EXPECTATION PROPAGATION FOR DISTRIBUTED ESTIMATION IN SENSOR NETWORKS

*John MacLaren Walsh**

Dept. of Electrical & Computer Eng.
Drexel University
Philadelphia, PA 19104

Phillip A. Regalia

Dept. of Electrical Eng. & Computer Sci.
Catholic University of America
Washington, DC 20064

ABSTRACT

We show that the expectation propagation (EP) family of algorithms constitute a natural choice for distributed estimation and detection in sensor networks. In particular, random sleep strategies, which are commonly chosen to ensure robustness, equal power dissipation across the network, and ease of deployment, espouse a sparse dependence structure among the parameters to be estimated. This sparse dependence structure mimics the structure which belief and expectation propagation exploited in the decoding of turbo and low density parity check (LDPC) codes to bring the performance of physical layer communications systems to the fundamental limits set out by Shannon. We provide examples of practical sensor network tasks which fall into the framework set out in this paper. By applying extensions of the extrinsic information transfer (EXIT) chart theory to EP in these distributed estimation applications, we can predict the performance and convergence of the distributed estimation algorithm in very large networks with an easy to obtain plot.

1. INTRODUCTION

Recent increasing interest in distributed algorithms which can perform simple statistical calculations such as network wide averaging [1], together with belief propagation's message passing formulation have begun to attract researchers' attention to belief propagation as a distributed algorithm for use in wireless networks and sensor networks [2, 3, 4, 5, 6]. In [2] the convergence of belief propagation is studied for the distributed calculation of averages within a network, while [3] more generally discusses issues arising when applying belief propagation to distributed fusion in sensor networks. [6] considered belief propagation as a candidate algorithm for multi-base station detection in the cellular network. [4] proposes a particle-filtering like modification of belief propagation suited for message probability densities that are not discrete or Gaussian, and applies it to sensor localization. Finally, [5] points out the advantages of the message passing formulation of belief propagation in a sensor network, and investigates the effects of several issues arising in sensor networks on the performance of belief propagation in an Monte Carlo simulations based empirical manner. It is becoming clear that belief propagation is a possible candidate for several aspects of sensor network algorithm design.

Following and generalizing upon this recent trend, in this paper we examine the possibility of using expectation or belief propagation for distributed estimation in a sensor network. In particular, we consider a situation in which each sensor node has a different parameter vector (from the other sensor nodes) that it wants to infer. Each sensor's parameter statistically depends a posteriori on many or all of the observations across the entire network. Unique to our development is the discussion that certain common sensor design techniques, namely random sleep strategies[7, 8], which are often employed in order to guarantee ease of sensor placement, robustness, and equal power dissipation across the network, can build a sparse dependence structure among the parameters to infer that still retains this high amount of dependence. One important benefit of this sparse dependence is that, as the size of the network grows, one can then show that density evolution provides a viable method for analyzing the performance of EP based distributed estimation. We then exploit this fact, together with an additional strong Gaussian approximation, to apply extrinsic information transfer (EXIT) chart theory to study the performance and convergence of EP for these sensor network distributed estimation problems. One of the key aims of this paper is to give examples of common sensor network problems which fall neatly into this framework. We demonstrate how to use these newly extended tools to predict the large network performance and convergence behavior of EP, which will ultimately be important for algorithm designers considering EP as a candidate algorithm for distributed estimation in sensor networks. Such an approach using EXIT charts will help designers graphically predict the performance of expectation propagation in the large network limit, allowing for increasingly less reliance on brute force monte-carlo simulations to guide design choices and only for design verification.

*This research supported in part by Drexel University. Contact jwalsh@ece.drexel.edu for inquiries regarding this paper.

