Asymptotic Optimality Theory for Decentralized Sequential Hypothesis Testing in Sensor Networks

Yajun Mei

***Abstract*— The decentralized sequential hypothesis testing problem is studied in sensor networks, where a set of sensors receive independent observations and send summary messages to the fusion center, which makes a final decision. In the scenario where the sensors have full access to their past observations, the first of asymptotically Bayes sequential tests is developed, and the proposed test has same asymptotic performance as the optimal centralized test that has access to all sensor observations. Next, in the scenario where the sensors do not have full access to their past observations, a simple but asymptotically Bayes sequential tests is developed, in which sensor message functions are what we call *tandem quantizer*, where each sensor only uses two different sensor quantizers with at most one switch between these two quantizers. Moreover, a new minimax formulation of finding optimal stationary sensor quantizers is proposed and is studied in detail in the case of additive Gaussian sensor noises. Finally, our results show that the feedback from the fusion center does not improve asymptotic performance in the scenario with full local memory, however, even a one-shot one-bit feedback can significantly improve asymptotic performance in the scenario with limited local memory.**

***Index Terms*— Asymptotically Bayes, distributed detection, multi-sensor, quantization, sensor networks, sequential detection, tandem quantizer**

I. Introduction

SENSOR networks were originally motivated by their applications in military surveillance [19], and now they have many other important applications, including mobile and wireless communication, internet or computer network monitoring, and urban disaster prevention and response.

There are various possible configurations for sensor networks. Figure 1 illustrates the general setting of a widely used configuration. In such a network, at time n, each of a set of K sensors receives an observation $X_{k,n}$, and then sends a sensor message $U_{k,n}$ to a central processor, called the *fusion center*, which makes a final decision when observations are stopped. If necessary, the fusion center can also send feedback to the local sensors.

In many interesting applications, sensors are assumed to be *intelligent* and the information produced by the senors has been transmitted and fused in a fashion that provides reliable summary information while using the minimum amount of network resources. A standard mathematical formulation that can meet these requirements is to limit the sensor messages $U_{k,n}$ within a finite alphabet (perhaps binary). While this approach exhibits information loss as compared to the centralized setting of sending all raw data to the fusion center, it can often provide a formulation that is better suited for practical applications, due to the need for data compression, notable decrease in communication bandwidth, significant reduction in computational complexity at the fusion center, and privacy of raw data.

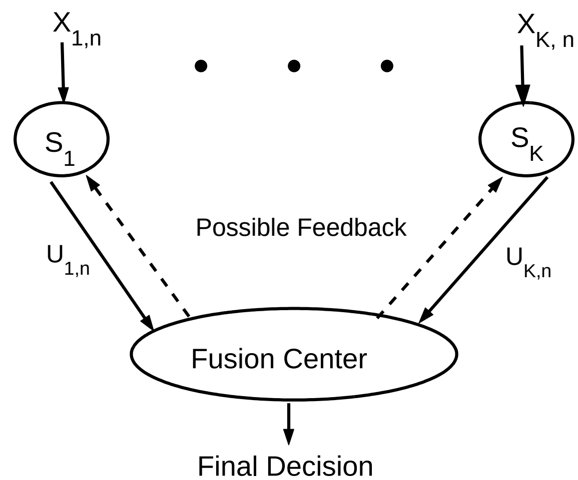

Fig. 1. General setting for sensor networks

In sensor networks, three of the most fundamental problems are: (1) how should each sensor send messages $U_{k,n}$'s to the fusion center; (2) how should the fusion center combine the messages from different sensors to make a final decision; and (3) if necessary, what kind of feedback should be sent from the fusion center to the sensors? It is worth emphasizing that these three problems are closely related with each other, e.g., the "best" decision rule at the fusion center will depend on how the sensors send messages to the fusion center. Ideally one wants to find optimal solutions to all these three problems simultaneously so as to achieve the "best" network performance (under some suitable definition).

As in the classical or centralized setting, there are two kinds of decision problems in sensor networks: *static* and *dynamic*. In static decision problems, the number of sensor observations (sample size) is fixed before the data are taken, and the fusion center makes only one decision. Static decision problems in sensor networks have been studied over the past twenty years.

Manuscript received September 1, 2006; revised October 1, 2007. This work was supported in part by NIH Grant R01 AI055343. The material in this article was presented in part at the IEEE International Symposium on Information Theory, Seattle, WA, USA, 2006.

The author is with the School of Industrial and Systems Engineering, Georgia Institute of Technology, Atlanta, Georgia, USA (email: ymei@isye.gatech.edu).

Publisher Item Identifier

For a review, see [23] and the references therein for hypothesis testing in sensor networks, and see [12], [27] for estimation problems in sensor networks.

However, there is a large body of problems where the static setting is not useful. In particular, the nature of sensor networks is *dynamic*, as information is updated over time at the both local sensor and fusion center levels. In the statistical literature, dynamic decision problems are also known as sequential decision problems, which were first introduced by Wald [24]. In the classical or centralized setting, sequential decision problems have been the focus of the mature fields of sequential analysis and change-point problems, see, [2], [9] and there references therein. On the other hand, research on sequential decision problems in sensor networks is rather limited, see [3], [21] for a review.

In this article, we will study a *decentralized* sequential hypothesis testing problem, one of standard sequential decision problems in sensor networks, see [3], [23] for many important applications including distributed signal detection, recognition, and surveillance. In the decentralized sequential hypothesis testing problem, it is assumed that there are two simple hypotheses H_0 and H_1 on the distributions of sensor observations. In order to make a quick but good decision on which of these two hypotheses is true, the optimization needs to be done jointly over sensors and fusion center policies, as well as over time. This makes the optimization highly problematic. In fact, it is well-known [22] that dynamic programming or Bayes formulations generally become intractable, except in a special case where it has been an open-problem to derive the properties of Bayes solutions.

The goal of this article is to develop asymptotic optimality theory for the Bayes formulation of the decentralized sequential hypothesis testing problem, and to offer two classes of fairly simple decentralized sequential tests which are asymptotically Bayes in their respective scenarios. As a consequence, several open problems raised in [21], [22] for decentralized sequential hypothesis testing problems will be addressed asymptotically.

Throughout this article, we make the following assumptions, which are standard:

(A1) Conditioned on each of hypotheses, H_0 and H_1, the sensor observations are independent over time as well as from sensor to sensor.

(A2) For each $1 \leq k \leq K$, the density of the sensor observations $X_{k,1}, X_{k,2}, \cdots$ at sensor S_k is f_k under the null hypothesis H_0, and is g_k under the alternative hypothesis H_1. For each $1 \leq k \leq K$, the Kullback-Leibler information numbers (or relative entropy)

$$I(g_k, f_k) = \int \log\Big(\frac{g_k(x)}{f_k(x)}\Big) g_k(x) dx$$

and

$$I(f_k, g_k) = \int \log\Big(\frac{f_k(x)}{g_k(x)}\Big) f_k(x) dx$$

are finite and positive. Moreover,

$$\int \Big(\log \frac{g_k(x)}{f_k(x)}\Big)^2 g_k(x) dx < \infty,$$

and

$$\int \Big(\log \frac{f_k(x)}{g_k(x)}\Big)^2 f_k(x) dx < \infty.$$

Now let us briefly review some open problems raised in an influential paper by Veeravalli, Basar and Poor [22] and summary our main results. In [22], the authors considered two different scenarios of sensor networks, depending on how local information is used to produce the sensor messages $U_{k,n}$'s at the sensors. The first scenario is the system with *full local memory*, where the sensors have full access to their past observations. This scenario has the following two possible cases, which correspond to Cases B, and D in [22].

*Case (***i***) System with no Feedback and Full Local Memory:*

$$U_{k,n} = \phi_{k,n}(X_{k,[1,n]}), \tag{1}$$

where $X_{k,[1,n]} = (X_{k,1}, X_{k,2}, \ldots, X_{k,n})$ are all local sensor observations observed up to time n at sensor S_k.

*Case (***ii***) System with Full Feedback and Full Local Memory:*

$$U_{k,n} = \phi_{k,n}(X_{k,[1,n]}; \mathcal{E}_{n-1}), \tag{2}$$

where $\mathcal{E}_{n-1}$ denotes the past sensor message information given by

$$\mathcal{E}_{n-1} = \{U_{1,[1,n-1]}, U_{2,[1,n-1]}, \ldots, U_{K,[1,n-1]}\}, \tag{3}$$

with $U_{k,[1,n-1]} = (U_{k,1}, U_{k,2}, \ldots, U_{k,n-1})$, and $\mathcal{E}_0$ is the null set.

It is well-known [22] that it is untractable to find Bayes solutions under this scenario via the dynamic programming approach, and it has been an open problem to find any asymptotically optimal sequential tests in the system with full local memory. By using the asymptotic optimality approach, we will offer the first of asymptotically Bayes sequential tests in the system with full local memory. In our proposed tests, each sensor sends its local "sensor decisions" $U_{k,n}$ to the fusion center, which then combines all local sensor decisions by using an "AND" rule to decide which of the hypotheses is true.

Note that a criticism of the system with full local memory often made is that it will require sensors to have unlimited memory and computational capacities as the time goes to ∞. In our proposed tests, however, local sensor decisions (or sensor messages) are based on local sufficient statistics which can be calculated recursively, and thus at every time n, each senor only requires a memory of a data set of size 2 and only involves $O(1)$ computations.

The second scenario is the system with *limited local memory*, where the sensors do not have access to their past observations. The following three possible cases have been considered in [22].

*Case (***iii***) System with Neither Feedback from the Fusion Center nor Local Memory:*

$$U_{k,n} = \phi_{k,n}(X_{k,n}). \tag{4}$$

*Case (***iv***) System with no Feedback and Local Memory Restricted to Past Sensor Messages:*

$$U_{k,n} = \phi_{k,n}(X_{k,n}; U_{k,[1,n-1]}). \tag{5}$$

*Case (**v**) System with Full Feedback, but Local Memory Restricted to Past Sensor Messages:*

$$U_{k,n} = \phi_{k,n}(X_{k,n}; \mathcal{E}_{n-1}), \qquad (6)$$

where $\mathcal{E}_{n-1}$ is the past sensor messages defined in (3).

As illustrated in [22], in the system with limited local memory, the Bayes formulation is tractable only in the case with full feedback specified in (6), but it has been an open problem to derive the properties of the Bayes solution due to its complicated structure. In order to develop simple decentralized sequential tests with good performances, another closely related open problem raised in [22] is whether *stationary quantizers*, i.e., stationary sensor message functions, can lead to decentralized sequential tests that are asymptotically optimal in the system with limited local memory. That is, can we find asymptotically Bayes sequential tests from the case specified in (4) where the sensor message functions $\phi_{k,n} \equiv \phi_k$ do not change over time n?

We will provide asymptotic answers to both of these two open problems in the system with limited local memory. For the second open problem, we will show that tests with stationary quantizers are generally asymptotically suboptimal except in some extreme situations. Moreover, to address the optimality properties within the class of stationary quantizers, we also propose a new minimax formulation of finding *optimal* stationary quantizers, and study in detail in the case of additive Gaussian sensor noises.

For the first open problem, it is very difficult to study the properties of the Bayes solution *directly* although its structure was characterized in [22]. Thus we adopt an *indirect* approach by establishing sharp upper and lower bounds on the properties of the Bayes solution, where the leading terms in the upper and lower bounds are the same. A key step is to develop simple but asymptotically optimal decentralized sequential tests in the system with limited local memory. To accomplish this, we will consider the following new case where the fusion center *quantizes* the past sensor messages $\mathcal{E}_{n-1}$ in (3) before sending the feedback to the sensors.

*Case (**vi**) System with Quantized Feedback, and Local Memory Restricted to Past Sensor Messages:*

$$U_{k,n} = \phi_{k,n}(X_{k,n}; V_{k,n-1}), \qquad (7)$$

where the quantized feedback

$$V_{k,n-1} = \psi_{k,n-1}(\mathcal{E}_{n-1})$$

is chosen from a finite (possibly binary) list and $\mathcal{E}_{n-1}$ defined in (3) is the sensor messages available before time n.

Note that this new case can be thought of as a generalization of three cases considered in [22]. Specifically, this new case becomes the cases defined in (4), (5) and (6) when the (quantized) feedback $V_{k,n-1}$ is null, past sensor messages at each sensor, and full feedback, respectively. From the practical point of view, this new case is more appealing than the case with full feedback specified in (6) since the full feedback assumes the fusion center send all past message information $\mathcal{E}_{n-1}$ back to the sensors even for large n, and that will require large communication bandwidth between sensors and fusion center. To take into account limited computational capabilities at the sensor level, a more practice-based view of quantized feedback is that the fusion center determines and sends each sensor their own threshold values to be used in their sensor message functions.

It turns out that it will be sufficient for the fusion center to send one-shot one-bit feedback in order to construct simple sequential tests which are asymptotically Bayes in the system with limited local memory. Motivated by Abramson [1], and Kiefer and Sacks [6], our proposed tests will divide the decision making into two stages. In the first stage, preliminary samples are taken at the sensors until the fusion center makes a preliminary decision on which of H_0 and H_1 is true. The fusion center then sends the preliminary decision back to all sensors so that the sensors can optimize their respective quantizers according to the preliminary decision of the fusion center. In the second stage, all sensors use the new "optimal" quantizers to send all future sensor messages until the fusion center makes a final decision, which is based only on the sensor messages in the second stage. The number of time steps taken in the first stage (when the cost c is small) is large but is small relative to that in the second stage.

Another way to look at our proposed tests is from the sensor's point of view. In our proposed tests in the system with limited local memory, at each sensor, the sensor message function is what we call a *tandem quantizer*, where each sensor only uses two different sensor quantizers with at most one switch between these two quantizers. Note that the sensor quantizers in the Bayes solution need to be changed at each time step, whereas one can only obtain suboptimal decentralized tests in general if the sensor quantizers do not change over time, i.e., sensor quantizers are stationary. Hence, from the viewpoint of the number of switches, a tandem quantizer is the simplest possible candidate to construct asymptotically Bayes (or optimal) decentralized tests in the system with limited local memory.

