Distributed Adaptive Quantization for Wireless Sensor Networks: A Maximum Likelihood Approach

Jun Fang and Hongbin Li
Department of Electrical and Computer Engineering
Stevens Institute of Technology, Hoboken, NJ 07030
Emails: {Jun.Fang, Hongbin.Li@stevens.edu}

***Abstract*— We consider the problem of distributed parameter estimation in wireless sensor networks (WSNs), where due to bandwidth/power constraints, each sensor quantizes its local observation into one bit of information that is sent to a fusion center (FC) to form a global estimate. Conventional fixed quantization (FQ) approaches, which utilize a fixed threshold for all sensors, incurs an estimation error growing exponentially with the difference between the threshold and the unknown parameter to be estimated. To overcome this difficulty, we propose a distributed adaptive quantization (AQ) approach, where, under the condition that sensors *successively* broadcast their quantized data, each sensor adaptively adjusts its quantization threshold using prior transmissions from other sensors. Specifically, our strategy here is to let each sensor choose its quantization threshold as the maximum likelihood (ML) estimate of the unknown parameter based on the quantized data sent from other sensors. The Cramér-Rao bound (CRB) analysis shows that our proposed one-bit AQ approach asymptotically attains an estimation variance that is only $\pi/2$ times that of the clairvoyant sample-mean estimator using *unquantized* observations.**

***Index Terms*— Adaptive quantization, distributed estimation, wireless sensor networks.**

I. Introduction

Wireless sensor networks (WSNs) have attracted much attention over the past few years. Consisting of a large number of small, low-cost sensors with integrated sensing, processing, and communication abilities, WSNs can accomplish a variety of tasks including environment monitoring, battlefield surveillance, target localization and tracking, and many more [1], [2]. Bandwidth and power constraints are two primary issues that need to be addressed in WSNs, as limited communication bandwidth is shared across the entire network and, meanwhile, the sensors are often powered by irreplaceable batteries. As such, a major challenge of the WSN research is to design bandwidth and power efficient signal processing techniques.

In this paper, we consider distributed estimation of an unknown deterministic parameter in a bandwidth and power constrained WSN. Suppose we have N spatially distributed sensors, with each sensor making a noisy observation of an unknown parameter θ

$$x_n = \theta + w_n, \quad n = 1, 2, \ldots, N, \tag{1}$$

where w_n denotes additive Gaussian observation noise with zero mean and variance σ_w^2, and the noise is assumed independent and identically distributed (i.i.d.) with respect to n. Due to limited bandwidth and power constraints, all sensors quantize their observations $\{x_n\}$ into one-bit binary data $\{b_n\}$ and send the quantized data to the fusion center (FC). The problem of interest is to estimate θ from the quantized data $\{b_n\}$ received at the FC.

Conventional fixed quantization (FQ) approach employs a common quantization threshold for all sensors [3], [4]. The optimum choice of τ, however, is identical to θ which is unknown. It is also found that if τ is set away from θ, the best achievable estimation performance at the FC has an estimation error exponentially increasing with $|\tau - \theta|/\sigma_w$. An alternative strategy is to use a set of thresholds $\{\tau_k\}$, and each τ_k is used in a fraction ρ_k of the N sensors [4], in the hope that some of the thresholds are close to the unknown θ. However, to find solutions of $\{\tau_k, \rho_k\}$, a prior probability distribution of θ is required. Recently, an adaptive quantization (AQ) approach with *fixed* step size was introduced in [5]. The proposed scheme [5] is in essence a distributed Delta modulation technique, whereby the threshold of sensor n, τ_n, is obtained as the previous threshold τ_{n-1} plus or minus a fixed increment (step size) Δ. It is shown that the AQ scheme [5] demonstrates a certain extent of robustness to the unknown parameter θ and presents a performance advantage over the FQ approach.

In this paper, we present an AQ approach with *variable* step size. Specifically, each sensor computes the maximum likelihood (ML) estimate of the unknown parameter based on the previous transmissions as its quantization threshold. As compared with [5], the proposed AQ approach with variable step size presents a stronger robustness to the unknown parameter θ. Also, our algorithm is shown to asymptotically have the same estimation variance as that of the FQ approach with an optimum choice of threshold, i.e. $\tau = \theta$. This indicates that our AQ scheme adaptively finds the best threshold by learning from transmission from previous sensors.