2. SENSOR NETWORK MODEL

Consider a sensor network of nodes $\{n_{\mathtt{j}} | \mathtt{j} \in \{1, \ldots, \mathtt{S}\}\}$, with each sensor node $n_{\mathtt{j}}$ trying to determine information concerning a parameter $\boldsymbol{\phi}_{\mathtt{j}}$. In order to ensure energy efficiency, the sensor network employs a random sleep strategy, under which at each discrete time instant $\mathtt{k}$ a randomly selected collection of $\mathtt{d_f}$ nodes with indices denoted by $\mathcal{A}(\mathtt{k})$ are awake. Only nodes that are awake may observe the environment and communicate with each other. The random sleep strategy is implemented using a pseudo-random number generator which repeats after a certain number of time steps $\mathtt{W}$, so that $\mathcal{A}(\mathtt{k}) = \mathcal{A}(\mathtt{k} \mod \mathtt{W})$. In order to ensure constant average power dissipation across the network, the pseudo-random number generator is chosen such that over the time interval $\mathtt{W} - 1 \geq \mathtt{k} \geq 0$ each sensor node is awake the same number of times. The observations made at time $\mathtt{k}$ by the awake nodes are denoted by $\mathbf{r}_{\mathtt{k}}$. Given the parameters $\boldsymbol{\phi}_{\mathtt{j}} = \phi_{\mathtt{j}} \forall \mathtt{j} \in \{1, \ldots, \mathtt{S}\}$, the observations between different time periods before the repetition of the pseudo-random number generator controlling the sleep cycle are statistically independent, so we can write the density for all of the observations up until time $\mathtt{k}$, $\mathbf{r}_{0\ldots\mathtt{k}}$ given all of the parameters $\boldsymbol{\theta} := [\boldsymbol{\phi}_{\mathtt{j}} | \mathtt{j} \in \{1, \ldots, \mathtt{S}\}]$ as

$$\mathsf{p}_{\mathbf{r}_{0,\ldots,\mathtt{k}}|\boldsymbol{\theta}}(\mathbf{r}_{0\ldots\mathtt{k}}|\boldsymbol{\theta}) := \prod_{\mathtt{c}=0}^{\mathtt{k}} \mathsf{p}_{\mathbf{r}_{\mathtt{c}}|\boldsymbol{\theta}_{\mathtt{c}}}(\mathbf{r}_{\mathtt{c}}|\boldsymbol{\theta}_{\mathtt{c}}) \tag{1}$$

where $\boldsymbol{\theta}_{\mathtt{c}} := [\boldsymbol{\phi}_{\mathtt{j}} | \mathtt{j} \in \mathcal{A}(\mathtt{c})]$ for any $\mathtt{k} < \mathtt{W}$.

Because we will be studying the amount of communication employed by our proposed method to spread the information accrued during a single complete cycle of the pseudo-random number generator, we will assume that the sensor readings are only taken during the first complete cycle. We will then study the performance of the estimates obtained by our distributed estimation scheme as the number of complete cycles of the pseudo-random number generator grows to determine the amount of communication used per observation. Thus, for any time $\mathtt{k} > \mathtt{S}$, it suffices to consider the conditional probability density for $\mathbf{r}_{0\ldots\mathtt{S}-1} := \mathbf{r}$ given $\boldsymbol{\theta}$.

$$\mathsf{p}_{\mathbf{r}|\boldsymbol{\theta}}(\mathbf{r}|\boldsymbol{\theta}) := \prod_{\mathtt{k}=0}^{\mathtt{W}-1} \mathsf{p}_{\mathbf{r}_{\mathtt{k}}|\boldsymbol{\theta}_{\mathtt{k}}}(\mathbf{r}_{\mathtt{k}}|\boldsymbol{\theta}_{\mathtt{k}})$$

Here, and from now on, the probability "density" we will be working with should be understood to be Radon Nikodym derivative of the associated probability measure with respect to an appropriate dominating measure $\mathsf{d}\boldsymbol{\theta}$ which will usually be the counting measure if $\boldsymbol{\theta}$ is discrete or Lebesgue measure if $\boldsymbol{\theta}$ is continuous. We further assume a certain symmetry in the observation processes so that

$$\mathsf{p}_{\mathbf{r}_{\mathtt{k}}|\boldsymbol{\theta}_{\mathtt{k}}}(\mathbf{r}|\boldsymbol{\theta}) = \mathsf{p}_{\mathbf{r}_{\mathtt{c}}|\boldsymbol{\theta}_{\mathtt{c}}}(\mathbf{r}|\boldsymbol{\theta}) \quad \mathtt{c} \neq \mathtt{k} \, \forall \mathbf{r}, \boldsymbol{\theta} \tag{2}$$