The remainder of this article is organized as follows. In Section II, we provide a formal mathematical formulation of decentralized sequential hypothesis testing problems and present some well-known results on the optimal centralized sequential test to provide a benchmark for the decentralized sequential tests. Section III studies the decentralized sequential detection problems in the system with full local memory, and Section IV considers the system with limited local memory. Each of Sections III and IV develops asymptotic theory and offers asymptotically Bayes tests which are easy to implement. Section V focuses on tests with stationary sensor message functions, and proposes a new minimax formulation for the optimal stationary sensor message functions. In Section VI, we give a concrete example to illustrate our asymptotic results. The proofs of most theorems are included in the Appendix.

II. Problem Formulation and Notations

Suppose there are K sensors in a sensor network. At time n, an observation $X_{k,n}$ is made at each sensor S_k. Assume there are two simple hypotheses H_0 and H_1. Under the hypothesis H_0, the observations at sensor S_k, $X_{k,1}, X_{k,2}, \cdots$

are independent and identically distributed (i.i.d.) with density function f_k, and under the hypothesis H_1, they have density g_k. Denote by $\mathbf{P}_0$ and $\mathbf{P}_1$ respectively the probability measures under hypotheses H_0 and H_1. Let $\mathbf{E}_0$ and $\mathbf{E}_1$ be the corresponding expectations. Note that this notation is slightly different than the conventional notation for conditional probabilities and expectations under the Bayes formulation, but it has advantages in the asymptotic optimality approach.

Based on the information available at S_k at time n, a sensor message $U_{k,n}$, specified in (1) or (2) for the system with full local memory, and in (4) - (7) for the system with limited local memory, is chosen from a finite alphabet and sent to the fusion center. Without loss of generality, it is assumed that $U_{k,n}$ is chosen from the finite set $\{0, 1, \cdots, D_k - 1\}$. At time n, the fusion center may take one of three possible actions: (1) lets all sensors continue taking observations and sending messages to the fusion center; (2) informs all sensors to stop taking observations and decides the null hypothesis H_0 is true; or (3) informs all sensors to stop taking observations and decides the alternative hypothesis H_1 is true.

To define measures of performance for a decentralized sequential test δ, denote by τ the corresponding stopping time when the fusion center decides to stop taking observations. There are four quantities that are useful and appropriate to evaluate decentralized sequential tests: (1) probability of Type I error, $\mathbf{P}_0(\text{reject } H_0)$, (2) probability of Type II error, $\mathbf{P}_1(\text{accept } H_0)$, (3) expected decision-making time under H_0, $\mathbf{E}_0(\tau)$, and (4) expected decision-making time under H_1, $\mathbf{E}_1(\tau)$. The last two performance measures are included since τ, the number of time steps needed to make a final decision, depends on observations and is thus a random variable. Ideally we want these four quantities as small as possible.

Veeravalli, Basar and Poor [22] considered the following Bayes formulation of the decentralized sequential detection problem. Assume the two hypotheses H_0 and H_1 have known prior probabilities, say, $\pi = \mathbf{P}\{H_0 \text{ is true}\} = 1 - \mathbf{P}\{H_1 \text{ is true}\}$ for some $0 \leq \pi \leq 1$, and each time step taken for decision making costs a positive amount c. Also for $i = 0, 1$ let $W_i (> 0)$ be the cost of falsely rejecting H_i. Then the total expected cost or Bayes risk of a decentralized sequential test δ with a stopping time τ is

$$\begin{aligned} \mathcal{R}_c(\delta) &= \pi[c\mathbf{E}_0(\tau) + W_0\mathbf{P}_0\{\text{reject } H_0\}] \\ &\quad +(1-\pi)[c\mathbf{E}_1(\tau) + W_1\mathbf{P}_1\{\text{reject } H_1\}], \end{aligned} \tag{8}$$

where the notation $\mathcal{R}_c$ is used to emphasize that the cost for each time step is c. The Bayes formulation of decentralized sequential hypothesis testing problems can then be stated as follows.

Problem (P1): Minimize the Bayes risk $\mathcal{R}_c(\delta)$ in (8) over all possible choices of sensor message functions and over all possible policies at the fusion center.

Unfortunately, it is nearly impossible to find exactly optimal solutions in sensor networks (for some special cases see [22]), and only "asymptotic optimality" results seem to be working. In the asymptotic optimality approach, we first construct an asymptotic lower bound of $\mathcal{R}_c(\delta)$ in (8) as c goes to 0. Then we show that a given class of decentralized sequential tests attains the lower bound asymptotically. We will establish asymptotic optimality theorems for both scenarios of decentralized decision systems: full local memory (specified in (1) and (2)) and limited local memory (specified in (4)-(7)).

We now introduce some notation. Let D be a positive integer. Consider a random variable X whose density function is either f or g with respect to some σ-finite measure, and assume that the Kullback-Leibler information number $I(g, f)$ is finite. For a (*deterministic* or *random*) measurable function ϕ from the range of X to a finite alphabet of size D, say $\{0, 1, \ldots, D-1\}$, denote by f_ϕ and g_ϕ respectively the probability mass function of $\phi(X)$ when the density of X is f and g. Let

$$Z_\phi = \log \frac{g_\phi(\phi(X))}{f_\phi(\phi(X))},$$

and define

$$I_D(g, f) = \sup_\phi \mathbf{E}_g(Z_\phi) \tag{9}$$

and

$$V_D(g, f) = \sup_\phi \mathbf{E}_g(Z_\phi^2). \tag{10}$$

It is well known [20] that $I_D(g, f) \leq I(g, f)$, i.e., the reduction of the data from X to $\phi(X)$ cannot increase the information. Tsitsiklis [20] showed that the supremum $I_D(g, f)$ is achieved by a Monotone Likelihood Ratio Quantizer (MLRQ) φ of the form

$$\varphi(X) = d \quad \text{if and only if} \quad \lambda_d \leq \frac{g(X)}{f(X)} < \lambda_{d+1},$$

where $0 = \lambda_0 \leq \lambda_1 \leq \cdots \leq \lambda_{D-1} \leq \lambda_D = \infty$ are constants. Similarly, by switching the roles of f and g, the supremum $I_D(f, g)$ is achieved by a MLRQ with another (likely different) set of constant λ's. These optimal MLRQ's are not easily calculated, but we follow the standard practice in the literature of developing procedures that assume sensor messages are constructed optimally in the sensor. The definition of $V_D(g, f)$ was first proposed in [10], [11]. Some of our theorems assume that $V_D(g, f) < \infty$, and sufficient conditions for finiteness of $V_2(g, f)$ are given in Theorem 2 and Corollary 2 of [11].

Using these notations, define the information numbers

$$I_{\mathbf{D}} = \sum_{k=1}^K I_{D_k}(f_k, g_k) \text{ and } J_{\mathbf{D}} = \sum_{k=1}^K I_{D_k}(g_k, f_k), \tag{11}$$

where $\mathbf{D} = (D_1, D_2, \ldots, D_K)$. Moreover, define

$$I_{tot} = \sum_{k=1}^K I(f_k, g_k) \text{ and } J_{tot} = \sum_{k=1}^K I(g_k, f_k). \tag{12}$$

These four information numbers are key to our theorems.

In the remainder of this section, let us consider the optimal centralized sequential test, i.e., the Bayes solution to Problem (P1) in the centralized setting when the fusion center has access to all sensor observations, or equivalently, when all sensors send their raw data to the fusion center ($U_{k,n} = X_{k,n}$). The optimal centralized sequential test not only provides a benchmark for all decentralized sequential tests, but also sheds light on developing asymptotically optimal decentralized sequential tests. Denote by $\delta^*_{cen}(c)$ the optimal centralized sequential test. It is well-known [24] that $\delta^*_{cen}(c)$ is in the form of Wald's sequential probability ratio tests (SPRT) defined as follows. Choose two constant $a > 0$ and $b > 0$, and define the log-likelihood ratio

$$\mathbf{L}_n = \sum_{i=1}^{n} \Big(\sum_{k=1}^{K} \log \frac{g(X_{k,i})}{f(X_{k,i})} \Big). \tag{13}$$

Define a stopping time N = first $n \geq 1$ such that $\mathbf{L}_n \notin (-b, a)$. In other words,

$$N = \inf\{n \geq 1 : \mathbf{L}_n \notin (-b, a)\}. \tag{14}$$

The SPRT will stop taking observations at time N and

$$\begin{array}{ll} \text{Reject } H_0 & \text{if } \mathbf{L}_N \geq a, \\ \text{Accept } H_0 & \text{if } \mathbf{L}_N \leq -b. \end{array}$$

The appropriate values of the thresholds a and b in the optimal centralized test $\delta^*_{cen}(c)$ are determined by the *a priori* probability π of H_0 and the costs W_0, W_1 and c in (8). From the asymptotic viewpoint, however, the thresholds a and b in $\delta^*_{cen}(c)$ satisfy $a \approx b \approx |\log c|$ as $c \to 0$, as shown in [4], [24]. To see this, note that as $c \to 0$, both a and b are large and it is well-known [24] that SPRT in the centralized setting satisfies

$$\begin{array}{l} \mathbf{P}_0(\text{reject } H_0) \approx e^{-a}, \quad \mathbf{P}_1(\text{reject } H_0) \approx e^{-b}, \\ \mathbf{E}_0(N) \approx b/I_{tot}, \quad \text{and} \quad \mathbf{E}_1(N) \approx a/J_{tot}, \end{array}$$

where I_{tot} and J_{tot} are defined in (12). Hence, the Bayes risk $\mathcal{R}_c$ of an SPRT can be approximated by

$$\pi \Big[c\frac{b}{I_{tot}} + W_0 e^{-a} \Big] + (1-\pi) \Big[c\frac{a}{J_{tot}} + W_1 e^{-b} \Big].$$

Minimizing this approximation to $\mathcal{R}_c$ we have

$$\begin{array}{l} a \approx -\log c + \log[J_{tot} W_0 \pi/(1-\pi)] \approx |\log c|, \\ b \approx -\log c + \log[I_{tot} W_1 (1-\pi)/\pi)] \approx |\log c|, \end{array} \tag{15}$$

as $c \to 0$. Thus an SPRT with threshold values $a = b = |\log c|$ is asymptotically Bayes to Problem (P1) in the centralized setting. Since we are interested in asymptotic results in this article, to simplify our notation, we refer to the optimal centralized sequential tests as the SPRT whose stopping time is defined in (14) with thresholds $a = b = |\log c|$. Using the approximations for the SPRT again, it is easy to see that the risk corresponding the optimal centralized test $\delta^*_{cen}(c)$ is

$$\mathcal{R}_c(\delta^*_{cen}(c)) \approx c|\log c| \big(\frac{\pi}{I_{tot}} + \frac{1-\pi}{J_{tot}} \big). \tag{16}$$

By (15) and (16), from the asymptotic point of view, the optimal centralized sequential test and its Bayes risk depend mainly on c, I_{tot}, J_{tot} and π and are insensitive to the costs W_0 and W_1 of making the wrong decision. Moreover, the Bayes risk of the optimal centralized sequential test $\delta^*_{cen}(c)$ is mainly the cost of time step for decision making.

III. System with Full Local Memory

Let $\delta^*_A(c)$ denote a Bayes solution to the decentralized sequential hypothesis testing problem in the system with full local memory, specified in (1) or (2). The main result in this section is to develop a decentralized test $\delta_A(c)$ in the case specified in (1) such that, under some general conditions, $\lim_{c\to 0} \mathcal{R}_c(\delta^*_A(c))/\mathcal{R}_c(\delta_A(c)) = 1$, i.e., $\{\delta_A(c)\}$ is "asymptotically Bayes" in the system with full local memory as the cost of time step taken for decision making goes to 0. Here $\mathcal{R}_c$, of course, denotes the Bayes risk defined in (8) when c is the cost of time step taken for decision making.

A. The Structure of Our Proposed Tests $\delta_A(c)$

In the system with full local memory when $D_k \geq 3$ for each k, our proposed test $\delta_A(c)$ is defined as follows (see Remark 1 below for the extension of our proposed test in case of binary sensor messages).

1) Sensor Policy: Each sensor S_k uses its local sensor observations to conduct an SPRT test with *appropriate* but fixed boundaries, and sends its local decision to fusion center. Specifically, each sensor S_k calculates its local log-likelihood ratio statistic $L_{k,n}$ recursively by

$$L_{k,n} = L_{k,n-1} + \log \frac{g_k(X_{k,n})}{f_k(X_{k,n})} \tag{17}$$

for $n \geq 1$ and $L_{k,0} = 0$. Then each sensor uses a quantizer on $L_{k,n}$ to send messages to the fusion center, i.e., the sensor message sent from the sensor S_k to the fusion center is defined by

$$U_{k,n} = \begin{cases} 0 & \text{if } L_{k,n} \leq -r_k |\log c|, \\ 1 & \text{if } L_{k,n} \geq \rho_k |\log c|, \\ \text{NULL} & \text{otherwise.} \end{cases}$$

where

$$r_k = \frac{I(f_k, g_k)}{\sum_{k=1}^{K} I(f_k, g_k)} = \frac{I(f_k, g_k)}{I_{tot}}, \tag{18}$$

and

$$\rho_k = \frac{I(g_k, f_k)}{\sum_{k=1}^{K} I(g_k, f_k)} = \frac{I(g_k, f_k)}{J_{tot}}. \tag{19}$$

Here r_k and ρ_k can be thought of as the weight of sensor S_k in the overall final decision under the hypotheses H_0 and H_1, respectively. The message "NULL" is a special sensor symbol to indicate that the sensor has not reached a local decision yet. For example, "NULL" could be represented by the situation when the sensor does not send any sensor messages to the fusion center, e.g., the sensor is silent.

2) Fusion Center Policy: After receiving the local "sensor decisions" $U_{k,n}$'s from the sensors, the fusion center then combines all local sensor decisions by using an "AND" rule. That is, the fusion center will

$$\begin{cases} \text{stop and accept } H_1 & \text{if } U_{k,n} = 1 \text{ for all } 1 \leq k \leq K, \\ \text{stop and accept } H_0 & \text{if } U_{k,n} = 0 \text{ for all } 1 \leq k \leq K, \\ \text{continue sampling} & \text{otherwise.} \end{cases}$$

Note that since the local log-likelihood ratio statistic $L_{k,n}$ can be calculated recursively, our proposed test $\delta_A(c)$ reduces the local memory requirements at every time n from the full local memory $\{X_{k,1}, \cdots, X_{k,n}\}$ to the data set of size 2, i.e., $\{L_{k,n-1}, X_{k,n}\}$.

