A Stochastic Sensor Selection Scheme for Sequential Hypothesis Testing With Multiple Sensors

Cheng-Zong Bai, *Student Member, IEEE*, Vaibhav Katewa, Vijay Gupta, *Member, IEEE*, and Yih-Fang Huang, *Fellow, IEEE*

***Abstract*—We study the problem of binary sequential hypothesis testing using multiple sensors with associated observation costs. An off-line randomized sensor selection strategy, in which a sensor is chosen at every time step with a given probability, is considered. The objective of this work is to find a sequential detection rule and a sensor selection probability vector such that the expected total observation cost is minimized subject to constraints on reliability and sensor usage. First, the sequential probability ratio test is shown to be the optimal sequential detection rule in this framework as well. Efficient algorithms for obtaining the optimal sensor selection probability vector are then derived. In particular, a special class of problems in which the algorithm has complexity that is linear in the number of sensors is identified. An upper bound for the average sensor usage to estimate the error incurred due to Wald's approximations is also presented. This bound can be used to set a safety margin for guaranteed satisfaction of the constraints on the sensor usage.**

***Index Terms*—Hypothesis testing, sensor scheduling, sensor selection, sequential detection, sequential probability ratio test, SPRT.**

I. Introduction

THE design and implementation of hypothesis testing procedures have significantly evolved with the advancements in communication and computation technologies. Hypothesis testing typically involves collecting and processing multiple measurements in order to decide one among many possible hypotheses. In particular, sequential hypothesis tests are of great interest since they do not impose the constraint of collecting a pre-specified number of measurements and allow the flexibility of stopping the test when the measurements are informative enough to guarantee the desired probability of decision error. Hence, they offer a trade-off between the accuracy of the decision and the cost of taking more observations. A well known example of this class of tests is the Sequential Probability Ratio Test (SPRT) developed by Wald [3] for binary hypothesis testing. SPRT is optimal in the sense that for a given performance specification in terms of the error probabilities, it requires, on average, the minimum number of observations among all possible sequential tests.

In this paper, we focus on the problem of sensor scheduling for sequential binary hypothesis testing with multiple sensors, when each sensor has a usage constraint and an associated observation cost. Without loss of generality, we assume that only one sensor can be used at every time step. The goal is to obtain a sensor selection schedule that minimizes the expected total observation cost while satisfying constraints on the usage frequency for every sensor as well as on the reliability of the resulting decision.

While sensor scheduling is a well studied problem in the context of linear estimation (see, e.g., [7]–[9] and the references therein) and of fixed-sample-size hypothesis testing (see, e.g., [10]–[12] and references therein), it has received relatively lesser attention for sequential hypothesis testing, especially if constraints on the usage of sensors are imposed. Sensor usage constraints arise naturally in scenarios where the sensors have limited resources for taking measurements. For example, in a cooperative spectrum sensing application (see e.g., [4]–[6]), one among many mobile devices is chosen to transmit its measurements to the base station which acts as a fusion center that wants to detect whether the spectrum is free or occupied. This problem fits into our framework with the limitations on the battery life of sensors being incorporated as sensor usage constraints. We focus on "off-line" sensor selection strategies, in which the rules for choosing which sensor to use at any time step are independent of the previous observations [7], [10], [18]. Although "on-line" or adaptive strategies [13]–[17] may perform better, such strategies usually require smart sensors and/or communication from the fusion center to the sensors. Moreover, the analysis of these on-line strategies seems to be not trivial for the case when the sensor usage constraints are imposed. Following works such as [7], [16]–[18], we propose stochastic sensor selection strategies. In particular, we assume that every sensor has a given probability of being selected in an independent and identically distributed (i.i.d.) manner at every time step, and the solution aims to optimize the observation cost with respect to the sequential decision rule and sensor selection probability vector.

A notable work on sequential detection with off-line sensor selection is [18], in which the authors invoke the multi-hypothesis SPRT and aim at minimizing the average decision time. However, several issues remain to be explored. First, the results in [18] are valid only under the assumption that the expected number of measurements is asymptotically large. Moreover, [18] concludes that the optimal sensor selection scheme

Manuscript received July 07, 2014; revised December 15, 2014 and March 30, 2015; accepted April 13, 2015. Date of publication April 23, 2015; date of current version June 08, 2015. The associate editor coordinating the review of this manuscript and approving it for publication was Prof. Stefano Marano. This work was supported in part by AFOSR grant No. FA9550-10-1-0348 for the first three authors. Some preliminary results in this work were presented at the 2012 International Conference on Acoustics, Speech, and Signal Processing (ICASSP), Kyoto, Japan, and the American Control Conference 2012.