and thus that the difference among the factors $\mathsf{p}_{\mathbf{r}_{\mathtt{k}}|\boldsymbol{\theta}_{\mathtt{k}}}(\mathbf{r}_{\mathtt{k}}|\boldsymbol{\theta}_{\mathtt{k}})$ is that they are functions of different subsets of parameters $\boldsymbol{\theta}_{\mathtt{k}}$. Finally, a priori, the parameters $\{\boldsymbol{\phi}_{\mathtt{j}} | \mathtt{j} \in \{1, \ldots, \mathtt{S}\}\}$ are independent and identically distributed, so that

$$\mathsf{p}_{\boldsymbol{\theta}}(\boldsymbol{\theta}) := \prod_{\mathtt{j}=1}^{\mathtt{S}} \mathsf{p}_{\boldsymbol{\phi}_0}(\boldsymbol{\phi}_{\mathtt{j}}) \tag{3}$$

where $\mathsf{p}_{\boldsymbol{\phi}_0}$ is a minimal exponential family probability density [9, 10, 11], so that it can be written as

$$\mathsf{p}_{\boldsymbol{\phi}_0}(\boldsymbol{\phi}) := \exp\left(\mathbf{v}(\boldsymbol{\phi}) \cdot \boldsymbol{\lambda} - \psi(\boldsymbol{\lambda})\right) \tag{4}$$

for some real vector $\boldsymbol{\lambda}$, a (possibly vector valued) function $\mathbf{v}$, and appropriately chosen normalizing constant function $\psi(\boldsymbol{\lambda})$. The function $\mathbf{v}$ together with the dominating measure $\mathsf{d}\boldsymbol{\phi}$ specifies the type of exponential family density (e.g. specifies if $\boldsymbol{\phi}$ will be Gaussian, discrete with N outcomes, exponential, Poisson, etc), while λ specifies which probability distribution (i.e. specifies the hyper-parameters for the probability distribution) among that type. We will also assume for simplicity within the context of this paper that we have a conjugate prior situation, so that the a posteriori distribution for $\boldsymbol{\phi}_{\mathtt{j}}$ given $\mathbf{r}_{\mathtt{k}}$ for $\mathtt{j} \in \mathcal{A}(\mathtt{k})$ is also an exponential family distribution with the form of (4), but with a different $\boldsymbol{\lambda}$ [9].

The sort of model we have specified arises, for instance, when the observations are of parameters, which are a priori independent, through additive noise which is correlated across the sensors. We give some examples of practical sensor networks which satisfy (1) in Section 4, but first we show in Section 3 how the expectation propagation [12, 13, 14] family of algorithms can be applied to yield a distributed iterative estimation/detection method in sensor networks following this model.

3. EP FOR DISTRIBUTED ESTIMATION

The goal of distributed estimation in this context within sensor networks is to provide each node $n_{\mathtt{j}}$ with enough of the information concerning what observations happened while it was asleep (i.e. $\mathbf{r}_{\mathtt{i}}$ with $\mathtt{j}$ not in $\mathcal{A}(\mathtt{i})$), so as to give it knowledge of the a posteriori distribution $\mathsf{p}_{\boldsymbol{\phi}_{\mathtt{j}}|\mathbf{r}}(\cdot|\mathbf{r})$, while still obeying the constraints of the given sleep strategy. Expectation propagation (EP) [12, 13, 14] is a natural solution for this problem. To see this, consider the following factorization of the joint distribution

$$\mathsf{p}_{\mathbf{r},\boldsymbol{\theta}}(\mathbf{r},\boldsymbol{\theta}) := \prod_{\mathtt{k}=0}^{\mathtt{W}-1} \mathsf{p}_{\mathbf{r}_{\mathtt{k}}|\boldsymbol{\theta}_{\mathtt{k}}}(\mathbf{r}_{\mathtt{k}}|\boldsymbol{\theta}_{\mathtt{k}}) \prod_{\mathtt{j}=1}^{\mathtt{S}} \mathsf{p}_{\boldsymbol{\phi}_0}(\boldsymbol{\phi}_{\mathtt{j}}) \tag{5}$$

