An Efficient Message-Passing Algorithm for Optimizing Decentralized Detection Networks

O. Patrick Kreidl and Alan S. Willsky

Abstract*— A promising feature of emerging wireless sensor networks is the opportunity for each spatially-distributed node to measure its local state and transmit only information relevant to effective global decision-making. An equally important design objective, as a result of each node's finite power, is for measurement processing to satisfy explicit constraints on, or perhaps make selective use of, the distributed algorithmic resources. We formulate this dual-objective design problem within the Bayesian decentralized detection paradigm, modeling resource constraints by a directed acyclic network with low-rate, unreliable communication links. Existing team theory establishes when necessary optimality conditions reduce to a convergent iterative algorithm to be executed *offline* (i.e., before measurements are processed). Even so, this offline algorithm has exponential complexity in the number of nodes and its distributed implementation assumes a fully-connected communication network. We state conditions by which the offline algorithm admits an efficient *message-passing* interpretation, featuring linear complexity in the number of nodes and a natural distributed implementation. We experiment with a simulated network of binary detectors, applying the message-passing algorithm to optimize the achievable tradeoff between global detection performance and network-wide online communication. The empirical analysis also exposes a design tradeoff between constraining in-network processing to preserve resources (per online measurement) and then having to consume resources (per offline reorganization) to maintain effective detection performance.

I. INTRODUCTION

The vision of "collaborative self-organizing wireless sensor networks," a confluence of emerging technology in both miniaturized devices and wireless communications, is of growing interest in a variety of scientific fields and engineering applications e.g., geology, biology, surveillance, fault-monitoring [1]–[4]. Their promising feature is the opportunity for each spatially-distributed node to receive measurements from its local environment and transmit information that is relevant for effective global decision-making. The finite power that drives each node creates incentives for prolonging operational lifetime, motivating measurement processing strategies that satisfy explicit resource constraints (on e.g., communication, computation, memory) in the network layer. One also anticipates intermittent reorganization by the network to stay connected (due to e.g., node dropouts, link failures), implying resource constraints can change accordingly and creating an incentive for intermittent re-optimization in the application layer. So, unless the *offline* optimization algorithm is itself amenable to efficient distributed implementation, there is little hope for maintaining application-layer decision objectives without also rapidly diminishing the network-layer resources that remain for actual *online* measurement processing.

This work supported by U.S. ARO under contract W911NF-05-1-0207 and by U.S. AFOSR under contract FA9550-04-1-0351.

The authors are with the Department of Electrical Engineering and Computer Science, Massachusetts Institute of Technology, 77 Massachusetts Ave, Cambridge, MA 02139, USA {opk,willsky}@mit.edu

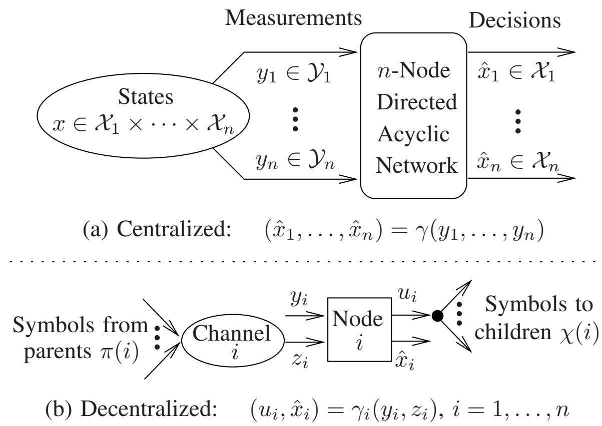


Fig. 1. Our detection model assuming (a) a centralized and (b) a decentralized strategy for processing measurements. Both assume (i) the hidden state x and observable measurement y take their values in, respectively, a discrete vector space $\mathcal{X} = \mathcal{X}_1 \times \cdots \times X_n$ and Euclidean vector space $\mathcal{Y} = \mathcal{Y}_1 \times \cdots \times \mathcal{Y}_n$; and (ii) the network is to generate an estimate $\hat{x} \in \mathcal{X}$ based on the distributed measurement vector $y \in \mathcal{Y}$. A decentralized strategy generates the components of $\hat{x}$ sequentially in the forward partial order of a given n-node directed acyclic graph $\mathcal{G}$, each edge (i, j) indicating a (perhaps unreliable) low-rate communication link from node i to node j. Each node i, observing only the component measurement y_i and the symbol(s) z_i received on incoming links with all parents $\pi(i) = \{j \mid \text{edge } (j, i) \text{ in } \mathcal{G}\}$ (if any), is to decide upon both its component estimate $\hat{x}_i$ and the symbol(s) u_i transmitted on outgoing links with all children $\chi(i) = \{j \mid \text{edge } (i, j) \text{ in } \mathcal{G}\}$ (if any). The collections of transmitted symbols u and received symbols z take their values in discrete vector spaces $\mathcal{U} = \mathcal{U}_1 \times \cdots \times \mathcal{U}_n$ and $\mathcal{Z} = \mathcal{Z}_1 \times \cdots \times \mathcal{Z}_n$, respectively.

We explore these design challenges assuming the decision-making objective is optimal Bayesian detection and the dominant measurement processing constraints arise from the underlying communication medium. Our main model is illustrated in Fig. 1, extending the canonical team-theoretic decentralized detection problem [5]–[7] in ways motivated by the vision of wireless sensor networks. Firstly, each spatially-distributed node i receives a noisy measurement $Y_i = y_i$ related only to its local state process X_i, which can be correlated with the states local to all other nodes, $X_{-i} = \{X_j; j \neq i\}$ [8]. Secondly, the network topology can be defined on any directed acyclic graph $\mathcal{G}$, each edge (i, j) representing a feasible point-to-point, low-rate communication link from node i to child $j \in \chi(i)$ [8], [9]. Thirdly, each node can employ a selective, or censored, transmission scheme and the multipoint-to-point channel into node i from parents $\pi(i)$ can be unreliable (due to e.g., uncoded interference, packet loss) [10]–[13].

The problem of optimal decentralized Bayesian detection is known to be NP-hard [5], in general. Also known is a team-theoretic approximation (satisfying necessary but not sufficient optimality conditions) that, under certain model assumptions, analytically reduces to a convergent iterative algorithm to be executed offline [5]–[7]. The algorithm strives to couple all nodes' local rules such that there is minimal performance loss from the in-network processing constraints. However, it requires that all nodes are initialized with common knowledge of global statistics and iterative per-node computation scales exponentially with n.

We identify a class of models for which the convergent offline algorithm admits an efficient *message-passing* interpretation, equivalent to a sequence of purely-local computations interleaved with only nearest-neighbor communications [14], [15]. In each offline iteration, every node adjusts its local rule (for subsequent online processing) based on incoming messages from its neighbors and, in turn, sends adjusted outgoing messages to its neighbors. Some of the messages received by each node define, in the context of its local objectives, a likelihood function for the symbols it may receive online (e.g., "what does the information from my neighbors mean to me") while the other messages define, in the context of all other nodes' objectives, a cost-to-go function for the symbols it may transmit online (e.g., "what does the information from me mean to my neighbors"). Each node need only be initialized with local statistics and iterative per-node computation becomes invariant to n (but scales exponentially with the number of neighbors, so the algorithm is well-suited for sparsely-connected networks).

