Asymptotic Optimality Theory For Decentralized Sequential Multihypothesis Testing Problems

Yan Wang, Yajun Mei

Abstract—The Bayesian formulation of sequentially testing $M \geq 3$ hypotheses is studied in the context of a decentralized sensor network system. In such a system, local sensors observe raw observations and send quantized sensor messages to a fusion center which makes a final decision when stopping taking observations. Asymptotically optimal decentralized sequential tests are developed from a class of “two-stage” tests that allows the sensor network system to make a preliminary decision in the first stage and then optimize each local sensor quantizer accordingly in the second stage. It is shown that the optimal local quantizer at each local sensor in the second stage can be defined as a maximin quantizer which turns out to be a randomization of at most $M - 1$ unambiguous likelihood quantizers (ULQ). We first present in detail our results for the system with a single sensor and binary sensor messages, and then extend to more general cases involving any finite alphabet sensor messages, multiple sensors, or composite hypotheses.

Index Terms—Asymptotic optimality, maximin quantizer, multihypotheses testing, sequential detection, two-stage tests, unambiguous likelihood quantizer (ULQ).

I. INTRODUCTION
Sequential detection or sequential hypothesis testing has many important real-world applications such as target detection in multiple-resolution radar (Marcus and Swerling [15]), serial acquisition of direct-sequence spread spectrum signals (Simon et al. [19]) and statistical pattern recognition (Fu [7]). The centralized version, in which all observations are available at a single central location, has been well studied. For example, when testing $M = 2$ hypotheses, a well-known optimal centralized test is the sequential probability ratio test (SPRT) developed by Wald [29], also see Wald and Wolfowitz [30]. When testing $M \geq 3$ hypotheses, i.e., in the sequential multihypothesis testing problem, there is no tractable closed-form expression for the optimal centralized sequential tests, although various asymptotically optimal sequential tests have been proposed and investigated in the literature, see, for example, Kiefer and Sacks [10], Lorden [14], Draglin, Tartakovsky and Veeravalli [5], [6].

In recent years, the decentralized version of sequential hypothesis testing problems has gained a great amount of attention and has been applied into a wide range of applications such as military surveillance (Tenney and Sandell [21]), target tracking and classification (Li et al. [13]), and data filtering (Ye et al. [31]). Under a widely used decentralized setting, raw data are observed at a set of geographically deployed sensors, whereas the final decision is made at a central location, often called the fusion center. The key feature here is that raw observations at the local sensors are generally not directly accessible by the fusion center, and the local sensors need to send quantized summary messages (generally belonging to a finite alphabet set) to the fusion center. This is due to limited communication bandwidth and requirements of high communication robustness.

Unfortunately, decentralized sequential hypothesis testing problems are very challenging, and to the best of our knowledge, existing research is restricted to testing two simple hypotheses, for example, see Veeravalli [26], Veeravalli, Basar and Poor [27], Nguyen, Wainwright and Jordan [18], and Mei [17]. It has been an open problem to find any sort of asymptotically optimal solutions for the decentralized sequential testing problem when testing $M \geq 3$ hypotheses. This is not surprising, because even in the centralized version, it requires sophisticated mathematical and statistical techniques and only asymptotic optimality results are available.

The primary goal of this paper is to develop a class of asymptotically optimal decentralized sequential procedures for testing $M \geq 3$ hypotheses. To do so, a major challenge we need to overcome is finding the “optimal quantizers” that can best send quantized summary sensor messages from the local sensors to the fusion center so as to lose as little information as possible. Intuitively, such a quantizer should depend on the true distribution of the raw data, which is unknown, and thus stationary quantizers are generally not optimal. In addition, since a quantizer can be any measurable function as long as its range is in the given finite alphabet set, it resides in an infinite dimensional functional space. Hence it is essential to investigate the form of the “optimal quantizers” so that one can reduce the infinite dimensional functional space to a finite-dimensional parameter space for the purpose of theoretical analysis and numerical computation. Note that when testing $M = 2$ hypotheses, Tsitsiklis [23] and Veeravalli et al. [27] showed that the optimal quantizers can be found from the family of monotone likelihood ratio quantizers (MLRQ), whose form is defined up to a finite number of parameters. Unfortunately, such a result does not apply to the case of testing $M \geq 3$ hypotheses. To find the form of the optimal quantizers for multi-hypotheses, we propose to combine three existing methodologies together: two-stage tests in Stein [20] and Kiefer and Sacks [10] (or equivalently, tandem quantizers in Mei [17]), unambiguous likelihood quantizers (ULQ) in Tsitsiklis [23], and randomized quantizers (see Chernoff [3],...
for a closely related topic on randomized experiments).

The remainder of the paper is organized as follows. Section II gives a rigorous formulation of decentralized sequential multihypothesis testing problems under a Bayesian framework. Section III provides a general definition of two-stage tests and discusses their implementation issues, especially those of the randomized quantizers. To highlight our main ideas, Section IV states our main results for a simplified sensor network system with a single sensor and binary sensor messages: Subsection IV-A develops asymptotically optimal decentralized sequential tests by considering two-stage tests when the local quantizers are the proposed “maximin quantizers,” and Subsection IV-B characterizes maximin quantizers and discusses their numerical computation issues. Section V extends our main results to three more general cases: (A) systems with finite alphabet sensor messages; (B) systems with conditionally independent multiple sensors; and (C) testing composite hypotheses. Numerical simulation results are presented in Section VI, and concluding remarks are included in Section VII. The technical details are provided in the appendices.

II. NOTATION AND PROBLEM FORMULATION

As illustrated in Fig. 1 in a widely used configuration, a sensor network consists of $K$ local sensors labeled by $S^1, \ldots, S^K$ and a fusion center which makes a final decision when stopping taking observations. At each time step $n = 1, 2, \ldots$, each local sensor $S^k$ observes raw data $\{X^k_n\}$ and sends quantized summary messages $\{U^k_n\}$ to the fusion center. Here the quantized messages $\{U^k_n\}$ are required to belong to a finite alphabet, say, $\{0, 1, \ldots, l^k - 1\}$, due to limited communication bandwidth or requirements of high communication robustness. In other words, the fusion center does not have direct access to the raw data $\{X^k_n\}$, and have to utilize the quantized sensor messages $\{U^k_n\}$ to make a final decision. If necessary, the fusion center can send feedback $\{V^k_n\}$ to the local sensors so as to improve the system efficiency.

To be more rigorous, we need to further specify the form of the sensor message functions. In this paper, we focus on systems with full feedback, but local memories restricted to past decisions, e.g., Case E of Veeravalli et al. [27]. Mathematically, at time $n$, for each $k = 1, 2, \ldots, K$, the quantized sensor message at the $k$th local sensor is assumed to be of the form

$$U^k_n = \phi^k_n(X^k_n; V^k_{n-1}) \in \{0, 1, \ldots, l^k - 1\}$$  

where the feedback $V^k_{n-1}$ is defined by

$$V^k_{n-1} = \psi^k_n(U^1_{1,n-1}, \ldots, U^K_{1,n-1})$$  

and $U^k_{1,n-1} = (U^1_{1,n-1}, \ldots, U^K_{1,n-1})$ denotes all past local sensor messages. That is, the quantizer $\phi^k_n$ is a function used by sensor $S^k$ to map the local raw data $X^k_n$ into $\{0, 1, \ldots, l^k - 1\}$, and the choice of $\phi^k_n$ can depend on the feedback $V^k_{n-1}$ and can be a randomized function (to be discussed later).

In decentralized sequential multihypothesis testing problems, there are $M$ hypotheses regarding the distribution $P$ of the raw data $\{X^k_n\}$:

$$H_m : P = P_m, m = 0, 1, \ldots, M - 1.$$  

Under each $P_m$, the raw data $X^k_n$ at local sensor $S^k$ are i.i.d. with density $f^k_m(\cdot)$ with respect to a common underlying measure, and the raw data $\{X^k_n\}$ are assumed to be independent among different sensors. Hence the distributions of the raw data under $P_m$ are completely determined by the $K$ densities: $f^1_m, \ldots, f^K_m$. Below we simply state that the true state of nature is $m$ or $P_m$ if the hypothesis $H_m$ is true. A decentralized sequential test $\delta$ consists of a rule to determine the sensor messages, a stopping time $N$ used by the fusion center and a final decision rule $D \in \{0, 1, \ldots, M - 1\}$ that chooses one of the $M$ probability measures $P_m$'s based on the information up to time $N$ at the fusion center. As in Wald [29], Veeravalli et al. [27], and Veeravalli [26], let $c > 0$ be the cost per time step until stopping, and let $W(m, m')$ be the loss of making decision $D = m'$ when the true state is $P_m$. It is standard to assume that $W(m, m) = 0$ but $W(m, m') > 0$ for any $m \neq m'$, i.e., no loss occurs if and only if a correct decision is made. Then when the true state of nature is $P_m$, the total expected cost of a decentralized test $\delta$ is

$$R_c(\delta; m) = cE_m(N) + \sum_{m'} W(m, m')P_m\{D = m'\}$$

where $E_m$ is the expectation operator under $P_m$. In a Bayesian formulation, we assign prior probabilities $\pi = (\pi_0, \ldots, \pi_{M-1})$ to the $M$ hypotheses $H_0, \ldots, H_{M-1}$. Hence, the Bayes risk of the decentralized test $\delta$ is

$$R_c(\delta) = \sum\pi_m R_c(\delta; m).$$

The Bayes formulation of the decentralized sequential multihypothesis testing problem can then be stated as follows.

**Problem (P1):** Minimize the $R_c(\delta)$ in (1) among all possible decentralized sequential multihypothesis test procedures $\delta$.

Denote by $\delta^*_B(c)$ a Bayes solution to (P1). In Veeravalli et al. [27], $\delta^*_B(c)$ is obtained through dynamic programming for the simplest case of testing binary hypotheses, i.e., $M = 2$. Unfortunately, in a general multihypothesis setting, when $M \geq 3$, it is impossible to find $\delta^*_B(c)$, since the problem is intractable even for the centralized version, see, for example, Dragalin, Tartakovsky and Veeravalli [3]. This prompts us to adopt the following asymptotic optimization approach in which the cost $c$ per time step goes to 0.

