compression can be used (either with or without routing) to enable energy savings.

In Section 2 we will present an argument for viewing data compression as a tool for achieving trade-offs in sensor networks among three important sensor network parameters: energy-efficiency, estimation accuracy, and latency. In Section 3 we will discuss the difficulties that arise when multiple estimation tasks must be performed and propose an approach that addresses this issue. Section 4 will apply the ideas of Section 3 to the specific problem of estimating the location of an RF emitter.

2. RATE-ENERGY-ACCURACY (R-E-A) TRADE-OFFS FOR COMPRESSION

Classical data compression theory relies on trade-offs between rate (R) and distortion (D) in terms of a R-D function. Rate is usually measured in terms of bits/sample and distortion is often measured as a mean-square error between the original and reconstructed signal. In the classical view, rate impacts latency and distortion impacts the accuracy of the signal reconstruction. However, as mentioned above, in sensor networks the rate can also impact energy efficiency. Thus, for sensor networks we propose the use of a 3-D extension of the R-D function: the Rate-Energy-Accuracy (R-E-A) function. The Energy axis assesses the amount of energy needed to move the collected data to the desired destination. Accuracy is related to distortion but is intended to better capture the effect of the compression error on the final use of the data – namely, the making of estimates. A good accuracy measure is the RMS estimation error.

Clearly, decreasing the rate via compression decreases the amount of transmission energy spent, but this comes at the expense of additional energy spent to perform the compression. These energy trade-offs depend on the computational efficiency of the compression algorithm, the energy efficiency of the computational architecture, and the energy efficiency of the transmission hardware; the goal is to achieve an overall reduction in energy consumption as shown in the top box in Figure 1. Likewise, decreasing the rate through compression decreases the transmission time needed to send the data to its destination, but it comes at the expense of additional time spent to perform the compression. These time trade-offs depend on the computational efficiency of the compression algorithm and the speed of the processor; the goal is to achieve an overall reduction in time delay as shown in the middle box in Figure 1. Finally, an additional cost of reducing the rate is an increase in the estimation error; the goal is to achieve a negligible increase in estimation error as shown in the bottom box in Figure 1.

The scenario shown in Figure 1 is also shown in the R-E-A space view of Figure 2, where the circle symbol shows the operating point without compression and the star symbol shows a desired operating point after compression: energy and time have been decreased (an improvement) but error has been increased (a degradation). Thus we can specify a desired operating point in R-E-A space and develop compression algorithms (as well as low-power computing & transmitting architectures) to attempt to achieve it. In classical R-D function theory, there are two dual R-D goals: (i) minimize distortion while obeying a rate constraint, and (ii) minimize rate while obeying a distortion constraint. In the R-E-A viewpoint the goal of sensor network compression can be specified in many ways; for example: (i) minimize energy subject to constraints on rate and accuracy – this would be optimizing along the line marked "(*i*)" in Figure 2, (ii) maximize accuracy subject to constraints on energy and rate – this would be optimizing along the line marked "(*ii*)" in Figure 2, (iii) jointly minimize energy and rate subject to a constraint on accuracy – this would be optimizing on the plane defined by the lines marked "(*i*)" and "(*iii*)" in Figure 2, etc.

In order to attack this kind of problem requires developments on many fronts, including (i) realistic energy models for transmission/reception, (ii) realistic energy models for computation, (iii) characterization of the computational complexity of the compression algorithms, and (iv) characterization of compression algorithms in terms of rate vs. accuracy. Although our research aims to address all of these issues we limit our presentation here to (iv), and in particular the case when there are several conflicting estimates to be made. One of the keys to addressing compression for multiple inference tasks is to use distortion measures that accurately reflect the ultimate performance on the tasks. To design compression algorithms suitable for use under conflicting inference goals it is essential to have appropriate, useable metrics that measure the impact of reducing the rate on the inference performance. For estimation tasks, the impact of compression should be assessed by its impact on the variance of the estimation error (at least in the unbiased estimate case). For decision tasks (e.g., detection, recognition, identification, etc.) the impact of compression should be assessed by its impact on the probability of an error in the decision. In addition, there may be the need for an end-user to view image data collected in a sensor network – the distortion measure for high-fidelity reconstruction is often a version of the mean-square error (MSE) measure. We have explored[7] trade-offs between estimation, detection, and reconstruction; here we explore a further issue that arises when multiple estimates are required that have differing levels of importance.