This paper is organized as follows. Section II reviews the theory of decentralized Bayesian detection in the generality implied by Fig. 1. Section III defines the class of models for which the offline iterative algorithm discussed in Section II admits its efficient message-passing interpretation. In Section IV, we consider a simulated network of binary detectors and apply the message-passing algorithm to quantify the tradeoff between online communication and detection performance. The empirical analysis also exposes a design tradeoff between constraining in-network processing to preserve resources (per online measurement) and then having to consume resources (per offline reorganization) to maintain effective detection performance. Conclusions and questions for future research are discussed in Section V.

II. DECENTRALIZED DETECTION NETWORKS

This section reviews the theory of decentralized Bayesian detection [5]–[7] in the generality implied by Fig. 1.

A. Dual-Objective Bayesian Formulation

For the model in Fig. 1, let distribution $p(x,y)$ jointly describe the hidden state process X and the noisy measurement process Y. Fig. 1(a) depicts how the *centralized* decision process $\hat{X} = \gamma(Y)$ is derived from Y by any function of the form $\gamma : \mathcal{Y} \to \mathcal{X}$. Fig. 1(b) depicts how the *decentralized* decision process unfolds node-by-node in the forward partial order of network topology $\mathcal{G}$. By constraint, the global decision process $\hat{X}$ is generated in a component-wise fashion, each node i generating both U_i and $\hat{X}_i$ upon observing both Y_i and Z_i. Let conditional distribution $p(z_i|x,y,u_{\pi(i)})$ describe the information Z_i received by node i based on information $U_{\pi(i)} = \{U_j | j \in \pi(i)\}$ transmitted by all parents $\pi(i)$. Altogether, any particular *strategy* $\gamma : \mathcal{Y} \times \mathcal{Z} \to \mathcal{U} \times \mathcal{X}$ induces a global decision process $(U,\hat{X}) = \gamma(Y,Z)$. Denote by Γ the set of all such strategies and by $\Gamma^{\mathcal{G}} \subset \Gamma$ the *admissible* subset of decentralized strategies in a given network topology $\mathcal{G}$.

The Bayesian criterion assigns a numeric "cost" $c(u,\hat{x},x) \geq 0$ to every possible realization of the joint process $(U,\hat{X},X)$, penalizing any fixed strategy $\gamma \in \Gamma$ by its expected cost i.e.,

$$J(\gamma) = E\left[c(U,\hat{X},X)\right] = E\left[E\left[c(\gamma(Y,Z),X|\,Y,Z\right]\right]. \tag{1}$$

We focus on a *dual-objective* specialization of (1), assuming

$$c(u,\hat{x},x) = c(\hat{x},x) + \lambda c(u,x)$$

where the constant $\lambda \geq 0$ specifies the unit conversion between detection-related costs $c(\hat{x},x)$ and communication-related costs $c(u,x)$. Formally, our decentralized design problem is to find the strategy $\gamma^* \in \Gamma^{\mathcal{G}}$ such that

$$J(\gamma^*) \quad = \quad \min_{\gamma\in\Gamma} J_d(\gamma) + \lambda J_c(\gamma) \text{ subject to } \gamma \in \Gamma^{\mathcal{G}}, \tag{2}$$

where the functions $J_d : \Gamma \to \mathbb{R}$ and $J_c : \Gamma \to \mathbb{R}$ quantify the *detection penalty* and *communication penalty*, respectively. The rest of this subsection further develops the underlying probabilistic model.

1) Centralized Strategies: The problem in (2) specializes to the centralized design problem when online communication is both unconstrained and unpenalized i.e., when $\Gamma^{\mathcal{G}} = \Gamma$ and $\lambda = 0$. Classical detection theory [16] states $J_d(\gamma)$ is minimized over all Γ by[1]

$$\bar{\gamma}^*(Y) = \arg\min_{\hat{x}\in\mathcal{X}} \sum_{x\in\mathcal{X}} \underbrace{p(x)c(\hat{x},x)}_{\bar{\theta}^*(\hat{x},x)\in\mathbb{R}} p(Y|x). \tag{3}$$

Note that (i) the likelihood function $p(Y|x)$ provides a sufficient statistic of Y and (ii) the optimal parameter values $\bar{\theta}^*(\hat{x},x) = p(x)c(\hat{x},x)$ can be computed offline. Minimizing $J_d(\gamma)$ over all Γ is thus equivalent to minimizing

$$J_d(\bar{\gamma}) = \sum_{x\in\mathcal{X}} p(x) \sum_{\hat{x}\in\mathcal{X}} c(\hat{x},x) \int_{y\in\mathcal{Y}} p(y|x) p\left(\hat{x}|y;\bar{\theta}\right) dy$$

over the space of distributions defined by parameters $\bar{\theta} \in \mathbb{R}^{|\mathcal{X}|\times|\mathcal{X}|}$ according to

$$p\left(\hat{x}|y;\bar{\theta}\right) = \begin{cases} 1 & , \text{ if } \hat{x} = \bar{\gamma}(y) \text{ in (3)} \\ 0 & , \text{ otherwise} \end{cases}. \tag{4}$$

[1] Equality of two random variables denotes "equal with probability one."

Of course, these sums and integrals over vector spaces become difficult to compute, especially as n grows large.

2) Myopic Strategies: The centralized strategy in (3) is clearly infeasible given the constraints of Fig. 1(b), as it assumes the global sufficient statistic $p(Y|x)$ is known at an individual processor and total computation scales exponentially with n. A simplest feasible strategy is for each node i to decide upon $\hat{x}_i$ as if in isolation i.e., given distribution $p(x_i, y_i)$ and costs $c(\hat{x}_i, x_i)$ involving only the local environment, the rule at node i is

$$\bar{\gamma}_i(Y_i) = \arg\min_{\hat{x}_i \in \mathcal{X}_i} \sum_{x_i \in \mathcal{X}_i} \underbrace{p(x_i)c(\hat{x}_i, x_i)}_{\bar{\phi}_i(\hat{x}_i, x_i) \in \mathbb{R}} p(Y_i|x_i). \quad (5)$$

We denote all such local rules by Γ_i and the set of all such strategies by $\Gamma^n = \Gamma_1 \times \cdots \times \Gamma_n$. Each member in $\Gamma^n \subset \Gamma$ is thus a collection of local rules $\bar{\gamma} = (\bar{\gamma}_1, \ldots, \bar{\gamma}_n)$ with parameters $\bar{\theta} = (\bar{\phi}_1, \ldots, \bar{\phi}_n)$ in block-diagonal form: no node transmits (receives) information on its the outgoing (incoming) links and total computation scales linearly with n. It follows that, for every $\bar{\gamma} \in \Gamma^n$, we have $J_c(\bar{\gamma}) = 0$ in (2) and $p(\hat{x}|y; \bar{\theta}) = \prod_i p(\hat{x}_i|y_i; \bar{\phi}_i)$ in (4).

