Decentralized Sequential Change Detection using Physical Layer Fusion

Leena Zacharias and Rajesh Sundaresan, *Senior Member, IEEE*

Abstract—The problem of decentralized sequential detection with conditionally independent observations is studied. The sensors form a star topology with a central node called fusion center as the hub. The sensors make noisy observations of a parameter that changes from an initial state to a final state at a random time where the random change time has a geometric distribution. The sensors amplify and forward the observations over a wireless Gaussian multiple access channel and operate under either a power constraint or an energy constraint. The optimal transmission strategy at each stage is shown to be the one that maximizes a certain Ali-Silvey distance between the distributions for the hypotheses before and after the change. Simulations demonstrate that the proposed analog technique has lower detection delays when compared with existing schemes. Simulations further demonstrate that the energy-constrained formulation enables better use of the total available energy than the power-constrained formulation in the change detection problem.

Index Terms—Ali-Silvey distance, change detection, correlation, Markov decision process, multiple access channel, sequential detection, sensor network.

I. INTRODUCTION

CONSIDER the use of a wireless sensor network for detection of a disruption or a change in environment. The change is required to be detected with minimum delay subject to a false alarm constraint. The standard medium access control and physical layer design for such a network (e.g., IEEE 802.15.4 standard) is one where sensors quantize their observations and send them to a fusion center via random access over a wireless Gaussian multiple-access channel (GMAC). The transmitted data are typically quantized individual log-likelihood ratios (LLR) of the hypotheses representing the environment before and after the change. The fusion center collects each sensor's LLR and adds them to get a fused statistic, if observations at sensors are independent conditioned on the state of the environment; this would be the case when the observation noises are additive and independent from sensor to sensor[1]. Such a design has a few drawbacks.

1) It does not exploit the spatial correlation in observations across sensors.
2) It does not exploit the superposition available on the GMAC.
3) It employs an ad hoc separation between quantization or compression on one hand, and transmission across the channel on the other; the latter requires adequate coding for noiseless reception and correct further processing at the fusion center.
4) It requires sufficient time slots for sensors to resolve all channel contentions[2].

Our goal in this paper is to detect change in environment in a manner that addresses the aforementioned drawbacks. Specifically, we consider a "star" topology of sensors. Sensors make an affine transformation of the observed data and transmit the output in an analog fashion over the GMAC. Given that observations at sensors at any instant are spatially correlated, only the sum of the LLRs is relevant to the decision maker, i.e., it is a sufficient statistic to decide on the change. By making the sensors simultaneously transmit an affine function of their LLRs in an analog fashion, and via distributed transmit beamforming, we exploit the spatial correlation in sensor data and the superposition available on the GMAC – the channel computes the required sum. Moreover, the analog data is in loose terms *matched* to the channel and does not require explicit channel coding. Finally, the sum is available at the fusion center in a single transmit duration unlike the situation in the random access case.

The biggest challenge in our proposed technique is the practicality of distributed transmit beamforming. The transmitters' clocks should be synchronized to some extent, so that carrier, phase, and symbol ticks align. A technique similar to the master-slave architecture proposed by Mudumbai, Barriac & Madhow [1] can be used to achieve this synchronization. The scheme exploits channel reciprocity in a time-division duplex (TDD) system.

1) Organization and preview of main results: In Section II, we formulate and solve a change detection problem under a power-constrained setting[3]. We arrive at a Markov decision problem framework and show that parameters of the affine transformation should minimize the variance of the combined observation and GMAC noises, which turns out to be a non-

[1]As we will see later, conditional independence notwithstanding, sensor observations are correlated.

[2]Alternatively, a time-division multiplexing protocol needs as many slots as there are sensors, and does not scale with the number of sensors.

[3]Sensors are usually powered by batteries with a fixed energy. The power-constrained model arises when this energy is evenly split over the desired life time of the sensor (in samples). An energy-constrained model arises when there is flexibility in how this energy is expended from sample to sample (subject to, of course, constraints imposed by the power amplifier).

convex optimization problem. We then provide an explicit algorithm to compute the optimal control parameters. Section III considers an energy-constrained setting. Section IV compares the simulation performance of our scheme with a previously known scheme. It also compares the energy-constrained formulation of Section III with the power-constrained formulation of Section II. Appendix A contains a new characterization of optimal control: maximize a certain Ali-Silvey distance [2] between the distributions of the fusion center's observation before and after the change. This is used to arrive at the minimum variance criterion of Section II.

2) Prior work: Change detection problems were solved in a centralized setting by Page [3], Lorden [4], and Shiryayev [5]. Shiryayev considered a Bayesian setting which is of relevance to our work. Veeravalli [6] solved the decentralized version of this problem with parallel error-free bit pipes of limited capacity from the sensors to the fusion center and identified the optimal stopping policy and quantizer structure. These results are analogous to those for hypothesis testing and sequential hypothesis testing (Tsitsiklis [7], Veeravalli et al. [8]). Prasanthi [9] considered access and decision delays in sequential detection over a random access channel, as it would be practically implemented using, for example, the IEEE 802.15.4 wireless personal area network standard. (See also [10]). Our work differs from those of Prasanthi and Veeravalli because we propose an analog transmission strategy.

Analog transmissions are optimal for transmission of a single Gaussian source over a Gaussian channel (Berger [11, p.100]) and a bivariate Gaussian source over a GMAC for a certain range of signal-to-noise ratios (SNR) (Lapidoth and Tinguely [12]), when a running estimate is required. Analog transmission via waveform design was considered by Mergen and Tong [13]. They used "type-based" multiple access to estimate a parameter over a GMAC. Their scheme, as does ours, exploits the superposition available in the GMAC. (See also [14], [15], [16], [17], [18], and [19] for analog transmission in other settings). Ertin and Potter [20] considered generalized cost functions which is mathematically analogous to our energy-constrained formulation.

II. PHYSICAL LAYER FUSION FRAMEWORK

A. Mathematical Formulation

$X \sim \mathcal{N}(\theta, \sigma^2)$ indicates that X is a Gaussian random variable with mean θ and variance σ^2.

(1) The state of nature is described by $\{\theta_k : k \in \mathbb{Z}_+\}$, a two-state discrete-time Markov chain taking values in $\{m_0, m_1\}$, with transition probabilities as described in Fig. 1(a)-(b). The quantities m_0 and m_1 denote, for example, the mean level of the observations before and after the disruption. The initial distribution for this Markov chain is obtained from $\Pr\{\theta_0 = m_1\} = \nu$. The change time Γ is $\mathbb{Z}_+$-valued, and given the event $\{\Gamma > 0\}$, Γ has the geometric distribution.

(2) The network has L sensors. At time k, sensor S_l makes an observation $X_{l,k} \sim \mathcal{N}(\theta_k, \sigma^2_{\text{obs},l})$, i.e., $X_{l,k} = \theta_k + Z_{l,k}$, where $Z_{l,k} \sim \mathcal{N}(0, \sigma^2_{\text{obs},l}), l = 1, \ldots, L$.

(3) The observations at each sensor are independent, conditioned on θ_k. Furthermore, the observations are independent
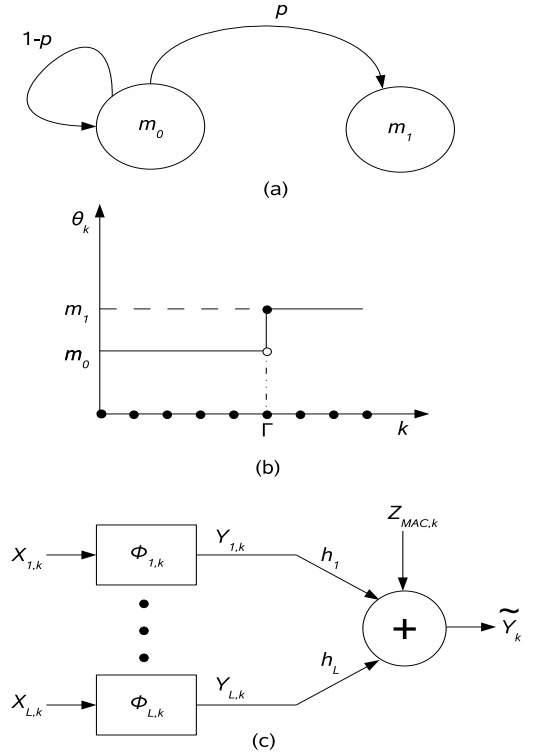

Fig. 1.   Problem set-up.

from sensor to sensor, conditioned on θ_k. Despite these conditional independence assumptions, we remark that $X_{l,k}, l = 1, \ldots L$, are correlated.