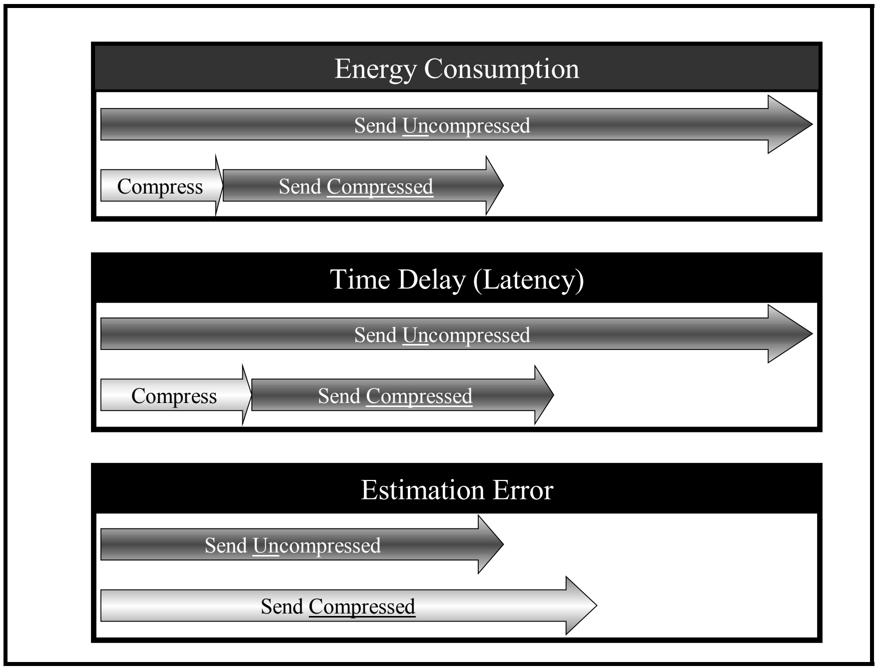

Figure 1: Gains and Costs in R-E-A viewpoint.

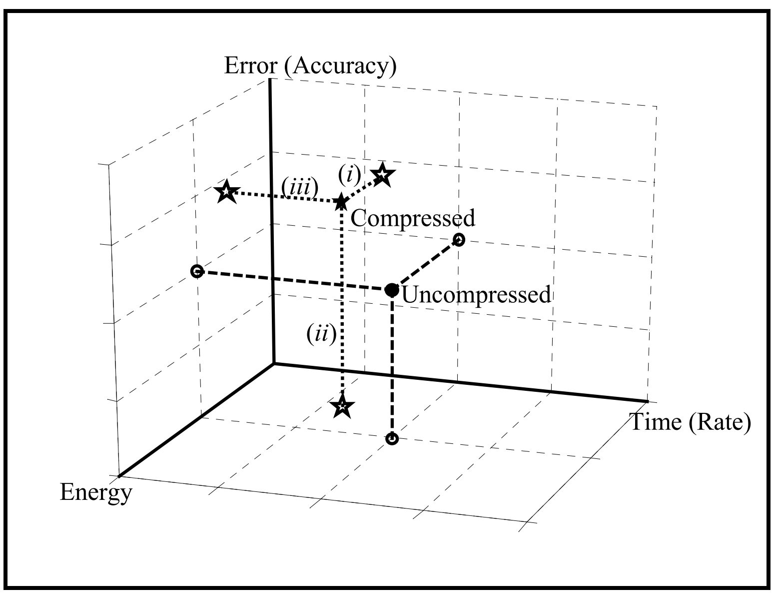

Figure 2: Illustration of R-E-A space; "hollow" symbols show projections on to subspaces for easier viewing.

3. GENERAL ESTIMATION-CENTRIC DISTORTION MEASURES

We model the collected sensor data as a deterministic signal plus additive white Gaussian noise having variance of σ^2. We assume that the estimation processing and compression processing are not jointly designed – this is motivated by our belief that sensor networks are likely to be called on to provide data to other systems/sensors that are independently designed (e.g., "legacy" systems and other "system of systems" scenarios); we assume that the estimation processing uses methods that are optimal in the absence of compression. 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[8] to quantify the degrading effect of compression on estimation accuracy.