Here $\mathsf{p}_{\mathbf{r}_{\mathtt{k}}|\boldsymbol{\theta}_{\mathtt{k}}}(\mathbf{r}_{\mathtt{k}}|\boldsymbol{\theta}_{\mathtt{k}})$ is the likelihood function for the measurements $\mathbf{r}_{\mathtt{k}}$ taken by the sensors awake at time interval $\mathtt{k}$. To this factored form (5), one may associate a bipartite factor graph [15], as in Figure 1. The $\mathtt{j}$-th left node represents the $\mathtt{j}$-th sensor, and the $\mathtt{k}$-th right node of degree greater than one represents the conditional pdf $\mathsf{p}_{\mathbf{r}_{\mathtt{k}}|\boldsymbol{\theta}_{\mathtt{k}}}(\mathbf{r}_{\mathtt{k}}|\boldsymbol{\theta}_{\mathtt{k}})$ at time $\mathtt{k}$. Thus,

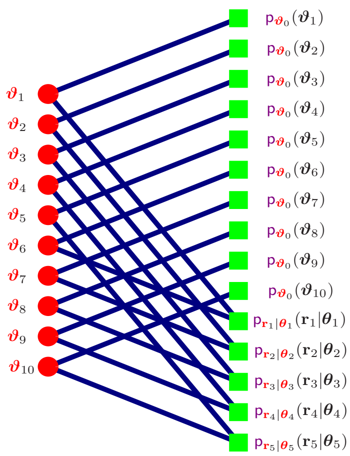

Fig. 1. An example of a factor graph.

sensor node $\mathtt{j}$ connects to factor node $\mathtt{k}$ of degree greater than one if it is awake at time $\mathtt{k}$, i.e., $\mathtt{j} \in \mathcal{A}(\mathtt{k})$. The $\mathtt{j}$-th right node of degree one represent the prior probability distribution $p_{\phi_0}(\phi_{\mathtt{j}})$.

The expectation propagation algorithm may then be applied as a message passing algorithm on edges of the factor graph [14]; the messages passed correspond to vectors $\boldsymbol{\lambda}$ which parameterize exponential distributions of the form (4). Specifically, at time $\mathtt{k}$, given the incoming messages $\mathbf{m}_{\mathtt{j}\to\mathtt{k}}$ [with $\mathtt{j} \in \mathcal{A}(\mathtt{k})$] to factor node $\mathtt{k}$, the outgoing message returned to sensor node $\mathtt{j}$ is

$$\mathbf{n}_{\mathtt{k}\to\mathtt{j}} := -\mathbf{m}_{\mathtt{j}\to\mathtt{k}} + \boldsymbol{\Lambda}^{-1}\left(\frac{\int \mathbf{v}(\phi_{\mathtt{j}}) f_{\mathtt{k}}(\boldsymbol{\theta}_{\mathtt{k}}) \prod_{\mathtt{i}\in\mathcal{A}(\mathtt{k})} \exp\left(\mathbf{v}(\phi_{\mathtt{j}}) \cdot \mathbf{m}_{\mathtt{j}\to\mathtt{k}}\right) d\phi_{\mathtt{j}}}{\int f_{\mathtt{k}}(\boldsymbol{\theta}_{\mathtt{k}}) \prod_{\mathtt{i}\in\mathcal{A}(\mathtt{k})} \exp\left(\mathbf{v}(\phi_{\mathtt{j}}) \cdot \mathbf{m}_{\mathtt{j}\to\mathtt{k}}\right) d\phi_{\mathtt{j}}}\right) \quad (6)$$

where $f_{\mathtt{k}}(\boldsymbol{\theta}_{\mathtt{k}}) = p_{\mathbf{r}_{\mathtt{k}}|\boldsymbol{\theta}_{\mathtt{k}}}(\mathbf{r}_{\mathtt{k}}|\boldsymbol{\theta}_{\mathtt{k}})$ and $\boldsymbol{\Lambda}^{-1}$ is the inverse of the transformation

