An efficient Monte Carlo approach for optimizing communication constrained decentralized estimation networks

AN EFFICIENT MONTE CARLO APPROACH FOR OPTIMIZING COMMUNICATION CONSTRAINED DECENTRALIZED ESTIMATION NETWORKS

Murat Üney[†‡], *Müjdat Çetin*[†]

[†]Faculty of Engineering and Natural Sciences, Sabancı University
Orhanlı-Tuzla, 34956, İstanbul, Turkey
[‡] Department of Electrical and Electronics Engineering, Middle East Technical University
06531, Ankara, Turkey
phone: + (90) 216 483-9594, fax: + (90) 216 483-9550, email: {muratuney, mcetin}@sabanciuniv.edu

ABSTRACT

We consider the design problem of a decentralized estimation network under communication constraints. The underlying low capacity links are modeled by introducing a directed acyclic graph where each node corresponds to a sensor platform. The operation of the platforms are constrained by the graph such that each node, based on its measurement and incoming messages from parents, produces a local estimate and outgoing messages to children. A Bayesian risk that captures both estimation error penalty and cost of communications, e.g. due to consumption of the limited resource of energy, together with constraining the feasible set of strategies by the graph, yields a rigorous problem definition. We adopt an iterative solution that converges to an optimal strategy in a person-by-person sense previously proposed for decentralized detection networks under a team theoretic investigation. Provided that some reasonable assumptions hold, the solution admits a message passing interpretation exhibiting linear complexity in the number of nodes. However, the corresponding expressions in the estimation setting contain integral operators with no closed form solutions in general. We propose particle representations and approximate computational schemes through Monte Carlo methods in order not to compromise model accuracy and achieve an optimization method which results in an approximation to an optimal strategy for decentralized estimation networks under communication constraints. Through an example, we present a quantification of the trade-off between the estimation accuracy and the cost of communications where the former degrades as the later is increased.

1. INTRODUCTION

The fundamental motivation of distributed estimation under communication constraints is provided by sensor networks (SNs) which are composed of networked platforms with limited sensing, communication, and computation capability operating together to obtain some useful information using the possibly high volume of observations collected at various locations and involving uncertainities. This nature suggests the necessity of some communications to take place over bandwidth (BW) limited links for a reasonable inference performance as well as that of performing the processing in a distributed and collaborative manner.

The canonical approach considering a static estimation task with a performance measure such as mean squared error (MSE) in accordance with the BW limitations is to collect quantized values due to observations at a center node which produces the required estimate. In this so-called star topology setting, the design problem involves finding quantization schemes for peripheral nodes addressing the limited BW and a fusion rule for the center node such that a reasonable estimation accuracy is achieved [1],[2]. Problem settings suitable to sensor networks that differ in the reflected domain knowledge such as the noise distribution and quantization level constraints have been studied as well as the case in which samples of a field are to be estimated (see e.g. [3], [4] and the references therein). Altough these treatments consider keeping the communication demand as low as possible, they are limited in capturing certain aspects of the problem. First of all, the topologies for which results can be produced for are restricted to star-shaped directed graphs. The cost of transmissions from peripherals to the fusion center which could vary considering the multi-hop nature is not explicitly accounted for. In the case of multiple random variables, a computational bottleneck is of concern since inference is performed only at the fusion center. Furthermore, the peripheral nodes do not cooperate with each other and exploit possible correlation structures that the problem might exhibit.

The framework of graphical models has proved to be useful for distributed inference problems arising in various SN applications including the estimation of a random field [5]. In this approach, the so called "information structure" of an inference task is represented with graphs revealing the correlation properties of the problem and inference is performed through message passing algorithms (MPAs) on them. After mapping the information structure onto the set of platforms, some messages correspond to real communications and provided that the required transmissions are supported by the underlying communication structure, a collaborative and distributed processing scheme is achieved. On the other hand, the cost of communications is of concern because application scenarios often involve cases in which no infrastructure can be provided so that the platforms have to rely on stored energy which is primarily consumed by the transmissions. Although it is possible to analyze the effects of errors induced by transmissions to MPAs (see e.g. [6]), it is hard to utilize this framework within a design problem in which the communication constraints are severe and the trade-off between estimation accuracy and cost of communications is explicitly of concern.

We consider the estimation of a random vector that takes values from an N-dimensional Euclidean space through a system with a communication and computation structure that better matches the underlying communication topology and exhibits collaborative processing. This scenario captures, e.g. the estimation of a scalar parameter (as in e.g.[7]) as well as samples of a field (as in e.g.[4]) with an ad-hoc sensor network. Unlike the canonical inference approaches mentioned above, we employ a design perspective in which the cost of communications and estimation errors are considered explicitly in a Bayesian setting as well as the constraints including the availability and capacity of links. Similar challenges are of concern in decentralized detection for which a general treatment has been presented in [8] (see also [9]). In this setting, the available links between sensor platforms render a directed acyclic graph (DAG) $\mathcal{G}=(\mathcal{V},\mathcal{E})$ where nodes and edges correspond to platforms and uni-directional links between two platforms respectively. The inference task is distributed through associating random variables with sensor platforms. Each node evaluates its local rule, given the incoming messages and

THIS WORK WAS PARTIALLY SUPPORTED BY THE SCIENTIFIC AND TECHNOLOGICAL RESEARCH COUNCIL OF TURKEY UNDER GRANT 105E090, BY THE EUROPEAN COMMISSION UNDER GRANT MIRG-CT-2006-041919 AND WITH A TURKISH ACADEMY OF SCIENCES YOUNG SCIENTIST AWARD. THE AUTHORS WOULD LIKE TO THANK O. PATRICK KREIDL FOR HIS HELP AND SUPPORT THROUGHOUT MANY DISCUSSIONS.

its own measurement, producing an inference on the associated random variable(s) and outgoing messages. The design problem involves finding the set of local rules, which is referred to as the strategy, that minimizes an expected cost which captures contributions of both cost of communications and detection errors in a Bayesian setting with the set of feasible strategies constrained by $\mathcal{G}$. Decentralized detection is NP-hard in general, nevertheless necessary (but not sufficient) optimality conditions yield nonlinear Gauss-Seidel iterations which converge to a person-by-person optimal strategy [10]. In [8], this treatment is utilized for a directed acyclic topology and an iterative solution together with conditions under which the iterations admit a message passing interpretation that is scalable with the number of nodes are established.