We present first the case of a single estimation task, which we have previously considered[9] but include here for completeness and for use in the development here. Let the real data vector $\mathbf{x}$ be drawn from a probability density function (PDF) $p(\mathbf{x};\theta)$ that is parameterized by θ, which is to be estimated; modifications to handle the complex case are straightforward[8]. The Fisher information of this estimation problem is defined to be

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

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 3) is to transform the original data into some appropriate set of coefficients $\boldsymbol{\chi} = \{\chi_n \mid n = 1, 2, \ldots, N\}$, only some of which are then selected (Ω is set of indices for the selected coefficients) and quantized to give the set $\hat{\boldsymbol{\chi}} = \{\hat{\chi}_n \mid n \in \Omega\}$. The resulting set of transform coefficients $\hat{\boldsymbol{\chi}}$ then has Fisher information $J(\hat{\boldsymbol{\chi}}_1) \le J(\mathbf{x})$, 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 χ_n is a coefficient of the "signal+noise" while ξ_n and ω_n are the signal coefficient and noise coefficient, respectively.

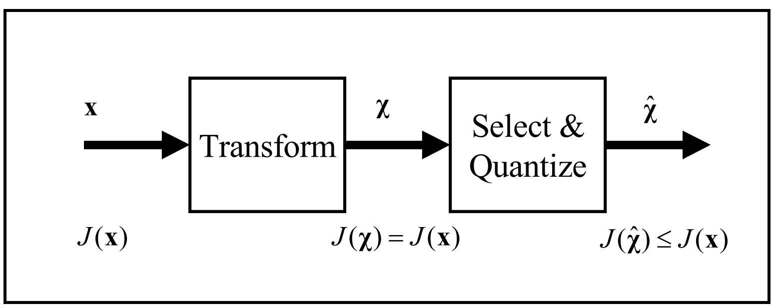

Figure 3: Single-estimation-task compression processing, the data vectors, and their corresponding Fisher informations.

We have previously shown[9] 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^{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 ξ_n. However, because we must compute $J(\hat{\chi})$ from our *data* coefficients χ_n, we must instead use a noisy version given by

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

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

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

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 $B = \{b_n \geq 0 \mid n \in 1, 2, \ldots, N\}$ that solves

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

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

When there are multiple parameters to be estimated from the data set **x** obviously it is desirable to maximize each Fisher information; Figure 4 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, …, K be the Fisher informations for the K desired estimation tasks and each has an importance weight α_k with

$$\sum_{k=1}^{K} \alpha_k = 1; \tag{6}$$

then we strive to find the bit allocation that maximizes

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

By using (7) we can easily apply our single-estimate results[9] to the multi-estimate case.

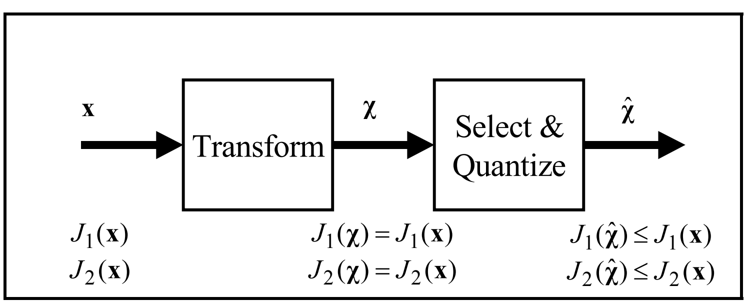

Figure 4: Double-estimation-task compression processing, the data vectors, and their corresponding Fisher informations.

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 α_k are specified; then (7) is maximized using an efficient Lagrange multiplier approach[10]. This will be illustrated for a specific application in the next section.

4. APPLICATION TO TDOA/FDOA EMITTER LOCATION

In this section we apply these results for the specific sensor network task of locating an RF or acoustic emitter. Such processing is commonly done using time-difference-of-arrival (TDOA) and frequency-difference-of-arrival (FDOA) methods[11,12]. The accuracy in *X-Y* location of the emitter depends on the accuracies of TDOA and FDOA estimates – thus the network has a two-parameter estimation task. However, the exact dependence on the *X-Y* location accuracy of these two intermediate estimation tasks depends highly on the geometry of the sensors[11]. Figure 5 illustrates three cases for two pairs of sensors. Each of the top plots shows sensor positions with hexagons and target position with an "X'd" circle (velocity of the target is not shown); sensor pairs (shown connected by a line) share data to estimate TDOA/FDOA and the TDOA/FDOA estimates from the two pairs of sensors are used to locate the emitter. Each of the bottom plots shows the resulting error ellipses from a theoretical covariance analysis for the cases of computing the location estimate (i) using TDOA measurements only, (ii) using FDOA measurements only, and using both TDOA and FDOA. In the left-hand pair of plots we see that TDOA and FDOA are both important since the corresponding error ellipse is much smaller than the ellipse for either TDOA alone or for FDOA alone. Likewise, in the other two cases we need only one of the two types of estimates. Thus, in some cases TDOA accuracy would be more important than FDOA, and vice versa in other scenarios. This is a clear example of the trade-offs that must be made in compression methods for multiple estimation tasks sensor networks. In this section we develop a compression algorithm to address this scenario.