Furthermore, it is also easy to see that in single-sensor systems, i.e., $K = 1$, our proposed test $\delta_A(c)$ coincides with the optimal centralized SPRT. However, our proposed test $\delta_A(c)$ requires that each sensor shall continue sending the local sensor decisions to the fusion center even after the log-likelihood ratio statistic exceeds the local thresholds. This essential feature can be seen from the following heuristic argument, which provides the motivation of $\delta_A(c)$.

Consider the optimal centralized SPRT whose stopping time is defined in (14) with thresholds $a = b = |\log c|$. By (17), the log-likelihood ratio statistics at sensor S_k is

$$L_{k,n} = \sum_{i=1}^{n} \log \frac{g_k(X_{k,i})}{f_k(X_{k,i})}.$$

Combining this with (13) yields that $\mathbf{L}_n = \sum_{k=1}^{K} L_{k,n}$ for all n. That is, at any given time, the log-likelihood ratio statistic at the fusion center in the centralized version is the sum of the local log-likelihood ratio statistics at all sensors. Thus, if each sensor were able to send its local log-likelihood ratio statistic (a sufficient statistic) to the fusion center, then the fusion center would have been able to perform the optimal centralized SPRT. However, under the restriction that the sensor messages belong to a finite alphabet, each sensor may not be able to send its local log-likelihood ratio statistic *directly* to the fusion center. Fortunately, the idea can be salvaged by the strong law of large numbers (SLLN). Observe that when H_0 is true, $\mathbf{L}_n = \sum_{k=1}^{K} L_{k,n}$ will go to $-\infty$ as n goes to ∞. Hence, under the null hypothesis H_0, the stopping rule of the optimal centralized SPRT can be approximated by

$$\{\sum_{k=1}^{K} L_{k,n} \le -|\log c|\} \tag{20}$$

for sufficiently small c. Now the SLLN implies that $L_{k,n}/n \to I(f_k, g_k)$ with probability 1 under the null hypothesis H_0. Hence, the weight of $L_{k,n}$ in the sum is roughly $I(f_k, g_k)/\sum_{k=1}^{K} I(f_k, g_k) = r_k$, and thus (20) can be further approximated by $\{L_{k,n} \le -r_k|\log c|$ for all $1 \le k \le K\}$, which is exactly the stopping rule when our proposed test $\delta_A(c)$ decides that H_0 is true. In other words, under the null hypothesis H_0, the stopping rule of the optimal centralized SPRT can be approximated by that of our proposed test $\delta_A(c)$. Similarly, the conclusion also holds under the alternative hypothesis H_1. Since the optimal centralized SPRT can be approximated by our proposed test $\delta_A(c)$ under both hypotheses H_0 and H_1, it is not surprising to see from the next subsection that our proposed test $\delta_A(c)$ is asymptotically optimal not only in the system with full local memory, but also in the centralized setting.

B. Asymptotic Optimality

We first establish asymptotic lower bounds on the Bayes risk $R_c(\delta)$ in (8) for any decentralized tests in the system with full local memory. Later these bounds will be used to prove the asymptotic optimality properties of $\delta_A(c)$. Note that such a lower bound can be established by the optimality of SPRT in the centralized version, and it turns out this bound will be sufficient for our purpose in the system with full local memory. The following theorem was a slight modification of Theorem 2 in [4]. While this theorem follows at once from the optimality of the centralized SPRT and its asymptotic property established in (16), an alternative proof, due to Chernoff [4], is given in the Appendix so that parts of the argument can be applied conveniently in the proof of Theorem 4.

Theorem 1: For any sequential or fixed-sample-size tests δ in the decentralized or centralized setting, if $R_c(\delta) \le Mc|\log c|$ for some constant M, then as $c \to 0$,

$$R_c(\delta) \ge (1 + o(1))c|\log c|\Big(\frac{\pi}{I_{tot}} + \frac{1-\pi}{J_{tot}}\Big).$$

where $o(1) \to 0$ does not depend on δ, and I_{tot} and J_{tot} are defined in (12).

Next, we consider the behavior of $\delta_A(c)$ proposed in the previous subsection for small values of c. For this purpose, we need to estimate the probabilities of making incorrect decision when $\delta_A(c)$ is used and the performance of $\hat{T}(c)$, the number of time steps required by $\delta_A(c)$ to make decisions. The following theorem, whose proof is in the Appendix, summarizes the asymptotic properties of our proposed test $\delta_A(c)$ in the system with full local memory.

Theorem 2: For any $0 < c < 1$, the probabilities that our proposed test $\delta_A(c)$ makes incorrect decision satisfy

$$\mathbf{P}_0\{\text{reject } H_0\} \le c \quad \text{and} \quad \mathbf{P}_1\{\text{rejects } H_1\} \le c. \tag{21}$$

Moreover, if we let $\hat{T}(c)$ denote the number of time steps required by $\delta_A(c)$ to make decisions, then as $c \to 0$,

$$\begin{aligned} \mathbf{E}_0(\hat{T}(c)) &\le (1 + o(1))|\log c|/I_{tot}, \\ \mathbf{E}_1(\hat{T}(c)) &\le (1 + o(1))|\log c|/J_{tot}, \end{aligned} \tag{22}$$

where I_{tot} and J_{tot} are defined in (12).

Now we are in a position to show that our proposed test $\delta_A(c)$ is asymptotically optimal in the system with full local memory.

Theorem 3: Under the assumptions of (A1) and (A2) and $D_k \ge 3$ for all $1 \le k \le K$, $\{\delta_A(c)\}$ is asymptotically Bayes in the system with full local memory, i.e., $\lim_{c\to 0} \mathcal{R}_c(\delta_A^*(c))/\mathcal{R}_c(\delta_A(c)) = 1$ where $\delta_A^*(c)$ is a Bayes solution in the system with full local memory when c is the cost per time step for decision making.

Proof: From Theorem 2, for our proposed test $\delta_A(c)$, we have

$$\limsup_{c\to 0} \frac{1}{c|\log c|}\mathbf{P}_0\{\text{reject } H_0\} = 0,$$
$$\limsup_{c\to 0} \frac{1}{c|\log c|}\mathbf{P}_1\{\text{reject } H_1\} = 0,$$
$$\limsup_{c\to 0} \frac{1}{|\log c|}\mathbf{E}_0(\hat{T}(c)) \le \frac{1}{I_{tot}},$$
$$\limsup_{c\to 0} \frac{1}{|\log c|}\mathbf{E}_1(\hat{T}(c)) \le \frac{1}{J_{tot}},$$

where $\hat{T}(c)$ is the number of time step required by $\delta_A(c)$ to make decisions. Combining these with the definition of $\mathcal{R}_c$ in (8) yields

$$\limsup_{c\to 0} \frac{\mathcal{R}_c(\delta_A(c))}{c|\log c|} \le \frac{\pi}{I_{tot}} + \frac{1-\pi}{J_{tot}}. \quad (23)$$

If our proposed test $\delta_A(c)$ is *not* asymptotically Bayes, we would have

$$\liminf_{c\to 0} \mathcal{R}_c(\delta_A^*(c))/\mathcal{R}_c(\delta_A(c)) < 1-\epsilon$$

for some positive constant ϵ which implies, due to (23), that there is a sequence $\{c_i\}$ with $c_i \to 0$ such that

$$\mathcal{R}_{c_i}(\delta_A^*(c_i)) < (1-\epsilon)c_i|\log c_i|\Big(\frac{\pi}{I_{tot}} + \frac{1-\pi}{J_{tot}}\Big). \quad (24)$$

Hence $\delta_A^*(c_i)$ satisfies the sufficient condition of Theorem 1 with $M = (1-\epsilon)\Big(\frac{\pi}{I_{tot}} + \frac{1-\pi}{J_{tot}}\Big)$, and thus Theorem 1 implies that

$$\mathcal{R}_{c_i}(\delta_A^*(c_i)) \ge (1+o(1))c_i|\log c_i|\Big(\frac{\pi}{I_{tot}} + \frac{1-\pi}{J_{tot}}\Big),$$

which is a contradiction of (24) as $\epsilon > 0$ is a constant. So the theorem holds. □

Observe that our proposed test $\delta_A(c)$ does not use feedback from the fusion center, but it is asymptotically optimal in the system with full local memory. This fact proves the following interesting result.

Corollary 1: Under the conditions of Theorem 3, feedback from the fusion center does not improve asymptotic performance in the system with full local memory, specified in (1)-(2).

C. Additional Remarks

Remark 1: Although we do not know the structure of Bayes solution $\delta_A^*(c)$ in the system with full local memory, its asymptotic properties can be easily established based on our results. By Theorems 1 - 3,

$$\mathcal{R}_c(\delta_A^*(c)) = (1+o(1))c|\log c|\Big(\frac{\pi}{I_{tot}} + \frac{1-\pi}{J_{tot}}\Big).$$

as $c \to 0$. Moreover, if we denote by $T_A^*(c)$ the stopping time (the sample size) of the Bayes solution $\delta_A^*(c)$, then the proof of Theorem 1 shows that

$$\begin{aligned} \mathbf{E}_0(T_A^*(c)) &= (1+o(1))|\log c|/I_{tot}, \\ \mathbf{E}_1(T_A^*(c)) &= (1+o(1))|\log c|/J_{tot}, \end{aligned}$$

as $c \to 0$.

Remark 2: The proof of Theorem 3 actually shows that our proposed test $\{\delta_A(c)\}$ is also asymptotically Bayes in the centralized setting, i.e., $\lim_{c\to 0} \mathcal{R}_c(\delta_{cen}^*(c))/\mathcal{R}_c(\delta_A(c)) = 1$, where $\delta_{cen}^*(c)$ is the optimal centralized SPRT. In other words, our proposed test $\delta_A(c)$ has same asymptotic performance as the optimal centralized test.

Remark 3: Our proposed test $\delta_A(c)$ does not involve π, the prior probability of the null hypothesis, but it is asymptotically Bayes in the system with full local memory regardless of the value of the prior probability π. This is very attractive as one does not need to worry about how to choose a prior distribution when using our proposed tests.

Remark 4: Since our results are first-order asymptotic, the efficiency of our proposed test $\delta_A(c)$ (with respect to the optimal centralized test) under nonasymptotic scenarios depends on the speed of convergence of its performance function to the corresponding asymptotic values. From the proof of Theorem 2, as $c \to 0$, we have $\mathcal{R}_c(\delta_A(c))/\mathcal{R}_c(\delta_{cen}^*(c)) = 1 + \frac{D+o(1)}{\sqrt{|\log c|}}$, where the constant $D > 0$ can be derived explicitly under additional reasonable conditions from nonlinear renewal theory ([17], [26]). Similar results (including the constant D) were reported both theoretically and experimentally in [11] in the context of decentralized sequential change-point detection problems. If the number of sensors K is large in the system, the value of D can be very large. In that situation, when the value c is only moderately small, the value of $D/\sqrt{|\log c|}$ can be large, implying that the nonasymptotic performance of $\delta_A(c)$ can be very poor.

Remark 5: In our proposed test $\delta_A(c)$, a special symbol "NULL" is used to denote that sensors have not yet reached local decisions. When the local sensor decision is "NULL," we do not send anything. From a practical point of view, this is quite desirable, as it actually reduces the transmission rate below 1 bit. From a theoretical point of view, however, our proposed test implicitly assumes that $D_k \ge 3$ for each k (i.e., the sensor symbols are $0, 1$ and "NULL"). One referee asked what happens in case of only binary sensor messages ($D_k = 2$), say, with the symbol pairs being either $(0, 1)$ or (1, "NULL"). It turns out that such asymptotically Bayes tests can be still constructed. To illustrate our idea explicitly, without loss of generality, we assume that the binary senor symbols are 0 and 1. If necessary, we can treat 0 as the special symbol "NULL" and do not send anything if the sensor message is 0. The key idea is to extend our proposed test $\delta_A(c)$ by using "blocks" with block length 2. In each block, "00" and "11" represents that the sensor decides H_0 or H_1 is true, respectively, whereas "01" or "10" means that a local decision has not been reached yet. Mathematically, our proposed test with binary sensor messages can be defined as follows. At each sensor, for all $n \ge 1$, define $U_{k,2n-1} = I\{L_{k,2n-1} \ge 0\}$, where $I\{A\}$ is the indicator function of the set A, and

$$U_{k,2n} = \begin{cases} 1 & \text{if } U_{k,2n-1} = 1 \text{ and} \\ & \qquad L_{k,2n} \ge \rho_k|\log c|, \\ 0 & \text{if } U_{k,2n-1} = 0 \text{ and} \\ & \qquad L_{k,2n} \le -r_k|\log c|, \\ 1 - U_{k,2n-1} & \text{otherwise.} \end{cases}$$

Then the fusion center will make decisions only at time $2n$, and

$$\begin{cases} \text{stop and accept } H_1 & \text{if } U_{k,2n-1} = U_{k,2n} = 1 \text{ for all } k, \\ \text{stop and accept } H_0 & \text{if } U_{k,2n-1} = U_{k,2n} = 0 \text{ for all } k, \\ \text{continue sampling} & \text{otherwise.} \end{cases}$$

By a fairly straightforward though tedious extension of the proof of Theorem 2, it can be shown that this test is asymptotically Bayes in the system with full local memory.

IV. System with Limited Local Memory

Let $\delta_B^*(c)$ denote a Bayes solution to the decentralized sequential hypothesis testing problem in the system with limited local memory, specified in (6) or (7). The main result in this section is to develop a decentralized sequential test $\delta_B(c)$ in the case specified in (7) such that, under certain restrictions, $\lim_{c\to 0}\mathcal{R}_c(\delta_B^*(c))/\mathcal{R}_c(\delta_B(c)) = 1$, i.e., that $\{\delta_B(c)\}$ is "asymptotically Bayes" in the system with limited local memory as the cost of time step taken for decision making goes to 0, and $\mathcal{R}_c$ denotes the Bayes risk defined in (8) when c is the cost of time step taken for decision making.

A. The structure of our proposed test $\delta_B(c)$

In the system with limited local memory, our proposed test $\delta_B(c)$ splits the decision making into two stages.

