BELIEF PROPAGATION DISTRIBUTED ESTIMATION IN SENSOR NETWORKS: AN OPTIMIZED ENERGY ACCURACY TRADEOFF

John MacLaren Walsh[*]

Drexel University
Dept. of Elec. & Comp. Eng.
Philadelphia, PA

Phillip A. Regalia[†]

Catholic University of America
Dept. of Elec. Eng & Comp. Sci.
Washington, DC

ABSTRACT

The estimation error performance of Gaussian belief propagation based distributed estimation in a large sensor network employing random sleep strategies is explicitly evaluated for a simple model using density evolution analysis. Both regular sleep strategies, in which the number of nodes awake at any time instant is fixed, as well as irregular sleep strategies, in which the number of awake nodes may vary, are analyzed. The calculated estimation error is used to study the tradeoff between estimation accuracy and energy consumption, as well as to dictate the optimal parameters for the random sleep strategy.

Index Terms— distributed estimation, sensor networks, belief propagation, expectation propagation, density evolution, energy efficiency

1. INTRODUCTION & SENSOR NETWORK DISTRIBUTED ESTIMATION MODEL

Consider a network $\mathcal{S}$ of N sensor nodes $\mathcal{S} := \{s_i | i \in \{1, \ldots, N\}\}$, each wishing to estimate its own parameter $\underline{\zeta}_i$. The sensor network employs a pseudo-random sleep strategy for energy efficiency, in which at discrete time instant k, only a subset of nodes $\mathcal{A}_k \subset \mathcal{S}$ are awake. Those sensors that are awake observe their parameters indirectly through noise that is correlated across the different sensor nodes, so that the observations are

$$\underline{r}_{k,i} := \underline{\zeta}_i + \underline{n}_{k,i}, \quad i \in \mathcal{A}_k$$

The goal of this distributed estimation problem is to coordinate the estimation of $\underline{\zeta}_i$ from the observations across the sensor network in a distributed manner across the sensor network. Note that this goal differentiates the results presented here from consensus propagation [1] style results for sensor networks, in which all nodes aim at inferring/estimating the same hypothesis/parameters, as well as many classic decentralized detection/estimation results in which the fusion and inference is done at a fusion center rather than in the network itself.

Although the results we provide may be generalized to a far broader context, for the sake of simplicity of exposition, suppose that a priori the parameters $\underline{\zeta}_i$ are independently and identically distributed (i.i.d.) according to Gaussian distributions with mean 0 and variance 1. We consider a physical regime in which the noise may be modeled as uncorrelated across different time instants, but correlated across different awake sensors, so that

$$\mathbb{E}[\underline{n}_{k,i}\underline{n}_{m,j}] := \begin{cases} \sigma^2 & k = m, i = j \\ \gamma\sigma^2 & k = m, i \neq j \\ 0 & k \neq m \end{cases}$$

As discussed in [2], we can associate with the distributed estimation algorithm a factor graph [3], whose left nodes correspond to different sensors in the sensor network, and whose right nodes correspond to different time instants in the cyclic pseudo-random sleep strategy. This abstraction allows expectation propagation to applied to perform distributed estimation in the sensor network as described in [2]. With the model setup we consider in this case, EP simplifies to Gaussian belief propagation (BP).

The coordination of the different estimation problems at different sensor nodes via Gaussian BP is performed as follows [2]. At a given time instant k, each node that is awake (s_i, $i \in \mathcal{A}_k$) broadcasts its observations and its saved estimated mean and variance of their parameter $\underline{\zeta}_i$ to the other awake nodes. Each awake node then computes (via a local Bayes rule) a new estimated a posteriori mean and variance of their parameter, saves the result, and goes to sleep. This corresponds to belief propagation in the associated factor graph in [2], where the messages passed in the factor graph correspond to marginal Gaussian densities, and thus are characterized by their mean and variance.