(4) Each sensor transmits $Y_{l,k} = \phi_{l,k}(X_{l,k})$; this being a function only of the observation at sensor l, our setting is a decentralized one. See Fig. 1(c). The function $\phi_{l,k}$ is affine:

$$\phi_{l,k}(x) = \alpha_{l,k}(x - c_{l,k}). \tag{1}$$

Quantities $\alpha_k = (\alpha_{1,k}, \ldots, \alpha_{L,k})$ and $c_k = (c_{1,k}, \ldots, c_{L,k})$ are parameters for optimal control. Transmission is done by setting the amplitude of an underlying unit-energy waveform to $Y_{l,k}$. All sensors use the same underlying waveform. The motivations for the analog amplify-and-forward transmissions in (1) are given in Section I: conditional independence of the observations given the state, and the Gaussian observation noise. If the latter does not hold, affine functions of LLRs instead of the direct observations could be sent ([21, Ch. 5]).

(5) The GMAC output at the fusion center when projected onto the common waveform yields

$$\widetilde{Y}_k = \sum_{l=1}^{L} h_l Y_{l,k} + Z_{\text{MAC},k},$$

where $Z_{\text{MAC},k} \sim \mathcal{N}(0, \sigma^2_{\text{MAC}})$ is independent and identically distributed (iid) across k, and is independent of all other quantities. The gain $h_l \in \mathbb{R}_+$ is the channel gain for the lth sensor and is deterministic. See Fig. 1(c). We assume perfect knowledge of the channel gains is available at the sensors and the fusion center. While this is not the case in practice, channel knowledge can be gleaned in time-division duplex (TDD)

systems that possess channel reciprocity (IEEE 802.15.4). See Mudumbai, Barriac & Madhow [1] for a suggested master-slave architecture. In a subsequent section, we study the effect of imperfect knowledge of these gains.

(6) At the fusion center, form $\widehat{Y}_k$ as follows:

$$
\begin{aligned}
\widehat{Y}_k &= \frac{1}{\sum_{l=1}^{L} h_l \alpha_{l,k}} \left(\widetilde{Y}_k + \sum_{l=1}^{L} h_l \alpha_{l,k} c_{l,k} \right) \\
&= \theta_k + \widehat{Z}_{\mathsf{MAC},k},
\end{aligned} \tag{2}
$$

where $\widehat{Z}_{\mathsf{MAC},k} \sim \mathcal{N}(0, \sigma_k^2)$ and

$$
\sigma_k^2 = \frac{\sum_{l=1}^{L} (\sigma_{\mathsf{obs},l} h_l \alpha_{l,k})^2 + \sigma_{\mathsf{MAC}}^2}{\left(\sum_{l=1}^{L} h_l \alpha_{l,k} \right)^2}. \tag{3}
$$

The quantity $\widehat{Y}_k$ in (2) is obtained from $\widetilde{Y}_k$ using a bijective mapping; so no information is lost. From (2), we also see that the distributed multi-sensor setting is equivalent to a centralized setting where the fusion center makes a direct (noisy) observation on θ_k with equivalent additive observation noise of variance σ_k^2 as given in (3). This is enabled by the affine nature of $\phi_{l,k}$. The centralized problem with constant σ_k^2 was studied by Shiryayev [5] with the aim of characterizing the stopping rule. The new aspect here is the dependence of σ_k^2 on the control parameters.

(7) The fusion center chooses an action $a_{k-1} \in \mathbb{A}$ at time $k-1$ from set $\mathbb{A}$ of actions (controls)

$$
\mathbb{A} = \{stop\} \cup \{(continue, \alpha, c) : \alpha \in \mathbb{R}_+^L, c \in \mathbb{R}^L\}.
$$

If $a_{k-1} = stop$, the fusion center stops. If $a_{k-1} = (continue, \alpha_k, c_k)$, the fusion center takes another sample (the kth), and all sensors transmit $\phi_{l,k}(X_{l,k})$ with parameters (α_k, c_k).

(8) As done by Veeravalli in [8], we assume a quasi-classical information structure, i.e., action a_{k-1} depends on

$$
i_{k-1} = \{a_0, \widehat{y}_1, a_1, \widehat{y}_2, \dots, a_{k-2}, \widehat{y}_{k-1}\}. \tag{4}
$$

Even though the sensors may have local memory of past observations, our framework does not make use of this additional information.[4] The fusion center feeds back the action parameters a_{k-1} to the sensors. (We use the following notation: the quantity i_{k-1} in (4) is a realization of the random variable I_{k-1} and takes values in the set $\mathbb{I}_{k-1}$. We set $\mathbb{I}_0 = \emptyset$.)

(9) Average power constraint at sensor l is

$$
\mathbb{E}\left[\alpha_{l,k}^2 (X_{l,k} - c_{l,k})^2 \,|\, I_{k-1} \right] \le P_l,
$$

i.e.,

$$
\alpha_{l,k}^2 \left[\sigma_{\mathsf{obs},l}^2 + \mathbb{E}\left[(\theta_k - c_{l,k})^2 \,|\, I_{k-1} \right] \right] \le P_l, \ l = 1, \dots, L. \tag{5}
$$

The set of feasible controls, given $I_{k-1} = i_{k-1}$, is denoted by

$$
\mathbb{A}(i_{k-1}) =
$$
$$
\{stop\} \cup \{(continue, \alpha, c) : (\alpha, c, i_{k-1}) \text{ satisfies (5)}\}. \tag{6}
$$

In Section III, we relax the constraint in (5) and impose an expected total energy constraint.

---

[4]Veeravalli [8, p.434] discusses other information structures and why they may be difficult to analyze.

(10) The fusion center policy π is a sequence of proposed (deterministic) actions $\pi = (\pi_{k-1}, k \ge 1)$, where π_{k-1} is a function $\pi_{k-1} : \mathbb{I}_{k-1} \to \mathbb{A}$. In particular, $\pi_{k-1}(i_{k-1}) = a_{k-1} \in \mathbb{A}(i_{k-1})$. Each policy π induces a probability measure. All expectations are with respect to this measure. The dependence of the expectation operation on π is understood and suppressed.

(11) τ is the first instant when the fusion center decides to stop.

The problem we wish to solve is the following:

Problem 1: (Change detection with delay penalty) Minimize over all admissible policies the expected detection delay, $E_{\mathsf{DD}} = \mathbb{E}\left[(\tau - \Gamma)^+ \right]$, subject to an upper bound on the probability of false alarm $P_{\mathsf{FA}} \le \delta$, where $x^+ = \max(0, x)$, and $P_{\mathsf{FA}} = \Pr\{\tau < \Gamma\}$.

The solution to Problem 1 is obtained via a solution to Problem 2 (below) for a particular $\lambda > 0$ (Shiryayev [5]). The quantity λ may be interpreted as the cost of unit delay.

Problem 2: (Change detection with a Bayes cost) Minimize over all admissible policies

$$
\begin{aligned}
R(\lambda) &= P_{\mathsf{FA}} + \lambda E_{\mathsf{DD}} = \Pr\{\Gamma > \tau\} + \lambda \mathbb{E}\left[(\tau - \Gamma)^+ \right] \\
&= \mathbb{E}\left[1\{\theta_\tau = m_0\} + \sum_{k=0}^{\tau-1} \lambda 1\{\theta_k = m_1\} \right]
\end{aligned} \tag{7}
$$

where $\lambda > 0$ and $\mathbb{E}$ is under the probability measure induced by the chosen policy.

The cost function is additive over time. The first term within the expectation in (7) is the terminal cost; the terms in the summation a running cost. At each stage the state θ_k evolves in a Markov fashion. The controller sees only a noisy version $\widehat{Y}_k$ of the state, but can control the observation noise variance σ_k^2 via α and c. It can also stop at any stage and pay a terminal cost. Any decision affects the future evolution of the cost process. Such problems are Markov decision problems (MDP) with partial observations. They can be analyzed by studying an equivalent complete observation MDP[5] with a reduced (posterior) state $\mu_k \stackrel{\triangle}{=} \mathbb{E}[1\{\theta_k = m_1\} \,|\, I_k] = \Pr\{\Gamma \le k \,|\, I_k\}$. The probability law for $\{\mu_k : k \ge 0\}$ is given as follows: $\mu_0 = \Pr\{\Gamma \le 0 \,|\, I_0\} = \nu$, and the law for μ_k, under $a_k = (continue, \alpha_{k+1}, c_{k+1})$, is (see Veeravalli [6, eqn. (9)])

$$
\begin{aligned}
\mu_{k+1} &= \frac{\beta_k f_{m_1, \alpha_{k+1}}\left(\widehat{Y}_{k+1}\right)}{\beta_k f_{m_1, \alpha_{k+1}}\left(\widehat{Y}_{k+1}\right) + (1 - \beta_k) f_{m_0, \alpha_{k+1}}\left(\widehat{Y}_{k+1}\right)} \\
&\stackrel{\triangle}{=} \frac{g\left(\widehat{Y}_{k+1}, \alpha_{k+1}, \mu_k\right)}{h\left(\widehat{Y}_{k+1}, \alpha_{k+1}, \mu_k\right)} \stackrel{\triangle}{=} \psi\left(\widehat{Y}_{k+1}, \mu_k, \alpha_{k+1}\right),
\end{aligned} \tag{8}
$$

where $\beta_k \stackrel{\triangle}{=} \Pr\{\Gamma \le k+1 | I_k\} = \mu_k + (1 - \mu_k)p$, and $f_{m_i, \alpha_{k+1}}$ is the density of an $\mathcal{N}(m_i, \sigma_{k+1}^2)$ random variable. The quantities h and g are as in (8); h is the density of $\widehat{Y}_{k+1}$