1) First Stage: The purpose of this stage is to obtain a preliminary result on which of the two hypotheses H_0 and H_1 is true. To do so, each sensor S_k uses any fixed (not necessarily optimal) stationary MLRQ ϕ_k, i.e., the sensor message function

$$U_{k,n} = \phi_k(X_{k,n}) = d \text{ if and only if}$$
$$\lambda_{k,d} \le \frac{g_k(X_{k,n})}{f_k(X_{k,n})} < \lambda_{k,d+1},$$

where $0 = \lambda_{k,0} \le \lambda_{k,1} \le \cdots \le \lambda_{k,D_k-1} \le \lambda_{k,D_k} = \infty$ are pre-specified constants.

Based on independent, identically distributed observations $\mathbf{U}_n = (U_{1,n},\cdots,U_{K,n})$, the fusion center calculates the log-likelihood ratio $\mathbf{L}_{u,n}$ of $\mathbf{U}_n$ recursively by

$$\mathbf{L}_{u,n} = \mathbf{L}_{u,n-1} + \sum_{k=1}^{K}\log\frac{g_{\phi,k}(U_{k,n})}{f_{\phi,k}(U_{k,n})},$$

for $n \ge 1$ and $\mathbf{L}_{u,0} = 0$, where $f_{\phi,k}$ and $g_{\phi,k}$ are the probability mass function induced on $U_{k,n}$ when the observations $X_{k,n}$ are distributed as f_k and g_k, respectively. Then the fusion center decides to stop the first stage at the time

$$M_c^{(1)} = \text{ first } n \ge 1 \text{ such that } |\mathbf{L}_{u,n}| \ge \log|\log c|, \tag{25}$$

and makes a preliminary decision

$$V = \begin{cases} 0, & \text{if } \mathbf{L}_{u,M_c^{(1)}} \le -\log|\log c|; \\ 1, & \text{if } \mathbf{L}_{u,M_c^{(1)}} \ge \log|\log c|. \end{cases}$$

2) Second Stage: In this stage, each sensor needs to switch to an "optimal" MLRQ with respect to the preliminary decision in the first stage, and the fusion center will use the sensor messages from the second stage to make a final decision on which of hypotheses H_0 and H_1 is true.

Specifically, after the fusion center stops the first stage, it will send its preliminary decision V to all sensors as a binary quantized feedback. When receiving the quantized feedback V from the fusion center, each sensor S_k needs to adjust the constants thresholds in the stationary MLRQ's to optimal values as follows. If $V = 1$, then the sensor message function at sensor S_k is switched to the stationary MLRQ ψ_k which maximizes the Kullback-Leibler information number $I(g_{\psi,k}, f_{\psi,k})$, where $f_{\psi,k}$ and $g_{\psi,k}$ are the probability mass function induced on $U_{k,n}$ when the observations $X_{k,n}$ are distributed as f_k and g_k, respectively. On the other hand, if $V = 0$, then each sensor message function should be switched to the stationary MLRQ ψ_k which maximizes $I(f_{\psi,k}, g_{\psi,k})$. To abuse the notation, the stationary sensor message function at sensor S_k in the second stage are denoted by ψ_k to emphasize that an optimal MLRQ should be used. However, the actual choice of ψ_k will depend on the quantized feedback V.

In the second stage, the fusion center begins a new cycle of making decisions, discarding all previous sensor messages and conclusions and starting afresh on the incoming i.i.d. vector sensor messages $\mathbf{U}_n = (U_{1,n},\cdots,U_{K,n})$ which are sent from the sensors using optimal stationary MLRQ ψ_k's. To be more specific, at time n of the second stage, the fusion center center calculates the log-likelihood ratio $\hat{\mathbf{L}}_{u,n}$ of $\mathbf{U}_n$ recursively by

$$\hat{\mathbf{L}}_{u,n} = \hat{\mathbf{L}}_{u,n-1} + \sum_{k=1}^{K}\log\frac{g_{\psi,k}(U_{k,n})}{f_{\psi,k}(U_{k,n})},$$

for $n \ge 1$ and $\hat{\mathbf{L}}_{u,0} = 0$. Then the fusion center decides to stop the second stage at the time

$$M_c^{(2)} = \text{ first } n \ge 1 \text{ such that } |\hat{\mathbf{L}}_{u,n}| \ge |\log c|, \tag{26}$$

and makes the following final decision:

$$\begin{cases} \text{decide } H_0 \text{ is true }, & \text{if } \hat{\mathbf{L}}_{u,M_c^{(2)}} \le -|\log c|; \\ \text{decide } H_1 \text{ is true }, & \text{if } \hat{\mathbf{L}}_{u,M_c^{(2)}} \ge |\log c|. \end{cases}$$

It is important to emphasize that the threshold of the log-likelihood ratio statistic is $\log|\log c|$ in the first stage and is $|\log c|$ in the second stage. These choices of thresholds make sure that the number of time steps taken in the first stage (when the cost c is small) is large but is small relative to that in the second stage.

The motivation of our proposed test $\delta_B(c)$ is simple: instead of jointly optimizing the sensor and fusion center policies, which is an extremely difficult problem, we divide the optimization problem into two stages: the first stage is used to optimize the policies at the sensors, and then the second stage uses the "optimal" sensor policies to develop optimal policy at the fusion center. Similar ideas have been applied in sequential experiment design, see, for example, [1], [6].

In our proposed test $\delta_B(c)$, each sensor uses what we call a *tandem quantizer*. A tandem quantizer is a sequence of two stationary quantizers with almost one switch from one quantizer to the other. On the one hand, as we will see in the next section, stationary quantizers at sensors will not lead to asymptotically optimal Bayes solutions in the cases specified in (6) or (7) except in some extreme situations. In other words, sensor quantizers generally need to be changed over time in order to develop asymptotically Bayes decision rules. On the other hand, in the Bayes solution derived in [22] for the case specified in (6), the sensor quantizers need to be updated at each time step. This precludes the use of Bayes solution in practice. In fact, it is nontrivial to do numerical simulations for the Bayes solution due to its high frequencies of switches on sensor quantizers. Hence, from the viewpoint of the number of switches, a tandem quantizer is the simplest possible candidate to construct asymptotically Bayes rules.

B. Asymptotic Optimality

We need to first establish asymptotic lower bounds on the Bayes risk $R_c(\delta)$ in (8) for any decentralized tests in the system with limited local memory. Of course the lower bounds in Theorem 1 still hold in the system in the limited local memory. Unfortunately, they are too crude in the system with limited local memory, although the previous section showed that they are sharp in the system with full local memory. The following theorem, whose proof is highly nontrivial and included in the Appendix, is of fundamental importance for proving asymptotic optimality in the system with limited local memory.

Theorem 4: Assume $V_{D_k}(g_k, f_k)$, defined in (10), and $V_{D_k}(f_k, g_k)$, defined similarly by switching the role of f_k and g_k, are finite for all $1 \leq k \leq K$. Then for any tests δ in the system with limited local memory, if $R_c(\delta) \leq Mc|\log c|$ for some constant M, then

$$R_c(\delta) \geq (1 + o(1))c|\log c|\Big(\frac{\pi}{I_{\mathbf{D}}} + \frac{1-\pi}{J_{\mathbf{D}}}\Big),$$

where $I_{\mathbf{D}}$ and $J_{\mathbf{D}}$ are defined in (11).

Next, we study the asymptotic performance of our proposed test $\delta_B(c)$ for small values of c. As in the previous section, we need to estimate the probabilities of making incorrect decision when $\delta_B(c)$ is used and the asymptotic performances of the number of time steps required by $\delta_B(c)$ to make decisions. The following theorem, whose proof is in the Appendix, summarizes the asymptotic properties of our proposed test $\delta_B(c)$ in the system with limited local memory.

Theorem 5: For any $0 < c < 1$, the probabilities that our proposed test $\delta_B(c)$ makes incorrect decision satisfy

$$\mathbf{P}_0\{\text{reject } H_0\} \leq c \quad \text{and} \quad \mathbf{P}_1\{\text{rejects } H_1\} \leq c. \tag{27}$$

Moreover, if we let M_c denote the number of time step required by $\delta_B(c)$ to make decisions, then as $c \to 0$,

$$\begin{aligned} \mathbf{E}_0(M_c) &\leq (1 + o(1))|\log c|/I_{\mathbf{D}}, \\ \mathbf{E}_1(M_c) &\leq (1 + o(1))|\log c|/J_{\mathbf{D}}, \end{aligned} \tag{28}$$

where $I_{\mathbf{D}}$ and $J_{\mathbf{D}}$ are defined in (11).

Now we are in a position to show that our proposed test $\delta_B(c)$ is asymptotically optimal in the system with limited local memory.

Theorem 6: Under the assumption of finiteness of $V_{D_k}(g_k, f_k)$ and $V_{D_k}(f_k, g_k)$ for all $1 \leq k \leq K$, $\{\delta_B(c)\}$ is asymptotically Bayes in the system with full local memory, i.e., $\lim_{c\to 0} \mathcal{R}_c(\delta_B^*(c))/\mathcal{R}_c(\delta_B(c)) = 1$ where $\delta_B^*(c)$ is a Bayes solution in the system with full local memory when c is the cost per time step for decision making.

Proof: The proof follows the same lines as that of Theorem 3, using Theorems 4 and 5 for the system with limited local memory. □

Remark 6: The Bayes test $\delta_B^*(c)$ in the case specified in (6) was actually found in [22], however, its performance is sensitive to the error in estimating the value of π, the prior probability of H_0. On the contrary, our proposed test $\delta_B(c)$ uses tandem quantizers or two stages to adaptively adjust itself and does not depend on the prior probability π.

In addition, it has been a long standing open problem to study the asymptotic properties of the Bayes test $\delta_B^*(c)$ in the literature, see [21], [22]. Our theorems provide an asymptotic approximation to its performance. To see this, Theorem 5 implies that

$$\limsup_{c\to 0} \frac{\mathcal{R}_c(\delta_B(c))}{c|\log c|} \leq \frac{\pi}{I_{\mathbf{D}}} + \frac{1-\pi}{J_{\mathbf{D}}}.$$

Then by Theorems 4 and 6, we have

$$\begin{aligned} &\mathcal{R}_c(\delta_B^*(c)) = (1 + o(1))\mathcal{R}_c(\delta_B(c)) \\ &= (1 + o(1))c|\log c|\Big(\frac{\pi}{I_{\mathbf{D}}} + \frac{1-\pi}{J_{\mathbf{D}}}\Big) \end{aligned}$$

as $c \to 0$. Moreover, if we denote by $T_B^*(c)$ the stopping time (the time step taken for decision) of the Bayes solution $\delta_B^*(c)$, then the proof of Theorem 4 shows that

$$\begin{aligned} \mathbf{E}_0(T_B^*(c)) &= (1 + o(1))|\log c|/I_{\mathbf{D}}, \\ \mathbf{E}_1(T_B^*(c)) &= (1 + o(1))|\log c|/J_{\mathbf{D}}, \end{aligned}$$

as $c \to 0$.

V. STATIONARY SENSOR MESSAGES

In the system with limited local memory, an interesting open problem raised in [22] is to investigate the asymptotic optimality properties of tests with stationary sensor message functions, i.e., tests whose sensor message functions $\phi_{k,n} \equiv \phi_k$ do not change over time n. Tests with stationary sensor message functions are attractive in both theory and application due to their simple structures. The main purpose in this section is to provide a negative answer to this open problem and also to develop asymptotic optimality theory within the classes of tests with stationary sensor messages.

For tests with stationary sensor message functions, a key observation is that since the sensor message functions are fixed, the sensor messages $\mathbf{U}_n = (U_{1,n}, \cdots, U_{K,n})$ are i.i.d. vectors (conditional on each hypothesis) and the fusion center is faced with a classical sequential hypothesis testing problem. Hence for stationary sensor message functions, the optimal decision policy at the fusion center is an SPRT based on i.i.d sensor message vectors $\mathbf{U}_n = (U_{1,n}, \cdots, U_{K,n})$. For this reason, to find optimal sequential tests with stationary sensor message functions, it is sufficient for us to focus on those with an SPRT used at the fusion center.

Suppose a stationary sensor message function ϕ_k is used at sensor S_k for $1 \leq k \leq K$, and denote by $f_{\phi,k}$ and $g_{\phi,k}$ the distribution induced on $U_{k,n} = \phi_k(X_{k,n})$ when the distribution of $X_{k,n}$ is f_k and g_k, respectively. To abuse the notation, define two new information numbers

$$I(f_\phi, g_\phi) = \sum_{k=1}^{K} I(f_{\phi,k}, g_{\phi,k}),$$

$$I(g_\phi, f_\phi) \;=\; \sum_{k=1}^{K} I(g_{\phi,k}, f_{\phi,k}). \tag{29}$$

These two information numbers are nothing but the relative entropy of the sensor message vector $\mathbf{U}_n = (U_{1,n}, \cdots, U_{K,n})$ with stationary sensor message functions.

Given that a stationary sensor message function ϕ_k is used at sensor S_k for $1 \le k \le K$, let $\phi = (\phi_1, \cdots, \phi_K)$ and denote by $\delta_\phi(c)$ the sequential test where an optimal SPRT policy is used at the fusion center so as to minimize the Bayes risk $\mathcal{R}_c$ in (8). By the classical results on SPRT, we have

$$\mathcal{R}_c(\delta_\phi(c)) \approx c|\log c|\Big(\frac{\pi}{I(f_\phi, g_\phi)} + \frac{1-\pi}{I(g_\phi, f_\phi)}\Big), \tag{30}$$

as $c \to 0$. Note that $I(f_\phi, g_\phi) \le I_{\mathbf{D}}$ and $I(g_\phi, f_\phi) \le J_{\mathbf{D}}$, and at least one of the inequalities is strict for stationary sensor message function ϕ, except when ϕ is a one-to-one function, i.e, when the sensor observations $X_{k,n}$'s are random variables chosen from a finite alphabet of size D_k for all k. Thus, by Theorems 4 - 6, we have

Corollary 2: Under the conditions of Theorem 6, for any stationary sensor message ϕ that is not one-to-one, $\delta_\phi(c)$ is asymptotically suboptimal in the case specified in (6) or (7) in the system with limited local memory, except in the extreme situation where $\pi = 0$ or 1 and where no testing is necessary since we can decide in favor of H_1 when $\pi = 0$ and in favor of H_0 for $\pi = 1$, without taking any observations.