**Problem (P2):** Find a family of decentralized sequential multihypothesis testing procedures $\{\delta_A(c)\}$ that is asymptotically
optimal in the sense that
\[ \lim_{c \to 0} \mathcal{R}_c(\delta_B(c))/\mathcal{R}_c(\delta_A(c)) = 1. \] (5)

Now let us discuss the concepts of quantizers and their Kullback-Leibler (K-L) divergences, both of which will be essential in our asymptotic optimality theory. A quantizer is either a deterministic measurable function or a randomization of some (possibly infinitely many) deterministic measurable functions that maps the raw data into a finite alphabet set, e.g., the function \( \phi_k \) in (1) is a quantizer. The quantizer is called a deterministic quantizer if the corresponding measurable function is deterministic. At a given local sensor \( S \) (here and below we miss the superscript \( k \) for simplicity), denote by \( \Phi \) the set of all possible local deterministic quantizers \( \phi \)'s and let \( f_m(\cdot; \phi) \) be the induced probability mass function of quantized message \( U_n = \phi(X_n) \) when the raw observation \( X_n \) is distributed according to \( f_m(\cdot) \) under \( P_m \), i.e.,
\[ f_m(u; \phi) = P_m(\phi(X_n) = u), \quad u = 0, 1, \ldots, l - 1. \] (6)

For the deterministic quantizer \( \phi \), it is easy to see that its K-L divergences are defined by
\[ I(m, m'; \phi) = \sum_{u=0}^{l-1} f_m(u; \phi) \log \frac{f_m(u; \phi)}{f_{m'}(u; \phi)} \] (7)
for all \( m \neq m' \). However, we need to be very careful when defining the K-L divergences of a randomized quantizer \( \bar{\phi} = \sum p^{j} \phi^j \) that assigns probability masses \( \{ p^j \} \) onto some countable subset of deterministic quantizers \( \{ \phi^j \} \subset \Phi \). On the one hand, one can directly substituting the \( \phi \) in (7) by \( \bar{\phi} \), i.e.,
\[ \bar{I}(m, m'; \bar{\phi}) = \sum_{u=0}^{l-1} f_m(u; \bar{\phi}) \log \frac{f_m(u; \bar{\phi})}{f_{m'}(u; \bar{\phi})} \] (8)
where
\[ f_m(u; \bar{\phi}) = P_m(\bar{\phi}(X) = u), \quad u = 0, 1, \ldots, l - 1. \]

This type of the K-L divergence has been defined for randomized quantizers in the engineering literature, e.g., Tsitsiklis [23]. On the other hand, one can also define the K-L divergence of the randomized quantizer \( \bar{\phi} \) by the weighted average of those of the deterministic quantizers it randomizes:
\[ I(m, m'; \bar{\phi}) = \sum p^j I(m, m'; \phi^j), \quad 0 \leq m \neq m' \leq M - 1. \] (9)

By Jensen’s inequality, we have \( \bar{I}(m, m'; \bar{\phi}) \leq I(m, m'; \bar{\phi}) \), i.e., the K-L divergence defined in (8) is dominated by that in (9), also see Appendix [A] for more discussions.

To the best of our knowledge, the K-L divergence in (9) has not been studied in the literature so far, and it turns out that it will play a central role in our asymptotic theory. The reason why our asymptotic theory involves the K-L divergence in (9) instead of that in (8) is due to our novel way of implementing randomized quantizers to minimize loss of information. Roughly speaking, when implementing randomized quantizers, it is essential for the fusion center to know which specific deterministic quantizer is going to be used at the local sensor at each time step, since otherwise the fusion center can be confused by randomized quantizers and its decision making will be less efficient. This issue will be discussed further in Subsection [III-B]. Also note that a deterministic quantizer can also be thought as a randomized quantizer that assigns probability one to itself. Denote by \( \tilde{\Phi} \) the set of all possible quantizers at the local sensor \( S \), deterministic or randomized.

Throughout our paper we make the following standard assumption to ensure the finiteness of the expectation of the raw data’s log-likelihood ratios.

**Assumption 1.** For any two different states \( 0 \leq m \neq m' \leq M - 1 \) and local sensor \( S^k \),
\[ 0 < E_m \left[ \log \frac{f_{m'}(X_n)}{f_{m}(X_n)} \right] < \infty. \]

In the literature, researchers often assume a uniform bound on the second moments of the log-likelihood ratio \( \log \frac{f_{m'}(x^n; \phi)}{f_{m}(x^n; \phi)} \) under \( P_m \). See, for example, Kiefer and Sacks [10] and Mei [17]. Here our assumption is much weaker, and it turns out that it will be sufficient for the first-order asymptotic optimality under our setting.

### III. Two-Stage Test Procedures

In this section, we introduce a class of “two-stage” decentralized sequential tests in which each local sensor uses two stationary (possibly randomized) local quantizers with at most one switch between these two quantizers. This type of tests are useful because they allow the fusion center to first make a preliminary guess about the true state of nature and then optimize the procedure accordingly.

To highlight our main ideas, in the present and next sections we assume that the sensor network system consists of a single local sensor, i.e., \( K = 1 \) and all quantized messages are binary, i.e., \( U_n \in \{0, 1\} \). Extensions to general cases are presented in Section [V]. To save notations, we drop all the superscripts denoting the sensors. That is, in this and next sections we assume that one observes raw data \( X_1, X_2, \ldots \), which are i.i.d. with density \( f_m(x) \) under the hypothesis \( H_m \). The final decision is based on quantized messages \( U_n = \psi_n(X_n; V_{n-1}) \in \{0, 1\} \) with the feedback \( V_{n-1} = \psi_{n-1}(U_1, \cdots, U_{n-1}) \). For a given (randomized) quantizer \( \phi \), the K-L divergence of \( P_m \) from \( P_{m'} \) is \( I(m, m'; \phi) \) defined in (9).

#### A. Our Proposed Test

Our proposed two-stage test \( \delta(c) \) can be defined as follows. In the first stage of \( \delta(c) \), the local sensor can use any “reasonable” stationary deterministic quantizer and the fusion center needs to make a preliminary guess about the true state of nature. The only requirement is that as the cost \( c \to 0 \), the probabilities of making incorrect preliminary guess go to zero but the time steps taken at this first stage become negligible as compared to those of the overall procedure (or the second stage).

To be more concrete, let \( u(c) \in (0, 1/2) \) be a function of \( c \) such that \( u(c) \to 0 \) and \( \log u(c)/\log c \to 0 \) when \( c \to 0 \), e.g., \( u(c) = 1/|\log c| \). Choose a deterministic quantizer \( \phi^0 \) such that \( I(m, m'; \phi^0) > 0 \) for any two states \( 0 \leq m \neq m' \leq M - 1 \), and let the local sensor use
the stationary quantizer $\phi^0$ to send i.i.d. sensor messages $U_n = \phi^0(X_n)$ to the fusion center. Then the fusion center faces a classical sequential detection problem with the i.i.d. sensor messages $U_n$’s as inputs, and thus it is intuitively appealing to make a preliminary decision based on posterior distributions. Specifically, at each time step $n = 0, 1, \cdots$, the fusion center updates recursively the posterior distribution $(\pi_0, n, \pi_1, n, \cdots, \pi_{M-1}, n)$ as follows:

$$
\pi_{m, n} = \frac{\pi_{m, n-1} f_m(U_n; \phi^0)}{\sum_{0 \leq m' \leq M-1} \pi_{m', n-1} f_{m'}(U_n; \phi^0)}.
$$ (10)

Then the fusion center will stop the first stage at time step

$$
N_0 = \min \{ n \geq 0 : \max_{0 \leq m \leq M-1} \{ \pi_{m, n} \} \geq 1 - u(c) \}
$$

and when stopped, the fusion center makes a preliminary decision

$$
D_0 = \arg \max_{0 \leq m \leq M-1} \pi_{m, N_0}.
$$

Note that the preliminary decision $D_0$ is well-defined because the maximum value of $\pi_{m, N_0}$ is attained at only one index $m$ due to the definition of $N_0$ and the fact that $u(c) < 1/2$. For the purpose of practical implementation, the preliminary decision $D_0$ can be transmitted to the local sensor through a feedback of $\log_2 M$ bits.

In the second stage of our proposed test $\delta(c)$, the local sensor will switch to another stationary (likely randomized) quantizer that may depend on the preliminary decision $D_0$. Without loss of generality, we assume that the local sensor uses the stationary quantizer $\phi_m$ when the preliminary decision at the first stage is $D_0 = m$ for $m = 0, 1, \ldots, M-1$. Here we put a bar over $\phi_m$ to emphasize that it is likely a randomized quantizer when optimized, and we will postpone the detailed discussion about how to implement randomized quantizers to the next subsection.

Now at the second stage, the fusion center shall ignore the preliminary decision $D_0$ and continue to update the posterior distribution $(\pi_0, n, \pi_1, n, \cdots, \pi_{M-1}, n)$ with the sensor messages generated from the new quantizer $\phi_m$ when $D_0 = m$ (how to update will be discussed in the next subsection). Then the fusion center will stop the second stage (hence the whole procedure) at time step

$$
N = \min \{ n \geq N_0 : \max_{0 \leq m \leq M-1} \{ \pi_{m, n} \} \geq 1 - c \} \quad (11)
$$

and when stopped, the fusion center makes a final decision

$$
D = \arg \max_{0 \leq m \leq M-1} \pi_{m, N}.
$$

From the asymptotic point of view, many other possible decision rules can also be used at the fusion center. For instance, let $r_{m, n} = \sum_{m' \neq m} \pi_{m', n} W(m', m)$ be the average posterior cost when making a decision $m$ at time $n$, and then the fusion center can stop the second stage at time

$$
N = \min \{ N_m : 0 \leq m \leq M - 1 \} \quad (12)
$$

where

$$
N_m = \min \{ n \geq N_0 : r_{m, n} \leq c \}, \quad m = 0, 1, \cdots, M - 1. \quad (13)
$$

Based on our experiences, the stopping time $N$ defined in (12) is slightly better than that in (11) in finite-sample numerical simulations, especially when the costs $W(m', m)$ are not a simple $0 - 1$ cost function. Moreover, at the second stage, our proposed test will continue to update the posterior distribution instead of starting afresh as required by the two-stage tests in Section V of Kiefer and Sacks [10] or in Section IV of Mei [17]. The main reason is to further utilize information gathered from the first stage so as to improve the efficiency in finite-sample simulations, although it also means extra treatments in asymptotic arguments.

B. Implementing Randomized Quantizers and Updating Posterior Distribution

When testing $M \geq 3$ hypotheses, randomized quantizers are likely needed in the second stage in order to develop the optimal two-stage tests, and thus it is necessary to determine the appropriate approach to implement them as well as how to update posterior distributions at the fusion center, especially at the second stage. Assume a randomized quantizer is given by $\phi = \sum p^i \phi^i$. The key requirements for randomization in our two-stage test is that the fusion center must know which deterministic quantizer is picked to quantize the raw observation, since otherwise the randomization can cause confusion at the fusion center. The most straightforward (though practically infeasible) implementation is to let the fusion center do the randomization directly. Specifically, at time step $n$ the fusion center will choose the deterministic quantizer $\phi^j$ with probability $p^j$, say choosing the deterministic quantizer $\phi^j(m)$. Through a feedback from the fusion center, the local sensor will then use the chosen deterministic quantizer $\phi^j(m)$ at time step $n$ to quantize the raw observation. After receiving the quantized sensor message $U_n$ at time step $n$, the fusion center then update the posterior distribution as follows:

$$
\pi_{m, n} = \frac{\pi_{m, n-1} f_m(U_n; \phi^j(m))}{\sum_{m' = 0}^{M-1} \pi_{m', n-1} f_{m'}(U_n; \phi^j(m))}, \quad (14)
$$

because the fusion center knows that $U_n$ comes from the deterministic quantizer $\phi^j(m)$ at time step $n$.