---

[5]See Shiryayev [5], Veeravalli [6] for results with stopping, Bertsekas & Shreve [22, Ch. 10] for discounted costs, and Bertsekas [23, Ch. V].

given (I_k, a_k), and g is a scaled density. The power constraint (5) when written for time $k+1$ simplifies to

$$\alpha_{l,k+1}^2 \left[\sigma_{\mathsf{obs},l}^2 + (m_0 - c_{l,k+1})^2(1 - \beta_k) \right.$$
$$\left. + (m_1 - c_{l,k+1})^2 \beta_k \right] \leq P_l. \qquad (9)$$

The set of feasible controls in (6) depends on i_k only through μ_k and can be simplified to

$$\mathbb{A}(\mu) = \{stop\} \cup \{(continue, \alpha, c) : (\alpha, c, \mu) \text{ satisfies } (9)\},$$

where $\mathbb{A}(\cdot)$ is re-used to denote the set of feasible controls for the equivalent complete observation MDP. Let $\mathbb{A}'(\mu) = \{(\alpha, c) : (continue, \alpha, c) \in \mathbb{A}(\mu)\}$ denote the set of control parameters when the action is to continue. Now consider the objective function. Taking conditional expectations with respect to the information process, (see Shiryayev [5, pp.195–196]), (7) reduces to

$$R(\lambda) = \mathbb{E} \left[(1 - \mu_\tau) + \sum_{k=0}^{\tau-1} \lambda \mu_k \right]. \qquad (10)$$

Minimization of (10) is done via dynamic programming. Some additional remarks are in order.

Remarks: 1. The variance σ_{k+1}^2 depends on α_{k+1} as shown in (3), and hence the dependence on α_{k+1} in (8). μ_{k+1} depends on c_{k+1} only through α_{k+1} because of the processing done in (2).

2. If the running cost is λ instead of $\lambda 1\{\theta_k = m_1\}$ in (7), every sample costs λ units, not just those beyond the change point that contribute to the delay. This is a minor variation to Problem 2 and has a similar solution.

3. Another variation is sequential hypothesis testing: set the transition probability $p = 0$, enhance the action *stop* to $(stop, \hat{\theta})$, where $\hat{\theta}$ is the decision (either m_0 or m_1), and set the terminal cost to $1\{\theta_\tau \neq \hat{\theta}\}$. The running cost is a constant λ for every sample.

B. Optimal Policy

As is usual with such problems, we first restrict the stopping time τ to a finite horizon T. Using Bertsekas's result [23, Ch.1, Prop.3.1], the cost-to-go function recursions are written as

$$J_T^T(\mu_T) = 1 - \mu_T,$$
$$J_k^T(\mu_k) = \min \left\{ 1 - \mu_k, \lambda \mu_k + A_k^T(\mu_k) \right\}, \ 0 \leq k < T,$$
$$A_k^T(\mu) = \min_{(\alpha,c) \in \mathbb{A}'(\mu)} \mathbb{E} \left[J_{k+1}^T \left(\psi \left(\hat{Y}, \mu, \alpha \right) \right) \right]$$
$$= \min_{(\alpha,c) \in \mathbb{A}'(\mu)} \int_{\mathbb{R}} J_{k+1}^T \left(\frac{g\left(\hat{y}, \alpha, \mu\right)}{h\left(\hat{y}, \alpha, \mu\right)} \right) h\left(\hat{y}, \alpha, \mu\right) d\hat{y}.$$

To solve Problem 2, let $T \to \infty$. From results in [8] and [6], the limit in (11) below exists, does not depend on k (i.e., the policy is stationary), and defines the infinite horizon cost-to-go function:

$$J(\mu) = \lim_{T \to \infty} J_k^T(\mu) = \min \left\{ 1 - \mu, \lambda \mu + A_J(\mu) \right\}, \qquad (11)$$

where

$$A_J(\mu) = \min_{(\alpha,c) \in \mathbb{A}'(\mu)} \mathbb{E} \left[J \left(\psi \left(\hat{Y}, \mu, \alpha \right) \right) \right]. \qquad (12)$$

The following lemma enables a characterization of the optimal stopping policy.

Lemma 1: The functions $J_k^T(\mu)$ and $A_k^T(\mu)$ are non-negative and concave functions of μ, for $\mu \in [0, 1]$. Moreover, $A_k^T(1) = J_k^T(1) = 0$. Similarly, the functions $J(\mu)$ and $A_J(\mu)$ are non-negative and concave functions of μ, for $\mu \in [0, 1]$, and $A_J(1) = J(1) = 0$.

The proof is the same as that in Bertsekas [23, p. 268] for sequential hypothesis testing. The concavity of $A_J(\mu)$ and (11) imply the following theorem (Shiryayev [5], Veeravalli [6]).

Theorem 2: An optimal fusion center policy has stopping time τ given by $\tau = \inf\{k : \mu_k \geq \mu^*\}$, where μ^* is the unique solution to $\lambda \mu + A_J(\mu) = 1 - \mu$.

To summarize, the optimal detection strategy at time k is as follows. Convert the received signal $\tilde{Y}_k$ into the posterior probability of change μ_k using (2) and (8). If μ_k exceeds a threshold, declare that a change has occurred. Otherwise, make the sensors transmit another sample using parameters α, c chosen optimally as described in the next subsection.

C. Parameters for Optimal Control

We begin this section with an algorithm that calculates the optimal α.

Algorithm 1: Let

$$\sigma_{\mathsf{obs},1}^2 h_1 \alpha_{\mathsf{max},1} \leq \cdots \leq \sigma_{\mathsf{obs},L}^2 h_L \alpha_{\mathsf{max},L},$$

where the quantity

$$\alpha_{\mathsf{max},l} = \left(P_l / \left(\sigma_{\mathsf{obs},l}^2 + (m_1 - m_0)^2 \beta(1 - \beta) \right) \right)^{1/2}$$

with $\beta = \mu + (1 - \mu)p$.

- **Step 1**: Find the unique $k \in \{1, \ldots, L-1\}$ that satisfies

$$\sigma_{\mathsf{obs},k}^2 h_k \alpha_{\mathsf{max},k} \leq \frac{\sum_{l=1}^k (\sigma_{\mathsf{obs},l} h_l \alpha_{\mathsf{max},l})^2 + \sigma_{\mathsf{MAC}}^2}{\sum_{l=1}^k h_l \alpha_{\mathsf{max},l}}$$
$$\leq \sigma_{\mathsf{obs},k+1}^2 h_{k+1} \alpha_{\mathsf{max},k+1} \qquad (13)$$

  if it exists. Otherwise, set $k = L$.
- **Step 2**: Set the optimal α as follows.

$$a^* = \sum_{l=1}^k h_l \alpha_{\mathsf{max},l} + \frac{\sum_{l=k+1}^L \sigma_{\mathsf{obs},l}^{-2}}{\sum_{l=1}^k h_l \alpha_{\mathsf{max},l}}$$
$$\cdot \left(\sum_{l=1}^k (\sigma_{\mathsf{obs},l} h_l \alpha_{\mathsf{max},l})^2 + \sigma_{\mathsf{MAC}}^2 \right), \qquad (14)$$

$$\alpha_m = \alpha_{\mathsf{max},m}, \ 1 \leq m \leq k,$$
$$\alpha_m = \frac{1}{\sigma_{\mathsf{obs},m}^2 h_m} \cdot \frac{a^* - \sum_{l=1}^k h_l \alpha_{\mathsf{max},l}}{\sum_{l=k+1}^L \sigma_{\mathsf{obs},l}^{-2}}, \ k < m.$$
$$(15)$$

The optimal choice sets amplitudes of the k sensors with the k least scaled observation noise variance $(\sigma_{\mathsf{obs},l}^2 h_l \alpha_{\mathsf{max},l})$ to $\alpha_{\mathsf{max},l}$. The remaining sensors' amplitudes are appropriately

chosen smaller values. Intuitively, sensors $l = k + 1, \ldots, L$ have so good a channel that scaling by $\alpha_{\max,l}$ for these sensors will amplify the observation noise leading to a larger overall noise variance. Note that when all channel gains, observation variances, and power constraints are equal, $\alpha_l = \alpha_{\max}$ for all sensors. This special case was earlier proved in [24].