Remark 7: Using the fact that $I_{\mathbf{D}} \le I_{tot}$ and $J_{\mathbf{D}} \le J_{tot}$, it is evident that any test with stationary message function ϕ is also asymptotically suboptimal in the system with full local memory, except when ϕ is one-to-one or when $\pi = 0$ or 1.

While stationary sensor message functions generally lead to a suboptimal test in the system with limited local memory, it is reasonable to ask which is the "optimal" stationary sensor message function that leads to "optimal" tests among all SPRTs with stationary sensor message functions. By (30), for an SPRT $\delta_\phi(c)$ with stationary sensor message function ϕ, minimizing the Bayes risk in (8) is equivalent to minimizing

$$\frac{\pi}{I(f_\phi, g_\phi)} + \frac{1-\pi}{I(g_\phi, f_\phi)}.$$

Thus we may use this minimization to define the "optimal" stationary sensor message function ϕ. Unfortunately, by this definition, the optimal choice of ϕ will depend heavily on the value of π, the prior probability of the null hypothesis. This may be undesirable in practice since different researchers may have different opinions on the value of π.

To overcome this problem, we will use a minimax formulation to define an "optimal" choice of stationary sensor message ϕ. As a motivation, for a sequential test $\delta_\phi(c)$ with stationary sensor message ϕ, one can define its asymptotic efficiency as

$$\frac{\mathcal{R}_c(\delta^*_{cen}(c))}{\mathcal{R}_c(\delta_\phi(c))}$$

where $\delta^*_{cen}(c)$ is the Bayes solution in the centralized version. Thus, it is reasonable to define the efficiency of stationary sensor message functions ϕ by

$$e(\phi) = \min_{0 \le \pi \le 1} \liminf_{c \to 0} \Big[\frac{\mathcal{R}_c(\delta^*_{cen}(c))}{\mathcal{R}_c(\delta_\phi(c))}\Big].$$

Note that $e(\phi)$ can be thought of as a measure which reflects the performance efficiency of using quantized sensor observation with respect to raw sensor observations.

By (16) and (30), we have

$$e(\phi) = \min_{0 \le \pi \le 1} \Big[\frac{\pi/I_{tot} + (1-\pi)/J_{tot}}{\pi/I(f_\phi, g_\phi) + (1-\pi)/I(g_\phi, f_\phi)}\Big].$$

A simple algebraic calculation shows that the efficiency of stationary sensor message functions ϕ is

$$e(\phi) = \min\Big(\frac{I(f_\phi, g_\phi)}{I_{tot}}, \frac{I(g_\phi, f_\phi)}{J_{tot}}\Big), \tag{31}$$

where I_{tot} and J_{tot} are defined in (12), and $I(f_\phi, g_\phi)$ and $I(g_\phi, f_\phi)$ are defined in (29).

A minimax formulation for the "optimal" choices of sensor message functions ϕ is then to seek ϕ which maximizes the efficiency $e(\phi)$ in (31). While the problem of finding the "optimal" quantizer seems to be of interest on its own in the information theory literature, to the best of our knowledge, a minimax formulation like ours has not been proposed so far.

Note that in the definition of $e(\phi)$ in (31), I_{tot} and J_{tot} can be replaced by $I_{\mathbf{D}}$ and $J_{\mathbf{D}}$, respectively. This new definition corresponds to the situation where the benchmark test is the Bayes solution in the system with limited local memory instead of the Bayes solution in the centralized version. Both definitions are useful, and either one can be used in the following arguments.

To further understand the ideas behind our minimax formulation, it is helpful to consider a concrete example. For that purpose, in the reminder of this section, we will focus only on the system with a single sensor and binary sensor message, i.e., $L = 1$ and $D_l = 2$, where the sensor observations are normally distributed. But we want to emphasize that our arguments can be easily extended to a general setting.

In a system with a single sensor, suppose $X_1, X_2, \cdots$ are independent normal random variable with unknown mean μ and variance 1, and we are interested in testing the null hypothesis $H_0 : \mu = -\theta$ against the alternative hypothesis $H_1 : \mu = \theta$ for some $\theta > 0$. Assume we need to quantize the data X's because of data compression and limitations of channel bandwidth, i.e., $Y_n = \phi_\lambda(X_n) = I(X_n \ge \lambda)$ for some constant λ, where $I(A)$ is the indicator function of the set A. Then based on the quantized data $Y_1, Y_2, \cdots$, we want to decide which of H_0 and H_1 is true. The problem is how to choose the best quantizer, i.e., how to choose the best value λ.

Intuitively one would like to define $Y_n = I(X_n > 0)$, i.e., $\lambda = 0$ by the symmetry argument. However, our results imply the answer might depend on what we mean by "best," or more precisely, what is a prior probability π of the null hypothesis.

To simplify our notation in the following discussion, define

$$h(a, b) = a\log(a/b) + (1-a)\log((1-a)/(1-b)), \tag{32}$$

and

$$h_\theta(\lambda) = h(\Phi(\lambda+\theta), \Phi(\lambda-\theta)), \tag{33}$$

where $\Phi(\cdot)$ is the distribution function of the standard normal distribution. Now if the prior probability $\pi = 0$, then the "best" quantizer is determined by the value of λ which maximizes $I(g_{\phi_\lambda}, f_{\phi_\lambda}) = h(\Phi(-\lambda+\theta), \Phi(-\lambda-\theta)) = h_\theta(-\lambda)$. In other words, if one strongly believes H_1 is true, then one will want to maximize the information contained in Y (in the sense of Kullback-Leibler information number) when the true underlying model is H_1. Similarly, if $\pi = 1$, then the "best" quantizer is determined by the value of λ which maximizes $I(f_{\phi_\lambda}, g_{\phi_\lambda}) = h(\Phi(-\lambda-\theta), \Phi(-\lambda+\theta)) = h_\theta(\lambda)$. Here we used the facts that $h(a,b) = h(1-a, 1-b)$ and $\Phi(u) + \Phi(-u) = 1$. That is, if one strongly believes H_0 is true, then one will want to maximize information contained in Y when the true underlying model is H_0.

For $\mu = 1$, numerical calculations show that the best values of λ in binary sensor quantizer $U_{l,n} = I(X_{l,n} > \lambda)$ are $0.6008, -0.6008$ when $\pi = 0, 1$, respectively. For any other values of $\pi \in (0,1)$, numerical calculations seem to suggest that the optimal choice of λ decreases from 0.6008 to -0.6008 as π increases from 0 to 1. In particular, if $\pi = 1/2$, then the corresponding optimal value is $\lambda = 0$.

Our minimax formulation tackles the optimal stationary binary quantizer in this example from a different viewpoint in the sense that we want Y to contain reasonably rich information under both H_0 and H_1. Note that for a quantizer $\phi_\lambda(X) = I(X > \lambda)$, the efficiency $e(\phi_\lambda)$ in (31) becomes

$$e(\phi_\lambda) = \frac{\min(h_\theta(\lambda), h_\theta(-\lambda))}{2\theta^2},$$

where $h_\theta(\lambda)$ is defined in (33). By Theorem 7 below, for a given $\theta > 0$, $e(\phi_\lambda)$ is maximized at $\lambda = 0$, and thus the best binary sensor quantizer under our minimax formulation is $U_{l,n} = I(X_{l,n} > 0)$, which is consistent with our intuition.

It is worth emphasizing that while our minimax formulation and the Bayes formulation with $\pi = 1/2$ lead to the same optimal stationary sensor message functions for normal distributions, they *may* lead to different answers for other distributions. For instance, if $X_1, X_2, \cdots$ are i.i.d. with exponential density function $f_\theta(x) = \theta \exp(-\theta x)$ for $x \geq 0$, and we want to test $H_0 : \theta = 1/2$ against $H_1 : \theta = 2$. Then the best choices of λ in stationary quantizer $Y = I(X > \lambda)$ are 2.3088 and 1.3949 under our minimax formulation and the Bayes formulation with $\pi = 1/2$, respectively. The difference mainly arises from the asymmetric properties of Kullback-Leibler information numbers, i.e., $I(g,f) \neq I(f,g)$. From our point of view, our minimax formulation is more reasonable than the Bayes formulation (with $\pi = 1/2$) as ours takes into account the difference in information contained in the raw data X's under H_0 and H_1.

To complete this section, we will state the following theorem, whose proof is in the Appendix.

Theorem 7: For given $\theta > 0$, $\min(h_\theta(\lambda), h_\theta(-\lambda))$ is maximized at $\lambda = 0$, where $h_\theta(\lambda)$ is defined in (33).

VI. EXAMPLE

The main goal of this section is to illustrate our theoretical results in previous sections through a specific example. Detailed numerical and simulation results will be presented elsewhere.

Suppose there are K sensors with each sensor sending binary message to the fusion center, i.e., $D_k = 2$. Assume that the observations at sensor S_k are i.i.d. normal random variables with mean 0 and variance σ_k^2 under H_0 and with mean μ_k and variance σ_k^2 under H_1. An interesting application of this model, as mentioned in [18], is to use K geographically separated sensors to detect a deterministic signal (or target), which is contaminated by additive white Gaussian noise at each sensor.

Denote by $\rho_k = \mu_k^2/(2\sigma_k^2)$ the signal-to-noise ratio (SNR) at sensor S_k, and by

$$\rho = \sum_{k=1}^K \rho_k = \sum_{k=1}^K \mu_k^2/(2\sigma_k^2) \tag{34}$$

the signal-to-noise ratio (SNR) at the fusion center in the centralized version of this example.

First, in the system with full local memory specified in (1) and (2), if we assume each sensor is also allowed to be silent to indicate that it has not reached any local decision, then our proposed test $\delta_A(c)$ is well-defined and is asymptotically Bayes. Alternatively, we can use the test defined in Remark 1 for binary sensor messages. It is easy to show that $I_{tot} = J_{tot} = \rho$, and by Theorems 1 - 3, the Bayes risk of our proposed test $\delta_A(c)$ satisfies

$$R_c(\delta_A(c)) \sim R_c(\delta_A^*(c)) \sim R_c(\delta_{cen}^*(c)) \sim c|\log c|/\rho \tag{35}$$

as $c \to 0$, where $\delta_A^*(c)$ and $\delta_{cen}^*(c)$ are Bayes solution in the system with full local memory and in the centralized version, respectively. Here and everywhere below $x_c \sim y_c$ as $c \to 0$ means that $\lim_{c\to 0}(x_c/y_c) = 1$.

Next, in the system with limited local memory, Theorem 2 and Corollary 2 in [11] show that both $V_{D_k}(f_k, g_k)$ and $V_{D_k}(g_k, f_k)$ are finite if f_k and g_k are Gaussian distributions. Thus the sufficient conditions in Theorems 4 and 6 holds, and our proposed test $\delta_B(c)$ is asymptotically Bayes in the system with limited local memory. Note that

$$I_{\mathbf{D}} = J_{\mathbf{D}} = \sum_{k=1}^K \sup_\lambda h(\Phi(\lambda - \mu_k/\sigma_k), \Phi(\lambda)),$$

and $h(a,b)$ is defined in (32). Moreover, by Proposition 2 in [11], if the SNR ρ_k's are small at all sensors, then

$$I_{\mathbf{D}} = J_{\mathbf{D}} \approx \frac{2}{\pi}\rho \approx \frac{2}{3.14}\rho.$$

Therefore, in the system with limited local memory in the case specified in (6) or (7), if all SNRs at local sensors are small, then the Bayes risks of our proposed test $\delta_B(c)$ and the Bayes solution $\delta_B^*(c)$ satisfy

$$R_c(\delta_B(c)) \sim R_c(\delta_B^*(c)) \sim 1.57c|\log c|/\rho, \tag{36}$$

as $c \to 0$.

Finally, at sensor S_k, by linear transformation and Theorem 7, it is easy to show that the optimal binary stationary sensor messages under our minimax formulation is $\phi_k(x) = I(x > \mu_k/2)$, which is consistent with our intuition. When this optimal binary stationary quantizer is used at sensor S_k for all k, the information numbers defined in (29) in this example become

$$I(f_\phi, g_\phi) = I(f_\phi, g_\phi) = \sum_{k=1}^{K} h\Big(\Phi(-\frac{\mu_k}{2\sigma_k}), \Phi(\frac{\mu_k}{2\sigma_k})\Big),$$

and $h(a,b)$ is defined in (32). Using Taylor expansions of $\Phi(x)$ and $h(a,b)$, it is straightforward to show that

$$\lim_{\max_k \rho_k \to 0} \frac{I(f_\phi, g_\phi)}{\rho} = \frac{1}{\pi} \approx \frac{1}{3.14}.$$

Therefore, if the SNRs ρ_k's are small at all sensors, then the Bayes risks of the test $\delta_\phi(c)$ with binary stationary quantizer satisfy

$$R_c(\delta_\phi(c)) \sim 3.14 c |\log c| / \rho, \tag{37}$$

as $c \to 0$.

A comparison of (35)-(37) leads to the following interesting comments. Let us choose the SPRT with the optimal binary stationary quantizer as the baseline decentralized sequential test, and denote by R_0 the corresponding Bayes risk. Then the Bayes solutions in the system with full and limited local memory will asymptotically reduce the Bayes risk to $R_0/3.14$ and $R_0/2$, respectively. Moreover, as demonstrated by our proposed test $\delta_B(c)$ in the system with limited local memory, a one-shot one-bit feedback from the fusion center to sensors can reduce the asymptotic risk of decentralized sequential tests by half, but more feedbacks will not improve the asymptotic performance further in the system with limited local memory. Finally, without any feedback from the fusion center, a test in the system with full local memory can achieve the same asymptotic performance as the optimal centralized test.

VII. Conclusions

We have studied a decentralized extension of sequential hypothesis testing problems in two different scenarios of sensor networks under a Bayes formulation. In the system with full local memory, we have developed the first of asymptotically optimal decentralized sequential tests. It fact, our proposed decentralized tests have the same asymptotic first-order performances as the optimal centralized tests, although it may perform poorly in some nonasymptotic settings due to the slow convergence rate. In the system with limited local memory, we have used the idea of tandem quantizers to offer decentralized sequential tests which are simple but asymptotically Bayes, and addressed a long-standing open problem on the asymptotic performance of the Bayes solution. We also clarified issues involving the optimal stationary sensor quantizers, and proposed a minimax formulation under which the optimal stationary quantizer is consistent with our intuition. It is interesting to note that the feedback from the fusion center does not improve asymptotic performance in the system with full local memory, but does so in the system with limited local memory, even with one-bit one-shot feedback.