3) Decentralized Strategies: The set of decentralized strategies $\Gamma^{\mathcal{G}} \subset \Gamma$ includes all myopic strategies Γ^n, but excludes the (optimal) centralized strategy $\bar{\gamma}^*$. More precisely, assign to each edge (i, j) in graph $\mathcal{G}$ an integer $d_{i\to j} \geq 2$, denoting the size of the symbol set supported by the link from node i to child $j \in \chi(i)$ (i.e., the link rate is $\log_2 d_{i\to j}$ bits per measurement). It follows that the information U_i transmitted by node i can take at most $|\mathcal{U}_i| = \prod_{j\in\chi(i)} d_{i\to j}$ distinct values. The ith channel model $p(z_i|x, y, u_{\pi(i)})$ specifies the number $|\mathcal{Z}_i|$ of distinct values taken by received information Z_i.[2] Let $\Gamma^{\mathcal{G}} = \Gamma^{\mathcal{G}}_1 \times \cdots \times \Gamma^{\mathcal{G}}_n$, where each $\Gamma^{\mathcal{G}}_i$ denotes the set of all feasible rules $\gamma_i : \mathcal{Y}_i \times \mathcal{Z}_i \to \mathcal{U}_i \times \mathcal{X}_i$ local to node i. Thus, by constraint, fixing a rule $\gamma_i \in \Gamma^{\mathcal{G}}_i$ is equivalent to specifying the distribution $p(u_i, \hat{x}_i|y_i, z_i; \gamma_i)$. It follows that fixing $\gamma \in \Gamma^{\mathcal{G}}$ specifies the distribution

$$p(u, z, \hat{x}|x, y; \gamma) = \prod_{i=1}^{n} p(z_i|x, y, u_{\pi(i)}) p(u_i, \hat{x}_i|y_i, z_i; \gamma_i) \quad (6)$$

and, in turn, the distribution to determine $J(\gamma)$ in (1):

$$p(u, \hat{x}, x; \gamma) = \sum_{z\in\mathcal{Z}} \int_{y\in\mathcal{Y}} p(x, y) p(u, z, \hat{x}|x, y; \gamma)\, dy. \quad (7)$$

B. Offline Iterative Algorithm

In general, it is not known whether the strategy γ^* in (2) lies in a finitely-parameterized subspace of $\Gamma^{\mathcal{G}}$. The team-theoretic approximation is to satisfy a set of necessary optimality conditions based on a simple observation: if a decentralized strategy $\gamma^* = (\gamma^*_1, \ldots, \gamma^*_n)$ is optimal over $\Gamma^{\mathcal{G}}$, then for each i and assuming rules $\gamma^*_{-i} = \{\gamma^*_j \in \Gamma^{\mathcal{G}}_j \mid j \neq i\}$ are fixed, the rule γ^*_i is optimal over $\Gamma^{\mathcal{G}}_i$ i.e., for each i,

$$\gamma^*_i \quad = \quad \arg\min_{\gamma_i\in\Gamma_i} J_d(\gamma^*_{-i}, \gamma_i) + \lambda J_c(\gamma^*_{-i}, \gamma_i). \quad (8)$$

[2]We focus on models where $|\mathcal{U}_i| \ll |\mathcal{Y}_i|$ and $|\mathcal{Z}_i| \leq |\mathcal{U}_{\pi(i)}|$ for all i.

Simultaneously satisfying (8) for all i is not a sufficient optimality condition because, in general, it does not preclude a decrease in J via joint minimization over multiple nodes.

Assumption 1 (Conditional Independence): Every node i measures a state-dependent deterministic signal corrupted by noise, where this local measurement noise is independent of both the local channel noise and the noise processes local to every other node $j \neq i$. i.e., for every i,

$$p(y_i, z_i|x, y_{-i}, z_{-i}, u_{\pi(i)}) = p(y_i|x) p(z_i|x, u_{\pi(i)}). \quad (9)$$

Lemma 1: Let Assumption 1 hold. For every strategy $\gamma \in \Gamma^{\mathcal{G}}$, the distribution in (7) specializes to

$$p(u, \hat{x}, x; \gamma) = p(x) \prod_{i=1}^{n} p(u_i, \hat{x}_i|x, u_{\pi(i)}; \gamma_i),$$

where for every i,

$$p(u_i, \hat{x}_i|x, u_{\pi(i)}; \gamma_i) = \sum_{z_i\in\mathcal{Z}_i} p(z_i|x, u_{\pi(i)}) \int_{y_i\in\mathcal{Y}_i} p(y_i|x) p(u_i, \hat{x}_i|y_i, z_i; \gamma_i)\, dy_i. \quad (10)$$

Proof: Substituting (9) into (6) and (7) results in

$$p(u, \hat{x}|x; \gamma) = \sum_{z\in\mathcal{Z}} \int_{y\in\mathcal{Y}} \prod_{i=1}^{n} p(y_i|x) p(z_i|x, u_{\pi(i)}) p(u_i \hat{x}_i|y_i, z_i; \gamma_i)\, dy.$$

Because only the ith factor in the integrand involves variables (y_i, z_i), global marginalization over (Y, Z) simplifies to n local marginalizations, each over (Y_i, Z_i). ■

Proposition 1: Let Assumption 1 hold. The ith component optimization in (8) reduces to

$$\gamma^*_i(Y_i, Z_i) = \arg\min_{(u_i,\hat{x}_i)\in\mathcal{U}_i\times\mathcal{X}_i} \sum_{x\in\mathcal{X}} \theta^*_i(u_i, \hat{x}_i, x; Z_i) p(Y_i|x) \quad (11)$$

where, for each $z_i \in \mathcal{Z}_i$ such that $p(Y_i, z_i; \gamma^*_{-i}) > 0$, the parameter values $\theta^*_i(z_i) \in \mathbb{R}^{|\mathcal{U}_i|\times|\mathcal{X}_i|\times|\mathcal{X}|}$ are given by (12).

Proof: See [17]. ■

It is instructive to note the similarity between a local rule γ^*_i in Proposition 1 and the centralized strategy $\bar{\gamma}^*$ in (3). Both process an $|\mathcal{X}|$-dimensional sufficient statistic of the available measurement with optimal parameter values to be computed offline. In rule γ^*_i, however, this offline computation is more than simple multiplication of probabilities $p(x)$ and costs $c(u, \hat{x}, x)$: parameter values $\theta^*_i \in \mathbb{R}^{|\mathcal{U}_i|\times|\mathcal{X}_i|\times|\mathcal{X}|\times|\mathcal{Z}_i|}$ in (12) now involve conditional expectations, taken over distributions that depend on the fixed rules γ^*_{-i} of all other nodes $j \neq i$. Each such fixed rule γ^*_j is similarly of the form in Proposition 1, where fixing parameter values θ^*_j specifies $p(u_j, \hat{x}_j|x, u_{\pi(j)}; \theta^*_j)$ local to node j through (10) and (11).