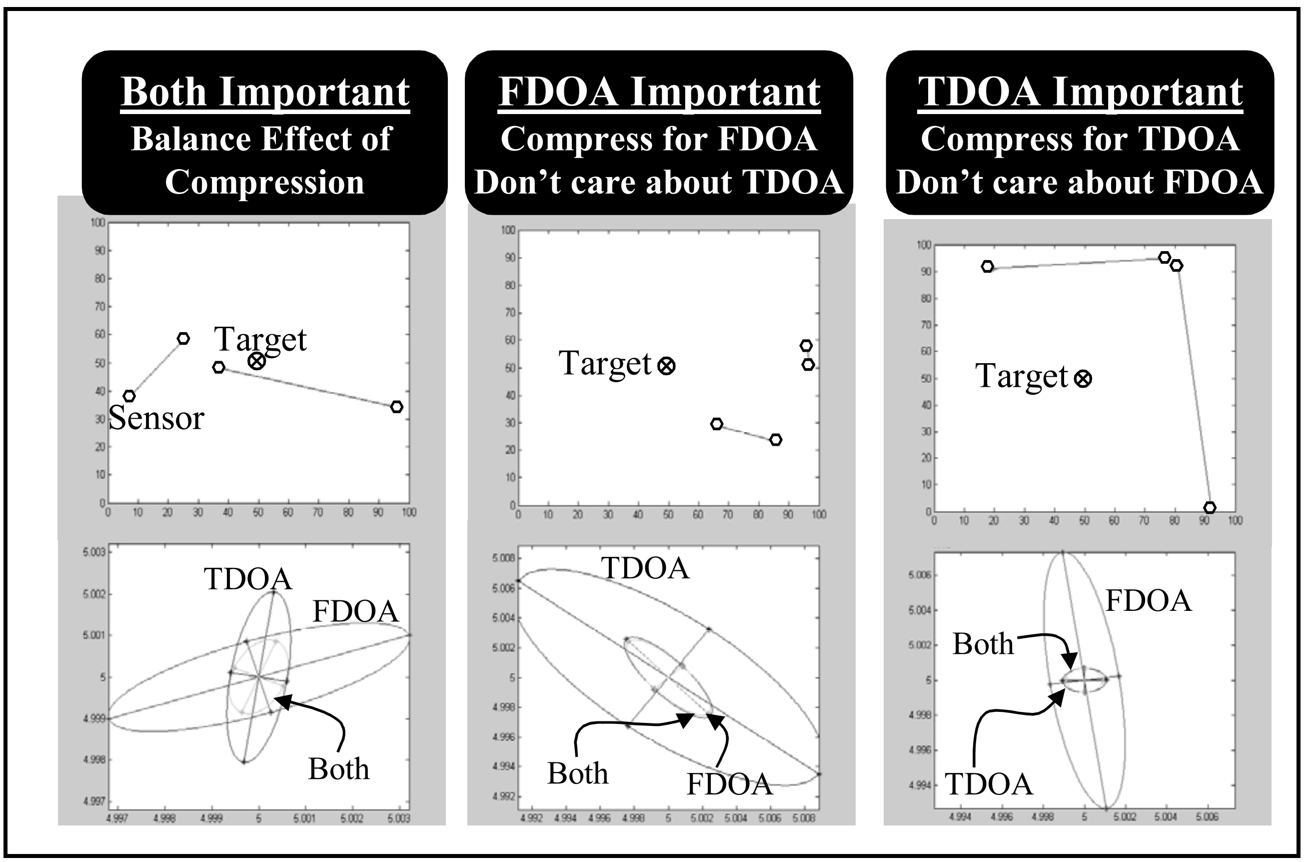

Figure 5: Effect of geometry on the relative importance of TDOA and FDOA accuracy.