We generalize this framework to decentralized estimation (DE), and address some of the limitations of the canonical distributed estimation algorithms mentioned above [1]-[3],[7]. However this approach leads to an iterative scheme that involves integral equations that have no closed form solutions in general. In order not to compromise model accuracy, we develop an approximation framework using Monte Carlo (MC) integration methods. In the resulting network, the platforms perform computations which correspond to approximations to an approximately person-by-person optimal rule. We maintain the scalability of the solution both in the number of nodes and sample sizes and we can produce results for any set of distributions as long as samples can be generated from them. Hence our main contribution is an efficient MC optimization algorithm for DE networks subject to communication constraints in a Bayesian setting. The algorithm can be carried out in a message passing fashion making it also suitable for self-organization.

2. DESIGNING DE NETWORKS

In this section we present the online structure we consider for processing the measurements collected by the platforms which is described with a DAG $\mathcal{G}$. Then we define the problem of decentralized estimation under communication constraints in Section 2.2, and present a team theoretic iterative solution in Section 2.3.

2.1 Online Processing Constrained With a DAG

A DAG $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ represents a communication and computation structure for a decentralized system where each platform is associated with a node $v \in \mathcal{V}$. An edge $(i, j) \in \mathcal{E}$ corresponds to the finite capacity communication link from platform i to j on which i can transmit a symbol $u_{i\to j}$ from a finite set $\mathcal{U}_{i\to j}$ where the number of elements $\left|\mathcal{U}_{i\to j}\right|$ is in accordance with the link capacity capturing the BW constraints. DAGs imply a partial ordering and without loss of generality, we assume that the nodes are labeled in accordance with a partial ordering using the reachability relation in which the parentless nodes have the smallest order.

Let $u_{\pi(j)}$ denote the incoming messages to node j from the parent nodes $\pi(j)$, given by $u_{\pi(j)} \triangleq \{u_{i\to j} | i \in \pi(j)\}$. Let $\mathcal{U}_{\pi(j)}$ denote the set from which $u_{\pi(j)}$ takes values from. This set is constructed through consecutive Cartesian products given by $\mathcal{U}_{\pi(j)} = \mathcal{U}_{\pi_1\to j} \times \ldots \times \mathcal{U}_{\pi_{P_j}\to j}$ where $\pi(j) = \{\pi_1, ..., \pi_{P_j}\}$ and $P_j = |\pi(j)|$. The set of outgoing messages from node j to child nodes $\chi(j)$, given by $u_j \triangleq \{u_{j\to k} | k \in \chi(j)\}$ takes values from the set $\mathcal{U}_j$ which can be defined in a similar way with that for $\mathcal{U}_{\pi(j)}$.

Each node j is associated with a random variable(s) X_j that takes values from the set $\mathcal{X}_j$.

The directed acyclic nature of $\mathcal{G}$ leads a causal online processing of observations when proceeded in accordance with the partial order. Starting from the parentless nodes, as node j measures $y_j \in \mathcal{Y}_j$ and receives $u_{\pi(j)} \in \mathcal{U}_j$, it evaluates a function, called its local rule and defined by $\gamma_j : \mathcal{Y}_j \times \mathcal{U}_{\pi(j)} \to \mathcal{U}_j \times \mathcal{X}_j$, which produces an estimate $\hat{x}_j \in \mathcal{X}_j$ as well as outgoing messages $u_j \in \mathcal{U}_j$ (Fig.(1a)). The space of rules local to node j is given by $\Gamma^{\mathcal{G}}_j \triangleq \{\gamma_j | \gamma_j : \mathcal{Y}_j \times \mathcal{U}_{\pi(j)} \to \mathcal{U}_j \times \mathcal{X}_j\}$ where the superscript $\mathcal{G}$ denotes that the definition relies on $\mathcal{G}$ together with the set $\{\mathcal{U}_{i\to j} | (i, j) \in \mathcal{E}\}$. In Section 4, we provide an example in which the online processing of a DE network is described by the DAG in Fig.(2a).

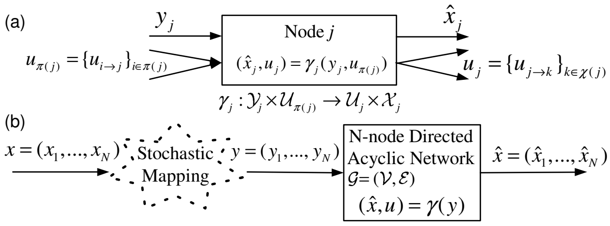

Figure 1: Online processing from (a) the viewpoint of node j, (b) the global viewpoint.

The aggregation of local rules $\gamma = (\gamma_1, \gamma_2, \ldots, \gamma_N)$ is called a strategy and takes values from the set of feasible strategies given by $\Gamma^{\mathcal{G}} = \Gamma^{\mathcal{G}}_1 \times \ldots \times \Gamma^{\mathcal{G}}_N$. The communication load of the system is the set of all transmitted symbols $u \triangleq \{u_{i\to j} | (i, j) \in \mathcal{E}\}$ and takes values from the set $\mathcal{U}$ which can be defined in a similar way with that for $\mathcal{U}_{\pi(j)}$. Similarly X takes values from the set $\mathcal{X} = \mathcal{X}_1 \times \ldots \times \mathcal{X}_N$. This global view is presented in Figure(1b).

2.2 Problem Definition

Since the online processing strategy γ is a function mapping Y to $\hat{X}$ and U, i.e. $(U, \hat{X}) = \gamma(Y)$, $(U, \hat{X}, X)$ is a random process with the joint density $p(u, \hat{x}, x; \gamma) = \int_{y\in\mathcal{Y}} \mathrm{d}y \;\; p(u, \hat{x} | x, y; \gamma) p(x, y)$ where "$;\gamma$" denotes that the distribution is specified by γ. The causal operation implies the coupling of local rules to $p(u, \hat{x}, x; \gamma)$ through $p(u, \hat{x} | x, y; \gamma) = \prod_{j=1}^{N} p(u_j, \hat{x}_j | y_j, u_{\pi(j)}; \gamma_j)$ where it is convenient to treat $p(u_j, \hat{x}_j | y_j, u_{\pi(j)}; \gamma_j)$ as a finite set of distributions parameterized on u_j and specified by γ_j as $p(u_j, \hat{x}_j | y_j, u_{\pi(j)}; \gamma_j) = p_{u_j}(\hat{x}_j | y_j, u_{\pi(j)}; \gamma_j)$ such that

$$p_{[\gamma_j(y_j, u_{\pi(j)})]_{\mathcal{U}_j}}(\hat{x}_j | y_j, u_{\pi(j)}; \gamma_j) = \delta(\hat{x}_j - \left[\gamma_j(y_j, u_{\pi(j)})\right]_{\mathcal{X}_j}) \tag{1}$$

where $[.]_{\mathcal{A}}$ selects the component of its N-tuple argument that takes values from $\mathcal{A}$, e.g. $[(\bar{u}_j, \hat{\bar{x}}_j)]_{\mathcal{U}_j} = \bar{u}_j$ and $[(\bar{u}_j, \hat{\bar{x}}_j)]_{\mathcal{X}_j} = \hat{\bar{x}}_j$. Therefore, γ_j and the distribution family $p_{u_j}(\hat{x}_j | y_j, u_{\pi(j)}; \gamma_j)$ specify each other accordingly. Moreover, the joint distribution $p(u, \hat{x}, x; \gamma)$ is constructed through these distributions.