Each ith minimization in (8) is thereby equivalent to minimizing

$$J(\gamma^*_{-i}, \gamma_i) = \sum_{x\in\mathcal{X}} p(x) \sum_{u\in\mathcal{U}} \sum_{\hat{x}\in\mathcal{X}} c(u, \hat{x}, x) p(u, \hat{x}|x; \theta^*_{-i}, \theta_i)$$

$$\theta_i^*(u_i, \hat{x}_i, x; z_i) = p(x) \sum_{u_{-i} \in \mathcal{U}_{-i}} p(z_i | x, u_{\pi(i)}) \sum_{\hat{x}_{-i} \in \mathcal{X}_{-i}} c(u, \hat{x}, x) \prod_{j \neq i} p(u_j, \hat{x}_j | x, u_{\pi(j)}; \gamma_j^*) \quad (12)$$

over the parameterized space of distributions defined by

$$p(u, \hat{x} | x; \theta_{-i}^*, \theta_i) = p(u_i, \hat{x}_i | x, u_{\pi(i)}, \theta_i) \prod_{j \neq i} p(u_j, \hat{x}_j | x, u_{\pi(j)}; \theta_j^*).$$

It follows that the simultaneous satisfaction of (8) at all nodes corresponds to solving for $\theta^* = (\theta_1^*, \ldots, \theta_n^*)$ in a system of nonlinear equations expressed by (10)-(12). Specifically, if we let $f_i(\theta_{-i}^*)$ denote the right-hand-side of (12), then offline computation of a *team-optimal* strategy reduces to solving the fixed-point equations

$$\theta_i = f_i(\theta_{-i}), \quad i = 1, \ldots, n. \quad (13)$$

Corollary 1: Initialize parameters $\theta^0 = (\theta_1^0, \ldots, \theta_n^0)$ and generate the sequence $\{\theta^k; k = 1, 2, \ldots\}$ by iterating (13) in any component-by-component order e.g., iteration k is

$$\theta_i^k := f_i(\theta_1^k, \ldots, \theta_{i-1}^k, \theta_{i+1}^{k-1}, \ldots, \theta_n^{k-1}), \quad i = 1, \ldots, n.$$

If Assumption 1 holds, then the associated sequence $\{J(\gamma^k)\}$ is non-increasing and converges.

Proof: By virtue of Proposition 1, each operator f_i is an analytical solution to the minimization of J over the ith coordinate function space $\Gamma_i^{\mathcal{G}}$. Any component-wise iteration of f is thus equivalent to a coordinate-descent iteration of J, implying $J(\gamma^k) \leq J(\gamma^{k-1})$ for every k [18]. Because the real-valued, non-increasing sequence $\{J(\gamma^k)\}$ is also bounded below by $J(\bar{\gamma}^*) \geq 0$, it has a limit point. ■

In the absence of additional technical conditions (e.g., J is convex, f is contracting [18]), Corollary 1 implies nothing about whether $\{J(\gamma^k)\}$ converges to the optimal performance $J(\gamma^*)$, whether the achieved performance is invariant to the choice of initial strategy γ^0, nor whether the associated sequence $\{\theta^k\}$ converges. Also note that the proof to Corollary 1 assumes every node i can *exactly* compute the local marginalization of (10). Some measurement models of practical interest lead to numerical or monte-carlo approximation of these marginalizations, and the extent to which the resulting errors affect convergence is not known.

The offline message-passing algorithm we develop in Section III (and experiment with in Section IV) is, in essence, an instance of Corollary 1 under some additional model assumptions. Thus, when these additional assumptions hold, it inherits the same theoretical convergence properties. Two interesting open problems are (i) whether the additional model assumptions allow for stronger theoretical convergence guarantees and (ii) whether the message-passing algorithm provides a satisfactory approximation even when the additional model assumptions are violated.

III. MESSAGE-PASSING ALGORITHM

Online measurement processing implied by Proposition 1 is, by design, well-suited for distributed implementation. However, a number of practical difficulties remain:

- convergent offline optimization requires global knowledge of probabilities $p(x)$, costs $c(u, \hat{x}, x)$ and statistics $\{p(u_i, \hat{x}_i | x, u_{\pi(i)}; \theta_i^k)\}$ in every iteration k;
- total (offline and online) memory/computation requirements scale exponentially with the number of nodes n.

In this section, we establish conditions so that convergent offline optimization can be executed in a recursive fashion: each node i starts with purely local knowledge (i.e., probabilities $p(x_{\pi(i)}, x_i)$ and costs $c(u_i, \hat{x}_i, x_i)$) and exchanges rule-dependent statistics, or *messages*, with only its neighbors $\pi(i) \cup \chi(i)$; yet, if generated appropriately, the sequence of messages received from its neighbors will converge to sufficient statistics of the required global knowledge. Moreover, total memory/computation requirements scale linearly with n. These results extend the computational theory discussed in [5], [8], [9] in the generality of Fig. 1.

A. Online Efficiency

The online computation implied by Proposition 1 scales exponentially with n due to the dependence of (11) on the global state process X.

Assumption 2 (State Locality): In addition to the conditions of Assumption 1, both the measurement model and channel model local to node i are invariant to all non-local state processes X_{-i} i.e., for every i,

$$p(y_i, z_i | x, y_{-i}, z_{-i}, u_{\pi(i)}) = p(y_i | x_i) p(z_i | x_i, u_{\pi(i)}). \quad (14)$$

Corollary 2: If Assumption 2 holds, then (11) and (12) in Proposition 1 specialize to

$$\gamma_i^*(Y_i, Z_i) = \arg \min_{(u_i, \hat{x}_i) \in \mathcal{U}_i \times \mathcal{X}_i} \sum_{x_i \in \mathcal{X}_i} \phi_i^*(u_i, \hat{x}_i, x_i; Z_i) p(Y_i | x_i) \quad (15)$$

and (16), respectively.

Proof: Recognizing (14) to be the special case of (9) with $p(y_i|x) = p(y_i|x_i)$ and $p(z_i|x, u_{\pi(i)}) = p(z_i|x_i, u_{\pi(i)})$ for every i, (10) in Lemma 1 similarly specializes to $p(u_i, \hat{x}_i | x_i, u_{\pi(i)}; \gamma_i) =$

$$\sum_{z_i \in \mathcal{Z}_i} p(z_i | x_i, u_{\pi(i)}) \int_{y_i \in \mathcal{Y}_i} p(y_i | x_i) p(u_i, \hat{x}_i | y_i, z_i; \gamma_i) \, dy_i$$

for every i. We then apply Proposition 1 with

$$\phi_i^*(u_i, \hat{x}_i, x_i; z_i) = \sum_{x_{-i} \in \mathcal{X}_{-i}} \theta_i^*(u_i, \hat{x}_i, x; z_i).$$