A theoretically equivalent but more feasible implementation in practice is to adopt a “pseudo-randomization” at the local level through the so-called “block design” (see Section V of Kiefer and Sacks [10]). To be specific, suppose $\phi$ randomizes a finite number (say $i$) of deterministic quantizers, and all $p^i$’s are (or can be approximated by) rational numbers with a common denominator. Then we divide the time steps into blocks of size $b$, and within each block, the raw data are quantized with deterministic quantizers $\{\phi^1, \cdots, \phi^i\}$ following a fixed order such that each $\phi^j$ is used for exactly $p^j b$ times. Under this implementation, the fusion center again knows which deterministic quantizer is used at each time step, and thus can update the posterior distribution as in (14).

We would like to point out that our implementation of randomized quantizers is very different from those existing implementations in the literature (see Tsitsiklis [23]). In the latter the randomization is done at the local level in the sense that the local sensor randomly picks one of the deterministic quantizer $\phi^j$’s, and the fusion center will only get
the quantized message $U_n$ without knowing exactly which deterministic quantizer is used to generate $U_n$. In this case, to update posterior distribution, the fusion center has to plug in $\bar{\phi}$ (instead of $\phi^{(n)}$) into $[14]$, i.e.,

$$\pi_{m,n} = \frac{\pi_{m-1,n}f_m(U_n; \bar{\phi})}{\sum_{n=1}^{M-1} \pi_{m',n-1}f_{m'}(U_n; \bar{\phi})}.$$ 

Since our proposed implementation and the local randomization implementation lead different likelihood ratios, it is not surprising that there are two different kinds of K-L divergences for a randomized quantizer in Section II, one defined in [3] and the other in [9].

IV. MAIN RESULTS

In the present section, we show that a two-stage test can be an asymptotic optimal solution to problem (P2) by carefully choosing the quantizers used in the second stage. We also give characterizations of these optimal quantizers as well as the corresponding numerical computation.

A. Maximin Quantizers and Asymptotic Theory

Let us begin with the definition of some useful information numbers. For a given (deterministic or randomized) quantizer $\bar{\phi} \in \Phi$, define

$$I(m; \bar{\phi}) = \min_{m' \neq m} I(m, m'; \bar{\phi}).$$

(15)

for each state $m = 0, 1, \ldots, M - 1$. That is, $I(m; \bar{\phi})$ characterizes the least divergence from the state $m$ to other states.

The following theorem, whose proof is presented in Appendix B, establishes the asymptotic properties of a two-stage test $\delta(c)$ as the cost $c$ goes to 0.

**Theorem 4.1.** Let $\delta(c)$ be a two-stage test with $\{\bar{\phi}_0, \ldots, \bar{\phi}_{M-1}\}$ being the set of (possibly randomized) quantizers used in its second stage. Assume each $\bar{\phi}_m$ randomizes a finite number of deterministic quantizers, and suppose that the prior probabilities $\pi_m > 0$ and $I(m; \bar{\phi}_m) > 0$ for all states $m = 0, 1, \ldots, M - 1$ and $m' = 0, 1, \ldots, M - 1$. Then as $c \to 0$, the time steps $N$ taken by the two-stage test $\delta(c)$ satisfies

$$E_m \{N\} \leq (1 + o(1)) \log c/I(m; \bar{\phi}_m), \quad m = 0, 1, \ldots, M - 1,$$

(16)

and the final decision $D$ of the two-stage test $\delta(c)$ satisfies

$$P_m \{D \neq m\} = O(c \log c), \quad m = 0, 1, \ldots, M - 1.$$ 

(17)

Thus, the Bayes risk of the two-stage test $\delta(c)$ is

$$R_c(\delta) = c \log c (1 + o(1)) \sum_{m=0}^{M-1} \pi_m / I(m; \bar{\phi}_m).$$ 

(18)

In light of Theorem 4.1, to asymptotically minimize the Bayes risk within the class of two-stage tests, it is clear that one should maximize the information numbers $I(m; \bar{\phi}_m)$ for $m = 0, 1, \ldots, M - 1$. This leads to a natural definition of the optimal quantizers that we should use in the second stage:

**Definition 4.1.** For $m = 0, 1, \ldots, M - 1$, define the maximin quantizer with respect to $P_m$ as

$$\bar{\phi}_m^\text{max} = \arg \sup_{\bar{\phi} \in \Phi} I(m; \bar{\phi})$$

and define the corresponding maximin information number by

$$I(m) = \sup_{\bar{\phi} \in \Phi} I(m; \bar{\phi})$$

As shown later in Theorems 4.2 and 5.1, the supremum of $I(m, \bar{\phi})$ is attainable, and the maximin quantizers not only exist, but also can be realized as randomization of a finite number of deterministic quantizers. Now we are ready to investigate the asymptotic optimality properties of the two-stage test when the maximin quantizers are used in the second stage. Denote by $\delta_A(c)$ such a two-stage test. Then by Theorems 4.1 we have

$$R_c(\delta_A(c)) = (1 + o(1)) \log c \sum_{m=0}^{M-1} \pi_m / I(m).$$ 

(19)

as $c \to 0$. What is surprising is that $\delta_A(c)$ is not only the best one within the class of two-stage tests, but also asymptotically optimal among all possible decentralized tests.

A key step in the proof is the following important theorem which establishes asymptotic lower bounds on the expected time steps of any decentralized tests with “suitably small” probabilities of making incorrect decisions.

**Theorem 4.2.** Assume that $\delta(c)$ is a decentralized (not necessarily a two-stage) test that makes a final decision $D$ and

$$P_m \{D \neq m\} = O(c \log c), \quad m = 0, 1, \ldots, M - 1,$$

(19)

as $c \to 0$. Then the time step $N$ taken by $\delta(c)$ satisfies

$$E_m \{N\} \geq (\log c - \log \log c + O(1)) / I(m) = (1 + o(1)) \log c / I(m)$$

(20)

for all $m = 0, 1, \ldots, M - 1$.

The proof of Theorem 4.2 is presented in Appendix C. The first-order asymptotic lower bound will be sufficient to prove the first-order asymptotic optimality of $\delta_A(c)$, and the reason why we present a higher order lower bounds is due to its potential usefulness in higher-order analysis in further research. By relation (19) and Theorem 4.2, we have

**Corollary 4.1.** The procedure $\delta_A(c)$ is asymptotically Bayes up to first-order.

**Proof:** Let $\delta^*_B(c)$ be the Bayes procedure. By definition, $R_c(\delta^*_B(c)) \leq R_c(\delta_A(c))$. Using the relation (19) and the definition of Bayes risk $R_c(\delta^*_B(c))$, the probabilities for the Bayes procedure $\delta^*_B(c)$ to make incorrect decisions are at most $O(c \log c)$. By Theorem 4.2, the stopping time $\tau^*_B(c)$ of the Bayes procedure $\delta^*_B(c)$ satisfies (20). Now using the definition of Bayes risk again, for any test, the cost of time steps taken to make the final decision is only portion of the Bayes risk. In particular,

$$R_c(\delta^*_B(c)) \geq c \sum_m \pi_m E_m \{\tau^*_B(c)\} \geq (1 + o(1)) c \log c \sum_m \pi_m / I(m).$$

Combing all arguments yields that $R_c(\delta^*_B(c)) / R_c(\delta_A(c)) \to 1$ as $c \to 0$, completing the proof of the corollary.
It is useful to point out that the test $\delta_A(c)$ is asymptotic Bayes mainly because the local sensor uses the maximin quantizers $\phi_m^{\text{max}}$s in the second stage. Since the maximin quantizers do not depend on the prior distribution $\{\pi_m\}$s, it is easy to see from (16) and (17) that the asymptotic optimality properties of $\delta_A(c)$ are actually robust with respect to $\{\pi_m\}$ as long as all prior probabilities are positive. Likewise, the asymptotic Bayes properties still hold if the stopping times of $\delta_A(c)$ at the fusion center are replaced by other efficient multi-hypotheses tests, e.g., those in Draglin, Tartakovsky and Veeravalli [5], [6].

B. Characterizing the Maximin Quantizers.

In this subsection, we provide a deeper understanding of the maximin quantizers $\{\phi_m^{\text{max}} : m = 0, 1, \ldots, M - 1\}$ and also illustrate how to compute them explicitly when the sensor messages are binary.

Let us first introduce the unambiguous likelihood quantizer (ULQ) which was first proposed in Tsitsiklis [23] as a generalization of Monotone Likelihood Ratio Quantizer (MLRQ). For notational convenience, here we give the definition of ULQ only for the case of binary sensor messages, and the general definition will be provided in Definition 5.1 in Subsection V-A.

**Definition 4.2.** A deterministic quantizer $\phi \in \Phi$ is said to be an unambiguous likelihood quantizer if there exist real numbers $\{a_m : m = 0, \ldots, M - 1\}$ such that

$$\phi(X) = I(\sum_{m=0}^{M-1} a_m f_m(X) > 0)$$

and for any $0 \leq m' \leq M - 1$, the set $\{a_m\}$ satisfies the following condition

$$P_{m'} \left\{ \sum_{m=0}^{M-1} a_m f_m(X) = 0 \right\} = 0.$$  (22)

When relation (22) holds for any set of $\{a_m\}$ that are not simultaneous zero, the set of pdf’s $\{f_m\}$ are said to be linearly independent. With the definition of ULQs, the following theorem characterizes the form of the maximin quantizers $\phi_m^{\text{max}}$. The proof is very technical and is deferred to Appendix A.

**Theorem 4.3.** For each $m = 0, 1, \ldots, M - 1$, the maximin quantizer $\phi_m^{\text{max}}$ exists and can be chosen as a randomization of at most $M - 1$ deterministic quantizers. Moreover, if the pdf’s $\{f_m\}$ are linearly independent, then it can actually be chosen as a randomization of at most $M - 1$ deterministic ULQ quantizers.

Clearly, when testing $M = 2$ simple hypotheses, the ULQs become MLRQs, and thus the maximin quantizers in the second stage is just the deterministic MLRQ, which is consistent with those results in Mei [17].

Note that Theorem 4.3 reduces the search of the maximin quantizers from an infinite dimensional function space to a parameter space of dimension $O(M^2)$. To see this, fix a state $m$ and define $M^2 - 1$ parameters as probability masses $\{p_{mj}^m : 1 \leq j \leq M - 1, p_{mj}^m \geq 0, \sum_{j=1}^{M-1} p_{mj}^m = 1\}$, and ULQ coefficients $\{a_{m,m'}^j : 1 \leq j \leq M - 1, 0 \leq m' \leq M - 1, \sum_{m'=0}^{M-1} (a_{m,m'}^j)^2 = 1\}$. Based on every combination of these parameters, define by $\phi$ the quantizer randomizing $M - 1$ ULQs: $\phi = \sum_{j=1}^{M-1} p_{mj}^m \phi_m$, where

$$\phi_m^j(X) = I(\sum_{m'=0}^{M-1} a_{m,m'}^j f_m(X) > 0).$$

The maximin quantizer $\phi_m^{\text{max}}$ can then be found as $\bar{\phi}$ that maximizes

$$\min_{\bar{\phi} \in \Phi} I(m; \bar{\phi})$$

among all possible combinations of

$$\{a_{m,m'}^j\}_{1 \leq j \leq M - 1, 0 \leq m' \leq M - 1}.$$

To further reduce computational complexity of the maximin quantizers, we can apply the following lemma which provides a sufficient condition that a deterministic MLRQ quantizer is the maximin quantizer.