Having built the joint distribution determined by γ, in the Bayesian setting, a cost function is selected such that an estimation error penalty for the pair $(x, \hat{x})$ and a cost due to the corresponding communication load u are assigned. Hence, in general $c : \mathcal{U} \times \mathcal{X} \times \mathcal{X} \to \mathbb{R}$ and there corresponds an objective value $J(\gamma) = E\{c(u, x, \hat{x}); \gamma\}$ where the expectation is over $p(u, \hat{x}, x; \gamma)$ for any selection of $\gamma \in \Gamma^{\mathcal{G}}$ following the discussion above. Therefore, the problem of finding the best strategy for estimation under communication constraints described by $\mathcal{G}$ and c turns to a constrained optimization problem given by

$$(\mathrm{P}): \qquad \min E\{c(u, x, \hat{x}); \gamma\} \qquad \text{subject to } \gamma \in \Gamma^{\mathcal{G}} \tag{2}$$

Problem (P) allows a broad range of settings to be expressed including the conventional star-topology treatment of DE networks. It is hard to find a global optimum in general, nevertheless the team theoretic approach which we present in the next section has proved to be useful in solving decentralized detection problems.

2.3 Team Theoretic Iterative Solution

In a team decision problem, N members taking actions $\gamma_j \in \Gamma_j$ with a single cost function $J(\gamma_1, \gamma_2, ..., \gamma_N)$ constitute a team. When it is hard to find $\gamma^*_j \in \Gamma_j$ for $j = 1, 2, ..., N$ such that $J(\gamma^*_1, \gamma^*_2, ..., \gamma^*_N)$ is minimum, a useful relaxation is to seek a Nash equilibrium satisfying

$$\gamma^*_j = arg \min_{\gamma_j \in \Gamma_j} J(\gamma_j, \gamma^*_{\backslash j}) \tag{3}$$

for all $j \in \{1, 2, ..., N\}$ where $\backslash j = \{1, 2, ..., N\} \backslash \{j\}$. Such a solution is also refered to be person by person (pbp) optimal [11]. It can easily be shown that Algorithm 1, starting initially from $\gamma^0 = (\gamma^0_1, ..., \gamma^0_N)$ where $\gamma^j \in \Gamma_j$ for $j \in \{1, 2, ..., N\}$ converges to a pbp optimal strategy.

For the Problem (P) given by Expression (2), provided that certain assumptions hold the optimality condition in Eq.(3) bears a structure such that the update step of Algorithm 1 scales with the number of nodes also admitting a message passing interpretation.

Algorithm 1 *Iterations converging to a pbp optimal strategy.*
0) (Initiate) $l = 0$, choose $\gamma^0 \in \Gamma$ where $\Gamma = \Gamma_1 \times ... \times \Gamma_N$;
1) (Update) $l = l + 1$;
For $j=1,...,N$, $\gamma_j^l = \arg\min_{\gamma_j \in \Gamma_j} J(\gamma_1^l, ..., \gamma_{j-1}^l, \gamma_j, \gamma_{j+1}^{l-1}, ..., \gamma_N^{l-1})$;
2) (Check) If $J(\gamma^{l-1}) - J(\gamma^l) < \varepsilon$, $\gamma^* = \gamma^l$, STOP; else GO TO 1;

Moreover, a scalable online processing is implied as well. Consider the following assumptions:
Assumption 1 (Conditional Independence): Sensor noise processes are mutually independent resulting $p(x, y) = p(x) \prod_{i=1}^N p(y_i|x)$.
Assumption 2 (Measurement Locality): We assume that, every node j observes y_j due to only x_j, i.e. $p(y_j|x) = p(y_j|x_j)$.
Assumption 3 (Cost Locality): The Bayesian cost function is additive over the nodes, i.e. $c(u, \hat{x}, x) = \sum_{j\in\mathcal{V}} c_j(u_j, \hat{x}_j, x_j)$.
Assumption 4 (Polytree Topology): $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ is a polytree, i.e. $\mathcal{G}$ is a DAG with an acyclic undirected counterpart.

Proposition 1: For Problem (P), under Assumptions 1-4, given a pbp optimal strategy $\gamma^* = (\gamma_1^*, ...\gamma_N^*)$ and fixing all local rules other than the j^{th}, i.e. $\gamma_{\setminus j} = \gamma_{\setminus j}^*$, the j^{th} optimal local rule given by Eq.(3) reduces to

$$\gamma_j^*(y_j, u_{\pi(j)}) = \underset{(u_j,\hat{x}_j)\in(\mathcal{X}_j,\mathcal{U}_j)}{\arg\min} \int_{\mathcal{X}_j} dx_j \, p(y_j|x_j)\phi_j^*(u_j, \hat{x}_j, x_j; u_{\pi(j)}) \qquad (4)$$

with probability 1 such that

$$\phi_j^*(u_j, \hat{x}_j, x_j; u_{\pi(j)}) \propto p(x_j)P_j^*(u_{\pi(j)}|x_j)\left[c_j(u_j, \hat{x}_j, x_j) + C_j^*(u_j, x_j)\right] \qquad (5)$$

holds where $P_j^*(u_{\pi(j)}|x_j) = 1$, if $\pi(j) = \emptyset$, and otherwise

$$P_j^*(u_{\pi(j)}|x_j) = \int_{\mathcal{X}_{\pi(j)}} dx_{\pi(j)} p(x_{\pi(j)}|x_j) \prod_{i\in\pi(j)} P_{i\to j}^*(u_{i\to j}|x_i) \qquad (6)$$

with terms regarding influence of $i \in \pi(j)$ on j given by

$$P_{i\to j}^*(u_{i\to j}|x_i) = \sum_{u_{\chi(i)\setminus j}\in\mathcal{U}_{\chi(i)\setminus j}} \sum_{u_{\pi(i)}\in\mathcal{U}_{\pi(i)}} P_i^*(u_{\pi(i)}|x_i)p(u_i|x_i, u_{\pi(i)}; \gamma_i^*) \qquad (7)$$