■

$$\phi_i^*(u_i,\hat{x}_i,x_i;z_i) = \sum_{x_{-i}} p(x) \sum_{u_{-i}} p(z_i|x_i,u_{\pi(i)}) \sum_{\hat{x}_{-i}} c(u,\hat{x},x) \prod_{j\neq i} p(u_j,\hat{x}_j|x_j,u_{\pi(j)};\gamma_j^*) \quad (16)$$

$$P_i^*(z_i|x_i) = \begin{cases} 1 & , \quad \pi(i) \text{ empty} \\ \sum_{x_{\pi(i)}} p(x_{\pi(i)}|x_i) \sum_{u_{\pi(i)}} p(z_i|x_i,u_{\pi(i)}) \prod_{j\in\pi(i)} P_{j\to i}^*(u_j|x_j) & , \quad \text{otherwise} \end{cases} \quad (17)$$

$$Q_{j\to i}^*(u_j,\hat{x}_j,x_j|u_i,x_i) = \sum_{x_{\pi(j)-i}} p(x_{\pi(j)},x_j|x_i) \sum_{u_{\pi(j)-i}} p(u_j,\hat{x}_j|x_j,u_{\pi(j)};\gamma_j^*) \prod_{m\in\pi(j)-i} P_{m\to j}^*(u_m|x_m) \quad (18)$$

It is instructive to note the similarity between γ_i^* in Corollary 2 and a local myopic rule $\bar{\gamma}_i$ in (5). Online computation is nearly identical, but with γ_i^* using parameters that reflect the composite decision space $\mathcal{U}_i \times \mathcal{X}_i$ and depend explicitly on the received information $Z_i = z_i$. This similarity is also apparent in the local computation of statistic $p(u_i,\hat{x}_i|x_i,z_i;\phi_i^*)$ for fixed parameters ϕ_i^* in (15), which per value $z_i \in \mathcal{Z}_i$ involves the same local marginalization over Y_i by which one computes the statistic $p(\hat{x}_i|x_i;\bar{\phi}_i)$ for fixed parameters $\bar{\phi}_i$ in (5).

B. Offline Efficiency

Efficiency in the offline iterative algorithm is grounded in the special probabilistic structure that results from the decentralized processing constraints in conjunction with Assumption 2. In particular, suppose the cost function $c(u,\hat{x},x)$ in (16) decomposes in a manner compatible with the structured form of distribution $p(u_{-i},z_i,\hat{x}_{-i}|x;\gamma_{-i}^*)$ at node i. Then, for each candidate decision $(u_i,\hat{x}_i)$, the sums taken over vectors u_{-i} and $\hat{x}_{-i}$ can be separated into a collection of $n-1$ partial sums, each jth such partial sum taken over $(u_{\pi(j)},u_j)$ and $\hat{x}_j$, respectively. These types of recursive computations are known to be maximally efficient in a *Markov tree* structure [14], [15].

Assumption 3 (Tree Topology): Graph $\mathcal{G}$ is a polytree i.e., there is at most one path between any pair of nodes.

Assumption 4 (Cost Locality): The Bayesian cost function is additive across the nodes of the network i.e.,

$$c(u,\hat{x},x) = \sum_{i=1}^{n} c(u_i,\hat{x}_i,x_i).$$

Proposition 2: If Assumptions 2-4 hold, then (15) applies with (16) specialized to the proportionality

$$\phi_i^*(u_i,\hat{x}_i,x_i;z_i) \propto p(x_i)P_i^*(z_i|x_i)C_i^*(u_i,\hat{x}_i,x_i),$$

where $P_i^*(z_i|x_i)$ is given by (17) and
$C_i^*(u_i,\hat{x}_i,x_i) =$

$$\begin{cases} c(u_i,\hat{x}_i,x_i) & , \quad \chi(i) \text{ empty} \\ c(u_i,\hat{x}_i,x_i) + \sum_{j\in\chi(i)} C_{j\to i}^*(u_i,x_i) & , \quad \text{otherwise} \end{cases} ;$$

at node i, the *forward message* from parent $j \in \pi(i)$ is

$$P_{j\to i}^*(u_j|x_j) = \sum_{z_j} P_j^*(z_j|x_j) \sum_{\hat{x}_j} p(u_j,\hat{x}_j|x_j,z_j;\gamma_j^*)$$

and the *backward message* from child $j \in \chi(i)$ is
$C_{j\to i}^*(u_i,x_i) =$

$$\sum_{x_j}\sum_{u_j}\sum_{\hat{x}_j} C_j^*(u_j,\hat{x}_j,x_j) Q_{j\to i}^*(u_j,\hat{x}_j,x_j|u_i,x_i),$$

where $Q_{j\to i}^*(u_j,\hat{x}_j,x_j|u_i,x_i)$ is given by (18).

Proof: See [17]. ■

Proposition 2 implies the parameters ϕ_i^* local to node i no longer (explicitly) depend on full knowledge of statistics $\{p(u_j,\hat{x}_j|x_j,u_{\pi(j)};\gamma_j^*); j \neq i\}$. Rather, it is sufficient for node i to know the messages $P_{\pi(i)\to i}^* = \{P_{j\to i}^*; j \in \pi(i)\}$ and $C_{\chi(i)\to i}^* = \{C_{j\to i}^*; j \in \chi(i)\}$ received from the parents and children, respectively. The forward messages $P_{\pi(i)\to i}^*$ define the likelihood function P_i^*, summarizing the rule statistics local to all ancestors of node i (i.e., the parents $\pi(i)$, the parents' parents $\pi(j)$ for $j \in \pi(i)$ and so on), while the backward messages $C_{\chi(i)\to i}^*$ define the cost-to-go function C_i^*, summarizing the rule statistics local to all nodes other than these ancestors. Eliminating P_i^*, C_i^* and $Q_{\chi(i)\to i}^*$ from the system of equations in Proposition 2 specializes (13) to the block-structured form

$$\begin{aligned} P_{i\to\chi(i)} &= g_i\left(\phi_i, P_{\pi(i)\to i}\right) \\ \phi_i &= f_i\left(P_{\pi(i)\to i}, C_{\chi(i)\to i}\right) \qquad i = 1,\ldots,n. \\ C_{i\to\pi(i)} &= h_i\left(\phi_i, P_{\pi(i)\to i}, C_{\chi(i)\to i}\right) \end{aligned} \quad (19)$$

Corollary 3: Initialize parameters $\phi^0 = (\phi_1^0,\ldots,\phi_n^0)$ and generate the sequence $\{\phi^k; k = 1,2,\ldots\}$ by iterating (19) in a repeated forward-backward pass through $\mathcal{G}$ e.g.,

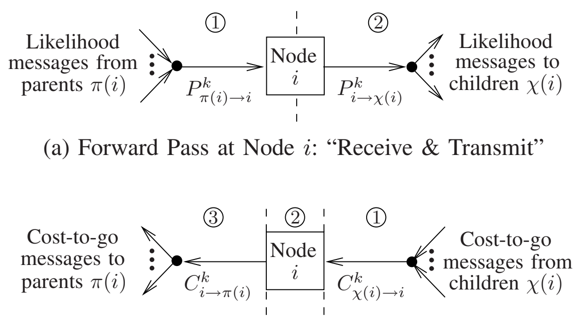
$$P_{i\to\chi(i)}^k := g_i(\phi_i^{k-1}, P_{\pi(i)\to i}^k)$$

from $i = 1,2,\ldots n$ and

$$\begin{aligned} \phi_i^k &:= f_i(P_{\pi(i)\to i}^k, C_{\chi(i)\to i}^k) \\ C_{i\to\pi(i)}^k &:= h_i(\phi_i^k, P_{\pi(i)\to i}^k, C_{\chi(i)\to i}^k) \end{aligned}$$

from $i = n, n-1, \ldots, 1$. If Assumptions 2-4 hold, then the associated sequence $\{J(\gamma^k)\}$ converges.