Theorem 3: The choice of $c_l = m_1 \beta + m_0 (1 - \beta)$, $l = 1, \ldots, L$, and α according to Algorithm 1 constitute the optimal controls that minimize (12).

Proof: **Step 1**: We prove that the optimal control minimizes the variance (3). Consider α and α' with resulting variances $\sigma^2 < \sigma'^2$. From the second equality in (2) we have

$$\widehat{Y}(\alpha) = \theta + \sigma Z, \qquad (16)$$

$$\widehat{Y}(\alpha') = \theta + \sigma' Z' \quad \sim \quad \theta + \sigma Z_1 + (\sigma'^2 - \sigma^2)^{1/2} Z_2, \quad (17)$$

where Z, Z_1, Z_2 are iid $\mathcal{N}(0,1)$ with $Z' = Z_1 + Z_2$. The time index k is understood.

From (16) and (17), $\widehat{Y}(\alpha')$ is a stochastically degraded version of $\widehat{Y}(\alpha)$ and is equivalent to an additional random processing on $\widehat{Y}(\alpha)$. Theorem 5 in Appendix A shows that

$$1 - \mathbb{E}_h \left[J \left(\frac{g\left(\widehat{Y}(\alpha), \alpha, \mu\right)}{h\left(\widehat{Y}(\alpha), \alpha, \mu\right)} \right) \right]$$

is an *Ali-Silvey distance* between two probability measures. In $\mathbb{E}_h$ the dependence of h on α is understood and suppressed. Ali-Silvey distances have a well-known monotonicity property: data processing, whether deterministic or random, cannot increase the dissimilarity measure between two distributions ([2], [25]). This property implies that

$$\mathbb{E}_h \left[J \left(\frac{g\left(\widehat{Y}(\alpha), \alpha, \mu\right)}{h\left(\widehat{Y}(\alpha), \alpha, \mu\right)} \right) \right] \leq \mathbb{E}_h \left[J \left(\frac{g\left(\widehat{Y}(\alpha'), \alpha', \mu\right)}{h\left(\widehat{Y}(\alpha'), \alpha', \mu\right)} \right) \right].$$

It follows that minimization of the variance in (3) is the criterion for getting the optimal α.

Step 2: We now identify the optimal c. The minimization mentioned in the previous step should be done subject to the power constraint given in (5), which can be rewritten as

$$\alpha_{l,k}^2 \leq P_l \cdot \left[\sigma_{\mathsf{obs},l}^2 + \mathbb{E}\left[(\theta_k - c_{l,k})^2 \,|I_{k-1}\right] \right]^{-1}. \qquad (18)$$

The constraint set is enlarged if the upper bound in (18) is higher. We should therefore choose the $c_{l,k}$ that minimizes $\mathbb{E}\left[(\theta_k - c_{l,k})^2 \,|I_{k-1}\right]$, i.e., $c_{l,k}$ is the minimum mean squared error (MMSE) estimate of θ_k given I_{k-1}. Clearly this is given by $c_{l,k} = \mathbb{E}\left[\theta_k | I_{k-1}\right] = m_1 \beta_{k-1} + m_0 (1 - \beta_{k-1})$, and is independent of l. Moreover,

$$\mathbb{E}\left[(\theta_k - c_{l,k})^2 \,|I_{k-1}\right] = \mathsf{Var}\left\{\theta_k | I_{k-1}\right\}$$
$$= (m_1 - m_0)^2 \beta_{k-1}(1 - \beta_{k-1}),$$

and (18) can be written as $\alpha_{l,k} \leq \alpha_{\max,l,k}$, where

$$\alpha_{\max,l,k} = \left(P_l / \left(\sigma_{\mathsf{obs},l}^2 + (m_1 - m_0)^2 \beta_{k-1}(1 - \beta_{k-1}) \right) \right)^{1/2}.$$
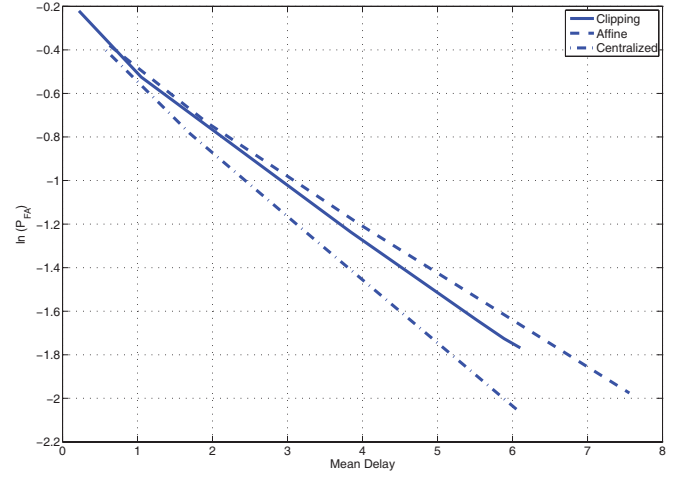

Fig. 2. Performance curves: 1) Clipped transmission via a sigmoidal function 2) Affine transformation 3) Centralized, where all sensor data is available without noise at the fusion center.

Step 3: Ignoring the time index k, the optimization problem to obtain the best α is:

Problem 3: Minimize

$$\left(\sum_{l=1}^{L} h_l \alpha_l \right)^{-2} \left[\sum_{l=1}^{L} (\sigma_{\mathsf{obs},l} h_l \alpha_l)^2 + \sigma_{\mathsf{MAC}}^2 \right],$$

where $\alpha_l \in [0, \alpha_{\max,l}]$ for $l = 1, \cdots, L$.

This is not a convex optimization problem. However, we can split it into two simpler convex optimization problems to get an explicit solution to Problem 3.

Lemma 4: Algorithm 1 solves Problem 3.

See Appendix B for a proof. This concludes the proof of Theorem 3. ∎

Under the restriction of affine controls, Theorem 3 describes the optimal choice. However, affine controls are not optimal in general. This is demonstrated in Fig. 2 where a piece-wise linear sigmoidal control outperforms the optimal affine control (see [21, Sec. 2.7]). It would be interesting to see if there are ranges of $\sigma_{\mathsf{obs},l}^2$ and σ_{MAC}^2 where the affine control is indeed optimal. We do not pursue this question in this work.

We now make some remarks on the complexity of overall detection. Theorem 3 says that the parameters for optimal control are obtained via a finite step procedure. Indeed, Algorithm 1 gives the output in time linear in the number of sensors, and is therefore easy to execute. The threshold calculation for a fixed set of parameters is a one time calculation and is obtained via the so-called *value iteration procedure* which yields an approximation. We now explore further simplifications with reduced feedback information.

D. A Simpler Suboptimal Policy

Let us now restrict the controls to be of the following form: the decision to stop or continue, say b_k, depends on I_k, but the parameters of the affine transformation at time $k + 1$ can only depend on I_0 and $b_k \in \{stop, continue\}$. I_0 denotes the

prior information before any observations are made and b_k is the decision of the fusion center at k. Note that this reduces the amount of feedback to simply the binary random variable b_k.

The structure of the controls is similar to that of the optimal policy of the previous section, but with

$$\beta_k = \Pr\{\Gamma \leq k + 1|I_0\} = 1 - (1 - \nu)(1 - p)^{k+1}$$

so that (α, c) depends on only I_0 and not on I_k. The stopping policy is chosen as in Theorem 2. As we see in simulation results presented in Section IV, the performance of this algorithm is close to optimal for the chosen parameters, yet requires feedback of only one bit at each stage.

III. ENERGY-CONSTRAINED FORMULATION

The energy-constrained problem is stated as follows.

Problem 4: Minimize the expected detection delay, E_{DD}, subject to an upper bound on the probability of false alarm, $P_{\mathsf{FA}} \leq \delta$, and an upper bound on the expected energy spent,

$$\mathbb{E}\left[\sum_{k=1}^{\tau} \mathbb{E}\left[\phi_{l,k}^2(X_{l,k})|I_{k-1}\right]\right] \leq E_l, \quad l = 1, 2, \ldots, L. \quad (19)$$

Let $\lambda = (\lambda_1, \ldots, \lambda_L, \lambda_{L+1})$. As before, to solve Problem 4, we set up the Bayes cost $R(\lambda)$ and minimize it over all admissible choices of stopping policy and the parameters $\alpha_{l,k}$ and $c_{l,k}$ of the affine transformation $\phi_{l,k}$. The Bayes cost can be written as

$$R(\lambda) = \mathbb{E}\Big[(1 - \mu_\tau) + \lambda_{L+1}\sum_{k=0}^{\tau-1}\mu_k$$
$$+ \sum_{k=1}^{\tau}\sum_{l=1}^{L}\lambda_l\mathbb{E}\left[\alpha_{l,k}^2(X_{l,k} - c_{l,k})^2|I_{k-1}\right]\Big].$$