The rest of the paper is organized as follows. We first briefly review the FQ approach [4] in Section II. The AQ approach with variable step size and the corresponding MLE and CRB are presented in Section III. Numerical results, comparisons, and discussions are contained in Section IV.

II. Fixed Quantization Approach

The fixed quantization approach is to apply a common threshold τ for all sensors and generate quantized data b_n as follows [3]:

$$b_n = \operatorname{sgn}(x_n - \tau), \quad n = 1, 2, \ldots, N, \tag{2}$$

which are sent to the FC. The probability mass function (PMF) of the binary random variable b_n is given by

$$P(b_n;\theta) = [F_w(\tau-\theta)]^{(1+b_n)/2}[1-F_w(\tau-\theta)]^{(1-b_n)/2}, \quad (3)$$

where $F_w(x)$ denotes the complementary cumulative density function (CCDF) of w_n. Since $\{b_n\}$ are i.i.d., the log-PMF or log-likelihood function is

$$\begin{aligned} L_{\text{FQ}}(\theta) &\triangleq \ln[P(b_1,\ldots,b_N;\theta)] \\ &= \sum_{n=1}^{N}\left\{\left(\frac{1+b_n}{2}\right)\ln[F_w(\tau-\theta)]\right. \\ &\quad \left. + \left(\frac{1-b_n}{2}\right)\ln[1-F_w(\tau-\theta)]\right\}, \quad (4) \end{aligned}$$

where the subscript FQ is used to denote fixed quantization. The MLE is given by [4]

$$\begin{aligned} \hat{\theta}_{\text{FQ}} &= \arg\max_{\theta} L_{\text{FQ}}(\theta) \\ &= \tau - F^{-1}\left(\frac{1}{N}\sum_{n=1}^{N}\frac{1+b_n}{2}\right). \quad (5) \end{aligned}$$

The CRB based on the above fixed quantization is ([3], [4]):

$$\text{CRB}_{\text{FQ}}(\theta) = \frac{F_w(\tau-\theta)[1-F_w(\tau-\theta)]}{Np_w^2(\tau-\theta)}, \quad (6)$$

where $p_w(x)$ denotes the probability density function (PDF) of w_n. Using the Chernoff bound, it can be easily shown [4] that the CRB increases exponentially with $|\tau-\theta|/\sigma_w$.

III. Adaptive Quantization Approach

A. AQ via MLE

We assume that the sensors broadcast their quantized data sequentially, i.e. sensor 1 transmits first, followed by sensor 2, and so on and so forth. The sensor's quantization threshold is computed as the ML estimate of the unknown parameter θ based on previous transmissions. Firstly, we use $\tau_1 = 0$ to generate b_1:

$$b_1 = \text{sgn}\{x_1\}. \quad (7)$$

After receiving b_1, sensor 2 computes $\tau_2 = \Delta b_1$, where Δ is chosen to be large enough such that $\Delta \gg |\theta|$. In doing this way, we can guarantee that $b_2 = \text{sgn}\{x_2 - \tau_2\}$ has a different sign from b_1, i.e. $b_1 = -b_2$. Based on the received $\{b_1, b_2\}$ and the pre-specified Δ, sensor 3 computes τ_3 as

$$\begin{aligned} \tau_3 = \hat{\theta} &= \arg\max_{\theta} L_3(\theta) \\ &= \arg\max_{\theta} \log P(b_1, b_2;\theta) \\ &= \arg\max_{\theta} \log P(b_1|\tau_1;\theta)P(b_2|\tau_2;\theta) \\ &= \arg\max_{\theta} \sum_{k=1}^{2}\left\{\left(\frac{1+b_k}{2}\right)\ln[F_w(\tau_k-\theta)]\right. \\ &\quad \left. + \left(\frac{1-b_k}{2}\right)\ln[1-F_w(\tau_k-\theta)]\right\}, \quad (8) \end{aligned}$$

and generates $b_3 = \text{sgn}\{x_3 - \tau_3\}$. Note that if b_1 and b_2 have the same sign, the ML estimate $\hat{\theta}$ goes to plus or minus infinity. This is the reason that we generate different signs of b_1, b_2.

For the Gaussian noise w_n, it can be readily shown that the above and the following MLE cost functions are concave [4]. Therefore, a gradient-based iterative algorithm is guaranteed to converge to the global maximum.