Proof: By virtue of Proposition 2, a sequence $\{\phi^k\}$ is the special case of a sequence $\{\theta^k\}$ considered in



(a) Forward Pass at Node i: "Receive & Transmit"

(b) Backward Pass at Node i: "Receive, Update & Transmit"

Fig. 2. The distributed message-passing interpretation of the kth iteration in the offline algorithm discussed in Corollary 3, each node i interleaving its purely-local computations with only nearest-neighbor communications.

Corollary 1. Each forward-backward pass in the partial-order implied by $\mathcal{G}$ ensures each iterate ϕ^k is generated in a component-wise fashion required for convergence. ■

Fig. 2 illustrates the iterative message-passing interpretation for which Corollary 3 applies. The simplification afforded by Proposition 2 is also apparent in the computation of the sequence $\{J(\gamma^k)\}$. Specifically, the global penalty associated to iterate ϕ^k is given by

$$J(\gamma^k) := \sum_i \sum_{x_i} p(x_i) \sum_{u_i} \sum_{\hat{x}_i} c(u_i, \hat{x}_i, x_i) p(u_i, \hat{x}_i | x_i; \phi^k)$$

with

$$p(u_i, \hat{x}_i | x_i; \phi^k) = \sum_{z_i} P_i^{k+1}(z_i|x_i) p(u_i, \hat{x}_i | x_i, z_i; \phi_i^k)$$

for every i. That is, given the statistic P_i^{k+1} is known local to each node i (which occurs upon completion of the forward pass in iteration $k+1$), the penalty $J(\gamma^k)$ is additive across the nodes so its computation scales linearly in n.

IV. NUMERICAL EXPERIMENTS

This section summarizes experiments with the offline message-passing algorithm presented in Section III. The setup involves a simulated network of linear-Gaussian binary detectors and discrete erasure channels. Our results quantify the extent to which the message-passing algorithm improves upon the achievable tradeoff between detection penalty J_d and communication penalty J_c in (2). As we should expect, this potential for improvement is a function of the correlation between the distributed states, the noise level in the measurements and the erasure probabilities in the channels. Each node employs a selective transmission scheme, where a decision to *not* transmit on any particular link incurs zero communication cost. Our results show the team strategy exploiting this cost-free silence: even when actual symbols can be transmitted without penalty (i.e., when $\lambda = 0$ in (2)), a node's selective silence can convey valuable information to its children. We also seek to quantify the offline communication overhead associated with team-optimal performance: we empirically measure the average number of iterations to convergence, recognizing that per iteration k each link (i,j) must reliably communicate messages $P^k_{i\to j}$ and $C^k_{j\to i}$, each a collection of up to $|\mathcal{X}_i \times \mathcal{U}_i|$ real numbers.

A. Simulation Model

We consider an instance of the model in Fig. 1 that both satisfies Assumptions 2-4 and depends on only a few parameters for ease of illustration (see Fig. 3): the signal-to-noise ratio in every node's measurement model is $r \in (0, \infty)$, the per-transmission erasure probability in every link's channel model is $q \in [0, 1]$, and the correlation between components of the spatially-distributed state process X is determined by $w \in (0, 1)$. The specific parents $\pi(i)$, children $\chi(i)$, likelihood $p(y_i|x_i)$ and channel $p(z_i|x_i, u_{\pi(i)})$ local to each node i are shown in Fig. 3(b)-(d). Recall the offline message-passing algorithm also assumes every node i is initialized with a local prior $p(x_{\pi(i)}, x_i)$, which in our experiments are obtained by appropriate marginalization of a global prior

$$p(x) \propto \prod_{(i,j)\in\mathcal{E}} \psi_{i,j}(x_i, x_j), \tag{20}$$

where $\mathcal{E}$ denotes the edge set of the *undirected* graph illustrated in Fig. 3(a) and each non-negative function

$$\psi_{i,j}(x_i, x_j) = \begin{cases} w & , \quad x_i = x_j \\ 1 - w & , \quad x_i \neq x_j \end{cases}$$

specifies the correlation (i.e., negative, zero, or positive when w is less than, equal to, or greater than 0.5, respectively) between neighboring states X_i and X_j.

The costs $c(u, \hat{x}, x)$ are chosen to specialize the detection penalty J_d to the *node-error-rate* and the communication penalty J_c to the *link-use-rate*. Local to each node i, we set $c(u_i, \hat{x}_i, x_i) = c(\hat{x}_i, x_i) + \lambda \sum_{j\in\chi(i)} c(u_{i\to j})$ with

$$c(\hat{x}_i, x_i) = \begin{cases} 0, & \hat{x}_i = x_i \\ 1, & \hat{x}_i \neq x_i \end{cases}, \quad c(u_{i\to j}) = \begin{cases} 0, & u_{i\to j} = 0 \\ 1, & u_{i\to j} \neq 0 \end{cases}.$$

Thus, in conjunction with the channel model in Fig. 3(d), the symbol $u_{i\to j} = 0$ represents node i foregoing the opportunity to transmit a binary-valued message (designated by $u_{i\to j} \in \{-1, +1\}$) to child j, which avoids the potential for both a transmission cost and a symbol erasure. In turn, assuming $q > 0$ and upon receiving the symbol $z_{i\to j} = 0$, node j cannot conclusively determine whether parent i chose to be silent or link (i,j) just experienced an erasure.

A final model consideration is the initial rule γ_i^0 local to each node i. It is worth mentioning that initializing to a myopic rule $\bar{\gamma}_i$ in (5) prohibits the offline algorithm from making progress—in our model, the team-optimality conditions discussed in Section II is satisfied by this myopic strategy! Fig. 4 illustrates our choice of initial rule γ_i^0, which is motivated by the class of monotone threshold rules for binary detection networks [5], [10], [11]. We observe the algorithm making reliable progress from this initialization as long as the induced statistics at every node i satisfy $p(u_{i\to j}|z_i; \gamma_i^0) > 0$ for all $(z_i, u_{i\to j})$ and $j \in \chi(i)$.