A result analogous to Theorem 2 in Section II-B holds, and the optimal control at time $k + 1$, given I_k, is such that c_{k+1} is independent of l, the sensor index. More precisely,

$$c_{k+1} = m_1\beta_k + m_0(1 - \beta_k), \quad l = 1, \ldots, L,$$

$$\alpha_{k+1} = \arg\min_{\alpha\in\mathbb{R}_+}\left[\sum_{l=1}^{L}\lambda_l\alpha_l^2\left(\sigma_{\mathsf{obs},l}^2 + (m_1 - m_0)^2\beta_k(1 - \beta_k)\right)\right.$$
$$\left. + \int_{\mathbb{R}} J\left(\frac{g(\hat{y}, \alpha, \mu_k)}{h(\hat{y}, \alpha, \mu_k)}\right)h(\hat{y}, \alpha, \mu_k)\,d\hat{y}\right],$$

where $J(\mu) = \min\{1 - \mu, \lambda_{L+1}\mu + A_J(\mu)\}$, is the infinite horizon cost-to-go function with

$$A_J(\mu) = \min_{\alpha\in\mathbb{R}_+^L}\Big[\sum_{l=1}^{L}\lambda_l\alpha_l^2\left(\sigma_{\mathsf{obs},l}^2 + (m_1 - m_0)^2\beta(1 - \beta)\right)$$
$$+ \int_{\mathbb{R}} J\left(\frac{g(\hat{y}, \alpha, \mu)}{h(\hat{y}, \alpha, \mu)}\right)h(\hat{y}, \alpha, \mu)\,d\hat{y}\Big].$$

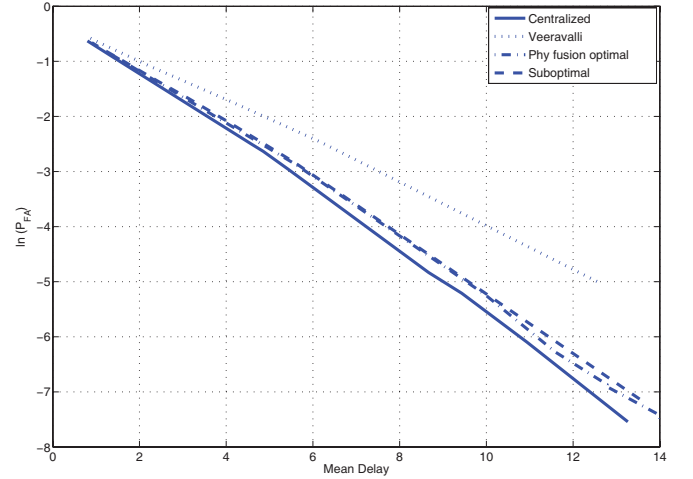A minimizing control α does exist as is shown in [21, Sec. 3.1].

Fig. 3. Comparison of our algorithms with Veeravalli's scheme. The "centralized" performance curve is for the case when all sensor data is available without noise at the fusion center.

IV. COMPARISONS AND PRACTICAL CONSIDERATIONS

A. Benefits from Exploiting Sensor Correlation

Veeravalli [6] addresses the structure of optimal D_l-level quantizer at sensor $S_l, l = 1, 2, \ldots, L$. His model is applicable to a system that allows $\log_2 D_l$ bits to be sent error-free from sensor S_l to the fusion center. For simplicity let $D_l = D, l = 1, 2, \ldots, L$. In order to show the benefit of exploiting correlation of observations when transmitting across the GMAC, we do the following. The quantized bits from the sensors in Veeravalli's scheme are transmitted using an optimal scheme designed for independent data streams over a coherent GMAC. If all sensors operate at the same transmission power, the SNR required to support such a transmission on the GMAC satisfies the sum rate constraint $L\log_2 D \leq (1/2)\log_2(1 + L \cdot \mathsf{SNR})$, and thus

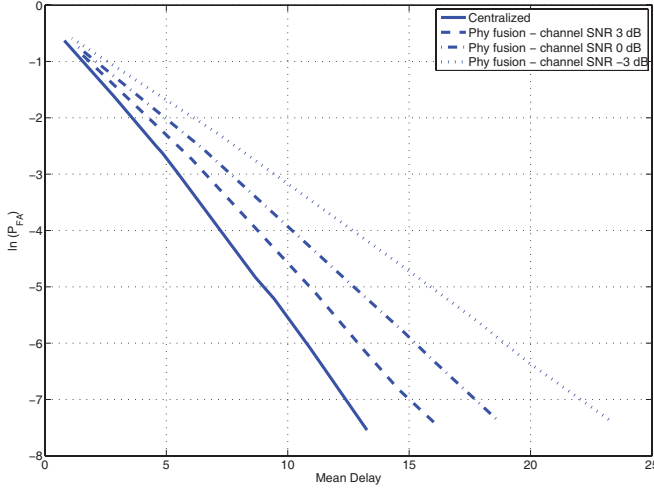$$\mathsf{SNR} \geq \frac{D^{2L} - 1}{L}. \quad (20)$$

For the simulations, we assume two sensors ($L = 2$) with equal gains, i.e., $h_l = 1$ for $l = 1, 2$. We also assume one-bit quantizers ($D = 2$). From (20) we get $\mathsf{SNR} \geq 7.5$. Algorithms operate at $\mathsf{SNR} = 7.5$ with $P_l = 7.5$ for $l = 1, 2$ and $\sigma_{\mathsf{MAC}}^2 = 1$. We now summarize the other simulation assumptions which will be used unless stated otherwise.

Simulation Setup 1: Consider $L = 2$ sensors with $\mathcal{N}(0, 1)$ and $\mathcal{N}(0.75, 1)$ observations before and after the change, respectively. The geometric parameter $p = 0.05$ and the initial probability of change $\nu = 0$. We obtain $J(\mu)$ via value iteration procedure until the difference between successive iterates falls below 0.0001 with 1000 points on the μ axis. All simulations assume $P_l = P$ and $\sigma_{\mathsf{obs},l}^2 = 1$ for $l = 1, 2$.

Fig. 3 shows that both our algorithms give lesser delays than Veeravalli's algorithm that is naively overlaid on the GMAC. Furthermore, the suboptimal policy of Section II-D degrades from that in Section II-B only for low false alarm probabilities.

In Veeravalli's algorithm, $D - 1$ thresholds ($\in \mathbb{R}^{D-1}$) and a decision to stop or continue are fed back to each sensor. Our scheme requires feedback of $\alpha_l \in \mathbb{R}_+, c_l \in \mathbb{R}$, and the binary

Fig. 4. Performance curves for channel SNR $= \infty, 3, 0, -3$ dB.

Fig. 5. Performance curves for observation SNR $= -1, -2.5, -4$ dB.

decision. Even simpler is the strategy in Section II-D; only a binary decision is fed back.

The network delay is independent of the number of sensors in both our algorithms; the performance improves with increasing number of sensors. Veeravalli's scheme on the other hand requires an exponential growth in **SNR** (with L, as in (20)) to maintain the same delay versus P_{FA} performance. Our algorithms need a higher level of time and frequency synchronization of the transmitters for beamforming. Section IV-D studies the effect of lack of perfect channel knowledge. Transmit beamforming can be achieved via uplink-downlink reciprocity in a static time-division duplex (TDD) system (see [1] for an example mechanism).

B. Performance Comparisons Under Different Channel and Observation SNRs

We now portray performance under three different settings.

- Fig. 4 shows performance for various channel SNRs $P/\sigma^2_{\mathsf{MAC}}$; the other parameters remain as in Simulation Setup 1.
- Fig. 5 shows performance for various observation SNRs $(m_1 - m_0)^2/\sigma^2_{\mathsf{obs}}$ when the channel SNR $P/\sigma^2_{\mathsf{MAC}}$ is fixed at 3 dB.
- Fig. 6 compares the symmetric and asymmetric channel gain cases. The symmetric curve is obtained with $h_l = 1$ for $l = 1, 2$, and the asymmetric one with $h_1 = 1$ and $h_2 = 0.75$. The weaker sensor is 2.5 dB lower than the stronger one.

The plots show graceful degradation with decreasing SNR with results along expected lines.