where

$$p(u_i|x_i, u_{\pi(i)}; \gamma_i^*) = \int_{\mathcal{X}_i} \mathrm{d}\hat{x}_i \int_{\mathcal{Y}_i} \mathrm{d}y_i \, p(u_i, \hat{x}_i|y_i, u_{\pi(i)}; \gamma_i^*)p(y_i|x_i) \qquad (8)$$

and

$$C_j^*(u_j, x_j) = \begin{cases} 0 & \text{, if } \chi(j) = \emptyset \\ \sum_{k\in\chi(j)} C_{k\to j}^*(u_{j\to k}, x_j) & \text{, if } \chi(j) \neq \emptyset \end{cases} \qquad (9)$$

with terms regarding the influence of j on $k \in \chi(j)$ given by

$$C_{k\to j}^*(u_{j\to k}, x_j) = \int_{\mathcal{X}_{\pi(k)\setminus j}} dx_{\pi(k)\setminus j} \int_{\mathcal{X}_k} dx_k \, p(x_{\pi(k)\setminus j}, x_k|x_j) \times \sum_{u_{\pi(k)\setminus j}} \prod_{j'\in\pi(k)\setminus j} P_{j'\to k}^*(u_{j'\to k}|x_{j'}) I_k^*(u_{\pi(k)}, x_k; \gamma_k^*) \qquad (10)$$

such that

$$I_k^*(u_{\pi(k)}, x_k; \gamma_k^*) = \int_{\mathcal{Y}_k} \mathrm{d}y_k \int_{\mathcal{X}_k} \mathrm{d}\hat{x}_k \sum_{u_k\in\mathcal{U}_k} [c_k(u_k, \hat{x}_k, x_k) + C_k^*(u_k, x_k)] \, p(u_k, \hat{x}_k|y_k, u_{\pi(k)}; \gamma_k^*)p(y_k|x_k) \qquad (11)$$

Proof: Due to lack of space the proof is not provided here but this proposition in the detection case[1] is proved in [9], and it is possible to obtain the above expressions from this version by replacing summations over $\mathcal{X}_j$s with integrations and changing the order of summations and integrations appropriately and also assuming that the links are errorneous.

Given a pbp optimal strategy γ^*, Proposition (1) expresses the j^{th} local rule in a variational form determined by ϕ_j which bears the influence of the rules local to the ancestors of node j as well as that of its descendants. Considering Eq.s(6) and (7) we note that $P_j^*(u_{\pi(j)}|x_j)$ is the likelihood of x_j given the incoming messages from parents, i.e. $u_{\pi(j)}$, and depends on the local rules of all ancestors. A similar treatment of Eq.s(9)-(11) reveals that $C_{k\to j}^*(u_{j\to k}, x_j)$ is the expected cost induced on the descendants of j on the branch starting with k if X_j actually takes the value x_j and node j sends $u_{j\to k}$ to node k. Hence $C_j^*(u_j, x_j)$ is the total expected cost induced on the descendants with u_j. In Eq.(5), this term is added to $c_j(u_j, \hat{x}_j, x_j)$ which reflects the cost of u_j. Hence, considering Eq.(4) with Expression (5) and realizing that under these assumptions $p(x_j)p(y_j|x_j)P(u_{\pi(j)}|x_j) \propto p(x_j|y_j, u_{\pi(j)})$, we conclude that the pbp optimal local rule for node j is to choose the communication and estimation variables with the minimum total expected cost, given the measurement $y_j \in \mathcal{Y}_j$ and incoming messages $u_{\pi(j)} \in \mathcal{U}_{\pi(j)}$.

It is possible to treat Eq.s(6)-(11) as operators that are valid for any choice of $\gamma_{\setminus j}$ (not necessarily the optimal one). Therefore, the "update" step in Algorithm 1 that determines γ_j^l can be expressed in terms of these operators by substituting $\gamma_1^l, ..., \gamma_{j-1}^l, ..., \gamma_{j+1}^{l-1}, \gamma_N^{l-1}$ instead of their optimal values. Then, utilizing Expression (5) together with Eq.(4) the corresponding local rule for node j is achieved. Using this perspective for specifying Algorithm 1 for Problem (P) yields Algorithm 2. The cost required in the "check" step, i.e. $J(\gamma^l)$, which is the expected risk for the strategy achieved at the l^{th} iteration is given by $J(\gamma^l) = \sum_{j\in\mathcal{V}} G_j(\gamma_j^l)$ with

$$G_j(\gamma_j^l) = \int_{\mathcal{X}_j} \mathrm{d}x_j p(x_j) \sum_{u_{\pi(j)}\in\mathcal{U}_{\pi(j)}} P_j^{l+1}(u_{\pi(j)}|x_j) \times \int_{\mathcal{Y}_j} \mathrm{d}y_j \int_{\mathcal{X}_j} \mathrm{d}\hat{x}_j \sum_{u_j\in\mathcal{U}_j} c_j(u_j, \hat{x}_j, x_j)p(u_j, \hat{x}_j|y_j, u_{\pi(j)}; \gamma_j^l)p(y_j|x_j) \qquad (12)$$

The update step of Algorithm 2 admits a message passing interpretation where in the first pass, starting from parentless nodes, $P_{i\to j}^l$s are computed and sent to child nodes $j \in \chi(i)$ for all $i \in \mathcal{V}$ in accordance with the implied ordering. Once all nodes have obtained incoming message likelihoods, in the second pass, starting from childless nodes, $C_{k\to j}^l$s are computed and sent to parent nodes $j \in \pi(k)$ for all $k \in \mathcal{V}$ in accordance with the implied backward ordering. Local rules are updated as soon as both terms are received.

3. MONTE CARLO APPROXIMATED ITERATIONS

In principle, Algorithm 2 provides a solution to Problem (P) given the distributions $p(x_{\pi(j)}, x_j)$ $\forall j \in \mathcal{V}$. However, the "update" step involves integral operators with no closed form solutions in general so we propose particle representations and approximate computation schemes through MC integration methods.