This paper focuses on the case where the sets $\mathcal{A}_k$ are chosen randomly, uniformly among the subsets of size c_k of $\mathcal{S}$. The random selection is performed using a pseudo-random number generator so that they repeat after M time instants; thus $\mathcal{A}_k = \mathcal{A}_{k|M}$. We thus call M consecutive time instants an entire sleep cycle. We consider both the *regular* case when the same number of nodes $c_k = c$ are awake at each time instant and each node s_i is awake the same number of time instants $d_i = d$ within an entire sleep cycle, as well as the case where c_k, d_i change for different i, k. In the latter case, the fraction of edges in the factor graph connected to variable nodes of degree $d_i = j$ is denoted by λ_j, and the fraction of edges in the factor graph connected to factor nodes of degree $c_k = m$ is denoted by ρ_m, and these fractions are collected into the *degree distribution polynomials* defined as

$$\lambda(z) := \sum_j \lambda_{j+1} z^j, \quad \rho(z) := \sum_m \rho_{m+1} z^m$$

A key benefit of this random sleep strategy setup is that, with probability $\to 1$ in the large network limit ($N \to \infty$) while the degree distributions remain fixed, for any finite number of iterations ℓ, BP provides the exact a posteriori density for $\underline{\zeta}_i$ given those observations in the factor graph no more than 2ℓ edges away from $\underline{\zeta}_i$. This

[*]J. M. Walsh is supported by the National Science Foundation under grant CCF-0728496.

[†]P. A. Regalia is supported by the National Science Foundation under grant CCF 0728521.

is because the set of all nodes $\mathcal{N}_\ell(\underline{\zeta}_i)$ no more than 2ℓ edges away from the variable node $\underline{\zeta}_i$, called the computation neighborhood $\underline{\zeta}_i$ of depth ℓ, is a tree with probability $\to 1$ as the size of the network $N \to \infty$ [4]. This fact allows density evolution [5] to be applied, in principle, to quantify the performance of BP based distributed estimation after ℓ iterations in the large network limit. We presently apply density evolution to quantify the performance of the BP based estimators, within the especially simple context of our sensor network distributed estimation problem, which allows for closed form exact calculations of the message densities involved. The goal, and ultimate novelty relative to many similar expositions [6, 7], of the paper is then the optimization of the random sleep strategy through the degree distribution polynomials, in order to have BP get the most accurate estimates it can while satisfying a limit on its energy consumption.

2. BP ESTIMATE ACCURACY CALCULUS VIA DENSITY EVOLUTION

We begin by using density evolution to calculate the estimate accuracy within this simple Gaussian BP case.

2.1. Regular Random Sleep Strategies

Let us first consider the case where the number of nodes that are awake at each discrete time instant in the sleep cycle is constant, and each node is awake the same number of times per sleep cycle, so that $d_i = d,\ \forall i \in \{1, \dots, N\}$ and $c_k = c,\ \forall k \in \{0, \dots, M-1\}$. Because all of the random variables involved are Gaussian, and there is a great deal of symmetry in the problem setup, we can calculate the estimated a posteriori distribution provided by Gaussian BP after ℓ iterations exactly. Furthermore, if we treat the mean of the estimated marginal distribution as the estimate of $\underline{\zeta}_i$, then the mean squared error (MSE) of this estimate will be the variance component of the estimated marginal distribution provided by BP. Additionally, due to the fact that the minimum mean squared error estimator is the conditional expectation, the estimated mean provided by BP will be the optimal (MMSE) estimator among any estimators which have access only to those observations in the computation tree rooted at $\underline{\zeta}_i$ of depth ℓ. Also, due to the fact that the factor graph is regular and the distributions are Gaussian, the variance portion of the messages passed by BP do not depend on the observations (as can be seen from the formula for conditioning in jointly Gaussian distributions). Thus, all of the variances in messages passed at the ℓth iteration will be the same deterministic number. Since our focus it to quantify performance, we presently exclusively calculate the evolution of these variance portions (and not the means) of the messages passed over iterations.

First we calculate the variance portion p of the outgoing message to variable node $\underline{\zeta}_i$ from a factor node, given the variances in the incoming messages are q. Relabeling the observations at the given factor node as $\underline{r}[1], \dots, \underline{r}[c]$, then the joint distribution of $\underline{\zeta}_i$ and $\underline{r}[1], \dots, \underline{r}[c]$ given by using the incoming messages as prior distributions will be a Gaussian distribution with covariance matrix

$$\begin{pmatrix} 1 & 1 & 0 & \cdots & 0 \\ 1 & (1+\sigma^2) & \gamma\sigma^2 & \cdots & \gamma\sigma^2 \\ 0 & \gamma\sigma^2 & (q+\sigma^2) & \gamma\sigma^2 & \cdots \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ 0 & \gamma\sigma^2 & \cdots & \gamma\sigma^2 & (q+\sigma^2) \end{pmatrix}$$