C. Comparison of Power- and Energy-Constrained Formulations

For $P_{FA} \le e^{-4}$, we first identify the minimum time to detect change as a function of the energy constraint. This yields a power constraint for the constant power formulation. We then compare the delays incurred by the optimal algorithm under the two formulations in Fig. 7. We use the parameters in Simulation Setup 1 and $h_l = 1$ for all sensors. For the same
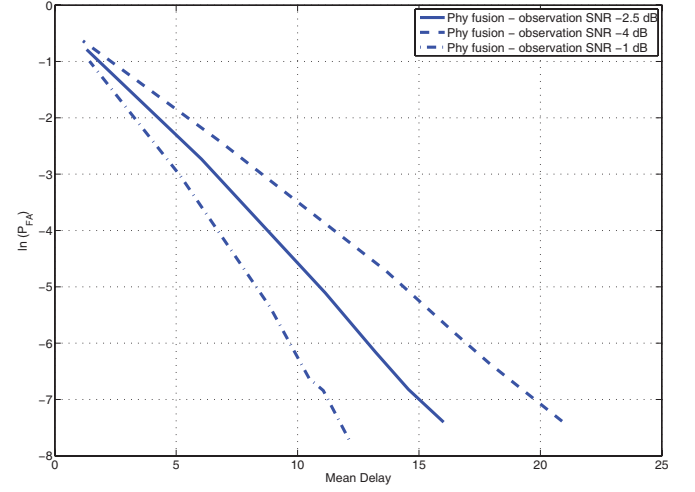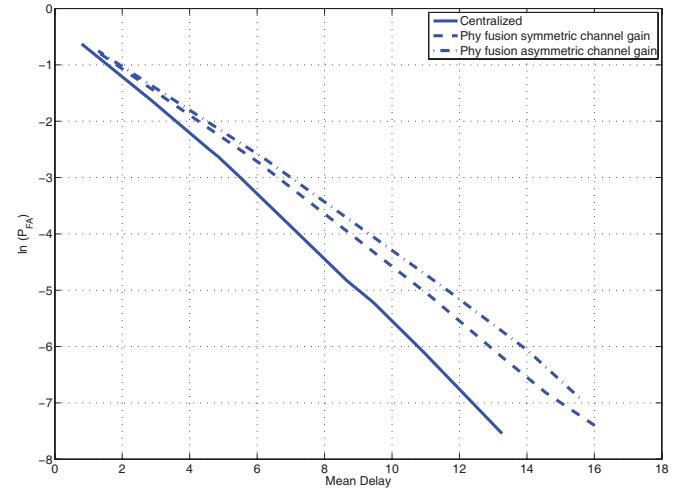

Fig. 6. Performance curves when 1) centralized (no channel noise) 2) symmetric channel gains 3) asymmetric channel gains with the weaker sensor 2.5 dB lower.

P_{FA}, the energy-constrained solution declares a change with lesser delay than the constant power solution.

As an illustration, we plot in Fig. 8 the variation of α^2, c, and μ with time in both the algorithms for a representative sample path. The change point is at 21 samples, shown using a dotted vertical grid line. The energy-constrained solution is more energy efficient because it uses lower energy (α^2) before and higher energy after the change point. Indeed, based on the prior information, the first few samples use negligible energy.

D. Channel Estimation Errors

Thus far we assumed a static channel with perfect knowledge available at both transmitter and receiver. Wireless channels, however, change with time. Only an estimate of the channel, based on signal processing on the pilots, beacons, or preambles, may be available. In this section, we study the effect of imperfect channel knowledge on the physical layer fusion algorithm.

To arrive at a model for channel errors, we consider complex channel gains over the GMAC with noise given by $Z_{\mathsf{MAC},k} \sim$
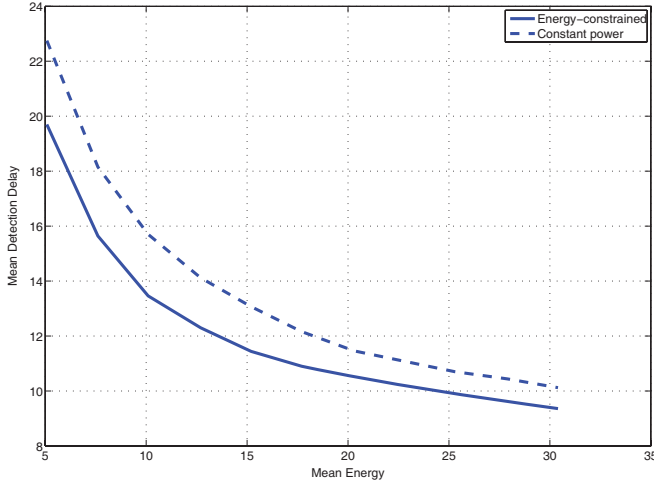
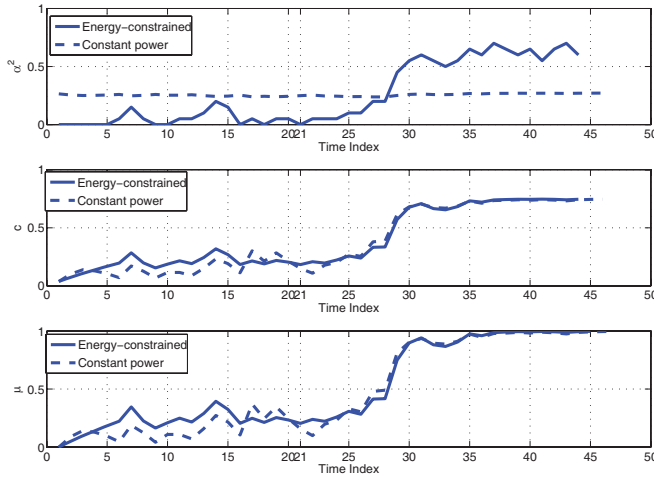Fig. 7. Comparison of constant power method and energy-constrained method.

Fig. 8. α^2, c, and μ of constant power method and energy-constrained method for a sample path.
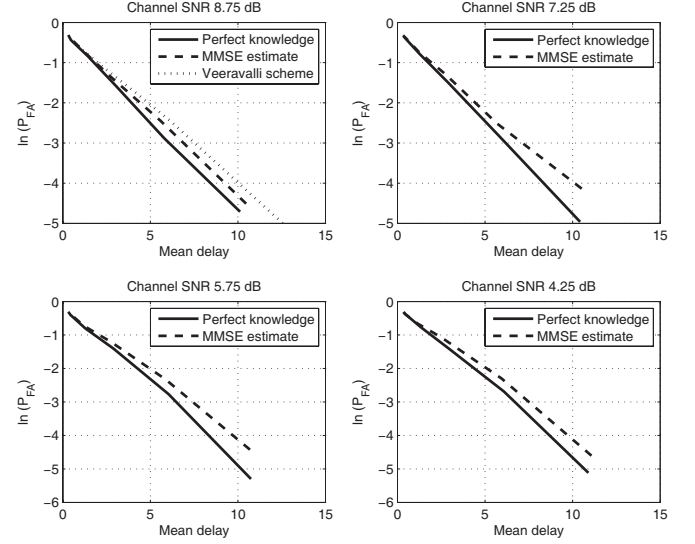

Fig. 9. Performance curves comparing the cases when 1) channel is perfectly known 2) MMSE estimates are used (Pilot SNR = Channel SNR).
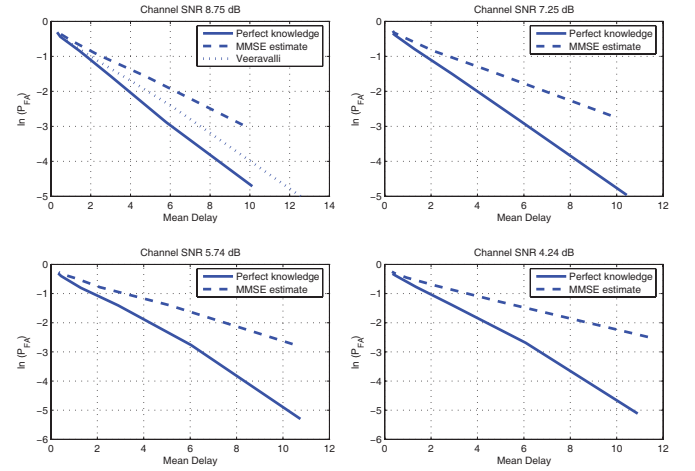

Fig. 10. Performance curves comparing the cases when 1) channel is perfectly known 2) MMSE estimates are used (Pilot SNR is 8.75 dB lower than Channel SNR).

$\mathbb{CN}(0, \sigma_{\mathsf{MAC}}^2)$, a circular symmetric complex Gaussian random variable. The observations are real-valued, but the complex baseband equivalent signal has two real-valued degrees of freedom per sample, leading to a bandwidth expansion factor of two. Suppose that the sensors use transmit beamforming[6], i.e., $\alpha_l = \frac{h_l^*}{|h_l|}\gamma_l$. Then it is sufficient to preserve only the real part of the received signal at the fusion center, and the problem reduces to that studied in the earlier parts of this paper with σ_{MAC}^2 replaced by $\sigma_{\mathsf{MAC}}^2/2$ in Section II. The quantity γ_l replaces α_l and $|h_l|$ replaces h_l in Algorithm 1. The output of the algorithm is γ_l.

Let $\{h_l\}$ be a sequence of $\mathbb{CN}(0,1)$ random variables that obey a block-fading model, i.e., the channel remains constant for T uses and then changes to an independent channel gain. If K of these T samples are available for channel estimation, then the MMSE estimate of the channel is $\hat{h}_l = (h_l + rZ)/(1+ r^2)$, where $r = \sigma_{\mathsf{MAC}}/\sqrt{KP_l}$, P_l is the power of sensor l and $Z \sim \mathbb{CN}(0,1)$. This is estimated at both ends (using TDD system's channel reciprocity).