3.1 Monte Carlo Integration

In the conventional MC method, we consider $i = \int_{\mathcal{X}} dx \, p(x)f(x)$ where $p(x)$ is a probability density for X which takes values from $\mathcal{X}$. Given M independent samples $\{x^{(m)}\}_{m=1}^M$ generated from $p(x)$, i.e. $x^{(m)} \sim p(x)$ for $m = 1, ..., M$, an estimate for i is given by $\hat{i}_M = \frac{1}{M}\sum_{k=1}^M f(x^{(k)})$ which converges to i almost surely. The Importance Sampling (IS) method is used if it is not possible to sample from $p(x)$ but from $g(x)$. Given $x^{(m)} \sim g(x)$ for $m = 1, ..., M$, the estimate $\hat{i}_M = \frac{1}{M}\sum_{k=1}^M \omega_{(k)} f(x^{(k)})$ where $\omega_{(k)} = p(x^{(k)})/g(x^{(k)})$ converges to i almost surely provided that the support of g is covered by the support of f. For cases where a small number of weights dominate the rest,

$$\hat{i}_M = \frac{1}{\sum_{k=1}^M \omega_{(k)}} \sum_{k=1}^M \omega_{(k)} f(x^{(k)}) \qquad (13)$$

is preferable altough it is slightly biased for small M [12].

3.2 Iterative MC Optimization Scheme

We utilize the MC methods presented in Section 3.1 in order to achieve a practically applicable version of Algorithm 2. Consider

Algorithm 2: *Iterations converging to a person by person optimal decentralized estimation strategy.*
0) (Initiate) $l = 0$, choose $\gamma^0 \in \Gamma^{\mathcal{G}}$;
1) (Update) $l = l + 1$;
For $j = 1, ..., N$
Using $\{P_{i\to j}^l(u_{i\to j}|x_i)\}_{i\in\pi(j)}$, Compute $\{P_{j\to k}^l(u_{j\to k}|x_j)\}_{k\in\chi(j)}$;
For $j = N, ..., 1$
Using $\{P_{i\to j}^l(u_{i\to j}|x_i)\}_{i\in\pi(j)}$ and $\{C_{k\to j}^l(u_{j\to k}, x_j)\}_{k\in\chi(j)}$
i) Update γ_j^l through ϕ_j^l
ii) Compute $\{C_{j\to i}^l(u_{i\to j}, x_i)\}_{i\in\pi(j)}$;
2) (Check) If $J(\gamma^{l-1}) - J(\gamma^l) < \epsilon$, $\gamma^* = \gamma^l$, STOP; else GO TO (1);

[1] For the detection problem $|\mathcal{X}_j| < \infty$ holds for all $j \in \mathcal{V}$ whereas this condition is not satisfied in the estimation setting.

dering Proposition (1), we perform MC approximations to γ_j^* keeping the rest fixed at their optimum. We proceed in three steps starting with the objective function of the variational form given in Eq.(4) together with Expression (5). Then we consider the components of this approximation which couple node-to-node terms and are given by Eq.s(6) and (9). Finally we approximate the node-to-node terms at the required sample points. As we express all the relevant computations utilizing particle representations and approximation schemes, we achieve an approximate pbp optimal local rule for node j, i.e. $\tilde{\gamma}_j^* \approx \gamma_j^*$. While applying MC methods we consider having the samples required to be generated independently from the marginal distributions of X and Y. This provides simplicity for application since we can avoid subtleties of sampling from a joint distribution using, e.g. Gibbs Sampling.

Step 1 Considering the objective of the variational form in Eq.(4) with Expression (5), an approximation through the classical MC method yields $(1/M)\sum_{m=1}^{M} p(y_j|x_j^{(m)}) P_j^*(u_{\pi(j)}|x_j^{(m)})[c_j(u_j,\hat{x}_j,x_j^{(m)}) + C_j^*(u_j,x_j^{(m)})]$ where $x_j^{(m)} \sim p(x_j)$ for $m = 1,2,...,M$. A one step approximate pbp optimal rule is achieved by replacing the objective function in Eq.(4) with this summation. The terms regarding the local likelihood and cost are known and given in the problem model. We approximate to the other required terms in the next step.

Step 2 $\{P_j^*(u_{\pi(j)}|x_j^{(m)})\}_{m=1}^M$ $\forall u_{\pi(j)} \in \mathcal{U}_{\pi(j)}$ and $\{C_j^*(u_j,x_j^{(m)})\}_{m=1}^M$ $\forall u_j \in \mathcal{U}_j$ are of concern. We consider Eq.(6) and assume that $\{P_{i\to j}^*(u_{i\to j}|x_i^{(m)})\}_{m=1}^M$ is known $\forall i \in \pi(j)$, $\forall u_{i\to j} \in \mathcal{U}_{i\to j}$ and $x_i^{(m)} \sim p(x_i)$ for $m = 1,2,...,M$. We note that exactly these samples are required to apply Step (1) for these nodes, i.e. for $i \in \pi(j)$. Also $x_{\pi(j)}^{(m)} = (x_i^{(m)})_{i\in\pi(j)} \sim \prod_{i\in\pi(j)} p(x_i)$ holds and hence, it is possible to apply the IS approximation given in Eq.(13) using weights $\omega_j^{(m)(m')} = p(x_{\pi(j)}^{(m')}|x_j^{(m)}) / \prod_{i\in\pi(j)} p(x_i^{(m')})$ and obtain

$$\tilde{P}_j^*(u_{\pi(j)}|x_j^{(m)}) = \frac{1}{\sum_{m'=1}^{M}\omega_j^{(m)(m')}} \sum_{m'=1}^{M} \omega_j^{(m)(m')} \prod_{i\in\pi(j)} P_{i\to j}^*(u_{i\to j}|x_i^{(m')}) \quad (14)$$

for $m = 1,2,...,M$ and $\forall u_{\pi(j)} \in \mathcal{U}_{\pi(j)}$. Similarly, consider Eq.(9) for the case where $\chi(j) \neq \emptyset$. Assuming that $\{C_{k\to j}^*(u_{j\to k},x_j^{(m)})\}_{m=1}^M$ is known $\forall k \in \chi(j)$ and $\forall u_{j\to k} \in \mathcal{U}_{j\to k}$, $\{C_j^*(u_j,x_j^{(m)})\}_{m=1}^M$ can be found without need for any approximation. In the next step, we approximate the node-to-node terms which are assumed to be known.