Thus, the conditional distribution of $\underline{\zeta}_i$ given all of the observations in its computation tree of size l will be a Gaussian distribution with variance

$$p := 1 - \begin{pmatrix} (1+\sigma^2) & \gamma\sigma^2 & \cdots & \gamma\sigma^2 \\ \gamma\sigma^2 & (\underline{q}+\sigma^2) & \gamma\sigma^2 & \cdots \\ \ddots & \ddots & \ddots & \vdots \\ \gamma\sigma^2 & \cdots & \gamma\sigma^2 & (\underline{q}+\sigma^2) \end{pmatrix}^{-1}_{1,1} \quad (1)$$

where $(\cdot)^{-1}_{1,1}$ denoted the element in the first row and first column of the inverse of the matrix $\cdot$. This variance can be rewritten with the help of the matrix inversion lemma, which states that if a matrix

$$\mathbf{A} = \mathbf{B}^{-1} + \mathbf{C}\mathbf{D}^{-1}\mathbf{C}^T$$

with $\mathbf{A}$,$\mathbf{B}$, and $\mathbf{D}$ positive definite, then

$$\mathbf{A}^{-1} = \mathbf{B} - \mathbf{B}\mathbf{C}\left(\mathbf{D} + \mathbf{C}^T\mathbf{B}\mathbf{C}\right)^{-1}\mathbf{C}^T\mathbf{B}$$

Taking $\mathbf{C} = \mathbf{1}$ the c dimensional column vector with all elements 1, $\mathbf{D}^{-1} = \gamma\sigma^2$, and $\mathbf{B}^{-1} = (1-q)\mathbf{e}_0\mathbf{e}_0^T + (q + (1-\gamma)\sigma^2)\mathbb{I}_c$, and introducing the intermediate variables $\delta = (1 + (1-\gamma)\sigma^2)^{-1}$, $\epsilon := (\gamma\sigma^2)^{-1}$, $\alpha = (1-\gamma)\sigma^2$ we can thus rewrite (1) as

$$p = 1 - \delta + \delta^2\left(\epsilon + \delta + \frac{c-1}{q+\alpha}\right)^{-1} \quad (2)$$

At a variable node $d-1$ such messages are combined to produce a new right-going message variance q

$$q = \frac{(p^{-1}-1)^{-1}}{d-1} = \frac{p}{(d-1)(1-p)} \quad (3)$$

Furthermore, the variance of the estimated a posteriori density is

$$\mathsf{mse} = \left(1 + \frac{d}{(p^{-1}-1)^{-1}}\right)^{-1} = \frac{p}{1+(d-1)(1-p)} \quad (4)$$

This variance also gives the mean squared error of the conditional mean (specified above), which is the minimum mean squared error estimator for $\underline{\zeta}_i$ given the observations in the computation tree. Solving (4) for p, and then substituting into (3) we have

$$q = \frac{d\mathsf{mse}}{d-1+(d-1)(d-2)\mathsf{mse}} \quad (5)$$

We can then put all of these equations together to get a recursion for the mean squared error mse_k after k entire sleep cycles by substituting (5) evaluated at mse_{k-1} into (2) and substituting the resulting expression into (4) to get mse_k. Of course, the prior distribution on $\underline{\zeta}_i$ gives an initial mean squared error $\mathsf{mse}_0 = 1$.

2.2. Irregular Random Sleep Strategies

This was all for a given d and c, but for the irregular factor graph case the node degrees will in general be random variables $\underline{d}, \underline{c}$. The degree distributions $\lambda(z)$ and $\rho(z)$ can in this case be interpreted as moment generating functions for $\underline{d}-1$ and $\underline{c}-1$.

To calculate the performance of BP based distributed estimation in this case, note that when the graph is irregular (1) becomes

$$\underline{p} := 1 - \begin{pmatrix} (1+\sigma^2) & \gamma\sigma^2 & \cdots & \gamma\sigma^2 \\ \gamma\sigma^2 & (\underline{q}_1+\sigma^2) & \gamma\sigma^2 & \cdots \\ \ddots & \ddots & \ddots & \vdots \\ \gamma\sigma^2 & \cdots & \gamma\sigma^2 & (\underline{q}_{c-1}+\sigma^2) \end{pmatrix}^{-1}_{1,1}$$

where $\underline{q}_1$ through $\underline{q}_{c-1}$ are i.i.d. random variables, and $c-1$ is an independent random variable whose probability mass function has moment generating function given by the factor node degree distribution $\rho(z)$. Thus (2) becomes

$$\underline{p} = 1 - \delta + \delta^2 \left(\epsilon + \delta + \sum_{i=1}^{c-1} \frac{1}{\underline{q}_i + \alpha} \right)^{-1} \tag{6}$$