There are a number of interesting problems which have not been addressed here. In practice, one may be interested in testing multiple hypotheses as in [5], [8]. The results developed here are for the binary hypothesis testing problems, but they provide benchmarks and ideas for the development of tests in multiple hypothesis testing problems. It is also of interest to study the system where the observations at the different sensors may be dependent. Moreover, while the tests developed are asymptotically optimal, they may perform poorly in some practical situations, especially in the system with large number of sensors, because of the slow asymptotic convergence. Thus finding fairly simple decentralized tests which are not only asymptotically optimal, but have good performance for practical values, will undoubtedly be of great importance. The ideas of tandem quantizers, or more generally quantized feedback, will provide a powerful tool to tackle these problems. Therefore, this article is just the beginning of further investigation.

Acknowledgement

The author would like to thank Gary Lorden for his enthusiastic and inspiring guidance of the author's thesis research, of which this work is an outgrowth. Venugopal V. Veeravalli and Alexander G. Tartakovsky are also thanked for helpful discussions, as well as two referees, whose comments improved the paper.

Appendix
Proof of Theorems and Lemmas

A. *Proof of Theorem 1*

The main tool is a lower bound on the sample size N of a test with Type I and II error probabilities α and β. By the well-known Wald's inequality [17], [24], for any (sequential or fixed-sample) test with Type I and II error probabilities α and β, its sampling size N satisfy

$$\mathbf{E}_0(N) \geq [(1-\alpha)\log\frac{1-\alpha}{\beta} + \alpha \log \frac{\alpha}{1-\beta}]/I_{tot},$$
$$\mathbf{E}_1(N) \geq [(1-\alpha)\log\frac{1-\alpha}{\beta} + \alpha \log \frac{\alpha}{1-\beta}]/J_{tot},$$

where I_{tot} and J_{tot} are defined in (12). As both α and β go to 0, these inequalities become

$$\mathbf{E}_0(N) \geq (1+o(1))|\log \beta|/I_{tot}, \tag{38}$$
$$\mathbf{E}_1(N) \geq (1+o(1))|\log \alpha|/J_{tot}.$$

Based on these inequalities, Chernoff [4] presented the following elegant proof of the theorem. Suppose a family of tests $\{\delta_c\}$ satisfies $R_c(\delta_c) \leq Mc|\log c|$ for some constant $M > 0$. Denote by α_c and β_c Type I and II error probabilities of δ_c. By the definition of $R_c(\delta_c)$ in (8) and the assumption that $R_c(\delta_c) \leq Mc|\log c|$, we have $\alpha_c \leq Mc|\log c|/(\pi W_0)$ and $\beta_c \leq Mc|\log c|/((1-\pi)W_1)$. By (38), the stopping time τ_c of the test δ_c satisfies

$$\begin{aligned}\mathbf{E}_0(\tau_c) &\geq (1+o(1))|\log \beta_c|/I_{tot} \\ &\geq (1+o(1))\frac{|\log((Mc|\log c|)/((1-\pi)W_1))|}{I_{tot}},\end{aligned}$$

which is simply $(1+o(1))|\log c|/I_{tot}$, as $c \to 0$, due to the fact

$$|\log(c|\log c|)| = |\log c + \log|\log c|| = (1+o(1))|\log c|.$$

Similarly

$$\mathbf{E}_1(\tau_c) \geq (1+o(1))\frac{|\log c|}{J_{tot}}.$$

Now using the definition of $R_c(\delta)$ in (8) again, for any test, the cost of time step taken for decision is only portion of the total expected cost or Bayes risk. Thus,

$$\begin{aligned} R_c(\delta_c) &\geq c[\pi \mathbf{E}_0(\tau_c) + (1-\pi)\mathbf{E}_1(\tau_c)] \\ &\geq (1+o(1))c|\log c|\Big(\frac{\pi}{I_{tot}} + \frac{1-\pi}{J_{tot}}\Big). \end{aligned}$$

Hence the theorem holds. □

B. *Proof of Theorem 2*

Note that the stopping time $\hat{T}(c)$ of our proposed test $\delta_A(c)$ can be written as $\hat{T}(c) = \min(T_0(c), T_1(c))$, where

$$T_0(c) = \inf\big\{n : L_{k,n} \leq -r_k|\log c| \text{ for all } k\big\}, \tag{39}$$

and

$$T_1(c) = \inf\big\{n : L_{k,n} \geq \rho_k|\log c| \text{ for all } k\big\}. \tag{40}$$

For our proposed test $\delta_A(c)$, when stopping taking observations, the fusion center will accept H_0 if $\hat{T}(c) = T_0(c)$, and will accept H_1 if $\hat{T}(c) = T_1(c)$. Equivalently, our proposed test $\delta_A(c)$ rejects H_0 if and only if $T_1(c) < T_0(c)$. Hence,

$$\begin{aligned} \mathbf{P}_0\{\delta_A(c) \text{ rejects } H_0\} &= \mathbf{P}_0(T_1(c) < T_0(c)) \\ &\leq \mathbf{P}_0(T_1(c) < \infty). \end{aligned}$$

For any stopping time τ, using Wald's likelihood ratio identity [17], [24], we have

$$\mathbf{P}_0(\tau < \infty) = \mathbf{E}_1 \exp(-\sum_{k=1}^{K} L_{k,\tau};\ \tau < \infty),$$

where $L_{k,n}$ is the log-likelihood ratio at sensor S_k defined in (17). By definition, for the stopping time $\tau = T_1(c)$, we have $L_{k,\tau} \geq \rho_k|\log c|$ for $k = 1, 2, \cdots, K$, and $\sum_{k=1}^{K}\rho_k = 1$. Thus

$$\begin{aligned} \mathbf{P}_0(T_1(c) < \infty) &\leq \mathbf{E}_1 \exp(-\sum_{k=1}^{K}\rho_k|\log c|; \\ &\qquad T_1(c) < \infty) \\ &= \mathbf{E}_1(\exp(-|\log c|;\ T_1(c) < \infty) \\ &\leq \exp(-|\log c|) = c \end{aligned}$$

since $0 < c < 1$. This proves the first inequality in (21). The second inequality in (21) can be proved similarly.

Now we will study the properties of the stopping time $\hat{T}(c)$ of our proposed test $\delta_A(c)$. It suffices to prove the second inequality in (22), as the proof of the first inequality in (22) is identical. To accomplish this, for $1 \leq k \leq K$, let

$$N_k = \inf\Big\{n \geq 1 : \sum_{i=1}^{n} \log\frac{g_k(X_{k,i})}{f_k(X_{k,i})} \geq \rho_k|\log c|\Big\},$$

and

$$\tau_k(N_k) = \sup\Big\{n \geq 1 : \sum_{i=N_k+1}^{N_k+n} \log\frac{g_k(X_{k,i})}{f_k(X_{k,i})} \leq 0\Big\}.$$

For simplicity, denote $\tau_k = \tau_k(0)$. It is well-known (e.g., Theorem D in [6]) that for any $1 \leq k \leq K$,

$$\mathbf{E}_1(\tau_k) < \infty \tag{41}$$

since $\log\big(g_k(X)/f_k(X)\big)$ has positive mean and finite variance under $\mathbf{P}_1$ by Assumption (*A2*).

By definition of N_k and $\tau_k(N_k)$, the stopping time $\hat{T}(c)$ of our proposed test $\delta_A(c)$ satisfies

$$\begin{aligned} \hat{T}_c &\leq \max_{1\leq k\leq K}\Big(N_k + \tau_k(N_k) + 1\Big) \\ &\leq \max_{1\leq k\leq K} N_k + \sum_{k=1}^{K}\tau_k(N_k) + 1. \end{aligned}$$

Now since $X_{k,1}, X_{k,2}, \ldots$ are i.i.d. under $\mathbf{P}_1$, we have $\mathbf{E}_1(\tau_k(N_k)) = \mathbf{E}_1(\tau_k)$, and thus

$$\mathbf{E}_1(\hat{T}_c) \leq \mathbf{E}_1\big(\max_{1\leq k\leq K} N_k\big) + \sum_{k=1}^{K}\mathbf{E}_1(\tau_k) + 1. \tag{42}$$

By renewal theory and Assumption (*A2*), under $\mathbf{P}_1$,

$$\mathbf{E}_1(N_k) = \frac{|\log c|}{J_{tot}} + O(1) \text{ and } Var_1(N_k) = O(|\log c|),$$

as $c \to 0$, see [16] and [17, p. 171]. Hence,

$$\begin{aligned} \big(\mathbf{E}_1\big|N_k - \frac{|\log c|}{J_{tot}}\big|\big)^2 &\leq \mathbf{E}_1(N_k - \frac{|\log c|}{J_{tot}})^2 \\ &= Var_1(N_k) + (\mathbf{E}_1 N_k - \frac{|\log c|}{J_{tot}})^2 \\ &= O(|\log c|), \end{aligned}$$

and so

$$\mathbf{E}_1\big|N_k - \frac{|\log c|}{J_{tot}}\big| = O(\sqrt{|\log c|}).$$

Thus

$$\begin{aligned} \mathbf{E}_1 \max_{1\leq k\leq K} N_k &= \frac{|\log c|}{J_{tot}} + \mathbf{E}_1 \max_{1\leq k\leq K}\big(N_k - \frac{|\log c|}{J_{tot}}\big) + 1 \\ &\leq \frac{|\log c|}{J_{tot}} + \sum_{k=1}^{K}\mathbf{E}_1\big|N_k - \frac{|\log c|}{J_{tot}}\big| + 1 \\ &= \frac{|\log c|}{J_{tot}} + O(\sqrt{|\log c|}). \end{aligned}$$

Combining this with relations (41) and (42) yields the first inequality in (22), completing the proof of the theorem. □

C. *Proof of Theorem 4*

The proof follows the same lines as that of Theorem 1. Assume a family of sequential tests $\{\delta_c\}$ in the system with limited local memory satisfies $R_c(\delta_c) \leq M(c|\log c|)$ for some constant $M > 0$. Then the definition of $R_c(\delta_c)$ in (8) implies that Type I and II error probabilities of the tests δ_c satisfy with $\alpha \leq Mc|\log c|/(\pi W_0)$ and $\beta \leq Mc|\log c|/((1-\pi)W_1)$. Now by a new lower bound established in Lemma 1 (below) for tests in the system with limited local memory, the stopping times τ_c of the tests δ_c satisfy

$$\begin{aligned}\mathbf{E}_1(\tau_c) &\geq (1+o(1))\frac{|\log(Mc|\log c|/(\pi W_0))|}{J_{\mathbf{D}}}\\ &= (1+o(1))\frac{|\log c|}{J_{\mathbf{D}}}\end{aligned}$$

as $c \to 0$. Similarly, switching the role of f_k and g_k, we have

$$\mathbf{E}_0(\tau_c) \geq (1+o(1))\frac{|\log c|}{I_{\mathbf{D}}}.$$

Then by the definition of $R_c(\delta)$ in (8), for any test, the cost of time step taken for decision is only portion of the total expected cost or Bayes risk. Thus,

$$\begin{aligned}R_c(\delta_c) &\geq c[\pi\mathbf{E}_0(\tau_c) + (1-\pi)\mathbf{E}_1(\tau_c)]\\ &\geq (1+o(1))c|\log c|\Big(\frac{\pi}{I_{\mathbf{D}}} + \frac{1-\pi}{J_{\mathbf{D}}}\Big),\end{aligned}$$

and the theorem holds.

To complete the proof, we need to prove the following lemma which establishes asymptotic lower bounds on the time steps taken for decision making by a (sequential or fixed sample size) test in the system with limited local memory.

Lemma 1: Assume $V_{D_k}(g_k, f_k)$, defined in (10), is finite for all $1 \leq k \leq K$. Denote by τ the stopping time (number of time step for decision making) of a test δ in the system with limited local memory satisfying

$$\mathbf{P}_0(\text{reject } H_0) \leq \alpha, \quad \mathbf{P}_1(\text{reject } H_1) \leq \beta, \tag{43}$$

where $0 < \alpha, \beta < 1$. Then as $(\alpha, \beta) \to (0, 0)$,

$$\mathbf{E}_1(\tau) \geq (1+o(1))\frac{|\log \alpha|}{J_{\mathbf{D}}}, \tag{44}$$

where $J_{\mathbf{D}}$ is defined in (11). Moreover, if $V_{D_k}(f_k, g_k)$ is also finite for all $1 \leq k \leq K$, then

$$\mathbf{E}_0(\tau) \geq (1+o(1))\frac{|\log \beta|}{I_{\mathbf{D}}}, \tag{45}$$

where $I_{\mathbf{D}}$ is defined in (11).

Proof: We only need to prove that (44) holds since the proof of (45) is identical. For a test δ satisfying (43), if $\mathbf{E}_1(\tau) = \infty$, then it is obvious that (44) holds. Thus it is suffices to show (44) holds if $\mathbf{E}_1(\tau) < \infty$.

Given the overwhelming difficulty of directly studying the hypothesis testing problem in the system with limited local memory, we need to prove the lemma by considering the corresponding "open-ended" hypothesis testing problems, developed by Robbins [13] and Robbins and Siegmund [14], [15]. For a decentralized test δ in the system with limited local memory satisfying (43), relation (44) will be proved through studying the properties of an open-ended test constructed from the test δ.

The motivation of the open-ended hypothesis testing problem is as follows. Assume that if H_0 is true, sampling costs nothing and the preferred action at the fusion center is just to take observations without stopping. On the other hand, if H_1 is true, each time step for decision making costs a fixed amount and the fusion center should stop taking observations as soon as possible and reject the null hypothesis H_0.

Since there is only terminal decision in an open-ended hypothesis testing problem, the policy at the fusion center is a stopping time N. The null hypothesis H_0 is rejected if and only if $N < \infty$.

Before we continue to prove the lemma, it is useful to mention two different viewpoints of open-ended hypothesis testing problems. In the above-mentioned motivation of open-ended hypothesis testing problems, the cost for each time step under the hypothesis H_0 is different than that under the alternative hypothesis H_1, whereas these costs are the same (or at least the same order) in the standard formulation of hypothesis testing problems. An alternative viewpoint is to think of open-ended hypothesis testing problems as standard hypothesis testing problems in which the tests are required to satisfy the constraint in (43) with $\beta = 0$. For that reason, an open-ended test is also called "power-one" test in the literature.