Step 3 We approximate to node-to-node terms that are required in Eq.(14) and for $\{C_j^*(u_j,x_j^{(m)})\}_{m=1}^M$ $\forall u_j \in \mathcal{U}_j$. Consider Eq.s(7) and (8) for any parent node $i \in \pi(j)$ and assume that $\{P_i^*(u_{\pi(i)}|x_i^{(m)})\}_{m=1}^M$ is known $\forall u_{\pi(i)} \in \mathcal{U}_{\pi(i)}$ where $x_i^{(m)} \sim p(x_i)$ for $m = 1,2,...,M$. Then, in order to evaluate Eq.(7) $\forall u_{i\to j} \in \mathcal{U}_{i\to j}$ and $x_i = x_i^{(m)}$ for $m = 1,2,...,M$ we need to evaluate Eq.(8) accordingly, i.e. $p(u_i|x_i^{(m)},u_{\pi(i)};\gamma_i^*)$. Also considering Eq.(1), it is possible to apply IS with $y_i^{(p)} \sim p(y_i)$ for $p = 1,2,...,P$ and obtain

$$\tilde{p}(u_i|x_i^{(m)},u_{\pi(i)};\gamma_i^*) = \frac{1}{\sum_{p=1}^{P}\omega_i^{(m)(p)}} \sum_{p=1}^{P} \omega_i^{(m)(p)} \delta_{u_i,[\gamma_i^*(y_i^{(p)},u_{\pi(i)})]_{\mathcal{U}_j}} \quad (15)$$

with $\omega_i^{(m)(p)} = p(y_i^{(p)}|x_i^{(m)})/p(y_i^{(p)})$. After substituting Eq.(15) in Eq.(7), we obtain $\tilde{P}_{i\to j}^*(u_{i\to j}|x_i^{(m)})$ and substituting them in Eq.(14) we obtain a two step approximation to $\{P_j^*(u_{\pi(j)}|x_j^{(m)})\}_{m=1}^M$.

Secondly, let us consider Eq.s(10) and (11). Substituting Eq.(1) in Eq.(11) yields

$$I^*(u_{\pi(k)},x_k;\gamma_k^*) = \int_{\mathcal{Y}_k} dy_k [\, c_k([\gamma_k^*(y_k,u_{\pi(k)})]_{\mathcal{U}_k}, [\gamma_k^*(y_k,u_{\pi(k)})]_{\mathcal{X}_k}, x_k) + C_k^*([\gamma_k^*(y_k,u_{\pi(k)})]_{\mathcal{U}_k}, x_k)\,] p(y_k|x_k)$$

and an IS approximation is immediately found assuming that $\{C_k^*(u_k,x_k^{(m)})\}_{m=1}^M$ is known for all $u_k \in \mathcal{U}_k$ and $y_k^{(p)} \sim p(y_k)$ for $p = 1,2,...,P$ by $\tilde{I}^*(u_{\pi(k)},x_k^{(m)};\gamma_k^*) =$

$$\frac{1}{M_k^{(m)}} \sum_{p=1}^{P} \omega_k^{(m)(p)} [\, c_k([\gamma_k^*(y_k^{(p)},u_{\pi(k)})]_{\mathcal{U}_k}, [\gamma_k^*(y_k^{(p)},u_{\pi(k)})]_{\mathcal{X}_k}, x_k^{(m)}) + C_k^*([\gamma_k^*(y_k^{(p)},u_{\pi(k)})]_{\mathcal{U}_k}, x_k^{(m)})\,]$$

where $\omega_k^{(m)(p)} = p(y_k^{(p)}|x_k^{(m)})/p(y_k^{(p)})$ and $M_k^{(m)} = \sum_{p=1}^{P}\omega_k^{(m)(p)}$ for $m = 1,2,...,M$ and $\forall u_{\pi(k)} \in \mathcal{U}_{\pi(k)}$.

Having obtained an approximate evaluation of Eq.(11) we consider Eq.(10) and note that likelihood terms for $j' \in \pi(k)\backslash j$ are required. Similar to the reasoning in Step 2, we assume that $\{P_{j'\to k}^*(u_{j'\to k}|x_{j'}^{(m)})\}_{m=1}^M$ are known $\forall u_{j'\to k} \in \mathcal{U}_{j'\to k}$ and $x_{j'}^{(m)} \sim p(x_{j'})$. Having $\{\tilde{I}^*(u_{\pi(k)},x_k^{(m)};\gamma_k^*)\}_{m=1}^M$ $\forall u_{\pi(k)} \in \mathcal{U}_{\pi(k)}$ and realizing that $x_{\pi(k)\backslash j}^{(m)} \sim \prod_{j'\in\pi(k)\backslash j} p(x_{j'})$ where $x_{\pi(k)\backslash j}^{(m)} = (x_{j'}^{(m)})_{j'\in\pi(k)\backslash j}$ IS weigths given by $\omega^{(m)(m')} = p(x_{\pi(k)\backslash j}^{(m')}, x_k^{(m')}|x_j^{(m)}) / p(x_k^{(m')}) \prod_{j'\in\pi(k)\backslash j} p(x_{j'}^{(m')})$ yield

$$\tilde{C}_{k\to j}^*(u_{j\to k},x_j^{(m)}) = \frac{1}{\sum_{m'=1}^{M}\omega^{(m)(m')}} \sum_{m'=1}^{M} \omega^{(m)(m')} \times \sum_{u_{\pi(k)\backslash j}} \prod_{j'\in\pi(k)\backslash j} P_{j'\to k}^*(u_{j'\to k}|x_{j'}^{(m')}) \tilde{I}^*(u_{\pi(k)},x_k^{(m')};\gamma_k^*)$$

for $m = 1,2,...,M$ and $\forall u_{j\to k} \in \mathcal{U}_{j\to k}$. After substituting these values in Eq.(9) we obtain an approximation to $\{C_j^*(u_j,x_j^{(m)})\}_{m=1}^M$. Utilizing all the above approximations in Step (1), we obtain an approximate pbp optimal local rule for node j denoted by $\tilde{\gamma}_j^* \approx \gamma_j^*$.

The MC framework provided through Steps (1)-(3) yields a MC optimization method given by Algorithm 3 in a similar way that Proposition (1) yields Algorithm 2. Starting from an initial strategy and applying Steps 1-3 for the local rules of all of the nodes, we obtain Algorithm 3 which corresponds to substituting the particle representations and the approximate computational schemes in Algorithm 2. We have provided scalability in the number of samples through employing IS such that sample sets from conditionals or joint distributions are not required. As soon as $\{\{x_j^{(m)}\}_{m=1}^M\}_{j=1}^N$ where $x_j^{(m)} \sim p(x_j)$ and $\{\{y_j^{(p)}\}_{p=1}^P\}_{j=1}^N$ where $y_j^{(p)} \sim p(y_j)$ are obtained and an initial strategy γ^0 is chosen, the iterations converge to a strategy which performs computations corresponding to an approximation to a pbp optimal one, i.e. $\tilde{\gamma}^* \approx \gamma^*$. Finally we similarly find $\tilde{J}(\tilde{\gamma}^l) \approx J(\gamma^l)$ by approximating to Eq.(12) using $\omega_k^{(m)(p)} = p(y_k^{(p)}|x_k^{(m)})/p(y_k^{(p)})$ with

$$\tilde{G}_j(\tilde{\gamma}_j^l) = \frac{1}{M} \sum_{m=1}^{M} \sum_{u_{\pi(j)}\in\mathcal{U}_{\pi(j)}} \tilde{P}_j^{l+1}(u_{\pi(j)}|x_j^{(m)}) \frac{1}{\sum_{p=1}^{P}\omega_k^{(m)(p)}} \sum_{p=1}^{P}\omega_k^{(m)(p)} \times c_j([\gamma_j(y_j^{(p)},u_{\pi(j)})]_{\mathcal{U}_j}, [\gamma_j(y_j^{(p)},u_{\pi(j)})]_{\mathcal{X}_j}, x_j^{(m)})$$