In general, for sensor n, it first recovers the previous thresholds $\{\tau_1, \tau_2, \ldots, \tau_{n-1}\}$ from the received quantized data $\{b_1, b_2, \ldots, b_{n-1}\}$, which can be computed straightforwardly by recursive calculation:

$$\begin{aligned} \tau_1 &= 0 \\ \tau_2 &= \Delta b_1 \\ \tau_3 &= \arg\max_{\theta} L_3(\theta; b_1, b_2, \tau_1, \tau_2) \\ &\vdots \\ \tau_{n-1} &= \arg\max_{\theta} L_{n-1}(\theta; b_1, \ldots, b_{n-2}, \tau_1, \ldots, \tau_{n-2}), \quad (9) \end{aligned}$$

where

$$\begin{aligned} L_{n-1}(\theta) &= \log P(b_1, \ldots, b_{n-2};\theta) \\ &= \log \prod_{k=1}^{n-2} P(b_k|\tau_k;\theta) \\ &= \sum_{k=1}^{n-2}\left\{\left(\frac{1+b_k}{2}\right)\ln[F_w(\tau_k-\theta)]\right. \\ &\quad \left. + \left(\frac{1-b_k}{2}\right)\ln[1-F_w(\tau_k-\theta)]\right\}. \quad (10) \end{aligned}$$

After obtaining $\{\tau_1, \tau_2, \ldots, \tau_{n-1}\}$, sensor n computes its current threshold τ_n as

$$\begin{aligned} \tau_n &= \arg\max_{\theta} L_n(\theta) \\ &= \arg\max_{\theta} \log P(b_1, \ldots, b_{n-1};\theta) \\ &= \arg\max_{\theta} \log \prod_{k=1}^{n-1} P(b_k|\tau_k;\theta) \\ &= \arg\max_{\theta} \sum_{k=1}^{n-1}\left\{\left(\frac{1+b_k}{2}\right)\ln[F_w(\tau_k-\theta)]\right. \\ &\quad \left. + \left(\frac{1-b_k}{2}\right)\ln[1-F_w(\tau_k-\theta)]\right\}, \quad (11) \end{aligned}$$

and generates $b_n = \text{sgn}\{x_n - \tau_n\}$.

Similarly, the ML estimator at the FC to find the final estimate of θ from the received quantized data $\{b_1, b_2, \ldots, b_N\}$ is given by

$$\begin{aligned} \hat{\theta} &= \arg\max_{\theta} L_{\text{AQ}}(\theta) \\ &= \arg\max_{\theta} \log P(b_1, \ldots, b_N;\theta) \\ &= \arg\max_{\theta} \log \prod_{k=1}^{N} P(b_k|\tau_k;\theta) \\ &= \arg\max_{\theta} \sum_{k=1}^{N}\left\{\left(\frac{1+b_k}{2}\right)\ln[F_w(\tau_k-\theta)]\right. \\ &\quad \left. + \left(\frac{1-b_k}{2}\right)\ln[1-F_w(\tau_k-\theta)]\right\}, \quad (12) \end{aligned}$$

where the thresholds $\{\tau_1, \ldots, \tau_N\}$ can be recovered from the quantized data $\{b_1, b_2, \ldots, b_{N-1}\}$ in a recursive way as in (9).

B. CRB

We analyze the performance of the proposed ML estimator of θ in (12). Noting that $F'_w(x) \triangleq \frac{\partial F_w(x)}{\partial x} = -p_w(x)$, we can quickly verify that the second-order derivatives of $L_{\text{AQ}}(\theta)$ are

$$\begin{aligned}
&\frac{\partial^2 L_{\text{AQ}}(\theta)}{\partial \theta^2} \\
&= \sum_{n=1}^{N} \Bigg\{ \left(\frac{1+b_n}{2}\right)\left(-\frac{p'_w(\tau_n-\theta)}{F_w(\tau_n-\theta)} - \frac{p_w^2(\tau_n-\theta)}{F_w^2(\tau_n-\theta)}\right) \\
&\quad - \left(\frac{1-b_n}{2}\right)\Bigg(-\frac{p'_w(\tau_n-\theta)}{\big[1-F_w(\tau_n-\theta)\big]} \\
&\quad + \frac{p_w^2(\tau_n-\theta)}{\big[1-F_w(\tau_n-\theta)\big]^2}\Bigg)\Bigg\} \\
&\triangleq \sum_{n=1}^{N} A(b_n,\tau_n,\theta), \qquad (13)
\end{aligned}$$