**Lemma 4.1.** Given $m' \neq m$, let $\phi_{m,m'}$ be the deterministic MLRQ quantizer that maximizes the K-L divergence of $m'$ from $m$, i.e.,

$$\phi_{m,m'} = \arg \sup_{\phi \in \Phi} I(m, m' ; \phi).$$

If there exists a state $m'' \neq m$ such that for any other state $m'' \neq m$:

$$I(m, m'' ; \phi_{m,m''}) \geq I(m, m' ; \phi_{m,m''})$$

then $\phi_{m,m'}$ is also the maximin quantizer for state $m$.

**Proof:** By definition,

$$I(m; \phi_{m,m'}) = \min_{m'' \neq m} I(m, m'' ; \phi_{m,m''}) = I(m, m' ; \phi_{m,m'}).$$

Take any $\phi \in \Phi$,

$$I(m; \phi) \leq I(m, m' ; \phi) \leq I(m, m' ; \phi_{m,m'}) = I(m; \phi_{m,m'})$$

and thus $\phi_{m,m'}$ is the maximin quantizer for state $m$. 

**V. Extensions**

Section IV deals with the simplest case when the network only has a single sensor with binary sensor messages. In this section, we extend our results to three more general scenarios: 1) the sensor messages belong to a finite alphabet (not necessarily binary); 2) there is more than one sensor in the network (though observations are independent between different sensors); and 3) the hypotheses are composite.

A. Sensor Messages Belonging to a Finite Alphabet

Suppose the network still consists of only one sensor, but now the sensor messages belong to a finite alphabet, say, $\{0, 1, \ldots, l - 1\}$ with $l \geq 2$. In this scenario, the definitions of two-stage tests (Subsection III-A) and maximin quantizers (Subsection IV-A) are still applicable, and Theorem 4.1 and Theorem 4.2 also hold. The only change is Theorem 4.3.
as we need to consider the following general definition of ULQ, which originally proposed in Tsitsiklis [23] and includes Definition 4.2 as a special case.

**Definition 5.1.** When the sensor messages belong to a finite alphabet \( \{0, 1, \ldots, l - 1\} \), a deterministic quantizer \( \phi \in \Phi \) is said to be an unambiguous likelihood quantizer (ULQ) if and only if there exist real numbers \( \{a_{i,m} : 0 \leq i \leq l - 1, 0 \leq m \leq M - 1\} \) such that

\[
\phi(X) = \arg \min_{0 \leq i \leq l - 1} \sum_{m=0}^{M-1} a_{i,m} f_m(X)
\]

and the probability of a tie is zero under every \( P_m \) for \( m = 0, 1, \ldots, M - 1 \).

With this definition, Theorem 4.3 can be generalized as follows.

**Theorem 5.1.** Suppose the sensor messages belong to a finite alphabet \( \{0, 1, \ldots, l - 1\} \) with \( l \geq 2 \). Then for \( m = 0, 1, \ldots, M - 1 \), the maximin quantizer \( \phi_m^{\max} \) can be realized as a randomization of at most \( M - 1 \) deterministic quantizers. Moreover, for every \( m \), there exists a sequence of quantizers \( \{\phi_{m,i}\} \) in the set of all (deterministic) quantizer vectors, and \( \phi_{m,i} \) is the maximin quantizer in the corresponding single-sensor case, and such results also hold for the multiple sensor cases except that the fusion center knows about which deterministic quantizer vector is picked, either letting the fusion center conduct the randomization directly or using the pseudo-randomization block design at the local sensor level.

Another benefit of Theorem 5.1 is that it can deal with the case when the sensor messages are binary but the pdf’s are not linearly independent. Such a case was not addressed by Theorem 4.3 and Theorem 5.1 shows that although the maximin quantizer \( \phi_m^{\max} \) may no longer be a randomization of at most \( M - 1 \) ULQs, it can still be approximated by a sequence of quantizers \( \{\phi_{m,i}\} \), each one randomizing at most \( M - 1 \) ULQs.

### B. Multiple Sensors

We now assume that there are \( K \geq 2 \) sensors in the system in which all raw observations are independent from sensor to sensor conditioned on each \( P_m \), \( m = 0, 1, \ldots, M - 1 \). In the following notation, we use the superscripts to denote different sensors as in Section III. For simplicity, we assume the sensor messages are binary, since the extension to the scenario with a finite alphabet sensor messages can be easily done as in Subsection V-A. The key to extend our results is to treat the quantizers in Sections III and V as vectors of quantizers. Specifically, a (deterministic) quantizer vector is \( \phi = (\phi^1, \ldots, \phi^K) \), where each local sensor \( \phi^k \) uses the deterministic quantizer \( \phi^k \) to quantize the raw data. Denote by \( \Phi^{(K)} \) the set of all (deterministic) quantizer vectors, and define a randomized quantizer vector

\[
\tilde{\phi} = \sum_j p^j \phi^j
\]

where \( \phi^j = (\phi^{1,j}, \ldots, \phi^{K,j}) \in \Phi^{(K)} \), and \( \{p^j\} \) are the probability masses assigned to the set of deterministic quantizer vectors \( \{\phi^j\} \subset \Phi^{(K)} \). Let the set of all quantizer vectors be \( \Phi^{(K)} \) (a deterministic quantizer can be viewed as a randomized one which assigns probability one to itself). The implementation of a randomized quantizer vector \( \tilde{\phi} \) is the same as that in Subsection III-B i.e., the fusion center knows about which deterministic quantizer vector is picked, either letting the fusion center conduct the randomization directly or using the pseudo-randomization block design at the local sensor level. Likewise, for a deterministic quantizer vector \( \phi = (\phi^1, \ldots, \phi^K) \), the K-L divergence of state \( m' \) from state \( m \) is defined as

\[
I(m, m'; \phi) = \sum_{k=1}^{K} I(m, m'; \phi^k)
\]

and for a randomized quantizer vector \( \tilde{\phi} = \sum_j p^j \phi^j \), the K-L divergence is a weighted average as in Section III

\[
I(m, m'; \tilde{\phi}) = \sum_j p^j I(m, m'; \phi^j).
\]

Now the maximin quantizer vectors \( \{\phi_m^{\max}\} \) and maximin information numbers \( \{I(m)\} \) for quantizer vectors can be defined in exactly the same way as in Subsection IV-A and the theories developed for single-sensor networks, i.e., Theorems 4.1-4.3 also hold for the multiple sensor cases except replacing the quantizers by quantizer vectors.

A special case is when the sensors are homogeneous, i.e., when the observations are independent and identically distributed among different sensors. In this case, the maximin quantizer vectors are simply replicates of the maximin quantizers in the corresponding single-sensor case, and such results are summarized in the following proposition.

**Proposition 5.1.** Assume that \( f^m_1 = \cdots = f^K_m = f_m \) for \( m = 0, 1, \ldots, M - 1 \). Fix a state \( m \), let \( \phi_m^{\max} = \sum_j p^j_m \phi_m^j \) be the maximin quantizer in the corresponding single sensor case where the system has only one sensor and the raw data are distributed according to \( \{f_m\} \). Define randomized quantizer vector \( \tilde{\phi}_m = \sum_j p^j_m \phi_m^j \) with each \( \phi_m^j \) being a \( K \)-time replication of \( \phi_m^j \), i.e., \( \phi_m^j = (\phi_{m,0}^j, \ldots, \phi_{m,K-1}^j) \). Then \( \phi_m^j \) is a maximin quantizer vector for the state \( m \).

**Proof:** The proof follows at once from (25) and (26).}

### C. Composite Multihypothesis Testing

Our theory can also be extended to the scenario of composite hypothesis with finitely many points. Suppose that there are \( B \) composite hypotheses, \( \mathbf{H}_0, \ldots, \mathbf{H}_{B-1} \), where

\[
\mathbf{H}_b = \{P_{i_b}, P_{i_b+1}, \ldots, P_{i_b+1-1}\}
\]
include $i_{b+1} - i_b$ points for $b = 0, 1, \ldots, B - 1$, and $i_0 = 0$. Without loss of generality, let us assume $M = i_B$. Then there are a total of $i_B = M$ simple hypotheses, and the decision maker is required to pick up one of the $B$ hypotheses that most likely includes the true state of nature $P_m$. Hence, the problem formulation is the same as that in Section II, except that the cost function $W(m, m')$ needs to be re-defined to reflect composite hypotheses in the multihypothesis testing problem. To simplify our notation, for $m = 0, 1, \ldots, M - 1$, denote by $[m]$ the hypothesis that $P_m$ is in, i.e., $[m] = H_b$ if and only if $P_m \in H_b$. In composite multihypothesis testing problem, the loss function $W$ has the form $\{W(m, [m'])\}$, where $W(m, [m'])$ indicates the loss caused by making a decision $D = [m']$ when the states of nature is $P_m$. We assume $W(m, [m']) \geq 0$ and $W(m, [m']) = 0$ if and only if $m \not\in [m']$, i.e., no loss in making a correct decision.

As in Section III, the total expected cost or risk of a test $\delta$ is when the true state of nature is $m$ is:
\[
R_c(\delta; m) = cE_m\{N\} + \sum_{m'} W(m, [m'])P_m\{D = [m']\}
\]
and the Bayes risk of $\delta$ is
\[
R_c(\delta) = \sum_m \pi_mR_c(\delta; m)
\]
(27)
where the prior probability of the hypothesis $H_b$ is $\pi_{i_b} + \ldots + \pi_{i_{b+1}}$.

In the scenario of composite hypotheses, the definition of the two-stage tests is similar except a slight modification of the stopping time $N$ and the final decision $D$ of the fusion center in the second stage. For simplicity, let us consider the simplest case of the single-sensor and binary sensor messages. At time step $n$ in the second stage, the fusion center computes
\[
r_{[m],n} = \sum_{m' \not\in [m]} \pi_{m',n}W(m', [m])
\]
which is the average loss if one makes a final decision $D = [m]$. Then the fusion center stops at time $N = \min\{N_{[m]}\}$, where
\[
N_{[m]} = \{n \geq N_0 : r_{[m],n} \leq c\}
\]
and $N_0$ is the stopping time for the first stage. When stopped, the fusion center makes a final decision $D = [m]$ if $N = N_{[m]}$. Note that we do not change the fusion center policies in the first stage, i.e., the preliminary decision $D_0$ at the fusion center still picks up the most promising state among the $M$ states instead of picking up one of the $B$ hypotheses.

To find the asymptotically optimal tests among the two-stage tests, we need to modify the definition of the information number $I(m; \bar{\phi})$ as follows:
\[
I(m; \bar{\phi}) = \min_{m' \in [m]} I(m, m'; \bar{\phi}), \quad m = 0, 1, \ldots, M - 1
\]
that is, when taking the minimum, we shall ignore those states grouped into the same hypothesis with $m$. With these new definitions, Theorems 4.1 and 4.2 remain valid, and we can still use Theorem 4.3 to numerically compute each maximin quantizer $\hat{\phi}_{m, \text{max}}^\text{m}$ by pretending $[m] = \{P_m\}$, i.e., by temporarily discarding other states in $[m]$.