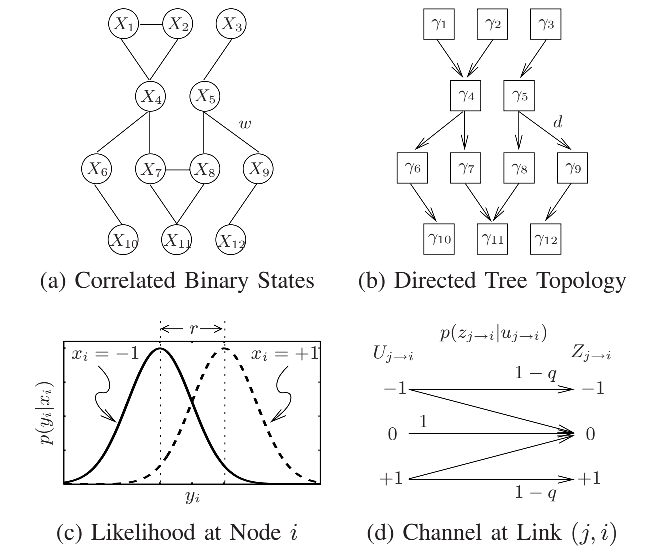


Fig. 3. The decentralized detection network used in our experiments: (a) the (undirected) graph upon which the global prior $p(x)$ is defined via (20) with edge weights $w \in (0,1)$; (b) a tree-structured network topology that spans the $n = 12$ vertices in (a) and defines the parents $\pi(i)$ and children $\chi(i)$ of each node i, each link supporting $d = 3$ symbols; (c) the ith node's likelihood function $p(y_i|x_i)$, defining a linear-Gaussian binary detector with signal-to-noise ratio $r \in (0,\infty)$; and (d) the ith node's multipoint-to-point channel $p(z_i|x_i, u_{\pi(i)}) = \prod_{j\in\pi(i)} p(z_{j\to i}|u_{j\to i})$, where each incoming link (j,i) has erasure probability $q \in [0,1]$. In relation to the main model illustrated in Fig. 1, we have $|\mathcal{X}_i| = 2, \mathcal{Y}_i = \mathbb{R}$, $|\mathcal{U}_i| = 3^{|\chi(i)|}$ and $|\mathcal{Z}_i| = 3^{|\pi(i)|}$ for every $i = 1,\ldots,12$.

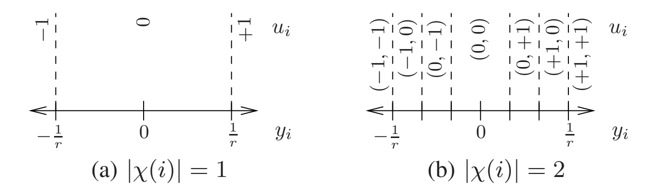


Fig. 4. The initial rule γ_i^0 used in our experiments. The measurement model in Fig. 3 yields a linear log-likelihood-ratio function $L_i(y_i) = ry_i$. For all z_i, we partition the real y_i-axis into a set of intervals with the shown $2(2^{|\chi(i)|}-1)$ thresholds to decide $u_i \in \{-1,0,+1\}^{|\chi(i)|}$ and a threshold of zero to decide $\hat{x}_i \in \{-1,+1\}$. In essence, each node i is initialized to (i) ignore all information received on the incoming links, (ii) myopically make a maximum-likelihood estimate of its local state and (iii) make a binary-valued decision per outgoing link (i,j), remaining silent (with $u_{i\to j} = 0$) or transmitting its local state estimate (with $u_{i\to j} = \hat{x}_i$).

B. Procedure and Results

Given values (w,r,q) in Fig. 3, one purpose of our experiments is to quantify the achievable tradeoff between node-error-rate J_d and link-use-rate J_c. Specifically, let $J^\lambda = J_d^\lambda + \lambda J_c^\lambda$ denote the penalty achieved by the offline message-passing algorithm for any fixed $\lambda \geq 0$. The optimized tradeoff is then expressed by the set of points $\{(J_c^\lambda, J_d^\lambda); \lambda \geq 0\}$. The starting point (J_d^0, J_c^0) defines both the minimum detection penalty and the maximum communication penalty. As λ increases, the curve will eventually reach and remain at the point $(J_d^\infty, 0)$ achieved by the myopic strategy in (5); let us denote by λ^* the smallest value of λ at this ending point. Finally, we compute a Monte-Carlo estimate of the detection penalty $J_d(\bar{\gamma}^*)$, sampling from environment $p(x,y)$ with parameters (w,r) and applying the optimal centralized strategy in (3).

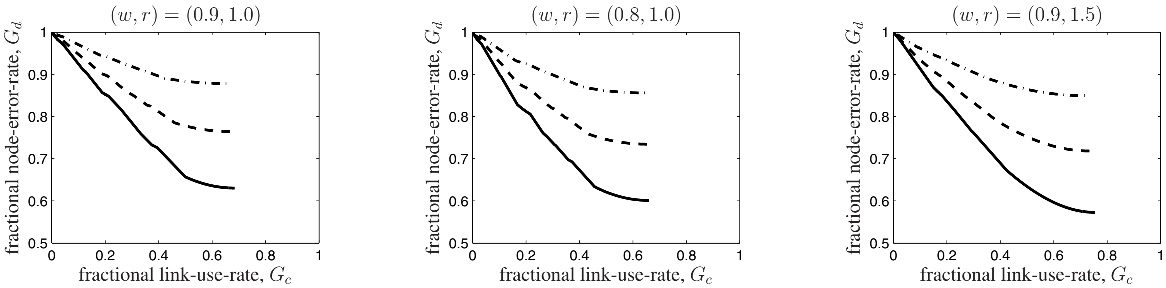
Another purpose of our experiments is to assess the offline communication overhead. We record the number of iterations to convergence each time we apply the message-passing algorithm on a distinct value of λ. Let k^* denote the empirical average number of iterations, taken over only the experimental runs with $\lambda < \lambda^*$ since the algorithm converges to myopic performance in typically two iterations otherwise. Per offline iteration, each link (i,j) must support the reliable transmission of $2|\mathcal{U}_i \times \mathcal{X}_i|$ real numbers: hence, per offline reorganization, the network in the model of Fig. 3 must reliably transmit on the order of $228k^*$ real numbers.

Fig. 5 displays the results over different environment parameters (w,r), each such environment over different channel parameters q. In every case, we increment λ in steps of size 3×10^{-4}, declare offline convergence at iteration k if $J(\gamma^{k-1}) - J(\gamma^k) < 10^{-3}$, and rely on 1000 samples in the Monte-Carlo estimate of $J(\bar{\gamma}^*)$. To compare results across different cases, we show the *normalized* tradeoff curve, or the set of points $\{(G_c^\lambda, G_d^\lambda); \lambda \geq 0\}$, where $G_c^\lambda = J_c^\lambda / \sum_i |\chi(i)|$ and $G_d^\lambda = 1 - [J_d^\infty - J_d^\lambda]/[J_d^\infty - J_d(\bar{\gamma}^*)]$ express the *fractional* link-use-rate (relative to full network capacity) and node-error-rate (relative to the gap between that of the myopic and centralized strategies), respectively.