Now let us go back to the proof of relation (44) for a test δ satisfying (43). The key idea is to use the stopping time τ of the test δ to construct an open-ended test in the system with limited local memory.

For that purpose, let $a = |\log \alpha|$ and we need to first define a pre-specified open-ended test $M(a)$ in the system with limited local memory as follows. Each sensor uses the optimal MLRQ φ_k :

$$U_{k,n} = \varphi_k(X_{k,n}) = d \text{ if and only if}$$
$$\lambda_{k,d} \leq \frac{g_k(X_{k,n})}{f_k(X_{k,n})} < \lambda_{k,d+1},$$

where $0 = \lambda_{k,0} \leq \lambda_{k,1} \leq \cdots \leq \lambda_{k,D_k-1} \leq \lambda_{k,D_k} = \infty$ are optimally chosen in the sense that the Kullback-Leibler information number $I(g_{\varphi,k}, f_{\varphi,k})$ achieves the supremum $I_{D_k}(g_k, f_k)$. Here $f_{\varphi,k}$ and $g_{\varphi,k}$ are the probability mass function induced on $U_{k,n}$ when the observations $X_{k,n}$ are distributed as f_k and g_k, respectively.

Based on the independent, identically distributed observations $\mathbf{U}_n = (U_{1,n}, \ldots, U_{K,n})$, the fusion center then uses the one-sided SPRT with log-likelihood ratio boundary a, i.e., the fusion center stops taking observations at time

$$\begin{aligned}M(a) &= \text{first } n \text{ such that}\\ &\sum_{i=1}^{n}\Big(\sum_{k=1}^{K}\log\frac{g_{\varphi,k}(U_{k,i})}{f_{\varphi,k}(U_{k,i})}\Big) \geq a,\end{aligned} \tag{46}$$

(and $M(a) = \infty$ if such n does not exist), and rejects the null hypothesis H_0 whenever stopping taking observations. It is evident that the stopping time $M(a)$ defines an open-ended test in the system with limited local memory. By classical results on SPRT, see Equations (8.3) and (8.4) of Siegmund [17], using the fact that $a = \log|\alpha|$, we have

$$\begin{aligned}&\mathbf{P}_0\big(M(a) < \infty\big) \leq \alpha, \quad \text{and}\\ &\mathbf{E}_1(M(a)) = |\log \alpha|/J_{\mathbf{D}} + O(1),\end{aligned} \tag{47}$$

where $J_{\mathbf{D}}$ is defined in (11).

Now for an arbitrary test δ in the system with limited local memory satisfying (43), we can construct the corresponding open-ended test by combining its stopping time τ with our pre-specified stopping time $M(a)$. Specifically, define a new stopping time $\hat{\tau}_\delta$ by

$$\hat{\tau}_\delta = \begin{cases} \tau & \text{if } \delta \text{ accepts } H_1 \\ \tau + M_1 & \text{if } \delta \text{ accepts } H_0 \end{cases},$$

where M_1 is the stopping time obtained by applying $M(a)$, defined in (46) with $a = |\log \alpha|$, to all sensors observations from time τ on, i.e., $X_{k,\tau+1}, X_{k,\tau+2}, \ldots$ for $1 \le k \le K$. Then the new stopping time $\hat{\tau}_\delta$ defines an open-ended test which will declare that the null hypothesis H_1 is true if and only if $\hat{\tau}_\delta$ is finite.

By definition, the open-ended test $\hat{\tau}_\delta$ is a test in the system with limited local memory since it is the combination of two tests, δ and $M(a)$, in the system with limited local memory. Note that if H_1 is true, both τ and M_1 will stop with probability 1, and hence $\hat{\tau}$ will also stop with probability 1 under H_1. In other words, if H_1 is true, then $\hat{\tau}_\delta$ will stop and reject the null hypothesis H_0 with probability 1. This suggests that $\hat{\tau}_\delta$ indeed is an open-ended test in the system with limited local memory.

Now we need to study the properties of our new open-ended test $\hat{\tau}_\delta$. Because M_1 is independent of τ, and as a copy of $M(a)$, M_1 also satisfies (47), we have

$$\begin{aligned} \mathbf{P}_0\Big(\hat{\tau}_\delta < \infty\Big) &\le \mathbf{P}_0\Big(\delta \text{ accept } H_1\Big) \\ &\quad + \mathbf{P}_0\Big(M_1 < \infty;\ \delta \text{ accepts } H_0\Big) \\ &\le \alpha + \mathbf{P}_0\big(M_1 < \infty\big) \le 2\alpha, \end{aligned}$$

and

$$\begin{aligned} \mathbf{E}_1(\hat{\tau}_\delta) &= \mathbf{E}_1(\tau;\ \delta \text{ accepts } H_1) + \\ &\quad + \mathbf{E}_1\big(\tau + M_1;\ \delta \text{ accepts } H_0\big) \\ &= \mathbf{E}_1(\tau) + \mathbf{E}_1(M_1)\mathbf{P}_1\Big(\delta \text{ accepts } H_0\Big) \\ &\le \mathbf{E}_1(\tau) + \beta\ \mathbf{E}_1(M_1) \\ &= \mathbf{E}_1(\tau) + \beta\Big(\frac{|\log\alpha|}{J_{\mathbf{D}}} + O(1)\Big). \end{aligned}$$

Meanwhile, using lemma 2 below, for any open-ended test N in the system with limited local memory satisfying

$$\mathbf{P}_0(N < \infty) \le \alpha_1 (< 1), \tag{48}$$

we have

$$\mathbf{E}_1(N) \ge (1 + o(1))\frac{\log|\alpha_1|}{J_{\mathbf{D}}}, \tag{49}$$

where $o(1) \to 0$ uniformly as $\alpha_1 \to 0$. In particular, this is true for $N = \hat{\tau}_\delta$, our open-ended tests constructed from the test δ. Combining the above three inequalities yields

$$\begin{aligned} \mathbf{E}_1(\tau) &\ge (1 + o(1))\frac{|\log(2\alpha)|}{J_{\mathbf{D}}} - \beta\ \Big(\frac{|\log\alpha|}{J_{\mathbf{D}}} + O(1)\Big) \\ &= (1 + o(1))\frac{|\log\alpha|}{J_{\mathbf{D}}}, \end{aligned}$$

where as $(\alpha, \beta) \to (0, 0)$, the term $o(1) \to 0$ and does not depend on τ. Relation (44) follows at once from the arbitrariness of the test δ and its corresponding stopping time τ, and thus Lemma 1 holds. □

To complete the proof of Lemma 1, we need to prove the following results for the open-ended tests, i.e., tests defined by a stopping time τ that stop taking observations only to reject H_0.

Lemma 2: Under the assumption of Lemma 1, for any stopping time τ in the system with limited local memory satisfying $\mathbf{P}_0(N < \infty) \le \alpha_1 (< 1)$, relation (48) implies (49).

Proof: A key idea in the proof of this lemma is to use conditional log-likelihood ratios to construct a martingale, and the proof follows the same argument as in Theorem 1 of [11]. Note that in the system with limited local memory, we can rewrite

$$U_{k,n} = \varphi_{k,n}(X_{k,n}),$$

where $\varphi_{k,n}$ may depend on $\mathcal{E}_{n-1}$, all past sensor messages defined in (3). Denote by $f^{\varphi}_{k,n}$ and $g^{\varphi}_{k,n}$ respectively the conditional density induced on $U_{k,n}$ given $\mathcal{E}_{n-1}$ when the density of $X_{k,n}$ is f_k and g_k. Denote by $Z_{k,n}$ the conditional log-likelihood ratio function of $U_{k,n}$, $\log(\Big(g^{\varphi}_{k,n}(U_{k,n})/f^{\varphi}_{k,n}(U_{k,n})\Big)$.

Since $X_{1,n}, \cdots, X_{K,n}$ are independent, so are $U_{1,n}, \cdots, U_{K,n}$ given $\mathcal{E}_{n-1}$. Thus in the fusion center, the conditional log-likelihood ratio of $(U_{1,n}, \cdots, U_{K,n})$ given $\mathcal{E}_{n-1}$ is $\mathbf{Z}_n = \sum_{k=1}^{K} Z_{k,n}$. By Theorem 1 (or Theorem 3) in Lai [8] (also see Theorem 1 of Lai [7]), to prove (49), it suffices to show that for any $\delta > 0$,

$$\limsup_{n\to\infty} \mathbf{P}_1\Big\{\max_{t \le n} \sum_{i=1}^{t} \mathbf{Z}_i \ge I_{\mathbf{D}}(1+\delta)n\Big\} = 0. \tag{50}$$

By the definition of $I_{D_k}(g_k, f_k)$,

$$\mathbf{E}_1(\mathbf{Z}_i) = \sum_{k=1}^{K} \mathbf{E}_1(Z_{k,i}) \le \sum_{k=1}^{K} I_{D_k}(g_k, f_k) = I_{\mathbf{D}},$$

thus the left-hand size of (50) is less than or equal to

$$\begin{aligned} &\limsup_{n\to\infty} \mathbf{P}_1\Big\{\max_{t\le n} \sum_{i=1}^{t}\sum_{k=1}^{K} \Big(Z_{k,i} - \mathbf{E}_1(Z_{k,i})\Big) \ge I_{\mathbf{D}}\delta n\Big\} \\ &\le \sum_{k=1}^{K} \limsup_{n\to\infty} \mathbf{P}_1\Big\{\max_{t\le n} \sum_{i=1}^{t} \Big(Z_{k,i} - \mathbf{E}_1(Z_{k,i})\Big) \ge \delta_1 n\Big\}, \end{aligned}$$

where $\delta_1 = I_{\mathbf{D}}\delta/K$.

Note that $\sum_{i=1}^{n}(Z_{k,i} - \mathbf{E}_1(Z_{k,i}))$ is a martingale. Doob's submartingale inequality (Theorem 14.6 of [25]) implies that

$$\begin{aligned} &\mathbf{P}_1\Big\{\max_{t\le n} \sum_{i=1}^{t} \Big(Z_{k,i} - \mathbf{E}_1(Z_{k,i})\Big) \ge \delta_1 n\Big\} \\ &\qquad \le \sum_{i=1}^{n} \mathbf{E}_1(Z_{k,i})^2/(\delta_1^2 n^2). \end{aligned}$$

Note that for any i, $\mathbf{E}_1(Z_{k,i})^2 \le V_{D_k}(g_k, f_k)$ by definition, and hence

$$\mathbf{P}_1\Big\{\max_{t\le n} \sum_{i=1}^{t} \Big(Z_{k,i} - \mathbf{E}_1(Z_{k,i})\Big) \ge \delta_1 n\Big\} \le \frac{V_{D_k}(g_k, f_k)}{\delta_1^2 n},$$

which implies (50) since $V_{D_k}(g_k, f_k)$ is finite. Relation (49) follows. □

D. Proof of Theorem 5

To prove the theorem, it is useful to think our proposed test $\delta_B(c)$ in the system with limited local memory as a combination of two SPRTs with stationary MLRQ at sensors. Denote by $\delta_B^{(1)}(c)$ and $\delta_B^{(2)}(c)$ the SPRT tests at the first and second stage, respectively. Now at the first stage, $\delta_B^{(1)}(c)$ is an SPRT with log-likelihood ratio threshold values $\pm \log|\log c|$, and by the classical results on SPRT, we have

$$\mathbf{P}_0(V=1) \le \frac{1}{|\log c|}, \quad \mathbf{P}_1(V=0) \le \frac{1}{|\log c|},$$

and

$$\mathbf{E}_0(M_c^{(1)}) \le \frac{\log|\log c|}{I_1}, \quad \mathbf{E}_1(M_c^{(1)}) \le \frac{\log|\log c|}{J_1},$$

where V is the decision of $\delta_B^{(1)}(c)$, i.e., the preliminary decision of our proposed test $\delta_B(c)$, $M_c^{(1)}$ is the time step taken at the first stage, and $I_1 = \sum_{k=1}^K I(f_{\phi,k}, g_{\phi,k})$ and $J_1 = \sum_{k=1}^K I(g_{\phi,k}, f_{\phi,k})$ are Kullback-Leibler information numbers of sensor messages corresponding to the stationary MLRQ ϕ_k's used in the first stage. By definition in (11), we have $I_1 \le J_{\mathbf{D}}$ and $J_1 \le I_{\mathbf{D}}$.

At the second stage, note that $\delta_B^{(2)}(c)$ only uses the observations in the second stage to make decisions, and thus conditional on the preliminary decision V, the SPRTs at two stages, $\delta_B^{(1)}(c)$ and $\delta_B^{(2)}(c)$, are independent. Note that $\delta_B^{(2)}(c)$ is an SPRT with log-likelihood ratio threshold values $\pm|\log c|$, by the classical results for SPRT, we have

$$\mathbf{P}_0(\delta_B^{(2)} \text{ reject } H_0|V) \le c, \quad \mathbf{P}_1(\delta_B^{(2)} \text{ reject } H_1|V) \le c,$$

regardless of the decision V at the first stage. However, if we denote by $M_c^{(2)}$ the time step taken at the second stage, i.e., $M_c^{(2)}$ is the stopping time of $\delta_B^{(2)}(c)$, then the properties of $M_c^{(2)}$ will depend on V. Specifically,

$$\begin{aligned}\mathbf{E}_0(M_c^{(2)}|V=0) &\le (1+o(1))|\log c|/I_{\mathbf{D}},\\ \mathbf{E}_1(M_c^{(2)}|V=0) &\le (1+o(1))|\log c|/J_2,\end{aligned}$$

where $I_{\mathbf{D}}$ and J_2 are Kullback-Leibler information numbers of sensor messages corresponding to optimal stationary MLRQs ψ_k's used at sensors, where the MLRQ ψ_k maximizes $I(f_{\psi,k}, g_{\psi,k})$ for all k. Similarly,

$$\begin{aligned}\mathbf{E}_0(M_c^{(2)}|V=1) &\le (1+o(1))|\log c|/I_2,\\ \mathbf{E}_1(M_c^{(2)}|V=1) &\le (1+o(1))|\log c|/J_{\mathbf{D}},\end{aligned}$$

where I_2 and $J_{\mathbf{D}}$ are Kullback-Leibler information number of sensor messages corresponding to optimal stationary MLRQs ψ_k's used at sensors, where the MLRQ ψ_k maximizes $I(g_{\psi,k}, f_{\psi,k})$ for all k.