VI. EXAMPLES

In this section we illustrate our theory via a numerical simulation study. Suppose we are interested in testing the mean of a normal distribution with unit variance in a network with a single sensor and binary sensor messages. That is, the raw data observed at the local sensor follows a normal distribution $P \sim N(\theta, 1)$. In the problem of testing three hypotheses regarding $\theta$, say, $H_0 : \theta = \theta_0$, $H_1 : \theta = \theta_1$ and $H_1 : \theta = \theta_2$, we assign the prior probability of 1/3 to each of these three hypotheses, and as in Draglin et al. [4], we also assume 0-1 loss for decision-making, i.e., $W(m, m') = 1$ if $m \neq m'$ and $= 0$ if $m = m'$. Two different scenarios will be considered:

1) Asymmetric (HT1): $(\theta_0, \theta_1, \theta_2) = (-0.5, 0, 1)$.
2) Symmetric (HT2): $(\theta_0, \theta_1, \theta_2) = (-0.5, 0, 0.5)$.

For our proposed asymptotic optimal decentralized test $\delta_A$ in these scenarios, it suffices to determine the local quantizers. The stationary quantizer in the first stage of $\delta_A$ is easy, as we can simply use $\phi_0^0(X) = I(X \geq 0)$, which satisfies the conditions in Subsection III-A. It is a little more challenging to characterize the maximin quantizers used in the second stage of $\delta_A$. For the asymmetric case (HT1), it is straightforward to show from Lemma 6.1 that the three maximin quantizers are all deterministic MLRQs. Numerical computations illustrate that the three maximin quantizers are $\phi_0^0 = I(X \geq -0.3963), \phi_1 = I(X \geq -0.1037), \phi_2 = I(X \geq 0.7941)$ and the corresponding maximin information numbers are $I_0 = 0.0796, I_1 = 0.0796, I_2 = 0.3186$, respectively.

The maximin quantizers of the symmetric case (HT2) are a little tricky. It is easy to check that Lemma 6.1 can be applied to state $m = 0$ and $m = 2$, yielding two maximin quantizers $\phi_0^0 = I(X \geq -0.1037)$ and $\phi_2 = I(X \geq 0.3963)$ with maximin information numbers $I_0 = I_2 = 0.07959$. However, we need to pay special attention to the maximin quantizer for the state $m = 1$ since the other two states $m = 0$ and $m = 2$ are symmetric with respect to $m = 1$. Since the three pdfs are obviously linearly independent as defined in Subsection IV-B by Theorem 4.3, the maximin quantizer for state $m = 1$ can be realized as a randomization of at most two ULQs. The following lemma, whose proof is straightforward and thus is omitted, gives more convenient descriptions of the ULQs in (HT2) when the observations are normally distributed.

Lemma 6.1. For the symmetric case (HT2), up to a permutation of the values it takes, a ULQ always takes one of the following two forms: $I(X \geq \lambda)$ or $I(\lambda_1 \leq X \leq \lambda_2)$, where $\lambda$ and $\lambda_1 \leq \lambda_2$ are real numbers.

This allows us to do numerical computation of the maximin quantizer for state $m = 1$ as in Subsection IV-B. Numerical computations turn out to show that the maximin quantizer for state $m = 1$ is also the deterministic quantizer defined by $\phi_1 = I(X > 0)$ up to the precision of 5 decimal digits, and $I_1 = 0.07928$.

For each of two scenarios, (HT1) and (HT2), we will consider two versions of our proposed tests: one is $\delta_A(c)$ for the system with a single sensor, and the other is $\delta_A'(c)$ for the system with two independent and identical sensors. As a comparison of our proposed tests, we also consider an
TABLE I: Expected values of time steps taken for each of the three tests.

| Test Type          | \( E_m(N) \) | \( \delta_a \) | \( \delta_A(c) \) | \( \delta'_A(c) \) |
|--------------------|--------------|---------------|-----------------|-----------------|
| Asymmetric (HT1)   |              |               |                 |                 |
| \( m = 0 \)        | 46.48        | 73.5±0.9      | 36.8±0.7        |                 |
| \( m = 1 \)        | 48.39        | 77.7±0.9      | 38.9±0.7        |                 |
| \( m = 2 \)        | 11.90        | 19.8±0.2      | 9.9±0.1         |                 |
| Symmetric (HT2)    |              |               |                 |                 |
| \( m = 0 \)        | 46.59        | 73.4±0.9      | 37.8±0.6        |                 |
| \( m = 1 \)        | 60.43        | 110.2±0.9     | 55.2±0.7        |                 |
| \( m = 2 \)        | 46.60        | 73.4±0.9      | 37.8±0.6        |                 |

Asymptotically optimal centralized test \( \delta_a \) proposed in Draglin et al. [5], [6] for the system with a single sensor (we omitted another family of asymptotically optimal centralized test \( \delta_b \) proposed in Draglin et al. [5], [6], since its performance is similar to that of \( \delta_a \)). For \( \delta_a \), the fusion center updates the posterior distribution \( \{\pi_{m,n}\} \) based on the raw data \( \{X_n\} \) and its stopping time is defined as \( N(a) = \min_{1\leq m\leq M} N_m(a) \), where \( N_m(a) = \inf\{n \geq 1 : \pi_{m,n} \geq A_m\} \). In other words, \( \delta_a \) steps as soon as one of the posterior probability \( \pi_{m,n} \) passes the threshold \( A_m \), which can take different values for different \( m \). In the numerical simulation given in [6], the values of these thresholds are as follows. For the asymmetric (HT1), \( A_0 = A_1 = 1 - 3.99 \times 10^{-3} \), \( A_2 = 1 - 5.33 \times 10^{-3} \). For the symmetric (HT2), \( A_0 = A_1 = A_2 = 1 - 3.99 \times 10^{-3} \). These particular values for the thresholds tune the overall probabilities of making incorrect decisions with test \( \delta_a \) to \( 1.0 \pm 0.1 \times 10^{-3} \).

In our simulations, the cost \( c = 3.6 \times 10^{-3} \), and the threshold \( u(c) \) at the first stage of our proposed tests \( \delta_A(c) \) and \( \delta'_A(c) \) is set as 0.1. Because of the selection of the parameters, \( \delta_A, \delta'_A, \) and \( \delta_a \) have similar probabilities of making incorrect decisions, i.e., \( 1.0 \pm 0.1 \times 10^{-3} \). Thus it suffices to report the simulated expected time steps \( E_m \{N\} \) under each of the three hypotheses \( H_m \) for \( m = 0, 1, 2 \), as smaller values of \( E_m \{N\} \) imply better performance of the test (in the sense of smaller Bayes risks). These results are reported in Table I.

The numerical results illustrate that the centralized test, \( \delta_a \), indeed performs better than the decentralized test \( \delta_A(c) \) that makes a final decision based on binary sensor messages instead of raw normal observations. However, Table I demonstrates that for (HT1) and (HT2), if we are able to deploy merely one extra identical sensor, the decentralized test \( \delta'_A(c) \) has smaller Bayes risk than the centralized test with a single sensor, not to mention other important benefits such as robustness and bandwidth saving capabilities. In other words, if designed appropriately, a decentralized test does not lose much information as compared to the centralized test, and in fact, a decentralized test with two sensors can outperform a centralized test with a single sensor.

VII. CONCLUSION

We have developed a family of asymptotically optimal decentralized sequential tests when testing \( M \geq 3 \) hypotheses. The main idea is to consider "two-stage" tests in which one first uses a small portion of total time steps to make a preliminary decision of the true state of nature, and then the local quantizer switches to the corresponding "maximin quantizers." Moreover, we show that each maximin quantizer can be realized (or approximated) as a randomization of at most \( M - 1 \) ULQs, and we also illustrate how to compute maximin quantizers numerically.

There are several theoretical issues in sequential multihypothesis testing problems that deserve further research. Instead of first-order optimality, it will be interesting to investigate higher-order asymptotic optimality. It is expected that we need to extend our two-stage test \( \delta_A(c) \) to more than two stages in order to achieve higher-order asymptotic optimality. In addition, it is interesting to see what happens if the sensor observations are no longer i.i.d., especially if they are dependent either over time or among different sensors.

APPENDIX A

PROOFS OF THEOREMS 4.3 AND 5.1

Since quantizers, especially randomized quantizers, play an important role in our theorems, we will gather some useful results for quantizers in this appendix, including the proofs of Theorems 4.3 and 5.1. Without loss of generality, we assume that the quantized messages belong to a finite alphabet, say, \( \{0, 1, \ldots, 1-1\} \). For a (deterministic or randomized) quantizer \( \phi \in \Phi \), define its distribution vector as a vector of \( M \) dimensions:

\[
q(\phi) = (q(i; m, \phi)) \quad 0 \leq i \leq 1, \quad 0 \leq m \leq M - 1
\]

where \( q(i; m, \phi) = P_{m}(\phi(X) = i) \). Now let us consider four subspaces induced by the distribution vectors \( q(\phi) \):

- Let \( Q \) be the set formed by the distribution vectors of all deterministic quantizers, i.e., \( Q = \{q(\phi) : \phi \in \Phi\} \);
- Let \( Q = \{q(\phi) : \phi \in \Phi\} \) be the set formed by the distribution vectors of all quantizers, deterministic or random;
- Denote by \( Q_U \subset Q \) the set of distribution vectors of all ULQs (see Definition 5.1);
- Denote by \( Q_{\alpha} \) the set of extreme points of \( Q \).

By Tsitsiklis [23], \( Q \) is compact and \( Q \) is the compact convex hull of \( Q \). By the Krein-Milman theorem, the compact convex set \( Q \) is also the convex hull of its extreme points. Thus it is useful to characterize \( Q_{\alpha} \). Tsitsiklis [23] showed that \( Q_U \subset Q_{\alpha} \subset Q \), and \( Q_U \) is a dense subset of \( Q_{\alpha} \). Moreover, it also studied in detail the case of testing \( M = 2 \) hypotheses. However, the case of \( M \geq 3 \) hypotheses is more challenging. Fortunately, below we are able to show that \( Q_{\alpha} = Q_U \) for \( M \geq 3 \) hypotheses under some reasonable additional assumptions.

Lemma A.1. If the sensor messages are binary (i.e., \( l = 2 \)) and the pdf’s \( f_0, \ldots, f_{M-1} \) are linearly independent (as defined in Subsection IV-B), then \( Q_{\alpha} = Q_U \).