where $p'_w(x) \triangleq \frac{\partial p_w(x)}{\partial x}$. The Fisher information for the estimation problem is given by (e.g., [6])

$$\begin{aligned}
J_{\text{AQ}}(\theta) &= -E\left\{\frac{\partial^2 L_{\text{AQ}}(\theta)}{\partial \theta^2}\right\} \\
&= -\sum_{n=1}^{N} E_{b_n,\tau_n}\{A(b_n,\tau_n,\theta)\}, \qquad (14)
\end{aligned}$$

where E_{b_n,τ_n} denotes the expectation with respect to the joint distribution of b_n and τ_n. Since

$$P(b_n,\tau_n;\theta) = P(\tau_n;\theta)P(b_n|\tau_n;\theta), \qquad (15)$$

we can write

$$\begin{aligned}
J_{\text{AQ}}(\theta) &= -\sum_{n=1}^{N} E_{\tau_n}\Big\{E_{b_n|\tau_n}[A(b_n,\tau_n,\theta)]\Big\} \\
&\overset{(a)}{=} \sum_{n=1}^{N} E_{\tau_n}\left[\frac{p_w^2(\tau_n-\theta)}{F_w(\tau_n-\theta)(1-F_w(\tau_n-\theta))}\right] \\
&\overset{(b)}{=} \sum_{n=1}^{N} \int P(\tau_n;\theta)G(\tau_n;\theta)d\tau_n \qquad (16)
\end{aligned}$$

where E_{τ_n} denotes the expectation with respect to the distribution $P(\tau_n;\theta)$, $E_{b_n|\tau_n}$ denotes the expectation with respect to the conditional distribution $P(b_n|\tau_n;\theta)$, (a) follows from the fact that b_n is a binary random variable with $P(b_n = 1|\tau_n,\theta) = F_w(\tau_n-\theta)$ and $P(b_n = -1|\tau_n,\theta) = 1-F_w(\tau_n-\theta)$, and we define $G(\tau_n;\theta) \triangleq \frac{p_w^2(\tau_n-\theta)}{F_w(\tau_n-\theta)(1-F_w(\tau_n-\theta))}$ in (b).

To compute the exact Fisher information (16), we need determine the distribution of $\{\tau_n\}$, i.e. $\{P(\tau_n;\theta)\}$. However, for our AQ approach, the number of possible thresholds increases exponentially with n, specifically, sensor n has 2^{n-1} possible thresholds with each threshold chosen with a certain probability. Hence the exact computation of $P(\tau_n;\theta)$ is cumbersome, especially when the number of sensors, N, is large. Instead, we examine the asymptotic performance of our proposed AQ approach. We have the following results.

Proposition 1: For the Gaussian noise w_n with zero-mean and variance σ_w^2, the performance of the proposed AQ approach converges to the following as $N \to \infty$

$$\text{CRB}_{\text{AQ}}(\theta) \to \frac{\pi\sigma_w^2}{2N} \approx 1.57\frac{\sigma_w^2}{N} \qquad (17)$$

Proof: See Appendix A. ■

From (17), we see that the estimation variance of our proposed AQ approach increases only by a factor of $\pi/2$ with respect to the clairvoyant estimator [6] that relies on *unquantized* observations. Note that the above estimation variance $\frac{\pi\sigma_w^2}{2N}$ is also achieved by the FQ approach with an optimum choice of threshold $\tau = \theta$ [4]. This indicates that our AQ scheme adaptively finds the best threshold by learning from transmission from previous sensors.

IV. Numerical Results

We illustrate the performance of our proposed AQ approach. The noise $\{w_n\}$ are i.i.d. Gaussian random variables with zero mean and variance $\sigma_w^2 = 1$. We set $\theta = 20$. We compare our AQ scheme (named as "AQ-VS") with the clairvoyant estimator [6] using unquantized data, the FQ approach [4], with the common threshold τ set to be 16 and 20 respectively, and the AQ approach with fixed step size [5] (named as "AQ-FS"), where the step size is set to be $\Delta = 5$.