The authors are with the Department of Electrical Engineering, University of Notre Dame, Notre Dame, IN 46556 USA (e-mail: cbai@nd.edu; vkatewa@nd.edu; vgupta2@nd.edu; huang@nd.edu).

Digital Object Identifier 10.1109/TSP.2015.2425804

uses only (at most) two sensors for binary hypothesis testing. Together, these facts imply that the results of [18] cannot readily include the sensor usage constraints. Second, the work in [18] assumes that the multi-hypothesis SPRT proposed in [19] is performed; however, the optimality of such a test with multiple sensors is not discussed. Third, [18] formulates the problem as sum-of-ratios linear fractional programing (LFP) over a probability simplex and exploits the properties of such a problem. However, in the presence of constraints on sensor usage, this class of programming problems is NP-complete [20], in general.

The main contributions of the paper are now summarized:

- We formally establish the optimality of SPRT among the class of sequential tests with multiple sensors and sensor usage constraints. This optimality holds without requiring that the number of observations be asymptotically large.
- For a wide class of problems, we derive an algorithm to compute the optimal sensor selection probability vector that satisfies the sensor usage constraints. The algorithm is computationally efficient with complexity only linear in the number of sensors. For the general problem, we propose a sub-optimal algorithm with computational complexity quadratic in the number of sensors. Numerical results show that the optimal solution is only marginally better than our proposed sub-optimal solution.
- We characterize the maximum deviation that may possibly result in the average number of measurements due to the use of standard Wald's approximations in the development of SPRT. Using this characterization, we provide a method that ensures that the sensor usage constraints are satisfied with a safety margin.

The paper is organized as follows. In Section II, we formulate the problem. Section III proves the optimality of SPRT with multiple sensors in the presence of reliability and sensor usage constraints. In Section IV-A, we prove some useful properties of the optimal sensor selection probability vector that allow us to compute it in an efficient manner. We propose two computationally efficient algorithms for this purpose in Sections IV-B and IV-C. In Section V, we analyze the error introduced in the constraint on the average sensor usage due to Wald's approximations and provide a method to ensure that these constraints are satisfied with a safety margin. In Section VI, we demonstrate the proposed algorithms with numerical examples. Section VII concludes the paper.

Notation: The sets of natural numbers, real numbers, and n-tuples of real numbers are denoted by $\mathbb{N}, \mathbb{R}, \mathbb{R}^n$, respectively. Vectors are denoted by bold case, e.g. $\boldsymbol{q}$. The notation $\boldsymbol{q}^T$ denotes the transpose of the vector $\boldsymbol{q}$. An n-tuple vector $\boldsymbol{q} \in \mathbb{R}^n$ is denoted by $\boldsymbol{q} = [q_1, q_2, \cdots, q_n]^T$, and the notation $\boldsymbol{q} \geq 0$ means that the inequality holds for every component of $\boldsymbol{q}$. We define the probability simplex in $\mathbb{R}^n$ Δ_{n-1} to be $\Delta_{n-1} \stackrel{\Delta}{=} \{\boldsymbol{p} \in \mathbb{R}^n : \sum_{i=1}^n p_i = 1, \boldsymbol{p} \geq 0\}$. Random variables are denoted by uppercase letters (e.g. X) and their realizations are denoted by lowercase letters (e.g. x). We let $X_i^j = \{X_k\}_{k=i}^j$ for $i \leq j$ denote a sequence of random variables and $x_i^j = \{x_k\}_{k=i}^j$ denote its realization. For conciseness, let $\mathbb{E}[\cdot|x]$ denote the conditional expectation given $X = x$. Further, we denote the conditional probability $P(\cdot|H_i)$ and the conditional expectation $\mathbb{E}[\cdot|H_i]$ by $P_i(\cdot)$ and $\mathbb{E}_i[\cdot]$, respectively.

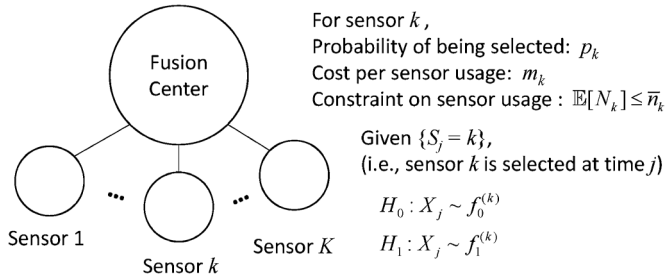

Fig. 1. An illustration of the sensor network.