$$\boldsymbol{\Lambda}(\boldsymbol{\lambda}) := \frac{\int \mathbf{v}(\phi) \exp(\boldsymbol{\lambda} \cdot \mathbf{v}(\phi)) d\phi}{\int \exp(\boldsymbol{\lambda} \cdot \mathbf{v}(\phi)) d\phi}$$

which is bijective whenever $\mathbf{v}(\cdot)$ is minimal[9]. The messages passed from the factor nodes of degree one (i.e. the prior density for $\phi_{\mathtt{j}}$) to their corresponding parameter node are simply the $\boldsymbol{\lambda}$ that specifies them in (4).

Reciprocally, given the incoming messages $\mathbf{n}_{\mathtt{k}\to\mathtt{j}}$ at sensor node $\mathtt{j}$, the outgoing message passed to the $\mathtt{k}$-th factor node of degree greater than one is

$$\mathbf{m}_{\mathtt{j}\to\mathtt{k}} := \sum_{\mathtt{c}\in\mathcal{F}(\mathtt{j})\setminus\{\mathtt{k}\}} \mathbf{n}_{\mathtt{c}\to\mathtt{j}} \quad (7)$$

where

$$\mathcal{F}(\mathtt{j}) := \{\mathtt{k} | \mathtt{j} \in \mathcal{P}(\mathtt{k})\}$$

are the indices of the parameters which neighbor the $\mathtt{j}$th parameter node, and $\mathcal{P}(\mathtt{k})$ are the indices of the factors to which $\phi_{\mathtt{j}}$ connects in the factor graph.

In a practical implementation, at time step $\mathtt{t}$, with $\mathtt{k} = \mathtt{t} \bmod \mathtt{W}$, the awake nodes broadcast the messages $\{\mathbf{m}_{\mathtt{j}\to\mathtt{k}} | \mathtt{j} \in \mathcal{A}(\mathtt{k})\}$ and the observations $\mathbf{r}_{\mathtt{k}}$ to each other; each node locally computes $\{\mathbf{n}_{\mathtt{k}\to\mathtt{j}} | \mathtt{k} \in \mathcal{F}(\mathtt{j})\}$, saves the result, and then returns to sleep.

This factor graph satisfies the homogeneity properties from [16, 17], with degree one factor nodes $p_{\phi_0}(\phi_{\mathtt{j}})$, and other common degree factor nodes $p_{\mathbf{r}_{\mathtt{k}}|\boldsymbol{\theta}_{\mathtt{k}}}(\mathbf{r}_{\mathtt{k}}|\boldsymbol{\theta}_{\mathtt{k}})$. The subgraph of this graph which drops the factor nodes of degree one and the edges connected to them is a regular bipartite graph, which was selected randomly. Thus we have all of the ingredients to allow for density evolution [16] to correctly model expectation propagation's behavior as a distributed estimation algorithm. Under an additional (strong Gaussian) approximation we can also graphically analyze the performance of expectation propagation based distributed data fusion using an EXIT chart [17], a possibility which we will investigate in a later section. First, it is important to discuss instances of tasks in practical wireless networks and sensor networks which follow the assumptions of our model (1,2,3).

4. EXAMPLES OF SENSOR NETWORKS FOLLOWING THE MODEL

One specific sensor network task suitable for our model is to determine the presence and location of intruders into the area that the sensor network is monitoring. Suppose for the sake of argument that the physical quantity being monitored is the temperature, and that the intruders are known to be at a higher temperature than the environment. The intruder detection algorithm at the sensors operates by comparing the local temperature with what it ought to be (based on the readings from other sensors) if there were no intruder. If there is a large difference between the local reading and this predicted value, then the intruder detection algorithm determines that an intruder is likely to be present. This situation falls neatly into the framework of the model (1,2,3). In particular, let the temperature observed by sensor node $n_{\mathtt{j}}$ at time $\mathtt{a}$ be $r_{\mathtt{a},\mathtt{j}}$, let $\zeta_{\mathtt{a},\mathtt{j}}$ be the value of that temperature if no intruder were present, and let $\phi_{\mathtt{j}}$ be zero if there is no intruder present and α if there is an intruder present, so that

$$r_{\mathtt{a},\mathtt{j}} = \phi_{\mathtt{j}} + \zeta_{\mathtt{a},\mathtt{j}} \quad (8)$$