V. CONCLUSION

A key challenge in modern sensor networks concerns the inherent design tradeoffs between application-layer decision performance and network-layer resource efficiency. We explored such tradeoffs for the decision-making objective of optimal Bayesian detection, assuming in-network processing constraints are dominated by a low-rate unreliable communication medium. Mitigating performance loss in the presence of such constraints demands an offline algorithm by which the processing rules local to all nodes are iteratively coupled in a manner driven by global problem statistics. We showed that, for a certain class of models, this offline algorithm admits an efficient message-passing interpretation: it can be implemented as a sequence of purely-local computations interleaved with only nearest-neighbor communications. The algorithm was successfully applied to a simulated network of binary detectors, emphasizing how design decisions to reduce online resource overhead by imposing explicit in-network processing constraints must be balanced with the offline resource expenditure to preserve satisfactory performance subject to such constraints.

It is interesting to speculate on the use of our offline algorithm for models in which not all required assumptions are satisfied (e.g., a non-tree-structured graph); while the elementary convergence proof presented here no longer applies, the algorithm may still yield quality approximations. A related open question is the robustness of the offline



q	λ^*	k^*
0.0	0.1368	4.5 ± 0.74
0.3	0.0957	4.3 ± 0.74
0.6	0.0549	3.9 ± 0.89

(a) Nominal Environment

q	λ^*	k^*
0.0	0.0774	4.4 ± 0.71
0.3	0.0543	4.2 ± 0.71
0.6	0.0312	3.9 ± 0.78

(b) Low State Correlation

q	λ^*	k^*
0.0	0.1056	4.7 ± 0.65
0.3	0.0762	4.1 ± 0.29
0.6	0.0462	4.0 ± 0.08

(c) Low Measurement Noise

Fig. 5. Normalized tradeoff curves for the model discussed in Fig. 3 given (a) a nominal environment, (b) low state correlation and (c) low measurement noise, each such environment with three different link erasure probabilities $q = 0$ (solid line), 0.3 (dashed line) and 0.6 (dash-dotted line). It is seen from e.g., the curve for $q = 0$ in (a) that a decentralized team strategy, constrained to selectively use at most eleven bits of communication (per online global measurement), can recover roughly 40% of the centralized detection performance lost by the purely myopic strategy, employing active transmission rates of roughly 70%. Altogether, we see the team strategy (i) gracefully degrades the node-error-rate as we increase the multiplier λ on the associated link-use-rate and (ii) consistently exploits the selective silence to convey up to an extra half-bit of information over each unit-rate link (i.e., "no news provides news," in the sense that active transmissions are employed only a fraction of the time even when $\lambda = 0$ in which case online communication is penalty-free). However, with k^* measuring between 4 and 5 iterations, achieving this satisfactory tradeoff depends upon the reliable communication of 912–1140 real numbers (per offline network reorganization). Note also that, all other things equal, λ^* is inversely related to q, quantifying the diminishing value of active transmission as link reliability degrades. Moreover, comparing the cases in (a) with those in (b) and (c), it appears lower state correlation or lower measurement noise similarly diminish the value of active transmission.

message-passing algorithm to inexact local computations or non-ideal nearest-neighbor communications.

VI. Acknowledgment

This work supported by the Army Research Office under contract W911NF-05-1-0207 and by the Air Force Office of Scientific Research under contract FA9550-04-1-0351. The authors are grateful for technical discussions with Professor John N. Tsitsiklis, Professor Pablo A. Parrilo, Jason L. Williams and Jerome Le Ny as well as the programming support provided by Michael Z. Chen.

References

[1] A. J. Goldsmith and S. B. Wicker, "Design challenges for energy-constrained ad hoc wireless networks," *IEEE Wireless Communications*, pp. 8–27, August 2002.

[2] F. Zhao, J. Liu, J. Liu, L. Guibas, and J. Reich, "Collaborative signal and information processing: an information-directed approach," *Proceedings of the IEEE: Special Issue on Sensor Networks and Applications*, vol. 91, no. 8, pp. 1199–1209, August 2003.

[3] C.-Y. Chong and S. P. Kumar, "Sensor networks: Evolution, opportunities, and challenges," *Proceedings of the IEEE: Special Issue on Sensor Networks and Applications*, vol. 91, no. 8, pp. 1247–1256, August 2003.

[4] A. M. Sayeed, D. Estrin, G. G. Pottie, and K. Ramchandran, "Guest editorial: Self-organizing distributed collaborative sensor networks," *IEEE Journal on Selected Areas in Communications*, vol. 23, no. 4, pp. 689–692, April 2005.

[5] J. N. Tsitsiklis, "Decentralized detection," in *Advances in Statistical Signal Processing*, H. V. Poor and J. B. Thomas, Eds. Greenwich, CT: JAI Press, 1993, vol. 2, pp. 297–344.

[6] P. K. Varshney, *Distributed Detection and Data Fusion*. New York, NY: Springer-Verlag, 1997.

[7] R. Viswanathan and P. K. Varshney, "Distributed detection with multiple sensors: Part I–Fundamentals," *Proceedings of the IEEE*, vol. 85, no. 1, pp. 54–63, January 1997.

[8] A. Pete, K. R. Pattipati, and D. L. Kleinman, "Optimization of decision networks in structured task environments," *IEEE Transactions on Systems, Man, and Cybernetics—Part A: Systems and Humans*, vol. 26, no. 6, pp. 739–748, November 1996.

[9] Z. Tang, K. Pattipati, and D. Kleinman, "Optimization of detection networks: Part II–Tree structures," *IEEE Transactions on Systems, Man and Cybernetics*, vol. 23, no. 1, pp. 211–221, January/February 1993.

[10] J. D. Papastavrou and M. Athans, "A distributed hypothesis-testing team decision problem with communications cost," in *Proceedings of the 25th IEEE Conference on Decision and Control*, December 1986, pp. 219–225.

[11] C. Rago, P. Willett, and Y. Bar-Shalom, "Censoring sensors: A low-communication-rate scheme for distributed detection," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 32, no. 2, pp. 554–568, April 1996.

[12] B. Chen, R. Jiang, T. Kasetkasem, and P. K. Varshney, "Channel aware decision fusion in wireless sensor networks," *IEEE Transactions on Signal Processing*, vol. 52, no. 12, pp. 3454–3458, December 2004.

[13] V. Saligrama, M. Alanyali, and O. Savas, "Distributed detection with packet losses and finite capacity links," *IEEE Transactions on Signal Processing*, 2006, prepint.

[14] S. L. Lauritzen, *Graphical Models*. New York, NY: Oxford University Press, 1996.

[15] A. S. Willsky, "Multiresolution markov models for signal and image processing," *Proceedings of the IEEE*, vol. 90, no. 8, pp. 1396–1458, August 2002.

[16] H. L. Van Trees, *Detection, Estimation, and Modulation Theory*. New York, NY: John Wiley & Sons, 1968, vol. 1.

[17] O. P. Kreidl and A. S. Willsky, "An efficient message passing algorithm for optimizing decentralized detection networks," MIT Laboratory for Information and Decision Systems, Cambridge, MA, Tech. Rep. LIDS-P-2726, December 2006.

[18] D. P. Bertsekas, *Nonlinear Programming*. Belmont, MA: Athena Scientific, 1995.