4. EXAMPLE

In this section, we consider an example scenario in which a DE network comprised of four platforms perform an estimation task. A random field $X = \{X_1,X_2,X_3,X_4\}$ is of concern and platform j is associated with X_j. We assume the underlying communication structure described by $\mathcal{G} = (\mathcal{V},\mathcal{E})$ in Figure (2a). We note that $\mathcal{G}$ includes partitions of a star topology (induced by nodes $\{1,2,3\}$), and series topologies (induced by nodes $\{1,3,4\}$ and $\{2,3,4\}$. We assume that the BW constraints render $\mathcal{U}_{1\to3} = \mathcal{U}_{2\to3} = \mathcal{U}_{3\to4} = \{0,1,2\}$. The online processing scheme operates as given in Section 2.1. Since nodes 1 and 2 are parentless, upon measuring y_1 and $y_2 \in \mathbb{R}$ induced by X_1 and X_2, they evaluate their local rules as $(u_{1\to3},\hat{x}_1) = \gamma_1(y_1)$ and $(u_{2\to3},\hat{x}_2) = \gamma(y_2)$ respectively. Upon receiving these messages and measuring $y_3 \in \mathbb{R}$ induced by X_3 node 3 evaluates its local rule as $(u_{3\to4},\hat{x}_3) = \gamma_3(y_3,u_{1\to3},u_{2\to3})$ and similarly node 4 evaluates $\hat{x}_4 = \gamma_4(y_4,u_{3\to4})$. The strategy $\gamma = (\gamma_1,...,\gamma_4)$ is subject to design and we utilize Algorithm 3.

Algorithm 3: *Iterations converging to a MC approximate person by person optimal decentralized strategy.*

0) (Initiate) $l = 0$; *choose* $\gamma^0 \in \Gamma^{\mathcal{G}}$;
1) (Update) $l = l + 1$;
For $j = 1,...,N$
Using $\{\{\tilde{P}_{i\to j}^l(u_{i\to j}|x_i^{(m)})\}_{m=1}^M\}_{i\in\pi(j)}$, *compute* $\{\{\tilde{P}_{j\to k}^l(u_{j\to k}|x_j^{(m)})\}_{m=1}^M\}_{k\in\chi(j)}$;
For $j = N,...,1$
Using $\{\{\tilde{P}_{i\to j}^l(u_{i\to j}|x_i^{(m)})\}_{m=1}^M\}_{i\in\pi(j)}$ *and* $\{\{\tilde{C}_{k\to j}^l(u_{j\to k},x_j^{(m)})\}_{m=1}^M\}_{k\in\chi(j)}$
i) Update $\tilde{\gamma}_j^l$;
ii) Compute $\{\{\tilde{C}_{j\to i}^l(u_{i\to j},x_i^{(m)})\}_{m=1}^M\}_{i\in\pi(j)}$;
3) (Check)If $|\tilde{J}(\tilde{\gamma}^{l-2}) - \tilde{J}(\tilde{\gamma}^{l-1})| - |J(\tilde{\gamma}^{l-1}) - \tilde{J}(\tilde{\gamma}^l)| > \varepsilon$ *GO TO (1)*;
else $\tilde{\gamma}^* = \tilde{\gamma}^l$, *STOP*;

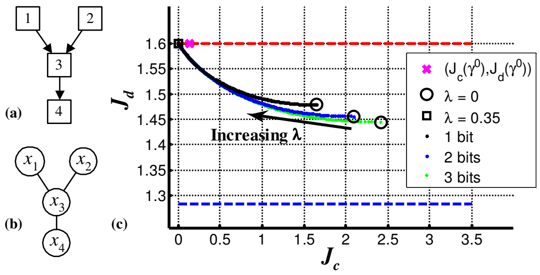

Figure 2: (a) Communication topology, (b) MRF representation of X, (c) Approximate points of the performance curves while λ is increased from 0, for the example scenario.

The cost function local to node j is in a separable form given by $c_j(u_j, \hat{x}_j, x_j) = c^d_j(x_j, \hat{x}_j) + \lambda c^c_j(u_j, x_j)$ where c^d_j and c^c_j penalize estimation errors and communication respectively. Therefore λ is a unit conversion coefficient admitting the interpretation of equivalent estimation penalty per unit cost of communication. $c^c_j(u_j, x_j) = \sum_{k\in\chi(j)} c^c_{j\to k}(u_{j\to k}, x_j)$ where $c^c_{j\to k}(u_{j\to k})$ is the cost of transmitting the symbol $u_{j\to k}$ on the link $(j, k) \in \mathcal{E}$. It is selected as $c^c_{j\to k}(u_{j\to k}, x_j) = 0$ for the case $u_{j\to k} = 0$ and $c^c_{j\to k}(u_{j\to k}, x_j) = 1$ otherwise, measuring the link use rate. Hence, $\mathcal{U}_{j\to k}$ together with $c^c_{j\to k}$ define a selective communication scheme where $u_{j\to k} = 0$ indicates no communications and otherwise transmission of a 1 bit message. The estimation error is penalized by $c^d_j(x_j, \hat{x}_j) = (x_j - \hat{x}_j)^2$. Hence the total cost of a strategy is $J(\gamma) = J_d(\gamma) + \lambda J_c(\gamma)$ where J_d is the MSE and J_c is the total link use rate.