While it is reasonable to assume that the presence, or lack thereof, of the intruder is independent of the ambient temperature process, the ambient temperature at two different sensors in the network ought to be highly correlated, given that everyday experience dictates that over a fairly large spatial range out of doors, the temperature remains relatively constant, and when it does vary, it varies in a spatially correlated manner. To state this more precisely, $\zeta_{\mathtt{a},\mathtt{j}}$ is best thought of as a sample from a spatial stochastic process $\zeta(t, \mathbf{x})$ of both time and position in space. The random variable $\zeta_{\mathtt{a},\mathtt{j}}$ can be considered to be a sample from this stochastic process at time $\mathtt{a}$ and at the sensor node $n_{\mathtt{j}}$'s position $\mathbf{x}_{\mathtt{j}}$. Based on the prominence of the central limit theorem and diffusion phenomena as modeling heat transfer, let us take $\zeta(t, \mathbf{x})$ to be a Gaussian random process with mean m and auto-correlation function $\mathsf{R}(t_1, \mathbf{x}_1, t_2, \mathbf{x}_2)$. In the regime in which we are interested, the autocorrelation function decays very quickly with $|t_1 - t_2|$ relative to the sample period, but for $t_1 = t_2$ is a slowly decaying function of $\|\mathbf{x}_1 - \mathbf{x}_2\|_2$ relative to the typical distance between any two sensors in the sensor network. A zeroth order model for such a situation specifies that

$$\begin{aligned} \mathsf{R}(t_1, \mathbf{x}_1, t_2, \mathbf{x}_2) \quad = \quad & \delta(t_1 - t_2) \left[\delta(\mathbf{x}_1 - \mathbf{x}_2)\eta^2 \right. \\ & \left. +(1 - \delta(\mathbf{x}_1 - \mathbf{x}_2))\rho\eta^2 \right] \end{aligned} \tag{9}$$

with a ρ close to one and δ either the Dirac point mass distribution in continuous time or the Kronecker delta function in discrete time, but such a specification should be simply understood as stating that the temperature process varies about its mean quickly with respect to the sampling period, and slowly with respect to the typical spatial scale on which the sensor network is operating. If we collect the $r_{\mathtt{a},\mathtt{j}}$ into vectors $\mathbf{r}_{\mathtt{a}} := \left[r_{\mathtt{a},\mathtt{j}} | \mathtt{j} \in \mathcal{A}(\mathtt{a}) \right]$ and the $\boldsymbol{\phi}_{\mathtt{j}}$ into vectors $\boldsymbol{\theta}_{\mathtt{a}} := \left[\boldsymbol{\phi}_{\mathtt{j}} | \mathtt{j} \in \mathcal{A}(\mathtt{a}) \right]$, then the resulting model satisfies (1,2,3). The distributed estimation with expectation propagation paradigm, then, offered in this paper can be seen as a rigorous distributed method for the intruder detection algorithm.

Alternate sensor network applications can now be recognized by keeping (8) and (9), and simply choosing $\boldsymbol{\phi}_{\mathtt{j}}$ to be another type of exponential family random variable, in which case one would aim to estimate, e.g., the effect of the intruder on the natural environment specified by $\boldsymbol{\zeta}_{\mathtt{a},\mathtt{j}}$. Alternatively, $\boldsymbol{\zeta}_{\mathtt{a},\mathtt{j}}$ could represent the best available current physical model for the phenomenon being monitored, and $\boldsymbol{\phi}_{\mathtt{j}}$ is a model error term which, when estimated, will be used to alter the sensor node if its deviation from the physical model is significant enough to warrant informing a command center or otherwise altering its behavior.