Now we are in a position to prove the theorem. Let us first consider the error probabilities of our proposed test $\delta_B(c)$. To prove the first inequality in (27), note that the final decision of $\delta_B(c)$ is just the decision of $\delta_B^{(2)}(c)$. Thus

$$\begin{aligned} & \mathbf{P}_0\Big(\delta_B(c) \text{ rejects } H_0\Big)\\ = & \mathbf{P}_0\Big(\delta_B^{(2)}(c) \text{ rejects } H_0\Big)\\ = & \mathbf{E}_0\Big(\mathbf{P}_0\big(\delta_B^{(2)}(c) \text{ rejects } H_0|V\big)\Big)\\ \le & \mathbf{E}_0(c) = c,\end{aligned}$$

where the second equality uses the conditional expectation on V, and the third inequality uses the property of $\delta_B^{(2)}(c)$. The second inequality in (27) can be proved similarly.

Next, let us consider the time step M_c taken by our proposed test $\delta_B(c)$. Observe that $M_c = M_c^{(1)} + M_c^{(2)}$, where $M_c^{(1)}$ and $M_c^{(2)}$ are time steps taken by $\delta_B^{(1)}(c)$ at the first stage and by $\delta_B^{(2)}(c)$ at the second stage, respectively. Hence,

$$\begin{aligned}\mathbf{E}_0(M_c) &= \mathbf{E}_0(M_c^{(1)}) + \mathbf{E}_0(M_c^{(2)})\\ &= \mathbf{E}_0(M_c^{(1)}) + \mathbf{E}_0(M_c^{(2)}|V=1)\mathbf{P}_0(V=1)\\ &\quad +\mathbf{E}_0(M_c^{(2)}|V=0)\mathbf{P}_0(V=0)\end{aligned}$$

Applying the results for $\delta_B^{(1)}(c)$ and $\delta_B^{(2)}(c)$, as $c \to 0$, we have

$$\begin{aligned}\mathbf{E}_0(M_c) &\le (1+o(1))\frac{\log|\log c|}{I_1}\\ &\quad +(1+o(1))\frac{|\log c|}{I_2} \times \frac{1}{|\log c|}\\ &\quad +(1+o(1))\frac{|\log c|}{I_{\mathbf{D}}} \times 1\\ &= (1+o(1))\frac{|\log c|}{I_{\mathbf{D}}}.\end{aligned}$$

This proved the first inequality in (28). The second inequality in (28) can be proved similarly. Thus the theorem holds. □

E. Proof of Theorem 7

To prove the theorem, by Lemma 7 below, for any given $\theta \ge 0$, $h_\theta(\lambda)$ defined in (33) is a decreasing function of $\lambda \in [0, \infty)$. Thus

$$\min(h_\theta(\lambda), h_\theta(-\lambda)) \le h_\theta(|\lambda|) \le h_\theta(0),$$

with equality holding for $\lambda = 0$. This shows that $\min(h_\theta(\lambda), h_\theta(-\lambda))$ is maximized at $\lambda = 0$, and so the theorem is proved.

To complete the proof, we need the following lemmas.

Lemma 3: Define

$$a(x) = (\phi(x))^2 + x\phi(x)\Phi(x) - (\Phi(x))^2.$$

Then $a(x) \le 0$ for all $x \in \mathbb{R}$.

Proof: Note that the first-order derivative of $a(x)$ is

$$\begin{aligned}a'(x) &= 2(-x)\phi(x)\phi(x) + \phi(x)\Phi(x) + x(-x\phi(x))\Phi(x)\\ &\quad +x\phi(x)\phi(x) - 2\phi(x)\Phi(x)\\ &= -x(\phi(x))^2 - (x^2+1)\phi(x)\Phi(x).\end{aligned}$$

On the one hand, if $x \geq 0$, it is easy to see $a'(x) \leq 0$. On the other hand, if $x \leq 0$, let $u = -x$, then $\phi(u) = \phi(-u)$ and

$$a'(x) = (\phi(u))^2(u^2+1)\Big[\frac{u}{u^2+1} - \frac{\Phi(-u)}{\phi(u)}\Big],$$

which is negative for all $u \geq 0$ by the well-known fact (e.g., proposition 14.8 on page 141 of [25]) that

$$\frac{\Phi(-u)}{\phi(u)} \geq \frac{1}{u+1/u} = \frac{u}{u^2+1}.$$

Thus, $a'(x) \leq 0$ for all $x \in \mathbb{R}$, and so $a(x)$ is a decreasing function of x over $\mathbb{R}$. Clearly, as $x \to -\infty$, both $\phi(x)$ and $\Phi(x)$ go to 0, and hence $a(x)$ goes to 0. Therefore, $a(x) \leq 0$ for all $x \in \mathbb{R}$. $\square$

Lemma 4: The function $b(x) = \frac{\phi(x)}{\Phi(x)} + x$ is an increasing function of $x \in \mathbb{R}$.

Proof: The lemma follows at once from the fact that

$$b'(x) = -\frac{a(x)}{(\Phi(x))^2},$$

which is positive for all $x \in \mathbb{R}$ by Lemma 3. $\square$

Lemma 5: For all $x \in \mathbb{R}$,

$$\frac{\phi(x)}{\Phi(-x)} - \frac{\phi(x)}{\Phi(x)} \leq \max(0, 2x). \tag{51}$$

Proof: If $x \leq 0$, then the conclusion is trivial since $\Phi(-x) \geq \Phi(x)$. If $x \geq 0$, then the left-hand side of (51) is equal to

$$(b(-x)+x) - (b(x)-x) = b(-x) - b(x) + 2x,$$

which is less than $2x$ since Lemma 4 implies $b(-x) \leq b(x)$ for all $x \geq 0$. $\square$

Lemma 6: Define

$$\begin{aligned} A(x,y) &= \phi(y)\Big[\log\frac{\Phi(y)}{\Phi(x)} - \log\frac{\Phi(-y)}{\Phi(-x)}\Big] \\ &\quad -\phi(x)\Big[\frac{\Phi(y)}{\Phi(x)} - \frac{\Phi(-y)}{\Phi(-x)}\Big]. \end{aligned} \tag{52}$$

Then $A(x,y) \leq 0$ for any $y \geq \max(x, -x)$.

Proof: A simple algebra calculation shows that

$$\frac{\partial A(x,y)}{\partial y} = \phi(y)B(x,y),$$

where

$$\begin{aligned} B(x,y) &= -y\Big[\log\frac{\Phi(y)}{\Phi(x)} - \log\frac{\Phi(-y)}{\Phi(-x)}\Big] \\ &\quad + \frac{\phi(y)}{\Phi(y)\Phi(-y)} - \frac{\phi(x)}{\Phi(x)\Phi(-x)}. \end{aligned}$$

Now it is easy to show that

$$\frac{\partial B(x,y)}{\partial x} = \frac{\phi(x)}{\Phi(x)\Phi(-x)}\Big[y + x + \frac{\phi(x)}{\Phi(x)} - \frac{\phi(x)}{\Phi(-x)}\Big],$$

which is positive for all $y \geq \max(x, -x)$ by Lemma 5. Since $B(y,y) = 0$, we know $B(x,y) \leq 0$ for all $y \geq \max(x, -x)$. Thus for each fixed x, $A(x,y)$ is a decreasing function of y if $y \geq \max(x,-x)$. Hence, $A(x,y) \leq A(x,|x|)$ for all $y \geq \max(x,-x)$. To prove the lemma, it now suffices to show that $A(x,|x|) \leq 0$ for all $x \in \mathbb{R}$.

If $x \geq 0$, then it is easy to see that $A(x,|x|) = A(x,x) = 0$. On the other hand, if $x \leq 0$, then $A(x,|x|) = A(x,-x)$, which can rewritten as

$$A(x,-x) = \phi(x)[2\log u - u + 1/u],$$

where

$$u = \frac{\Phi(-x)}{\Phi(x)} \geq 1$$

since $x \leq 0$. Let $w(u) = 2\log u - u + 1/u$, then

$$w'(u) = 2/u - 1 - 1/u^2 = -(1-1/u)^2 \leq 0.$$

Thus $w(u)$ is a decreasing function of u, and for all $u \geq 1$, $w(u) \leq w(1) = 0$. This shows that $A(x,-x) \leq 0$ for all $x \leq 0$. Therefore, $A(x,y) \leq 0$ for all $y \geq \max(x,-x)$. $\square$

Lemma 7: For any $\theta \geq 0$, the function $h_\theta(\lambda)$ defined in (33) is a decreasing function of $\lambda \in [0,\infty)$.

Proof: By (32) and (33), it is straightforward to show that

$$\frac{\partial}{\partial\lambda}h_\theta(\lambda) = A(\lambda-\theta, \lambda+\theta),$$

where $A(x,y)$ is defined in (52). Note that for any given $\theta \geq 0$, $\lambda+\theta \geq \max(\lambda-\theta, -(\lambda-\theta))$ for all $\lambda \geq 0$. Applying Lemma 6, we have $A(\lambda-\theta,\lambda+\theta) \leq 0$ for all $\lambda \geq 0$, i.e., the first-order derivative of $h_\theta(\lambda)$ (with respect to λ) is negative for $\lambda \in [0,\infty)$. Hence, the lemma is proved. $\square$

References

[1] L. R. Abramson, "Asymptotic sequential design of experiments with two random variables," *J. R. Stat. Soc. Ser. B Stat. methodol.*, vol. 28, No. 1, pp. 73-87, 1966.

[2] M. Basseville and I. Nikiforov, *Detection of Abrupt Changes: Theory and Applications.* Englewood Cliffs: Prentice-Hall, 1993.

[3] R. S. Blum, S. A. Kassam, and H. V. Poor, "Distributed detection with muliple sensors: part II- advanced topics," *Proceedings of the IEEE*, vol. 85, no. 1, pp. 64-79, 1997.

[4] H. Chernoff, "Sequential design of experiment," *Ann. Math. Statist.*, vol. 30, pp. 755-770, 1959.

[5] V. P. Dragalin, A. G. Tartakovsky, and V. V. Veeravalli, "Multihypothesis sequential probability ratio tests – part I: Asymptotic optimality," *IEEE Trans. Inform. Theory*, vol. 45, pp. 2448-2461, 1999.

[6] J. Kiefer and J. Sacks, "Asymptotically optimal sequential inference and design," *Ann. Math. Statist.*, vol. 34, pp. 705-750, 1963.

[7] T. L. Lai, "Information bounds and quick detection of parameter changes in stochastic systems," *IEEE Trans. Inform. Theory*, vol. 44, pp. 2917-2929, 1998.

[8] T. L. Lai, "Sequential multiple hypothesis testing and efficient fault detection-isloation in stochastic systems," *IEEE Trans. Inform. Theory*, vol. 46, pp. 595-608, 2000.

[9] T. L. Lai, "Sequential analysis: some classical problems and new challenges (with discussion)," *Statistica Sinica*, vol. 11, pp. 303-408, 2001.

[10] Y. Mei, "Asymptotically optimal methods for sequential change-point detection," *Ph.D. thesis*, California Institute of Technology, 2003. Available at <http:// resolver.caltech.edu/CaltechETD:etd-05292003-133431> .

[11] Y. Mei, "Information bounds and quickest change detection in decentralized decision systems," *IEEE Trans. Inform. Theory*, vol. 51, pp. 2669-2681, Jul. 2005.

[12] A. G. O. Mutambara, *Decentralized estimation and control for multi-sensor systems*, CRC Press, 1998.

[13] H. Robbins, "Statistical methods related to the law of the iterated logarithm," *Ann. Math. Statist.*, vol. 41, pp. 1397-1409, 1970.

[14] H. Robbins and D. Siegmund, "Boundary crossing probabilities for the Wiener process and partial sums," *Ann. Math. Statist.* vol. 41, pp. 1410-1429, 1970.

[15] H. Robbins and D. Siegmund, "Statistical tests of power one and the integral representation of solutions of certain partial differential equations," *Bull. Inst. Math., Academia Sinica* vol. 1, pp. 93-120, 1973.
[16] D. Siegmund, "The variance of one-sided stopping rules," *Ann. Math. Statist.*, vol. 40, pp. 1074-1077, 1969.
[17] D. Siegmund, *Sequential Analysis: Tests and Confidence Intervals*. New York: Springer-Verlag, 1985.
[18] A. G. Tartakovsky and V. V. Veeravalli, "An efficient sequential procedure for detecting changes in multichannel and distributed systems," *Proceedings of the 5th International conference on Information fusion*. Annapolis, Maryland, July 2002, vol. 2, Page 1-8.
[19] R. R. Tenney and N. R. Sandell Jr., " Detection with distributed sensors," *IEEE Trans. Aerospace Elect. Syst.* , vol. AES-17, pp. 501-510, 1981.
[20] J. N. Tsitsiklis, "Extremal properties of likelihood ratio quantizers," *IEEE Trans. Commun.*, vol. 41, pp. 550-558, 1993.
[21] V. V. Veeravalli, "Sequential decision fusion: Theory and applications," *J. Franklin Inst.*, vol. 336, pp. 301-322, Feb. 1999.
[22] V. V. Veeravalli, T. Basar, and H. V. Poor, "Decentralized sequential detection with a fusion center performing the sequential test," *IEEE Trans. Inform. Theory*, vol. 39, pp. 433-442, Mar. 1993.
[23] R. Viswanathan and P. K. Varshney, "Distributed detection with multiple sensors — Part I: Fundamentals," Proceedings of the IEEE, vol. 85, pp. 54-63, Jan 1997.
[24] A. Wald, *Sequential Analysis*. New York: Wiley, 1947.
[25] D. Williams, *Probabilty with Martingles*. Cambridge univeristy press, 1991.
[26] M. Woodroofe, *Nonlinear Renewal Theory in Sequential Analysis*. Society for Industrial and Applied Mathematics, Philadelphia, 1982.
[27] J. Xiao and Z. Luo, "Decentralized estimation in an inhomgeneous sensing environment," *IEEE Trans. Inform. Theory*, vol. 51, pp. 3564-3575, Oct. 2005.