Proof: Since \( Q_U \) is a dense subset of \( Q_{\alpha} \), it is sufficient to show that if \( q^0 \in Q_{\alpha} \), then \( q^0 \in Q_U \). Since \( Q_U \) is dense in \( Q_{\alpha} \), there is a sequence of ULQs \( \phi^j \), say, \( \phi^j = f(\sum_m \alpha^j_m f_m(X) > 0) \) with \( \sum_m (\alpha^j_m)^2 = 1 \), such that \( q(\phi^j) \to q^0 \). By Bolzano-Weierstrass theorem, each bounded sequence has a convergent subsequence. By passing to subsequences, we can simply assume that \( \alpha^j_m \) converges to...
$a_0^m$ for each state $m$, and so $\sum_m (a_0^m)^2 = 1$. By the condition of linear independence, $\tilde{q}(X) = \mathbb{I}(\sum_m a_0^m f_m(X) > 0)$ is a ULQ. It remains to show that $q^\ast = \tilde{q}(X)$, or equivalently, to show that for each state $m$, $\lim_{j \to \infty} \Delta_m(\Delta_j) = 0$, where $\Delta_j = \Delta_j^1 \cup \Delta_j^2$ and

$$\Delta_j^1 = \{ X : \sum_m a_m^j f_m(X) \leq 0 \text{ and } \sum_m a_0^m f_m(X) > 0 \}$$

and

$$\Delta_j^2 = \{ X : \sum_m a_m^j f_m(X) > 0 \text{ and } \sum_m a_0^m f_m(X) \leq 0 \}.$$

To prove this, without loss of generality, let us further assume that $0 \leq f_m(X) \leq 1$ for any state $m$, as we can always substitute $f_m(X)$ by $f_m(X)/\sum_i f_m(X_i)$. Define another sequence of sets $\{\Delta_j^1\}$ by $\Delta_j^1 = \{ X : \sum_m a_0^m f_m(X) \leq M \varepsilon_j \}$, where $\varepsilon_j = \max_m |a_0^j - a_0^m|$. We claim that $\Delta_j^1 \subset \Delta_j^1$, for each $j$. Indeed, if $X \in \Delta_j^1$, then $\sum_m a_0^m f_m(X) \leq 0$ and

$$\sum_m a_0^m f_m(X) \leq \sum_m (a_0^j - a_0^m) f_m(X) \leq \sum_m |a_0^j - a_0^m| = M \varepsilon_j,$$

where the second inequality uses the assumption that $0 \leq f_m(X) \leq 1$. Moreover, if $X \in \Delta_j^1$, then $\sum_m a_0^m f_m(X) > 0$, and thus $\Delta_j^1 \subset \Delta_j^1$. Similarly, $\Delta_j^2 \subset \Delta_j^2$. So $\Delta_j \subset \Delta_j^j$.

Let $\Delta_j^0 = \bigcap_{i=1}^\infty \bigcup_{m=0}^{M-1} \Delta_j^i$. Since $a_m^j$ converges to $a_0^m$ for each state $m$, we have $\varepsilon_j = \max_m |a_0^i - a_0^m| = 0$, and thus $\Delta_0^0 = \{ X : \sum_m a_0^m f_m(X) = 0 \}$. Because the pdf’s are assumed to be linearly independent, $\sum_m a_0^m f_m(X) = 0$ for any state $m$. Hence, $\lim_{j \to \infty} \Delta_m(\Delta_j) = 0$. So $\lim_{j \to \infty} \Delta_m(\Delta_j) = 0$, and the lemma is proved.

Now let us consider the K-L divergences for distribution vectors of quantizers. Given $q \in Q$, say $q = q(i; m, \phi)$, denote $q_{i,m} = q(i; m, \phi)$, where $i = 0, \ldots, M-1$ and $m = 0, \ldots, M-1$. For $0 \leq m \neq m' \leq M-1$, define the K-L divergence of the distribution vector $q$ of state $m'$ from state $m$ by

$$J(m, m'; q) = \sum_{i=0}^{M-1} q_{i,m} \log \frac{q_{i,m}}{q_{i,m'}}. \tag{28}$$

and

$$J^\ast(m; \mu) = \min_{m' \neq m} J^\ast(m, m'; \mu). \tag{30}$$

Then for a randomized quantizer $\hat{\phi} \in \Phi$, the K-L divergence defined in (9) is equivalent to $J^\ast(m, m'; \mu)$ for some suitably chosen $\mu$. To see this, note that $\hat{\phi}$ assigns probability masses to a finite or countable subset of $\Phi$, and thus induces a probability measure $\mu(\hat{\phi})$ on $Q$. Hence, $I(m, m'; \hat{\phi}) = J^\ast(m, m'; \mu(\hat{\phi}))$ and

$$I(m; \hat{\phi}) = J^\ast(m; \mu(\hat{\phi})). \tag{31}$$

Our next result is to provide an alternative representation of the maximin information number $I(m)$ defined in Definition 4.1 in Subsection IV-A.

**Lemma A.2.** The maximin information number $I(m) = \sup_{\mu \in \mathcal{M}} J^\ast(m; \mu)$ is continuous viewed as a function of $\mu \in \mathcal{M}$ under weak-convergence. Thus the lemma follows at once from the denseness of $\mathcal{M}^0$ (or $\mathcal{M}^0$) in $\mathcal{M}$ (or $\mathcal{M}$), provided that $I(m) \geq \sup_{\mu \in \mathcal{M}^0} J^\ast(m; \mu)$. Hence, it suffices to show that for each $\mu \in \mathcal{M}^0$, there exists a $\mu' \in \mathcal{M}^0$ such that $J^\ast(m, m'; \mu) \leq J^\ast(m, m'; \mu')$ for each $m' \neq m$. By linearity, we only need to prove it under the further assumption that $\mu \in \mathcal{M}^0$ is supported on a single point $q = q(\bar{\phi})$ for a randomized quantizer $\bar{\phi} \in \Phi$. In this case $J^\ast(m, m'; \mu) = J(m, m'; q) \leq I(m, m'; \hat{\phi})$. By our previous argument, $\hat{\phi}$ can be identified to a probability measure $\mu' = \mu(\hat{\phi}) \in \mathcal{M}^0$ with the property $I(m, m'; \hat{\phi}) = J^\ast(m, m'; \mu')$. Therefore $J^\ast(m, m'; \mu) \leq J^\ast(m, m'; \mu')$, completing the proof of the lemma.

Finally, we are in a position to prove Theorems 4.3 and 5.1.

**Proofs of Theorems 4.3 and 5.1**

Note that Theorem 4.3 is a special case of Theorem 5.1 and follows at once from Theorem 5.1 and Lemma A.2 under the assumption of binary sensor messages and linearly independent pdf’s in which $Q_U = Q_\alpha$. By symmetry and the fact that $Q_U$ is a dense subset in $Q_\alpha$, it is sufficient to show that under the assumption of Theorem 5.1 for the state $m = 0$, exists one maximin quantizer which is a randomization of at most $M-1$ quantizers with their distribution vectors in $Q_\alpha$.

Define two sets in $M-1$ dimensional space, $\mathcal{I} = \{ J(0, 1; q), \ldots, J(0, M-1; q) : q \in Q \}$, and $\mathcal{I}_{\alpha} = \{ J(0, 1; q), \ldots, J(0, M-1; q) : q \in Q_{\alpha} \}$. We define the same for $\mathcal{I}^\ast$ and $\mathcal{I}^\ast_{\alpha}$ when $J(0, 0; \mu)$ is replaced by $J^\ast(0, m; \mu)$ with $\mu \in \mathcal{M}$ and $\mu \in \mathcal{M}_{\alpha}$, respectively, where $\mathcal{M}_{\alpha}$ is the set of probability measures supported in $Q_{\alpha}$. As we have mentioned earlier, $J(0, m; q)$ is continuous if viewed as a function of $q \in Q$, so both $\mathcal{I}$ and $\mathcal{I}_{\alpha}$ are compact. Obviously,
\( J^* \) and \( J^*_* \) are convex hulls of \( J \) and \( J_n \), so they are compact as well. The main idea of the proof is to relate the maximin information number \( I(0) \) with the set \( J^*_* \).

First, we claim that \( I(0) = \sup_{J \in J^*_*} h(J) \), where \( h(\cdot) \) is a function on \( M - 1 \) dimensional space defined by \( h(x_1, \ldots, x_{M-1}) = \min \{ x_1, \ldots, x_{M-1} \} \). By Lemma A.2, we have \( I(0) = \sup_{J \in J^*_*} h(J) \). Since \( 0 \in J^*_* \), to prove the claim, we only need to show, for any \( J \in J^*_* \), there exists \( J' \in J^*_* \), such that each component of \( J' \) is less or equal to the corresponding component of \( J' \). By linearity, it is sufficient to prove for \( J \in J^*_* \), say, \( J = (J(0,1; q), \ldots, J(0, M-1; q)) \) for some \( q \in Q \). Decompose \( q \) as a convex combination of points in \( Q_\alpha \); then

\[
J(m, m'; q) = \sum p_j J(m, m'; q_j), \quad 0 \leq m \neq m' \leq M - 1.
\]

Let \( J' = (J^*(0, 1; \mu), \ldots, J^*(0, M-1; \mu)) \) with \( \mu \) assigns probability mass \( p' \) to \( q' \) for each \( j \), and our claim is justified.

Second, we will show that

\[
\sup_{J \in J^*_*} h(J) = \min_{1 \leq m \leq M - 1} J^*(0, m; \mu_0) \quad (32)
\]

for a probability \( \mu_0 \in M_\alpha \) whose support includes at most \( M - 1 \) points. To see this, note that \( J^*_* \) is a compact convex subset in \( M - 1 \) dimensional space. Thus \( h(\cdot) \) attains its maximum at a point \( J \) on the surface of \( J^*_* \) and \( J \) can be realized as a convex combination of at most \( M - 1 \) points in \( J^*_* \); see, for example, Homander [9]. Suppose that \( J = \sum_{j=1}^{M-1} p_j J^*(j) \), where \( p_j = 1 \) and \( J^* \in J^*_* \). For each \( j \), let \( J' = (J(0, 1; q_j^0), \ldots, J(0, M-1; q_j^0)) \), with \( q_j^0 \in Q_\alpha \). Define \( \mu_0 \in M_\alpha \) be a probability measure such that \( \mu_0(q_j^0) = p_j \) for \( j = 1, \ldots, M - 1 \), then \( (32) \) holds.

Finally, define the randomized quantizer \( \phi_0 \) as the one induced by the measure \( \mu_0 \) in \( (32) \). Then \( I(0) = \min_{\mu \neq 0} I(0; \mu; \phi_0) \) and \( \phi_0 \) can be rewritten as \( \sum_{j=1}^{M-1} p_j J^*(j) \) where \( \phi_j \) has \( q_j^0 \) as its distribution vector. Equivalently, \( \phi_0 \) is just the maximin quantizer \( \phi_0^{\text{max}} \), and it can be taken as a randomization of at most \( M - 1 \) quantizers with their distribution vectors in \( Q_\alpha \). This completes our proof.