Another application of these ideas, again with (8) and (9) being true, involves source separation. In this instance, $\boldsymbol{\phi}_{\mathtt{j}}$ could represent the signal being emitted by a transmitter in the vicinity of sensor node $n_{\mathtt{j}}$ and the noise $\boldsymbol{\zeta}_{\mathtt{j}}$ could represent the aggregate signal at sensor node $n_{\mathtt{j}}$ formed from all of the other communication signals in the network, which will be correlated across different sensors. The sensor node wishes to know the signal being emitted by the nearby transmitter, which before viewing the observations (a priori) are independently and identically distributed. It is this example that we will apply EXIT chart theory to in the next section.

5. EXIT PERFORMANCE OF EP BASED SENSOR FUSION

In this section we highlight one of the significant consequences of using expectation propagation in large sensor networks satisfying our model. In particular, in many instances one may use extrinsic information transfer (EXIT) charts, which were developed for the analysis of iterative decoding[18, 19, 20, 21], to predict the performance of the estimates provided by EP performing distributed estimation in a sensor network. This fact is perhaps best illustrated by an example. Let us continue with our last example sensor network application of distributed estimation with EP for source separation. To allow us to plot an example EXIT chart, let us additionally assume that the aggregate signal from all of the far away transmitters, and hence, $\mathsf{p}_{\mathbf{r}_{\mathtt{k}}|\boldsymbol{\theta}_{\mathtt{k}}}(\mathbf{r}_{\mathtt{k}}|\boldsymbol{\theta}_{\mathtt{k}})$ is Gaussian with mean $\boldsymbol{\theta}_{\mathtt{k}}$ and covariance matrix $\boldsymbol{\Sigma}$ whose diagonal elements are η^2 and whose off diagonal elements are $\rho\eta^2$ for all $\mathtt{k}$. Additionally, we choose the factor node degree to be $\mathtt{d}_{\mathtt{f}} = 2$, so that only two sensor nodes are on at any given time, and let the correlation coefficient ρ of the 2×2 covariance matrix be .999, so that the nodes in the sensor network stand to gain a lot from collaboration. Furthermore, again for the sake of demonstrating the theory, let us assume that $\boldsymbol{\phi}_{\mathtt{j}}$ is a priori a Gaussian random variable with mean μ and variance γ (so that it is a scalar quantity) for all $\mathtt{j}$, and thus $\mathbf{v}(\boldsymbol{\phi}) = [\boldsymbol{\phi}, \boldsymbol{\phi}^2]$. An EXIT chart for this situation is depicted in 2. From this graph, we can see that if the size of the network is large enough ($\mathtt{S} \geq 100000$) in this case, the EXIT chart technique correctly predicts both the convergence behavior and asymptotic performance of EP to be a mean squared error (MSE) of about -35 dB. Each "step" in the "staircase"-like part of the graph represents an entire sleep cycle for the network, so we see that in this instance the EP algorithm converges to this MSE after about eight iterations on average.

6. CONCLUSIONS

We have shown that, in sensor networks utilizing random sleep strategies for energy efficiency, expectation propagation is viable candidate for distributed estimation or detection. As the size of the network grows, the sparsity of the dependence structure espoused by the random sleep strategy allows for good convergence and performance behavior from expectation propagation which can be accurately predicted with the help of extrinsic information transfer (EXIT) charts. We provided several practical examples of distributed estimation or detection tasks in sensor networks which fall within our system model, to which expectation propagation is well applied. Finally, we demonstrated the utility of the EXIT chart technique by applying it to a practical source separation application in a sensor network.

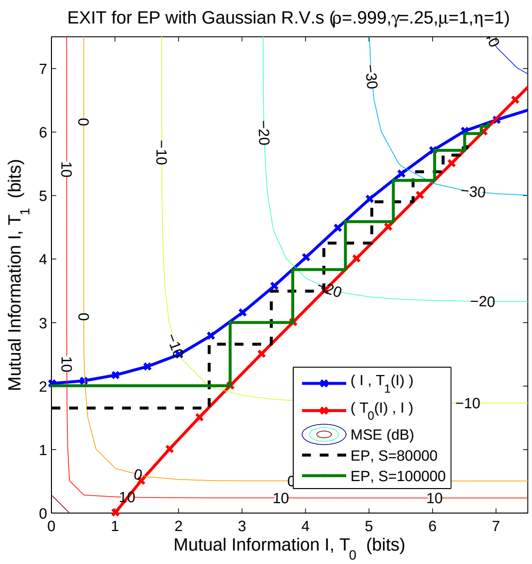

Fig. 2. An example EXIT chart from [17]. The solid and dashed lines are the average results of MC runs of actual EP for a network size S.