The signal model for two passively-received signals having an unknown TDOA of Δ is given by

$$
\begin{aligned}
x_1[n] &= s[n-(n_0+\Delta/2)] + w_1[n] \quad n = 0,1,\ldots,N-1 \\
x_2[n] &= s[n-(n_0-\Delta/2)] + w_2[n] \quad n = 0,1,\ldots,N-1,
\end{aligned}
\tag{8}
$$

where n_0 is an unknown nuisance parameter that can not be estimated, and $w_i[n]$ is complex Gaussian noise with variance of σ_i^2, with σ_1^2 assumed known. The most-natural transform for compressing in TDOA applications is the symmetrically-indexed DFT[9], which is given by

$$X_1[k] = S[k]\exp[-j2\pi k(n_0+\Delta/2)] + W_1[k] \quad k = -N/2, -N/2+1, \ldots, N/2-1\ , \tag{9}$$

where the $S[k]$ are the DFT coefficients (for negative and positive frequencies) of the non-delayed signal and $W_1[k]$ are the DFT coefficients of the noise. The resulting Fisher information form to use in (4) is given by[9]

$$\hat{J}_n(TDOA) = \frac{2\pi^2 n^2 |X_1[n]|^2}{N\sigma_1^2 + q_n^2}, \quad n = -N/2, -N/2+1, \ldots, N/2-1\ . \tag{10}$$

Likewise, the signal model for two passively-received signals having an unknown FDOA of Δ is given by

$$
\begin{aligned}
x_1[n] &= s[n]e^{j(v_0+\Delta/2)} + w_1[n] \quad n = -N/2, -N/2+1, \ldots, N/2 \\
x_2[n] &= s[n]e^{j(v_0-\Delta/2)} + w_2[n] \quad n = -N/2, -N/2+1, \ldots, N/2,
\end{aligned}
\tag{11}
$$

where v_0 is an unknown nuisance parameter that can not be estimated, and $w_i[n]$ is complex Gaussian noise with variance of σ_i^2, with σ_1^2 assumed known. Here the natural choice of transform is the identity transform and the resulting Fisher information form to use in (4) is given by[9]

$$\tilde{J}_n(FDOA) = \frac{4\pi^2 n^2 |x_1[n]|^2}{\sigma_1^2 + q_n^2}, \quad n = -N/2, -N/2+1, \ldots, N/2\ . \tag{12}$$

From this we see that TDOA estimation depends more strongly on high frequencies (both positive and negative) while FDOA estimation depends more strongly on early and late times. To *jointly* attack this will require using a joint time-frequency orthogonal representation – the wavelet packet[13] is one of the most flexible such transforms and is selected for use here. For our results here we use a filter bank using fully-cascaded two-channel stages where each channel of each stage gets decomposed into two sub-channels (see Fig. 12.28 of Porat[14]). To reduce the complexity of the algorithm we group the resulting wavelet coefficients into sub-blocks within which all coefficients are allocated the same number of bits. The frequency and time location of each coefficient is taken as the center of the sub-block in the time-frequency plane, which is easily calculated using standard results in wavelet packet theory[13]. For our simulations we used a 4096-sample record of a simulated radar pulse train signal, which after applying the 3-level full-cascade wavelet packet transform results in 8 channels with 512 time samples in each channel. Each channel's 512 samples are grouped into 8 sub-blocks of 64 samples each. For this scenario we find that the Fisher information measure in (7) becomes

$$\tilde{J} = \sum_{j=1}^{M}\left(\frac{\alpha \sum_{i\in j\ cell} f_j^{\,2}|c_i|^2 + (1-\alpha)\sum_{i\in j\ cell} t_j^{\,2}|c_i|^2}{W_j\sigma^2 + W_j q_j^2}\right). \tag{13}$$