II. Problem Formulation

Consider a binary hypothesis testing problem in which the objective is to determine which one of the two hypotheses $\{H_0, H_1\}$ is closer (in terms of better representing the data) to the true state of nature. We assume that the *a priori* probabilities of H_0 and H_1 are $1 - \pi_1$ and π_1, respectively, with $0 < \pi_1 < 1$. Let $K \in \mathbb{N}$ be the total number of sensors available in the sensor network. At each time step, a single sensor is chosen according to a constant probability vector $\boldsymbol{p} = [p_1, p_2, \cdots, p_K]^T \in \Delta_{K-1}$, where p_k is the probability of selecting sensor k at any time step. Further, we assume that the sensors are chosen independently at every time step. Denote the random variable that indicates the sensor selected at time step j by S_j. S_j takes value from the set $\mathcal{S} \stackrel{\Delta}{=} \{1, 2, \cdots, K\}$ for all $j \in \mathbb{N}$ with the event $\{S_j = s_j\}$ representing that the sensor s_j is selected at time j. Hence, S_1^∞ is an i.i.d. random sequence. As shown in Fig. 1, at each time step, an observation is taken by the selected sensor and transmitted to a fusion center. Let $X_1^\infty = \{X_j\}_{j=1}^\infty$ denote the random observation sequence collected by the fusion center from time $j = 1, \cdots, \infty$. The observations taken by different sensors at any time instant are assumed to be conditionally independent given either hypothesis. Let $f_i^{(k)}(x_j)$ denote the probability density function (pdf) of X_j if H_i is true and sensor k is selected where $i \in \{0, 1\}$ and $k \in \mathcal{S}$. We assume that $0 < D(f_0^{(k)} \| f_1^{(k)}) < \infty$ and $0 < D(f_1^{(k)} \| f_0^{(k)}) < \infty, \forall k \in \mathcal{S}$ where $D(f\|g) \stackrel{\Delta}{=} \int f \log \frac{f}{g}$ is the Kullback-Leibler divergence [21] (KLD) between two pdf's f and g.

Suppose that a sequential detection scheme is implemented at the fusion center. We follow the standard notation in the sequential detection literature (see, e.g., [22]) to define a sequential test. A sequential detection rule (φ, δ) consists of two elements: the stopping rule φ and the terminal decision rule δ. At each time j, the stopping rule φ decides whether we should stop or take more observations given the values of x_1^j and s_1^j. Therefore, we can define the stopping time N where the event $\{N = n\}$ indicates that the test is terminated at time n. Note that N is a random variable since the stopping time is a function of the observations and selected sensors. Moreover, the event $\{N = n\}$ depends on $\{X_1^n = x_1^n\}$ and $\{S_1^n = s_1^n\}$, while it is independent of both $\{X_{n+1}^\infty = x_{n+1}^\infty\}$ and $\{S_{n+1}^\infty = s_{n+1}^\infty\}$. When the test

is stopped, a decision between H_0 and H_1 is made by using the terminal decision rule δ, as a function of the available observations and the sensors used till the stopping time. The reliability of the sequential detection rule (φ, δ) is defined in terms of the conditional decision error probabilities, namely,

$$\alpha_i(\varphi, \delta, \boldsymbol{p}) = 1 - P_i(\text{Decide} H_i \text{at time} N), \quad i \in \{0, 1\},$$

where N is the stopping time. We set constraints on reliability of the form $\alpha_0(\varphi, \delta, \boldsymbol{p}) \leq \bar{\alpha}_0$ and $\alpha_1(\varphi, \delta, \boldsymbol{p}) \leq \bar{\alpha}_1$, where $0 < \bar{\alpha}_0 < 0.5$ and $0 < \bar{\alpha}_1 < 0.5$. If clear from context, we will suppress the dependence of α_i on $(\varphi, \delta, \boldsymbol{p})$ in the sequel.

A classical example of sequential detection is Wald's Sequential Probability Ratio Test (SPRT) [3]. For the case of multiple sensors with off-line sensor selection, an SPRT(a, b), where $a, b \in \mathbb{R}$ are the two thresholds satisfying $a < 0 < b$, is defined by the following sequential detection rule $(\hat{\varphi}, \hat{\delta})$,

$$\hat{\varphi}: \begin{cases} \text{Continue the test at time } j, & \text{if } a < L\left(x_1^j, s_1^j\right) < b, \\ \text{Stop the test at time } j, & \text{otherwise.} \end{cases}$$

$$\hat{\delta}: \begin{cases} \text{Decide} H_0 \text{at} N = n, & \text{if } L\left(x_1^n, s_1^n\right) \leq a, \\ \text{Decide} H_1 \text{at} N = n, & \text{if } L\left(x_1^n, s_1^n\right) \geq b, \end{cases}$$

where $L(x_1^j, s_1^j)$ is the log-likelihood ratio based on observations x_1^j and sensors s_1^j defined as

$$L\left(x_1^j, s_1^j\right) \triangleq \sum_{i=1}^{j} \log \frac{f_1^{(s_i)}(x_i)}{f_0^{(s_i)}(x_i)}. \tag{1}$$