Furthermore, (3) becomes

$$\underline{q} = \left(\sum_{i=1}^{\underline{d}-1} (\underline{p}_i^{-1} - 1) \right)^{-1} \tag{7}$$

where $\underline{p}_i$ are $\underline{d}$ i.i.d. random variables whose distribution matches the distribution for $\underline{p}$ determined by (6), and $\underline{d}-1$ is an independently distributed random variable with moment generating function $\lambda(z)$. To get this into a form amenable to density evolution, change variables to $\underline{\theta} = \underline{q}^{-1}$ and $\underline{\phi}_i = p^{-1} - 1$, then (7) becomes a familiar sum of a random number of random variables, ie

$$\underline{\theta} = \sum_{i=1}^{\underline{d}-1} \underline{\phi}_i$$

The characteristic function $Q(\Omega)$ of this sum can be found by composing the moment generating function of the variable $\underline{d}-1$ (which is just the left node degree distribution $\lambda(z)$) with the characteristic function $P(\Omega)$ for $\underline{\phi}_i = \underline{p}_i^{-1} - 1$, i.e. $Q(\Omega) = \lambda(P(\Omega))$.

The factor node update equation (6) can be transformed by introducing the intermediate variable

$$\underline{w} := \sum_{i=1}^{c-1} \frac{\underline{\theta}_i}{1 + \alpha \underline{\theta}_i} \tag{8}$$

to rewrite (6) as $\underline{p} = 1 - \delta + \delta^2 (\epsilon + \delta + \underline{w})^{-1}$, and thus the new distribution for the $\underline{\phi}_i$s is the distribution of the random variable

$$\underline{\phi} = \underline{p}^{-1} - 1 = \frac{(\epsilon + \underline{w})}{1 + \alpha(\epsilon + \underline{w})}$$

Define the operator $\mathcal{F}_\alpha$ which takes a characteristic function $Q(\Omega)$ for a non-negative random variable $\underline{\theta}$ and returns the characteristic function

$$\mathcal{F}_\alpha Q(\Omega) := \mathbb{E}\left[\exp\left(\imath \Omega \frac{\underline{\theta}}{1 + \alpha \underline{\theta}} \right) \right]$$

Then, the random variable $\underline{w}$ defined in (8) has the characteristic function

$$W(\Omega) := \rho(\mathcal{F}_\alpha Q(\Omega))$$

Next define the operator $\mathcal{G}_{\alpha,\epsilon}$ which takes the characteristic function $W(\Omega)$ for a random variable $\underline{w}$ and returns the characteristic function

$$\mathcal{G}_{\alpha,\epsilon} W(\Omega) := \mathbb{E}\left[\exp\left(\imath \Omega \frac{w + \epsilon}{1 + \alpha(\epsilon + w)} \right) \right]$$

Then the density evolution between $P_k(\Omega)$ and $P_{k+1}(\Omega)$ can be found as

$$P_{k+1}(\Omega) = \mathcal{G}_{\alpha,\epsilon} \rho(\mathcal{F}_\alpha \lambda(P_k(\Omega)))$$

or equivalently

$$Q_{k+1}(\Omega) = \lambda(\mathcal{G}_{\alpha,\epsilon} \rho(\mathcal{F}_\alpha Q_k(\Omega)))$$

Finally, letting $\underline{x}$ be the random variable whose characteristic function is $P_k(\Omega)\lambda(P_k(\Omega))$, the distribution of the MSE of the estimate provided at a randomly selected sensor node after k iterations has a characteristic function $M_k(\Omega)$ defined by

$$M_k(\Omega) := \mathbb{E}\left[\exp\left(\imath \frac{\Omega}{\underline{x}} \right) \right]$$

The expectation of this MSE (over the random sleep strategies) then gives the average accuracy of the estimate provided by Gaussian BP based distributed estimation.

3. OPTIMIZED ENERGY VS. ACCURACY TRADEOFF

The estimation error decreases with increasing numbers of iterations, as well as with increasing (average) degrees d or c, since all of these imply more observations are used in forming the (exact) marginal a posteriori density provided by Gaussian BP. However, this increase in estimation accuracy comes at the cost of more messages being passed, and thus greater energy expended. Thus, it is of interest to study to energy vs. accuracy tradeoffs within this simplified scenario.