**APPENDIX B**

**PROOF OF THEOREM 4.1**

At each stage of our proposed two-stage test \( \delta(c) \), since the local sensor uses stationary (though possibly randomized) quantizers, the sensor messages \( U_m \) are i.i.d. and the fusion center essentially faces the classical centralized sequential hypothesis testing problems. Thus Theorem 4.1 can be proved by standard arguments and by conditioning on the preliminary decision \( D_0 \) of the two-stage test \( \delta(c) \). In the following we will focus on the proof of \( (16) \) to highlight the associated technical mathematical problems that need special attention. Denote by \( N_0 \) and \( N_1 \) the total time steps of the first and second stages of the two-stage test \( \delta(c) \), respectively, then the total time step \( N \) taken by \( \delta(c) \) satisfies

\[
E_m \{ N \} = E_m \{ N_0 \} + E_m \{ N_1 \} = E_m \{ N_0 \} + E_m \{ N_1 | D_0 = m \} P_m \{ D_0 = m \} + E_m \{ N_1 | D_0 \neq m \}.
\]

By standard arguments for the classical centralized sequential multihypothesis testing problems, the stopping boundary of \( 1 - u(c) \) at the first stage guarantees that \( P_m \{ D_0 = m \} = 1 - O(u(c)) \) and \( E_m \{ N_0 \} = O(\log u(c)) \). Since \( u(c) \to 0 \) satisfies \( \frac{\log u(c)}{\log c} \to 0 \), e.g., \( u(c) = 1/\log c \), we have \( P_m \{ D_0 = m \} = 1 - o(1) \) and \( E_m \{ N_0 \} = o(\log c) \). Hence, equation \( (16) \) holds if we can further show that

\[
E_m \{ N_1 | D_0 = m \} = (1 + o(1)) \log c / I(m; \phi_m) \quad (33)
\]

and

\[
E_m \{ N_1 | D_0 \neq m \} = o(\log c) \quad (34).
\]

To prove \( (33) \) and \( (34) \), note that at time \( n \) of the second stage of our proposed two-stage test \( \delta(c) \), the log-likelihood ratio statistic of the latest sensor message at the fusion center is

\[
\Delta Z_n(m, m'; \phi^{(i)}(n)) = \log \frac{f_m(U_n; \phi^{(i)}(n))}{f_m(U_n; \phi^{(i)}(n))},
\]

where \( \phi^{(i)}(n) \) is the deterministic quantizer selected through the randomization at time step \( n \) and \( U_n = \phi^{(i)}(n) (X_n) \) is the quantized sensor message. Hence, for our proposed two-stage test, the log-likelihood ratio statistic of all available sensor messages up to time \( n \) is

\[
Z_n(m, m'; \hat{\phi}) = \sum_{i=1}^{\log c} \Delta Z_i(m, m'; \phi^{(i)}(n)).
\]

Furthermore, since \( \hat{\phi} \) is assumed to be a randomization of a finite number of deterministic quantizers, our implementation of randomized quantizers implies that \( \{ \Delta Z_n(m, m'; \phi^{(i)}(n)), n = 1, 2, \ldots \} \) is a sequence of i.i.d. random variables with mean \( I(m, m'; \hat{\phi}) \) in \( [9] \) and finite variance.

To prove \( (33) \), it is sufficient to show that

\[
E_m \{ N_1 | D_0 = m, \pi, N_0 \} = (1 + o(1)) \log c / I(m; \phi_m)
\]

where \( \bar{\pi}_N = (\pi_0, N_0, \ldots, \pi_{M-1}, N_0) \) denotes the posterior distributions at time \( N_0 \) and the \( o(1) \) term is uniform on the event \( \{ D_0 = m \} \) for any possible \( \pi, N_0 \). This relation itself follows at once from the fact that \( Z_n(m, m'; \hat{\phi}) \) is the sum of i.i.d. random variables with mean \( I(m, m'; \hat{\phi}) \) in \( [9] \) and finite variance, but we need some extra work to prove the uniformness of the \( o(1) \) term. For that purpose, given the state \( m, \) let \( B_c = | \log c (1 - c) | + | \log (1 - u(c)) | \) and consider the following stopping time:

\[
T(B_c; \phi_m) = \inf\{ n : \min_{m' : m' \neq m} Z_n(m, m', \phi_m) \geq B_c \} \quad (36)
\]

where \( Z_n(m, m'; \phi_m) \) is the log-likelihood ratio in \( (35) \) except that the quantizer \( \hat{\phi} \) is now replaced by \( \phi_m \) since we condition on \( D_0 = m \). Clearly, under the conditional distribution \( P_m \{ | D_0 = m, \pi, N_0 \} \), the stopping time \( N_1 \) is dominated by \( T(B_c; \phi_m) \), which does not depend on \( \pi, N_0 \). By the law of large numbers, we have \( E_m \{ T(B_c; \phi_m) \} / B_c \to 1 / I(m; \phi_m) \), also see Theorem 5.1 of Baum and Veeravalli [1]. Thus we can have a \( o(1) \) term with the \( \leq \) part of relation \( (33) \) due to the above arguments and the fact that \( \log u(c) = o(\log c) \).
The $\geq$ part of the relation can be proved similarly and thus relation (33) holds.

The proof of (34) involves more technical details. It suffices to show that $E_m \{ N_1 \{ D_0 = m' \} \} = o(\| \log c \|)$ for each $m' \neq m$. Now when $\{ D_0 = m' \}$, our proposed two-stage procedure $\delta(c)$ uses the stationary (likely randomized) quantizer $\phi_{m'}$ at the second stage. Hence, we can define $Z_m(m; m'; \phi_{m'})$ as in (35) except that we now use the stationary quantizer $\phi_{m'}$. Likewise, define $T(B'_c; \phi_{m'})$ as in (36) with $B'_c = | \log c/(1-c) | + | \log \pi_m |$, where $\pi_m = \pi_{m, N_0}$ is the posterior probability of the $m$th hypothesis at time $N_0$. Then

$$E_m \{ N_1 \{ D_0 = m' \} \} \leq E_m \{ (1 + o(1))B'_c 1 \{ D_0 = m' \} / I(m; \phi_{m'}) \}$$

$$= E_m \{ (1 + o(1))I(m; \phi_{m'}) \} \times E_m \{ (| \log c/(1-c) | + | \log \pi_m |) 1 \{ D_0 = m' \} \}$$

$$= O(|| \log c ||) P_m \{ D_0 = m' \} + O(1) E_m \{ | \log \pi_m | 1 \{ D_0 = m' \} \}$$

$$= o(|| \log c ||) + O(1) E_m \{ | \log \pi_m | 1 \{ D_0 = m' \} \}.$$

Thus, to prove (34), it remains to show that $E_m \{ | \log \pi_{m, N_0} | 1 \{ D_0 \neq m \} \} = o(1)$. By assumption, at time $N_0$, if $D_0 = m'$ then $\pi_{m, N_0} \geq 1 - u(c) > 1/2$. So $\pi_{m, N_0} < u(c) < 1/2$ and for all $L > 0$,

$$P_m \{ | \log \pi_{m, N_0} | > L, D_0 \neq m \} \leq P_m \{ \log \frac{1 - \pi_{m, N_0}}{\pi_{m, N_0}} > L - \log 2, D_0 \neq m \}$$

$$\leq P_m \{ \sup_{n \geq 1} \log \frac{1 - \pi_{m, n}}{\pi_{m, n}} > L - \log 2 \}$$

$$\leq P_m \left\{ \sup_{m', m: m' \neq m} \pi_{m', n}^{m/m'} \exp \left\{ -Z_m(m, m'; \phi^0) \right\} \exp \left\{ -\frac{e^L / 2}{m} \right\} \right\}$$

$$\leq P_m \left\{ \min_{m', m: m' \neq m} Z_m(m, m'; \phi^0) \right\}$$

$$< -L + \log \frac{2(M - 1)}{\pi_m} \right\}.$$

Assume for a moment that the minimum $Z^* = \min_{m': m' \neq m} Z_m(m, m'; \phi^0)$ is exponentially bounded in the sense that there exists a constant $C_1 > 0$ and $0 < \rho < 1$ such that for any $L > 0$,

$$P_m \{ Z^* \leq -L \} \leq C_1 \rho^L.$$

Then we have

$$P_m \{ | \log \pi_{m, N_0} | > L, D_0 \neq m \} \leq C_2 \rho^L$$

with the constant $C_2 = C_1 \exp \left\{ -\log \rho \log \frac{2(M - 1)}{\pi_m} \right\}$. Consequently,

$$E_m \{ | \log \pi_{m, N_0} | 1 \{ D_0 \neq m \} \} \leq E_m \{ | \log \pi_{m, N_0} | 1 \{ D_0 \neq m, | \log \pi_{m, N_0} | \geq | \log u(c) | \} \} \leq C_2 \int \rho^L dL$$

$$= C_2 \frac{\log u(c)}{\log \rho}.$$

which goes to zero as $c \rightarrow 0$. Thus (34) is proved and the theorem holds.

It remains to prove (37). Since the log-likelihood ratio statistic $Z_m(m, m'; \phi)$ in (35) is the sum of i.i.d. random variables with positive mean and finite variance under $P_m$, the minimum

$$Z_{m'}^* = \inf_{n \geq 0} Z_m(m, m'; \phi)$$

is a well-defined (non-positive valued) random variable under $P_m$. Moreover,

$$P_m \{ Z^* \leq -L \} \leq \sum_{m': m' \neq m} P_m \{ Z_{m'}^* \leq -L \}.$$

Thus, to prove (37), it suffices to show that $Z_{m'}^*$ is exponentially bounded for each $m'$. Define a stopping time $\tau_\alpha = \inf\{ n : Z_m(m, m'; \phi) < 0 \}$ and let $Y_1, Y_2, \ldots$ be i.i.d. random variables, where $Y_1 = Z_{m'}(m, m'; \phi)$ conditional on the event $\tau_\alpha < \infty$. Then it is well-known that $Z_{m'}$ has the same distribution as $\sum_{i=1}^{N_\alpha} Y_i$, where $N_\alpha$ is a geometric random variable independent of $Y_i$’s such that $P(N_\alpha = n) = (1-p)^n$ with $p = P_m \{ Z_{m'}^* = 0 \} > 0$, see Klass [11], or Lemma 11.3 and Remark 11.3 of Gut [8]. Now in our case, since $U_n$ is discrete and $\phi$ is randomization of a finite number of deterministic quantizer, $\Delta Z_m(m, m'; \phi)$ has a lower bound, say $-C$ for some $C > 0$. Thus $Y_1 = Z_{m'}(m, m'; \phi)$ also has a lower bound $-C$. So

$$P_m \{ Z_{m'}^* \leq -L \} = P(\sum_{i=1}^{N_\alpha} Y_i \leq -L)$$

$$= P(N_\alpha \geq L/C)$$

$$= (1-p)^{[L/C]}$$

where the last relation uses the fact that $N_\alpha$ is geometrically distributed. Hence $Z_{m'}^*$ is exponentially bounded and the theorem holds. It is also instructive to compare $Z_{m'}^*$ with Brownian motion. Let $B(t)$ denote standard Brownian motion with mean zero and variance parameter $1$. Then for all positive $L, \mu$ and $\sigma$,

$$P(\inf_{t \geq 0} \{ \sigma B(t) + \mu t \} \leq -L) = \exp(-2\mu \sigma^{-2}L).$$

**APPENDIX C**

**PROOF OF THEOREM 4.2**

To prove Theorem 4.2 the main idea is to construct a martingale based on log-likelihood ratios and then apply the optional stopping theorem and Wald’s inequalities. Since Theorem 4.2 deals with general decentralized sequential tests
that may or may not implement randomized quantizers as we proposed for the two-stage tests, denote by $\phi_n$ the quantizer used at time step $n$ to the best knowledge of the fusion center. For example, when a randomized quantizer $\tilde{\phi} = \sum p_i \phi_i$ is implemented and the fusion center knows that the deterministic quantizer $\phi_i$ is picked at time step $n$, then $\phi_n = \phi_i$. Meanwhile, if the randomization is done at the local sensor and the fusion center has no access about which deterministic quantizer is picked, then $\tilde{\phi}_n$ is $\tilde{\phi}$.