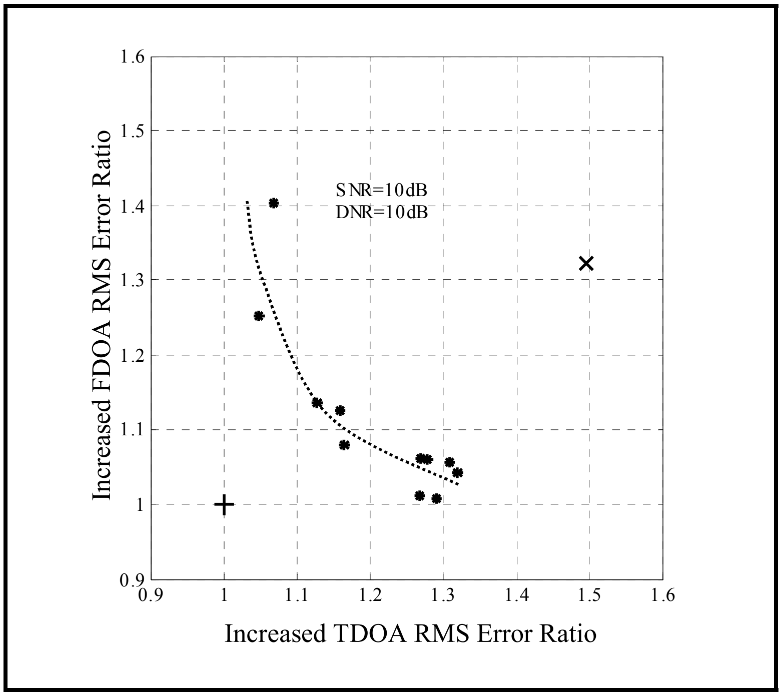

Figure 6: Results for SNR = 10 dB & DNR = 20 dB; symbol + denotes "without compression", symbol x denotes "standard MSE," and symbol * denotes "our method"; dashed curve helps visualize the trade-off.

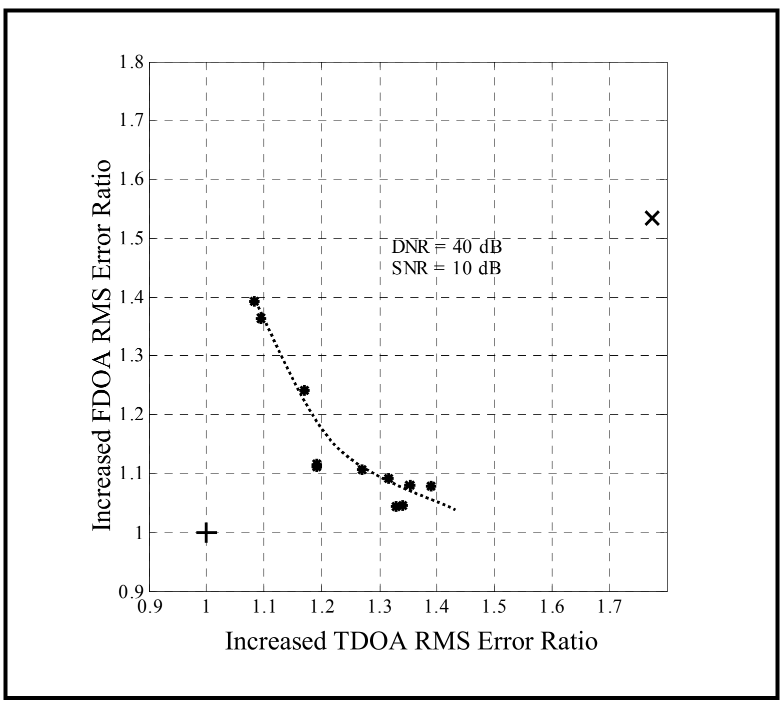

Figure 7: Results for SNR = 10 dB & DNR = 40 dB; symbol + denotes "without compression", symbol x denotes "standard MSE compression," and symbol * denotes "our method"; dashed curve helps visualize the trade-off.