3.1. Regular Random Sleep Strategies

Intuitively, the energy consumed by a node for estimation is proportional to the number of discrete time instants that it is awake and sends a message, and the amount of energy consumed per time instant is proportional to the number of nodes that are awake per time instant. This is because of limited communications resources amongst the nodes. In particular, as the number of nodes grows for a fixed time interval corresponding to a single discrete time instant in the sleep cycle, the nodes will have to expend more energy to communicate with each other. Furthermore, the number of discrete time instants a node must be awake and send a message for the estimation is the number of time instants it is awake per sleep cycle times the number of sleep cycles used for the estimation. The simplest model capturing these proportionalities takes the energy to be

$$E(\ell, d, c) = \ell c d$$

It is then of interest both to study the tradeoff between energy efficiency and estimation accuracy as dictated by the optimization problem

$$\mathsf{mse}^*(E^*) := \min_{\ell,c,d | E(\ell,d,c) \le E^*} \mathsf{mse}_\ell(d, c)$$

the optimal regular sleep strategies may also be determined as a result of this optimization problem according to

$$\ell^*(E^*), c^*(E^*), d^*(E^*) := \arg \min_{\ell,c,d | E(\ell,d,c) \le E^*} \mathsf{mse}_\ell(d, c)$$

Figure 1 shows the optimal sleep strategies and energy accuracy tradeoff for particular values of the parameters γ and σ^2.

3.2. Irregular Random Sleep Strategies

Just as in the regular case, the energy consumed by a node for estimation in a sensor network employing irregular random sleep strategies is proportional to the number of time instants it is awake times the energy consumed per time instant. The energy consumed per time instant, is in turn, determined by the number of nodes awake per time instant. A simple energy metric is then

$$E(\ell, \lambda, \rho) = \frac{\ell}{\left(\int_0^1 \lambda(z) dz \right) \left(\int_0^1 \rho(z) dz \right)}$$

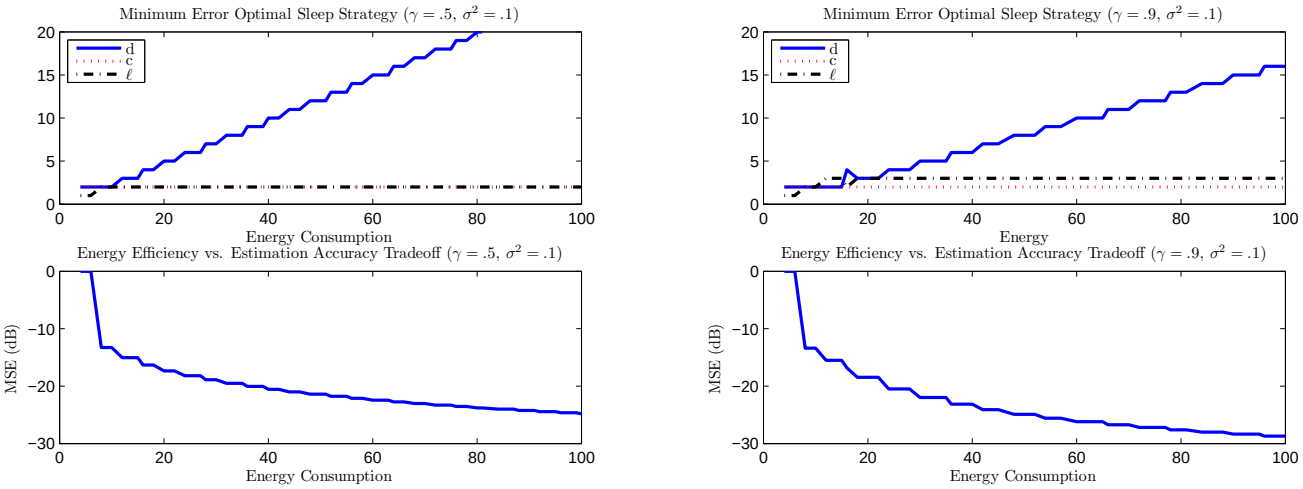

Fig. 1. Optimized sleep strategies and energy vs. accuracy tradeoff for $\gamma = .9$ and $\sigma^2 = .1$ and $\gamma = .5$ and $\sigma^2 = .1$.