The random field of concern is a multivariate Gaussian, i.e. $x \sim \mathcal{N}(\mathbf{0}, \mathbf{C}_x)$, and Markov with respect to the graph in Figure (2b) together with the covariance matrix

$$\mathbf{C}_x = \begin{bmatrix} 2 & 1.125 & 1.5 & 1.125 \\ 1.125 & 2 & 1.5 & 1.125 \\ 1.5 & 1.5 & 2 & 1.5 \\ 1.125 & 1.125 & 1.5 & 2 \end{bmatrix} \tag{16}$$

Altough the communication structure of the DE network is not related with the MRF representation of X and Algorithm 3 would produce results for any choice, for sake of simplicity we selected the graph in Fig.(2b) as the undirected counterpart of that in Fig.(2a).

The noise processes n_j for $j \in \mathcal{V}$ are additive, mutually independent and given by $n_j \sim \mathcal{N}(0, 0.5)$, so that the observation likelihoods are $p(y_j|x_j) = \mathcal{N}(x_j, 0.5)^2$.

Since separable local cost functions are utilized, Eq.(4) splits into two minimizations which define local estimation and communication rules respectively. We will denote these rules by $\hat{x}_j = \delta_j(y_j, u_{\pi(j)})$ and $u_j = \mu_j(y_j, u_{\pi(j)})$ and initiate as follows: Each node applies a myopic rule by performing local MMSE estimation regardless of incoming messages, i.e. $\delta^0_j(y_j, u_{\pi(j)}) = \int_{-\infty}^{\infty} dx\, x_j p(x_j|y_j)$. The initial communication rule for each node is a quantization of the observation y_j, such that $\mu^0_i(y_i, u_{\pi(i)}) = 1, 0$ and 2 for $y_i < -2\sigma_n$, $-2\sigma_n \leqslant y_i \leqslant 2\sigma_n$ and $y_i > 2\sigma_n$ respectively.

For different values of λ, the converged performance point (J_c, J_d) will be different. Moreover, after a certain value $\lambda = \lambda^*$, the communication cost λJ_c will dominate such that the decrease in the decision cost J_d with the contributions of the communicated symbols will not be enough to decrease J and symbol 0 will be the best choice. Consequently, the individual estimators will be the myopic rules, since myopic rules with no communications constitute a pbp optimal strategy. Hence, it is possible to interpret λ^* as the maximum price per bit that the system affords to decrease the estimation error. As we increase λ from 0 we obtain approximate points from the performance curve for Problem (P) which lets us to quantify the tradeoff between the cost of estimation errors and communication.

In Figure (2c) we present approximately computed pairs $(\tilde{J}_c, \tilde{J}_d)$ of the converged strategies for different choices of λ and $|\mathcal{U}_{i\to j}|$s, where J_c is the total link use rate and J_d is the total MSE. The upper and lower limits are MSEs corresponding to the myopic rule and the centralized optimal rule[3] respectively. Considering $(\tilde{J}_c, \tilde{J}_d)$ points for the 1-bit selective communication scheme, for $\lambda = 0$, the transmission has no cost but the link use rate is well below 75% of the total 3 bits. This indicates that the information of receiving no messages is successfully maintained in this perspective. Moreover, the communication stops for $\lambda^* \approx 0.355$. Similarly approximate points for 2-bits and 3-bits schemes indicate that, if λ is small enough, we can achieve less MSE for the same total communication load as we increase the link capacities.

[2] Considering $\mathbf{C}_x$, each sensor has an SNR of 6dB.

5. CONCLUSION

We have considered the design problem in decentralized estimation under communication constraints and adopted a recent approach for decentralized detection based on a team decision theoretic investigation in a Bayesian setting. With the merit of this framework the existing approach of quantization for estimation is extended in the sense that a broader range of constraints are considered. However, the iterative solution scheme which converges to a person by person optimal strategy involves integral operators that have no closed form solutions in general. In order not to compromise model accuracy, we have utilized approximations and proposed a Monte Carlo optimization method which requires scalable number of samples generated from only the marginal distributions without any restriction on their type. It is also possible to quantify the tradeoff between cost of communications and estimation accuracy through the approximate performance curves achieved.

REFERENCES

[1] J. A. Gubner, “Distributed Estimation and Quantization,” *IEEE Trans. on Info. Theory*, vol. 39, pp.1456–1459, July 1993.

[2] W.-M. Lam and A. R. Reibman ,“Design of Quantizers for Decentralized Estimation Systems,” *IEEE Trans. on Communications*, vol.41, pp.1602–1605, Nov. 1993.

[3] A. Ribeiro, I. D. Schizas, J.-J. Xiao, G. B. Giannakis, and Z.-Q. Luo, ”Distributed Estimation under Bandwidth and Energy Constraints,” in *Wireless Sensor Networks: Signal Processing and Communication Perspectives*, J. Wiley & Sons, 2007.

[4] Y. Wang and P. Ishwar ,“Distributed Field Estimation With Randomly Deployed, Noisy, Binary Sensors,” *IEEE Trans. on Signal Processing*, vol.57, pp.1177–1189, March 2009.

[5] M. Cetin, 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 Magazine*, vol. 23, pp. 42–55, July 2006.

[6] A. T. Ihler, J. W. Fisher III, and A. S. Willsky, “Loopy Belief Propagation: Convergence and Effects of Message Errors,” *JMLR*, vol. 6, pp. 905–936, May 2005.

[7] Z.-Q. Luo,“Universal Decentralized Estimation in a Bandwidth Constrained Sensor Network,” *IEEE Trans. on Info. Theory*, vol.51, pp.2210–2219, June 2005.

[8] O. P. Kreidl and A. S. Willsky, “An Efficient Message-Passing Algorithm for Optimizing Decentralized Detection Networks,” in *Proc. IEEE 45th Conf. on Dec. and Control*, Dec. 13-15. 2006, pp. 6776–6783.

[9] O. P. Kreidl and A. S. Willsky, “An Efficient Message Passing Algorithm for Optimizing Decenralized Detection Networks,” *Tech. Report LIDS-P-2726*, MIT LIDS: Dec. 2006.

[10] J. N. Tsitsiklis, “Decentralized Detection,” in *Advances in Statistical Signal Processing*, H.V. Poor and J.B. Thomas, Eds. Greenwich, CT: JAI Press, 1993.

[11] Y.-C. Ho, K.-C. Chu,“Team Decision Theory and Information Structures in Optimal Control Problems,” *IEEE Trans. on Auto. Control*, vol. 17, pp.15–22, Feb. 1972.

[12] C. P. Robert and G. Casella, *Monte Carlo Statistical Methods 2nd Ed.*. New York, NY, USA: Springer, 2004.

[3] For the squared error cost, optimal centralized rule is the mean vector of the joint posterior $p(x_1, ..., x_4|y_1, ..., y_4)$ and $J_c = 3Q$ where Q is the number of bits used to quantize y_j before transmitting to the fusion center.