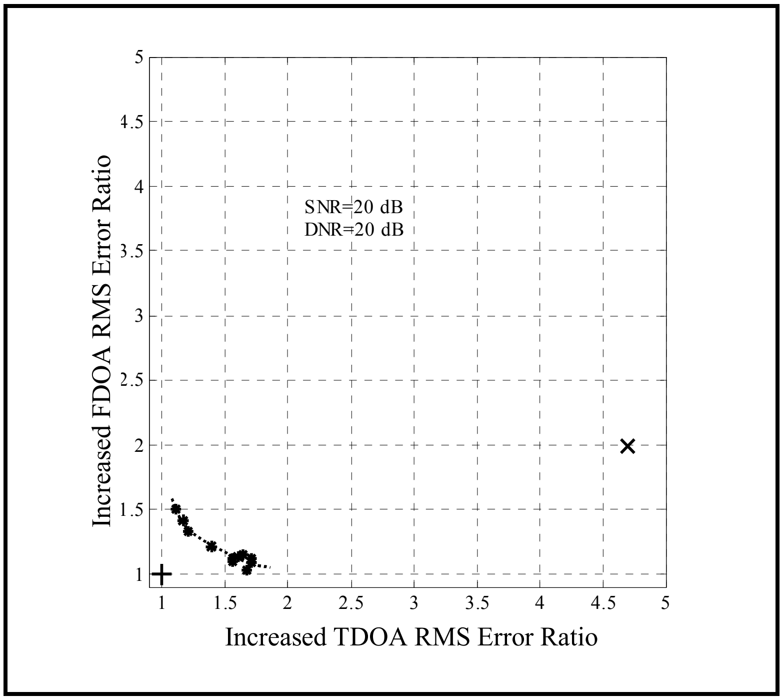

Figure 8: Results for SNR = 20 dB & DNR = 20 dB; symbol + denotes "without compression", symbol x denotes "standard MSE compression," and symbol * denotes "our method"; dashed curve helps visualize the trade-off.

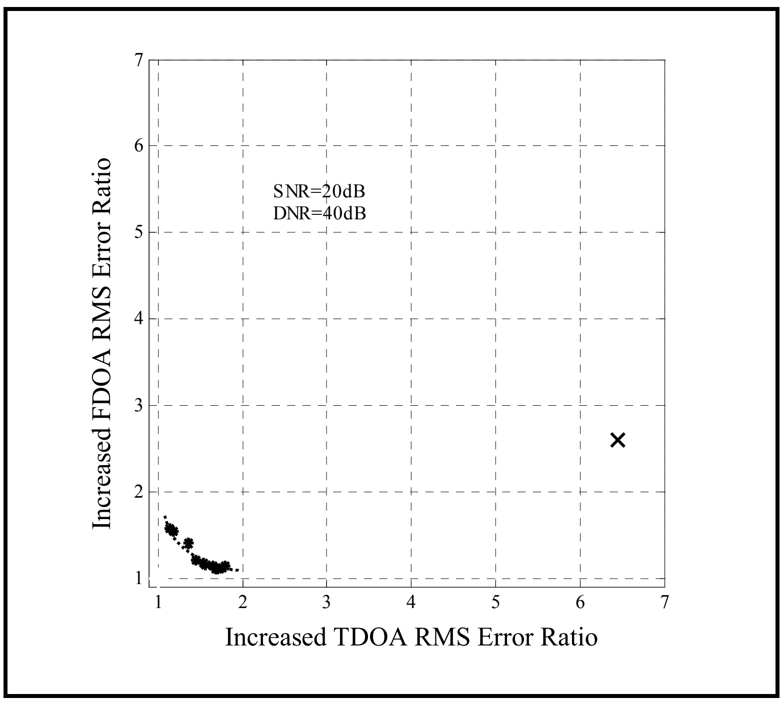

Figure 9: Results for SNR = 20 dB & DNR = 40 dB; symbol + denotes "without compression", symbol x denotes "standard MSE compression," and symbol * denotes "our method"; dashed curve helps visualize the trade-off.

where the c_i are the wavelet packet coefficients, $q_j^2 = \varepsilon_j^2 \sigma_j^2 2^{-2b_j}$ with ε_j^2 is chosen to be 9/2 (which gives us the best performance) with σ_j^2 being the estimated variance of the coefficients in the sub-block, the sub-block size is $W_j = 64, j = 1,\ldots,64$, and the frequency weights for sub-blocks can be assigned as $f_j = [1,2,3,4,5,6,7,8]/8$ from the bottom to top in the frequency axis, and the time weights for sub-blocks can be assigned as $t_j = [-4,-3,-2,-1,1,2,3,4]/4$. The trade-off control parameter $\alpha \in [0,1]$ allows the user to set the relative importance between TDOA accuracy and FDOA accuracy; setting α closer to 1 favors TDOA while setting α closer to 0 favors FDOA. For a given value α the quantity in (13) is maximized under the constraint of a total number of bits by using a Lagrange approach to constrained optimization[10]. The results for four different signal-to-noise-ratio (SNR) scenarios are given in Figure 6 - Figure 9, where *SNR* is the SNR for the signal being compressed and *DNR* is the SNR for the signal not compressed. All results are for the case of compressing to 2.5 bits per sample.

The axes in these figures show the relative increase in the estimation error. Note that in each case as we vary α over the range [0,1] we get a roughly convex trade-off between the impact of the compression on TDOA accuracy and FDOA accuracy that is shown by the * symbols in the figures. Note the significance improvement our algorithm provides relative to the standard MSE approach.