[6]The optimality of cooperative transmit beamforming by sensors remains an open question.

Figures 9 and 10 show performance of the policy of Section II-C with $\hat{h}$ used in place of actual h, across different channel SNRs. Simulation Setup 1 parameters are used. $K = 1$, i.e., only one sample pilot is used for channel estimation so that transmit beamforming is only loosely enabled. The pilot SNR equals the channel SNR in Fig. 9 and is 8.75 dB lower in Fig. 10. The top-left subplot in Fig. 9 shows that the transmit beamforming scheme with estimation errors is indeed superior to Veeravalli's scheme on a coherent GMAC. Fig. 10 shows no benefit because the pilot SNR is not sufficient. Other subplots show graceful degradation with decreasing SNR.

V. SUMMARY

We considered the use of an analog transmission strategy via an affine transformation in order to exploit correlation in the sensor observations. The goal was to detect a change with minimum expected detection delay given an upper bound on the false alarm rate. We modeled the problem as a Markov decision problem with partial observations. We characterized

the optimal control as one that maximizes an Ali-Silvey distance between the two hypotheses before and after the change (Appendix A). In the GMAC setting, the optimal strategy minimizes the error variance of an equivalent observation at the fusion center. We then gave an explicit algorithm to identify the optimal control parameters.

We also studied a suboptimal policy that traded performance for quantity of information fed back. We then demonstrated via simulation the performance gain achieved by our algorithm over another scheme that makes only a naive use of the GMAC. The latter is a multi-access strategy optimal for independent data coupled with an optimal distributed quantization scheme for change detection; it is suboptimal because it does not exploit the correlation in sensor observations. Our proposed algorithm exploits this correlation via transmit beamforming on the GMAC. Given the control feedback in our setting, optimal transmission strategies will change from channel to channel. Techniques based on separation principles are therefore likely to be suboptimal.

Distributed transmit beamforming is crucial to realize our proposed scheme. The master-slave architecture of Mudumbai, Barriac & Madhow [1] and associated channel sensing techniques can be used for frequency and phase synchronization. Simulations with channel estimation errors indicate that the degradation due to lack of perfect channel knowledge is tolerable, making this analog technique a viable option for implementation.

We then considered a constraint on the average energy expended instead of a power constraint. We demonstrated via simulation that this made better use of the scarce energy resource. Extensions to arbitrary but known distributions, in particular to the exponential family, and to M-ary hypotheses can be found in [21, Ch. 5].

APPENDIX A
A CHARACTERIZATION OF OPTIMAL CONTROL

The following characterization of $A_J(\mu)$ was used in identifying the optimal controls. The characterization refers to a quantification of dissimilarity between probability measures called Ali-Silvey distances ([2]). Relative entropy (Kullback-Leibler divergence) is one example. Such dissimilarity measures have a well-known monotonicity property: data processing, whether deterministic or random, cannot increase the dissimilarity measure between two distributions ([2], [25]). This characterization may be of interest in other sequential detection settings.

Theorem 5: The minimization in (12) is obtained via a maximization of an Ali-Silvey distance between the density functions $f_{m_1,\alpha}$ and $f_{m_0,\alpha}$.

Proof: We first show that the minimization in (12) can be expressed as the maximization of an Ali-Silvey distance $\mathbb{E}_{p_1}\left[C\left(\phi(\widehat{Y})\right)\right]$ between probability density functions (pdf) p_1 and p_2 where

$$\phi(\hat{y}) = \frac{p_2(\hat{y})}{p_1(\hat{y})} = \frac{f_{m_1,\alpha}(\hat{y})}{h(\hat{y}, \alpha, \mu)},$$

and C is a convex function. To see this, observe that both $p_1(.)$ and $p_2(.)$ are densities. The density p_1 is a mixture of

pdfs under the two hypotheses while p_2 is the pdf under H_1. Thus $g(\hat{y}, \alpha, \mu)/h(\hat{y}, \alpha, \mu) = \beta\phi(\hat{y})$, where $\beta = \mu + (1-\mu)p$. From (12), we have

$$
\begin{aligned}
A_J(\mu) &= \min_{\alpha,c} \mathbb{E}_{p_1}\left[J\left(\beta\phi(\widehat{Y})\right)\right] = \min_{\alpha,c} \mathbb{E}_{p_1}\left[G\left(\phi(\widehat{Y})\right)\right] \\
&= 1 - \max_{\alpha,c} \mathbb{E}_{p_1}\left[C\left(\phi(\widehat{Y})\right)\right],
\end{aligned}
\tag{21}
$$

where $G(x) \triangleq J(\beta x)$ and $C(x) \triangleq 1 - G(x)$. J is concave; so G is concave, C is convex, and (21) is obtained via a maximization of an Ali-Silvey distance between p_1 and p_2. Now,

$$
\begin{aligned}
\mathbb{E}_{p_1}\left[C\left(\phi(\widehat{Y})\right)\right] &= \mathbb{E}_{p_1}\left[C\left(\frac{p_2(\widehat{Y})}{p_1(\widehat{Y})}\right)\right] \\
&= \mathbb{E}_{p_2}\left[\frac{p_1(\widehat{Y})}{p_2(\widehat{Y})}C\left(\frac{p_2(\widehat{Y})}{p_1(\widehat{Y})}\right)\right] \\
&= \mathbb{E}_{p_2}\left[C_1\left(\phi'(\widehat{Y})\right)\right],
\end{aligned}
\tag{22}
$$

where $\phi'(\hat{y}) = p_1(\hat{y})/p_2(\hat{y})$, and $C_1(x) = xC(1/x)$. $C_1(x)$ is a convex function because $C(x)$ is convex and x is non-negative. Now, let $p_3(\hat{y}) = f_{m_0,\alpha}(\hat{y})$. Since

$$p_1(\hat{y})/p_2(\hat{y}) = \beta + (1-\beta)p_3(\hat{y})/p_2(\hat{y}),$$

it is clear that $C_2(x) \triangleq C_1(\beta + (1-\beta)x)$ is a convex function. Setting $\phi''(\hat{y}) = p_3(\hat{y})/p_2(\hat{y})$, the likelihood ratio between the original two hypotheses, (22) can be written as $\mathbb{E}_{p_2}\left[C_2\left(\phi''(\widehat{Y})\right)\right]$, an Ali-Silvey distance between $f_{m_1,\alpha}$ and $f_{m_0,\alpha}$, and the theorem follows. ∎

APPENDIX B
PROOF OF LEMMA 4

Here we solve Problem 3. Order the indices so that $\sigma^2_{\mathsf{obs},1}h_1\alpha_{\mathsf{max},1} \leq \cdots \leq \sigma^2_{\mathsf{obs},L}h_L\alpha_{\mathsf{max},L}$. Let us first add a constraint $\sum_{l=1}^{L} h_l\alpha_l = a$, where without loss of generality $a \in [0, a_{\mathsf{max}}]$, with $a_{\mathsf{max}} = \sum_{l=1}^{L} h_l\alpha_{\mathsf{max},l}$, and solve the convex optimization problem:

Problem 5: Minimize $\sum_{l=1}^{L} \sigma^2_{\mathsf{obs},l}h_l^2\alpha_l^2$ subject to $\alpha_l \in [0, \alpha_{\mathsf{max},l}]$, $\sum_{l=1}^{L} h_l\alpha_l = a \in [0, a_{\mathsf{max}}]$.

This problem is a special case of a separable convex optimization problem studied in Padakandla and Sundaresan [26]. Execution of [26, Algorithm 1] yields the following solution. Break $[0, a_{\mathsf{max}}]$ into L intervals $[a_k, a_{k+1}]$, $k = 0, 1, \ldots, L-1$, where $a_0 = 0$ and

$$a_k = \left(\sum_{l=1}^{k} h_l\alpha_{\mathsf{max},l} + \sigma^2_{\mathsf{obs},k}h_k\alpha_{\mathsf{max},k} \sum_{l=k+1}^{L} \sigma^{-2}_{\mathsf{obs},l}\right).$$

The ordering of $\sigma^2_{\mathsf{obs},l}h_l\alpha_{\mathsf{max},l}$ implies $a_m \leq a_{m+1}$ so that each interval is nonempty. With k such that $a \in [a_k, a_{k+1}]$, the optimal solution is:

$$\alpha_l = \alpha_{\mathsf{max},l}, \quad l = 1, \ldots, k, \tag{23}$$

$$\alpha_l = \frac{1}{\sigma^2_{\mathsf{obs},l}h_l} \cdot \frac{a - \sum_{m=1}^{k} h_m\alpha_{\mathsf{max},m}}{\sum_{m=k+1}^{L} \sigma^{-2}_{\mathsf{obs},m}}, \quad l > k. \tag{24}$$

The corresponding minimum value of Problem 5 for a given a, denoted by $V(a)$, is given by

$$V(a) = \sum_{l=1}^{k} \sigma_{\mathsf{obs},l}^2 h_l^2 \alpha_{\mathsf{max},l}^2 + \frac{\left(a - \sum_{l=1}^{k} h_l \alpha_{\mathsf{max},l}\right)^2}{\sum_{l=k+1}^{L} \sigma_{\mathsf{obs},l}^{-2}}.$$

We next look for an optimal a by solving

Problem 6: Minimize $f(a) = \frac{V(a) + \sigma_{\mathsf{MAC}}^2}{a^2}$ subject to $a \in [0, a_{\max}]$.