Fig. 1 shows the CRBs of the above approaches (note that to compute the CRB of the proposed AQ approach, we collect a number of realizations and find out its approximate pdf $P(\tau_n;\theta)$). For FQ approach, $\tau = \theta$ is the optimum choice, when $\tau = \theta = 20$, the performance of FQ achieves the best among those one-bit rate-constrained estimators. However, as we can see from the figure, the FQ approach is very sensitive to the value of τ; as the threshold τ become more apart from θ (even not too far apart), the performance of the FQ degrades significantly. Since θ to be estimated is unknown, the choice of τ is always a tricky problem. Our proposed AQ approach does not have the above problem. Its performance approaches that of FQ with optimum threshold ($\tau = \theta$) while without knowing the true θ. This illustrates the effectiveness of our proposed approach and corroborates our previous claim in Proposition 1. Furthermore, we also observe that the proposed AQ approach with variable step size yields an additional performance gain as compared with the AQ approach with fixed step size.

Appendix A
Proof of Proposition 1

Note that sensor m computes its threshold τ_m as

$$\begin{aligned}
\tau_m &= \arg\max_{\theta} L_m(\theta) \\
&= \arg\max_{\theta} \log P(b_1,\ldots,b_{m-1};\theta) \\
&= \arg\max_{\theta} \sum_{k=1}^{m-1} \Bigg\{ \left(\frac{1+b_k}{2}\right) \ln[F_w(\tau_k-\theta)] \\
&\quad + \left(\frac{1-b_k}{2}\right) \ln[1-F_w(\tau_k-\theta)] \Bigg\} \qquad (18)
\end{aligned}$$

It can be easily verified that $\log P(b_1, \ldots, b_{m-1}; \theta)$ satisfies the "regularity" conditions. Hence, for large data records, i.e. m is large, the ML estimate τ_m is consistent and asymptotically distributed according to $\tau_m \sim \mathcal{N}(\theta, J_m^{-1}(\theta))$ [6], where $J_m(\theta)$ is the corresponding Fisher information. Consequently, for any small $\epsilon > 0$ and $\varepsilon > 0$, we can find a sufficiently large m such that

$$P(|\tau_n - \theta| < \epsilon) > 1 - \varepsilon \qquad n \geq m \tag{19}$$

Considering (16), we express $J_{\text{AQ}}(\theta)$ as the summation of the following two terms

$$\begin{aligned} J_{\text{AQ}}(\theta) = &\sum_{n=1}^{m-1} \int P(\tau_n; \theta) G(\tau_n; \theta) d\tau_n \\ &+ \sum_{n=m}^{N} \int P(\tau_n; \theta) G(\tau_n; \theta) d\tau_n \end{aligned} \tag{20}$$

where m is chosen to satisfy (19). Noting that the function $G(\tau_n; \theta)$ is unimodal, positive, symmetric and achieves its maximum when $\tau_n = \theta$, the first term and the second term of (20) can then be bounded as follows, respectively

$$\begin{aligned} J_1(\theta) &\triangleq \sum_{n=1}^{m-1} \int P(\tau_n; \theta) G(\tau_n; \theta) d\tau_n \\ &< \sum_{n=1}^{m-1} G(\tau_n = \theta; \theta) = \frac{2(m-1)}{\pi \sigma_w^2} \end{aligned} \tag{21}$$

$$\begin{aligned} J_2(\theta) &\triangleq \sum_{n=m}^{N} \int P(\tau_n; \theta) G(\tau_n; \theta) d\tau_n \\ &= \sum_{n=m}^{N} \left[\int P(|\tau_n - \theta| < \epsilon) G(\tau_n; \theta) d\tau_n \right. \\ &\quad \left. + \int P(|\tau_n - \theta| \geq \epsilon) G(\tau_n; \theta) d\tau_n \right] \\ &> \sum_{n=m}^{N} \int P(|\tau_n - \theta| < \epsilon) G(\tau_n; \theta) d\tau_n \\ &> \sum_{n=m}^{N} (1 - \varepsilon) G(\tau_n = \theta - \epsilon; \theta) \\ &\overset{(a)}{>} (N - m + 1)(1 - \varepsilon) \frac{4}{\sqrt{2\pi}\sigma_w} p_w(\epsilon) \\ &\overset{(b)}{>} (N - m + 1) \frac{2}{\pi \sigma_w^2} (1 - \varepsilon) \left(1 - \frac{\epsilon^2}{2\sigma_w^2} \right) \end{aligned} \tag{22}$$