Let $U_n$ be the sensor message at time step $n$ and let $q(\tilde{\phi}_n)$ be the distribution vector of $\tilde{\phi}_n$. For $n = 1, 2, \ldots$, define $F_{n-1}$ as the $\sigma$-algebra generated by $U_1, \ldots, U_{n-1}$ and $q(\tilde{\phi}_1), \ldots, q(\tilde{\phi}_n)$. In other words, $F_{n-1}$ is all the past information available to the fusion center before the $n$th time step. Then at time step $n$, the log-likelihood ratio of state $m$ with respect to state $m'$ is $Z_n = \sum_{i=1}^{N} \Delta Z_i$, where

$$\Delta Z_i = \log \frac{f(m(U_i|F_{i-1}))}{f(m(U_i|F_{i-1}))}$$

and $f_m(\cdot|F_{i-1})$ is the conditional mass function induced on $U_i$. Since $U_i$ depends on $F_{i-1}$ only through $\phi_i$, $f_m(\cdot|F_{i-1})$ is simply $f_m(\cdot; \tilde{\phi})$ in (38), and thus $E_m[\Delta Z_i|F_{i-1}] = J(m, m'; q(\tilde{\phi}))$. Theorems 4.2 holds. Therefore,

$$M_n = \sum_{i=1}^{n} \left[ \Delta Z_i - J(m, m'; q(\tilde{\phi})) \right] = Z_n - \sum_{i=1}^{n} J(m, m'; q(\tilde{\phi}))$$

forms a martingale under $P_m$, with respect to $\{F_n\}$. Applying the optimal stopping theorem to the martingale $\{M_n; F_n\}$, for the stopping time $N$ of a decentralized test $\delta(c)$, we have $E_m[M_N] = 0$, or equivalently,

$$E_m[\{Z_N\}] = E_m\left\{ \sum_{i=1}^{N} J(m, m'; q(\tilde{\phi})) \right\}. \quad (38)$$

Now let us go back to the proof of Theorem 4.2. Obviously, for a decentralized test $\delta(c)$ satisfying the error probability assumption in Theorem 4.1, if the sample size $N$ satisfies $E_m[N] = \infty$, then Theorem 4.2 holds. Thus we only need to consider the case when $E_m[N] < \infty$. To derive the asymptotic lower bound on $E_m[N]$, we construct a new test $\delta'(c)$ that accepts $H_m$ if the final decision of $\delta(c)$ is $D = m$ but accepts $H_m'$ (for a given $m' \neq m$) if $D \neq m$. Then this new test $\delta'(c)$ is a well-defined sequential test in the problem of testing a simple hypothesis $H_m$ against a simple alternative $H_m'$. Moreover, the assumption of Theorem 4.2 guarantees that both error type I and type II errors of $\delta'(c)$ are less than $\epsilon_c = A_c \log c$, where $A > 0$ is a constant. Hence, $Z_N$ represents the log-likelihood ratio of the test $\delta(c)$ when stopped and by Wald’s inequalities (also see Theorem 2.39 of Siegmund [24]),

$$E_m[\{Z_N\}] \geq (1 - \alpha_c) \log \frac{1 - \alpha_c}{\alpha_c} + \alpha_c \log \frac{\alpha_c}{1 - \alpha_c} \geq (1 - \alpha_c) \log \alpha_c - \log 2 = \log c - \log |\log c| + O(1)$$

as $c \to 0$, where the $O(1)$ term depends only on $A$. Here the second inequality follows from the facts that $\alpha \log (1 - \alpha)^{-1}$ is nonnegative and that $(1 - \alpha)^{-1}(1 - \alpha) + \alpha \log \alpha$ attains minimum value $-\log 2$ when $\alpha = \frac{1}{2}$. By (38), we have

$$E_m\left\{ \sum_{i=1}^{N} J(m, m'; q(\tilde{\phi})) \right\} \geq |\log c| - \log |\log c| + O(1). \quad (39)$$

Now we claim that the left-hand side of (39) can be rewritten as $J^*(m, m'; \mu_m)E_m\{N\}$ for a suitably chosen probability measure $\mu_m$ on $Q$, where $J^*(m, m'; \mu_m)$ is defined as in (29). Then the theorem follows at once from this claim, relation (39), and Lemma A.2. It remains to prove this claim. To do so, define $\mu_m$ as a convex combination of a sequence of probability measures $\{\mu_{m,n,i}: i \leq n\}$ as follows.

$$\mu_m = \sum_{n=1}^{\infty} \sum_{i=1}^{n} P_m\{N = n\} \frac{\mu_{m,n,i}}{E_m\{N\}}.$$

Then let $\mu_{m,n,i}$ be the distribution of $q(\tilde{\phi}_i)$ under $P_m$ and conditioned on the event $N = n$. In other words, for any Borel set $A \subset Q$, $\mu_{m,n,i}(A) = P_m\{q(\tilde{\phi}_i) \in A|N = n\}$. We have

$$E_m\{N\} J^*(m, m'; \mu_m) = \sum_{n=1}^{\infty} \sum_{i=1}^{n} P_m\{N = n\} \frac{J(m, m'; \mu_{m,n,i})}{E_m\{N\}} = \sum_{n=1}^{\infty} \sum_{i=1}^{n} E_m\{J(m, m'; q(\tilde{\phi}_i)) \mid N = n\} = E_m\left\{ \sum_{i=1}^{N} J(m, m'; q(\tilde{\phi}_i)) \right\}.$$

REFERENCES

[1] C. W. Baum, V. V. Veeravalli, “A sequential procedure for multi-hypothesis testing”, IEEE Trans. Inf. Theory, vol. 40, pp. 1994-2007, Nov. 1994.
[2] R. S. Blum, S. A. Kassam, and H. V. Poor, “Distributed detection with multiple sensors: part II—advanced topics,” Proceedings of the IEEE, vol. 85, no. 1, pp. 64-79, Jan. 1997.
[3] H. Chernoff, “Sequential design of experiments,” Ann. Math. Statist., vol. 30, pp. 755-770, Sep. 1959.
[4] H. Chernoff, Sequential Analysis and Optimal Design. Philadelphia, PA: SIAM, 1972.
[5] V. P. Draganin, A. G. Tartakovsky, V. V. Veeravalli, “Sequential Probability Ratio Tests—Part I: Asymptotic Optimality”, IEEE Trans. Inf. Theory, vol. 45, pp. 2448-2461, Nov. 1999.
[6] V. P. Draganin, A. G. Tartakovsky, V. V. Veeravalli, “Sequential Probability Ratio Tests—Part II: Accurate Asymptotic Expansions for the Expected Sample Size”, IEEE Trans. Inf. Theory, vol. 46, pp. 1366-1383, Jul. 2000.
[7] K. S. Fu, Sequential Methods in Pattern Recognition and Learning, New York: Academic, 1968.
[8] A. Gut, Stopped Random Walks: Limit Theorems and Applications. New York: Springer-Verlag, 1988.
[9] L. Hornabrook, Notions Of Convexity, Chapter II.1. Boston: Birkhauser, 1994.
[10] J. Kiefer and J. Sacks, “Asymptotically optimal sequential inference and design,” Ann. Math. Statist., vol. 34, pp. 705-750, Sep. 1963.
[11] M. J. Klass, “On the maximum of a random walk with small negative drift,” Ann. Probab., vol. 11, No. 3, pp. 491-505, 1983.
[12] E. L. Lehmann, Testing Statistical Hypotheses. New York: Wiley, 1959.
[13] D. Li, K. D. Wong, Y. H. Hu, A. M. Sayeed, “Detection, classification and tracking of targets in distributed sensor networks”, IEEE Signal Processing Magazine, vol. 19, pp 17-29, Mar. 2002.
[14] G. Lorden, “Nearly-optimal sequential tests for finitely many parameter values,” Ann. Statist., vol. 5, No.1, pp. 1-21, 1977.

[15] M. B. Marcus and P. Swerling, “Sequential detection in radar with multiple resolution elements,” IRE Trans. Inform. Theory, pp. 237-245, Apr. 1962.

[16] Y. Mei, “Information bounds and quickest change detection in decentralized decision systems,” IEEE Trans. Inf. Theory, vol. 51, pp. 2669-2681, Jul. 2005.

[17] Y. Mei, “Asymptotic optimality theory for decentralized sequential hypothesis testing in sensor networks” IEEE Trans. Inf. Theory, vol. 54, pp. 2072-2089, May. 2008.

[18] X. Nguyen, M. J. Wainwright, M. I. Jordan, “On optimal quantization rules for some problems in sequential decentralized detection” IEEE Trans. Inf. Theory, vol. 54, pp. 3285-3295, Jul. 2008.

[19] M. K. Simon, J. K. Omura, R. A. Scholtz, and B. K. Levitt, Spread Spectrum Communications, vol. III. Rockville, MD: Comput. Sci., 1985.

[20] C. Stein, “A two-sample test for a linear hypothesis whose power is independent of the variance,” Ann. Math. Statist., vol. 16, No. 3, pp. 243-258, 1945.

[21] R. R. Tenney, N. R. Sandell Jr., “Detection with distributed sensors,” IEEE Trans. Aerospace Elect. Syst., vol. AES-17, pp.501-510, Jul. 1981.

[22] I. I. Tsitovich, “Sequential design of experiments for hypothesis testing” Theory Prob. Appl., vol. 29, pp. 814-817, Jan. 1985.

[23] J. N. Tsitsiklis, “Extremal properties of likelihood ratio quantizers”, IEEE Trans. Commun., vol. 41, pp. 550-558, Apr. 1993.

[24] D. Siegmund, Sequential Analysis, Tests and Confidence Intervals, New York: Springer-Verlag, 1985.

[25] H. L. Van Trees, Detection Estimation and Modulation Theory, vol. I. New York: Wiley, 1968.

[26] V. V. Veeravalli, “Sequential decision fusion: theory and applications”, J. Franklin Inst., vol. 336, pp. 301-322, Feb. 1999.

[27] V. V. Veeravalli, T. Basar, and H. V. Poor, “Decentralized sequential detection with a fusion center performing the sequential test,” IEEE Trans. Inf. Theory, vol. 39, pp. 433-442, Mar. 1993.

[28] R. Viswanathan, P. K. Varshney, “Distributed detection with multiple sensors; part I- Fundamentals,” Proceedings of the IEEE, vol. 85, no. 1, pp. 54-63, 1997.

[29] A. Wald, Sequential Analysis, New York: Wiley, 1947.

[30] A. Wald and J. Wolfowitz, “Optimal character of the sequential probability ratio test,” Ann. Math. Statist., vol. 19, pp. 326-339, Sep. 1948.

[31] F. Ye, H. Luo, S. Lu, L. Zhang, “Statistical en-route filtering of injected false data in sensor networks”, IEEE Journal on Selected Areas in Communications, vol 23, pp 839-850, Apt. 2005.