While this is not yet a convex optimization, the transformation $b = 1/a$ casts it into one. Define

$$
\begin{aligned}
g(b) &= f\left(\frac{1}{a}\right) \\
&= b^2 \left[\sum_{l=1}^{k} \sigma_{\mathsf{obs},l}^2 h_l^2 \alpha_{\mathsf{max},l}^2 + \sigma_{\mathsf{MAC}}^2 + \frac{\left(\sum_{l=1}^{k} h_l \alpha_{\mathsf{max},l}\right)^2}{\sum_{l=k+1}^{L} \sigma_{\mathsf{obs},l}^{-2}} \right] \\
&\quad - 2b \cdot \frac{\sum_{l=1}^{k} h_l \alpha_{\mathsf{max},l}}{\sum_{l=k+1}^{L} \sigma_{\mathsf{obs},l}^{-2}} + \frac{1}{\sum_{l=k+1}^{L} \sigma_{\mathsf{obs},l}^{-2}},
\end{aligned}
$$

for $b \in [1/a_{\max}, \infty)$, where k depends on b through the index of the interval in which $a = 1/b$ lies. The following observations on g are easy to verify:

- $g(b)$ is a convex parabola on each $[1/a_{k+1}, 1/a_k], k = L-1, \cdots, 1$, and on $[1/a_1, \infty)$;
- $g(b)$ is continuous in $[1/a_{\max}, \infty)$. This needs checking only at interval boundaries $1/a_k$;
- $g(b)$ is continuously differentiable in $(1/a_{\max}, \infty)$ with left continuous derivative at $1/a_{\max}$;
- $\lim_{b \to \infty} g'(b) = +\infty$, so that the derivative eventually becomes positive for large b.

Since g is convex and continuously differentiable, if we can find a b^* such that $g'(b^*) = 0$ and $b^* \in [1/a_{k+1}, 1/a_k]$ (or $[1/a_1, \infty)$) where k corresponds to $a^* = 1/b^*$, then b^* is a point of global minimum. This holds if the minimum point for a parabola defined in $[1/a_{k+1}, 1/a_k]$ (or $[1/a_1, \infty)$), which is easily verified to be

$$
\begin{aligned}
a^* = 1/b^* &= \sum_{l=1}^{k} h_l \alpha_{\mathsf{max},l} \\
&+ \frac{\sum_{l=k+1}^{L} \sigma_{\mathsf{obs},l}^{-2}}{\sum_{l=1}^{k} h_l \alpha_{\mathsf{max},l}} \cdot \left(\sum_{l=1}^{k} (\sigma_{\mathsf{obs},l} h_l \alpha_{\mathsf{max},l})^2 + \sigma_{\mathsf{MAC}}^2 \right),
\end{aligned}
$$

also belongs to that interval. This leads to the condition (13). If no such point occurs, $g'(b) \neq 0$ in $[1/a_{\max}, \infty)$, and since g' is eventually positive, it must be positive in the entire interval. In this latter case g is an increasing function on $[1/a_{\max}, \infty)$ and the minimum is attained at $b^* = 1/a_{\max}$ or $a^* = a_{\max}$ or $k = L$. Substitution of a^* in (23) and (24) completes the proof.

REFERENCES

[1] R. Mudumbai, G. Barriac, and U. Madhow, "On the feasibility of distributed beamforming in wireless networks," *IEEE Trans. Wireless Commun.*, vol. 6, no. 5, pp. 1754–1763, May 2007.

[2] S. M. Ali and S. D. Silvey, "A general class of coefficients of divergence of one distribution from another," *J. Royal Statist. Soc., Ser. B*, vol. 28, no. 1, pp. 131–142, 1966.

[3] E. S. Page, "Continuous inspection schemes," *Biometrika*, vol. 41, no. 1/2, pp. 100–115, June 1954.

[4] G. Lorden, "Procedures for reacting to a change in distribution," *Ann. Math. Statist.*, vol. 42, pp. 1897–1908, Dec. 1971.

[5] A. N. Shiryayev, *Optimal Stopping Rules*. New York: Springer-Verlag, 1978.

[6] V. V. Veeravalli, "Decentralized quickest change detection," *IEEE Trans. Inform. Theory*, vol. 47, no. 4, pp. 1657–1665, May 2001.

[7] J. N. Tsitsiklis, *Advances in Statistical Signal Processing*. Greenwich, CT: JAI Press, 1993, vol. 2, ch. Decentralized Detection, pp. 297–344.

[8] V. V. Veeravalli, T. Başar, and H. V. Poor, "Decentralized sequential detection with a fusion center performing the sequential test," *IEEE Trans. Inform. Theory*, vol. 39, no. 2, pp. 433–442, Mar. 1993.

[9] V. K. Prasanthi, "Towards a cross layer design of sequential change detection over ad hoc wireless sensor networks," Master's thesis, Indian Institute of Science, Bangalore, India, June 2006.

[10] V. K. Prasanthi and A. Kumar, "Optimizing delay in sequential change detection on ad hoc wireless sensor networks," in *Proc. IEEE SECON*, Reston, VA, Sept. 2006.

[11] T. Berger, *Rate Distortion Theory: A Mathematical Basis for Data Compression*. Englewood Cliffs, NJ: Prentice-Hall, 1971.

[12] A. Lapidoth and S. Tinguely, "Sending a bi-variate Gaussian source over a Gaussian MAC," in *Proc. 2006 IEEE International Symposium on Information theory*, Seattle, Washington, July 2006.

[13] G. Mergen and L.Tong, "Type based estimation over multiaccess channels," *IEEE Trans. Signal Processing*, vol. 54, no. 2, pp. 613–626, Feb. 2006.

[14] G. Mergen, V. Naware, and L. Tong, "Asymptotic detection performance of type-based multiple access over multiaccess fading channels," *IEEE Trans. Signal Processing*, vol. 55, no. 3, pp. 1081–1092, Mar. 2007.

[15] K. Liu and A. M. Sayeed, "Type-based decentralized detection in wireless sensor networks," *IEEE Trans. Signal Processing*, vol. 55, no. 5, pp. 1899–1910, May 2007.

[16] K. Liu, H. E. Gamal, and A. Sayeed, "Decentralized inference over multiple-access channels," *IEEE Trans. Signal Processing*, vol. 55, no. 7, pp. 3445–3455, July 2007.

[17] M. Gastpar and M. Vetterli, "Scaling laws for homogeneous sensor networks," in *Proc. 41st Ann. Allerton Conf. Commun., Contr., Comput.*, Monticello, IL, Oct. 2003.

[18] M. Gastpar, B. Rimoldi, and M. Vetterli, "To code, or not to code: Lossy source-channel communication revisited," *IEEE Trans. Inform. Theory*, vol. 49, pp. 1147–1158, May 2003.

[19] M. Gastpar and M. Vetterli, "Power, spatio-temporal bandwidth, and distortion in large sensor networks," *IEEE J. Select. Areas Commun.*, vol. 23, pp. 745–754, Apr. 2005.

[20] E. Ertin and L. C. Potter, "Decentralized detection with generalized costs," in *Proc. Second ARL Feder. Lab. Symp.*, College Park, MD, Feb. 1998.

[21] L. Zacharias, "Decentralized sequential detection using physical layer fusion," Master's thesis, Indian Institute of Science, Bangalore, India, June 2007.

[22] D. P. Bertsekas and S. E. Shreve, *Stochastic Optimal Control: The Discrete Time Case*. New York: Academic Press, 1978.

[23] D. P. Bertsekas, *Dynamic Programming and Optimal Control, Volume 1*. Athena Scientific, 1995.

[24] L. Zacharias and R. Sundaresan, "Decentralized sequential change detection using physical layer fusion," in *Proc. 2007 IEEE Int. Symp. on Inform. Theory*, Nice, France, June 2007.

[25] I. Csiszár, "Information-type measures of difference of probability distributions and indirect observations," *Studia Sci. Math. Hungar.*, vol. 2, pp. 299–318, 1967.

[26] A. Padakandla and R. Sundaresan, "Separable convex optimization problems with linear ascending constraints," submitted to *SIAM J. on Opt.* http://arxiv.org/abs/0707.2265, July 2007.