where (a) comes from the Chernoff bound: $F_w(x)(1 - F_w(x)) \leq \frac{1}{4} e^{-\frac{x^2}{2\sigma_w^2}}$, (b) follows from the Taylor expansion. Since $\frac{2(m-1)}{\pi \sigma_w^2} > J_1(\theta) > 0$, we can further write

$$J_1(\theta) = \frac{2(m-1)\eta}{\pi \sigma_w^2} \tag{23}$$

where $0 < \eta < 1$. Combining (20–23), we therefore have

$$\begin{aligned} J_{\text{AQ}}(\theta) &> (N - m + 1) \frac{2}{\pi \sigma_w^2} \xi + \frac{2(m-1)\eta}{\pi \sigma_w^2} \\ &= \frac{2N\xi}{\pi \sigma_w^2} - \frac{2(m-1)\eta'}{\pi \sigma_w^2} \end{aligned} \tag{24}$$

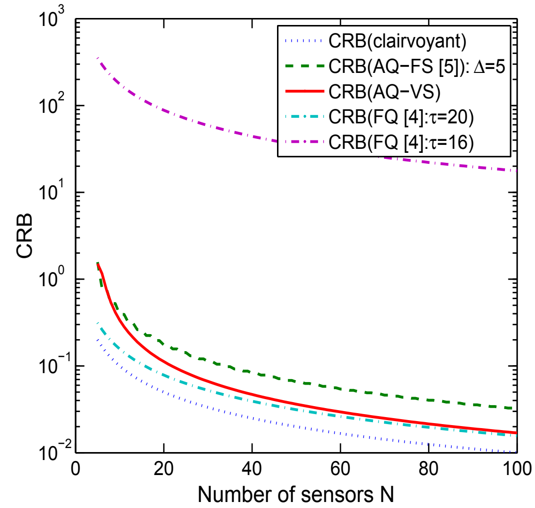

Fig. 1. CRBs of the respective approaches versus the number of sensors N.

where $\xi \triangleq (1 - \varepsilon) \left(1 - \frac{\epsilon^2}{2\sigma_w^2} \right)$ and $1 > \eta' \triangleq \xi - \eta$. The CRB is upper bounded by

$$\begin{aligned} \text{CRB}_{\text{AQ}}(\theta) &= \frac{1}{J_{\text{AQ}}(\theta)} \\ &< \frac{\pi \sigma_w^2}{2} \frac{1}{N\xi - (m-1)\eta'} \\ &= \frac{\pi \sigma_w^2}{2N} \frac{1}{\xi - \frac{(m-1)\eta'}{N}} \end{aligned} \tag{25}$$

Considering $N \to \infty$, we have $\xi \to 1$ and $\frac{(m-1)\eta'}{N} \to 0$, hence

$$\text{CRB}_{\text{AQ}}(\theta) \to \frac{\pi \sigma_w^2}{2N} \approx 1.57 \frac{\sigma_w^2}{N} \tag{26}$$

The proof is completed here.

References

[1] D. Li, K. D. Wong, Y. H. Hu, and A. M. Sayeed, "Detection, classification, and tracking of targets," *IEEE Signal Processing Magazine*, vol. 19, no. 2, pp. 17–29, Mar. 2002.

[2] I. F. Akyildiz, W. Su, Y. Sankarasubramaniam, and E. Cayirci, "A survey on sensor networks," *IEEE Communications Magazine*, pp. 102–114, Aug. 2002.

[3] H. Papadopoulos, G. Wornell, and A. Oppenheim, "Sequential signal encoding from noisy measurements using quantizers with dynamic bias control," *IEEE Trans. Inform. Theory*, vol. 47, no. 2, pp. 978–1002, Mar. 2001.

[4] A. Ribeiro and G. B. Giannakis, "Bandwidth-constrained distributed estimation for wireless sensor networks, Part I Gaussian PDF," *IEEE Trans. Signal Processing*, vol. 54, no. 3, pp. 1131–1143, Mar. 2006.

[5] H. Li and J. Fang, "Distributed adaptive quantization and estimation for wireless sensor networks," *IEEE Signal Processing Lett.*, vol. 14, no. 10, pp. 669–672, Oct. 2007.

[6] S. M. Kay, *Fundamentals of Statistical Signal Processing: Estimation Theory*. Upper Saddle River, NJ: Prentice Hall, 1993.