If the relative importance between TDOA and FDOA is known in advance then the user can specify an appropriate α value. As discussed via Figure 5, the relative importance of between TDOA and FDOA can not be known until the geometry is established – in the geometry for the left-hand side of Figure 5 we would want to use $\alpha \approx 0.5$, for the case shown in the middle of Figure 5 we would want $\alpha = 0$, and for the case shown on the right-hand side of Figure 5 we would want $\alpha = 1$. However, because our job is to locate the emitter there is no *a priori* knowledge of the geometry, so how does one deal with this. A simple approach is to send a very small amount of initial data (compressed with equal priority between TDOA and FDOA) to allow rough determination of the geometry through coarse location processing and then feed this back to determine the proper α value.

5. CONCLUSIONS

As demonstrated in the example applications, the use of a distortion measure designed specifically for a specific estimation problem can lead to compression methods that far outperform those using MSE-based distortion measures. While MSE distortion accurately captures the effect of the compression on the compressed signal's SNR, it fails to capture the true impact of compression on the estimation accuracy. This is similar to the scenario in image and audio compression, where MSE distortion fails to capture the impact of compression perceptual quality of the compressed data. In those areas researchers have proposed effective distortion measures based on the psychology of perception. Thus, by comparison we have, in a sense, captured the "psychology of estimation" through the use the Fisher information. By using the Fisher information we provide a measure that captures what statisticians view as the essence of the data that is useful for the estimation problem.

Benefits of our approach include:

- Effectively captures the impact of compression on estimation accuracy
- Formulation lends itself to the operational rate-distortion viewpoint
- Efficient means of optimizing the compression within an specified operational compression framework
- Provides insight into the choice of the proper operational compression framework (e.g., choice of transform)
- Applicable to a wide variety of estimation problems.
- Extendable to the case of multiple estimates as we have shown here

Despite these accomplishments, there are some directions for which further work is needed:

- Examination of the computational and implementation aspects.

Furthermore, an interesting issue is illuminated by the TDOA/FDOA application considered here – it may not be possible to set the *a priori* priority for multiple inferences. However, a promising approach was proposed here that involves sending a small amount of initial data – enough to simply determine the proper trade-offs – and then send the rest using compression with proper weighting factors.

6. REFERENCES

1. J. Pottie and W. J. Kaiser, “Wireless integrated network sensors,” *Communications of the ACM*, vol. 43, pp. 51 – 58, May 2000.
2. 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.
3. S. S. Pradhan, J. Kusuma, and K. Ramchandran, “Distributed compression in a dense microsensor network,” *IEEE Signal Processing Magazine*, pp. 51– 60, March 2002.
4. 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.
5. 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.
6. M. Chen and M. L. Fowler, “The importance of data compression for energy efficiency in sensor networks,” CD-ROM *Proceedings of the Conference on Information Sciences and Systems*, Johns Hopkins University, March 12-14, 2003.
7. M. Chen and M. L. Fowler, “Data compression trade-offs for multiple inferences in sensor networks,” CD-ROM *Proceedings of the Conference on Information Sciences and Systems*, Princeton University, March 17-19, 2004.
8. S. Kay, *Fundamentals of Statistical Signal Processing: Estimation Theory,* Englewood Cliffs, NJ: Prentice Hall, 1993.
9. M. L. Fowler and M. Chen, “Fisher-information-based data compression for estimation in sensor networks,” submitted to *IEEE Trans. Aerospace and Electronics Systems*.
10. Y. Shoham and A. Gersho, “Efficient bit allocation for an arbitrary set of quantizers,” *IEEE Trans. on Acoustics, Speech, and Signal Processing*, 36(9): 1445-1453, 1988.
11. D. Torrieri, “Statistical Theory of Passive Location Systems,” *IEEE Trans. on Aerospace and Electronic Systems*, pp. 183 - 198, March 1984.
12. S. Stein, “Differential Delay/Doppler ML Estimation with Unknown Signals,” *IEEE Trans. on Signal Processing*, August, pp. 2717 - 2719, 1993.
13. N. Hess-Nielsen and M. V. Wickerhauser, “Wavelet and time-frequency analysis,” *Proceeding of the IEEE*, vol. 84, No. 4, April 1996, PP, 523-540.
14. B. Porat, *A Course in Digital Signal Processing*, Wiley, New York, 1997.