$$\mathsf{mse}^*(E^*) := \min_{\lambda,\rho,\ell | E(\ell,\lambda,\rho) \le E^*} \mathsf{mse}_\ell(\lambda,\rho)$$

$$(\ell^*(E^*), \lambda^*(E^*), \rho^*(E^*)) := \arg \min_{\lambda,\rho,\ell | E(\ell,\lambda,\rho) \le E^*} \mathsf{mse}_\ell(\lambda,\rho)$$

Optimization can be carried out by adjusting the coefficients of $\lambda(z)$ or $\rho(z)$ subject to the constraints that

$$\sum_{j \ge 1} \lambda_{j+1} = 1 \quad \text{and} \quad \int_0^1 \lambda(z)\, dz = \sum_{j \ge 1} \frac{\lambda_{j+1}}{j+1} = \text{constant}$$

and similarly for $\rho(z)$, to ensure that $\lambda(z)$ or $\rho(z)$ captures a valid probability mass function, and that the energy measure is fixed during the adaptation. A random search direction η is produced for the coefficients of $\lambda(z)$ or $\rho(z)$, and projected along the constraints $\sum_j \eta_j = 0$ and $\sum_j \eta_j / j = 0$, with the better of $\lambda \pm \eta$ (or $\rho \pm \eta$) for the mean-square error retained for the next search step.

As an example, consider $\sigma^2 = 0.1$ and $\gamma = 0.5$ from the fixed degree case, and suppose that $\lambda(z)$ and $\rho(z)$ are both constrained to have degree 8 (or less). Using as initial values $\lambda(z) = z^3$, $\rho(z) = z$ and $\ell = 2$ found in the fixed degree case, optimized polynomials are found as

$$\begin{aligned} \lambda(z) &= 0.2817z^2 + 0.1905z^3 + 0.1734z^4 + 0.2514z^5 \\ &\quad + 0.0264z^6 + 0.0010z^7 + 0.0756z^8 \\ \rho(z) &= 0.996z + 0.0001z^6 + 0.0002z^7 + 0.001z^8 \end{aligned}$$

giving a mean-square error of -16 dB after 2 iterations of expectation propagation. This gives a slight, although not earth-shattering, improvement in estimation accuracy for the same energy expenditure, which is 20. We observe that the coefficient λ_2 has vanished, indicating the absence of degree 2 nodes from the optimization procedure, and that the optimization procedure has favored wake cycles of degree 7 or higher.

4. CONCLUSION

We have quantified the performance of BP based distributed estimation in large sensor networks, and have showed how to optimize random sleep strategies to trade estimate accuracy for energy efficiency. Future work will continue to study numerically energy accuracy tradeoffs using the established density evolution for irregular sleep strategies, searching for not just locally, but globally optimal degree distributions, quantifying their improvement over regular sleep strategies of similar average energy consumption.

5. REFERENCES

[1] C. C. Moallemi and B. Van Roy, "Consensus propagation," *IEEE Trans. Inform. Theory*, vol. 52, no. 11, pp. 4753–4766, Nov. 2006.

[2] J. M. Walsh and P. A. Regalia, "Expectation propagation for distributed estimation in sensor networks," in *8th IEEE International Workshop on Signal Processing Advances for Wireless Communications (SPAWC)*, 2007.

[3] F. R. Kschischang, B. J. Frey, and H.-A. Loeliger, "Factor graphs and the sum-product algorithm," *IEEE Trans. Inform. Theory*, vol. 47, pp. 498–519, Feb. 2001.

[4] T. J. Richardson and R. L. Urbanke, "The capacity of low-density parity-check codes under message-passing decoding," *IEEE Trans. Inform. Theory*, vol. 47, no. 2, pp. 599–618, Feb. 2001.

[5] J. M. Walsh, "Density evolution for expectation propagation," in *International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, Honolulu, Hawaii, 2007.

[6] V. Saligrama, M. Alanyali, and O. Savas, "Distributed Detection in Sensor Networks With Packet Losses and Finite Capacity Links," *IEEE Trans. Signal Processing*, vol. 54, no. 11, pp. 4118–4132, Nov. 2006.

[7] J. M. F. Moura, J. Lu, and M. Kleiner, "Intelligent sensor fusion: A graphical model approach," in *Proc. Int. Conf. Acoustics, Speech and Signal Processing*, Singapore, April 2003, vol. VI, pp. 733–736.