7. REFERENCES

[1] S. Boyd, A. Ghosh, B. Prabhakar, and D. Shah, "Randomized Gossip Algorithms," *IEEE Trans. Inform. Theory*, vol. 52, no. 6, pp. 2508–2530, June 2006.

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

[3] M. Çetin, L. Chen, J. W. Fisher III, A. T. Ihler, R. L. Moses, M. J. Wainwright, and A. S. Willsky, "Distributed Fusion in Sensor Networks," *IEEE Signal Processing Mag.*, pp. 42–55, July 2006.

[4] A. T. Ihler, III J. W. Fisher, R. L. Moses, and A. S. Willsky, "Nonparametric Belief Propagation for Sensor Network Self-Calibration," *IEEE J. Select. Areas Commun.*, vol. 23, april 2005.

[5] C. Crick and A. Pfeffer, "Loopy belief propagation as a basis for communication in sensor networks," in *Uncertainty in AI*, San Francisco, CA, 2003, pp. 159–16, Morgan Kaufmann.

[6] E. Aktas and J. Evans and S. Hanly, "Distributed decoding in a cellular multiple-access channel," in *Proc. International Symposium on Information Theory (ISIT)*, June 2004, p. 484.

[7] C. Hsin and M. Liu, ""Randomly Duty-cycled Wireless Sensor Networks: Dynamics of Coverage," *IEEE Trans. Wireless Commun.*, vol. 5, no. 11, pp. 3182–3192, Nov. 2006.

[8] O. Dousse, P. Mannersalo, and P. Thiran, "Latency of wireless sensor networks with uncoordinated power saving mechanisms," in *Mobihoc*, 2004, pp. 109–120.

[9] Lawrence D. Brown, *Fundamentals of Statistical Exponential Families with Applications in Statistical Decision Theory*, Institute of Mathematical Statistics, 1986.

[10] M. J. Wainwright and M. I. Jordan, "Graphical models, exponential families, and variational inference," Tech. Rep., Department of Statistics, University of California, Berkeley, 2003.

[11] S. Amari, *Methods of Information Geometry*, vol. 191, AMS Translations of Mathematical Monographs, 2004.

[12] T. P. Minka, *A Family of Algorithms for Approximate Bayesian Inference*, Ph.D. thesis, Massachusetts Institute of Technology, 2001.

[13] T. P. Minka, "Expectation propagation for approximate Bayesian inference," in *Uncertainty in AI'01*, 2001.

[14] J. M. Walsh, *Distributed Iterative Decoding and Estimation via Expectation Propagation: Performance and Convergence*, Ph.D. thesis, Cornell University, 2006.

[15] 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.

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

[17] J. M. Walsh, "EXIT and Density Evolution Analysis for Homogeneous Expectation Propagation," in *IEEE International Symp. Inform. Theory*, 2007, To Appear.

[18] D. Divsalar, S. Dolinar, and F. Pollara, "Iterative turbo decoder analysis based on density evolution.," *IEEE J. Select. Areas Commun.*, vol. 19, pp. 891–907, May 2001.

[19] H. El Gamal and A. R. Hammons, Jr., "Analyzing the turbo decoder using the gaussian approximation," *IEEE Trans. Inform. Theory*, vol. 47, pp. 671–686, Feb. 2001.

[20] S. ten Brink, "Convergence behavior of iteratively decoded parallel concatenated codes.," *IEEE Trans. Commun.*, vol. 49, pp. 1727–1737, Oct. 2001.

[21] S. ten Brink, G. Kramer, and A. Ashikhmin, "Design of low-density parity-check codes for modulation and detection," *IEEE Trans. Commun.*, pp. 670 – 678, Apr. 2004.