In the single sensor case $(K = 1)$, the Wald-Wolfowitz inequality [3], [23] proves the optimality of SPRT in the sense that, for appropriately designed thresholds (a, b), the test SPRT(a, b) minimizes both $\mathbb{E}_0[N]$ and $\mathbb{E}_1[N]$ subject to the reliability constraints $\alpha_0 \leq \bar{\alpha}_0$ and $\alpha_1 \leq \bar{\alpha}_1$. However, it is not clear if the optimality holds for the case of multiple sensors. We will show that the optimality indeed holds even when multiple sensors are present.

We now introduce the constraints on sensor usage. Let N_k be the number of times the sensor k is used up to the stopping time N. Thus, $N = \sum_{k=1}^{K} N_k$. In practice, there might be battery related or fairness constraints on how often a sensor should be used. Since the stopping time in a sequential test is not bounded deterministically, we consider probabilistic sensor usage constraints of the form $\mathbb{E}[N_k] \leq \bar{n}_k$, $\bar{n}_k > 0$, $\forall k \in \mathcal{S}$. Note that an alternative way to represent $\mathbb{E}[N_k]$ is as follows. Define for each $k \in \mathcal{S}$, the random sequence $\{\mathbf{1}_k(S_j)\}_{j=1}^{\infty}$, where $\mathbf{1}_k$ is the indicator function such that $\mathbf{1}_k(s_j) = 1$ if $s_j = k$ and $\mathbf{1}_k(s_j) = 0$ otherwise. Since S_1^{∞} is an i.i.d. random sequence, so is $\{\mathbf{1}_k(S_j)\}_{j=1}^{\infty}$. We may use Lemma 2 (see Appendix A) and write the usage for sensor k for any sequential detection rule with the sensor selection probability vector $\boldsymbol{p}$ as

$$\begin{aligned} \mathbb{E}[N_k] &= \mathbb{E}\left[\sum_{j=1}^{N} \mathbf{1}_k(S_j)\right] = \mathbb{E}\left[\mathbf{1}_k(S_j)\right] \mathbb{E}[N] \\ &= P(S_j = k)\mathbb{E}[N] = p_k \mathbb{E}[N]. \end{aligned} \tag{2}$$

Hence, the sensor usage constraints can be written as $p_k \mathbb{E}[N] \leq \bar{n}_k$ for all $k \in \mathcal{S}$.

The classical sequential detection formulation aims at minimizing the expected number of observations. With multiple sensors, we introduce a more general objective. We assume that using the sensor k to generate an observation incurs a cost $m_k \in \mathbb{R}$ where $m_k > 0$. Let $C_1^{\infty} = \{C_j\}_{j=1}^{\infty}$ denote the random observation cost sequence. The random variable C_j takes value from the set $\{m_1, m_2, \cdots, m_K\}$ with $P(C_j = m_k) = p_k$ for all $j \in \mathbb{N}$ and $k \in \mathcal{S}$. Note that C_1^{∞} is an i.i.d. random sequence which is independent of the true hypothesis. By defining the vector $\boldsymbol{m} = [m_1, m_2, \cdots, m_K]^T$, we can write the mean observation cost as $\mathbb{E}[C_j] = \boldsymbol{m}^T \boldsymbol{p}$. Therefore, we can write the expected total observation cost as the expectation of the cumulative observation cost up to the stopping time N, namely, $\mathbb{E}[\sum_{j=1}^{N} C_j]$. The main objective of this paper is to obtain the triplet $(\varphi, \delta, \boldsymbol{p})$ that minimizes the expected total observation cost while satisfying the constraints on reliability and sensor usage, i.e.,

$$\begin{aligned} &\min_{(\varphi, \delta, \boldsymbol{p})} \mathbb{E}\left[\sum_{j=1}^{N} C_j\right] \\ &\text{subject to } \alpha_i(\varphi, \delta, \boldsymbol{p}) \leq \bar{\alpha}_i, \quad \forall i \in \{0, 1\}, \\ &\qquad \mathbb{E}[N_k] \leq \bar{n}_k, \quad \forall k \in \mathcal{S}, \\ &\qquad \boldsymbol{p} \in \Delta_{K-1}, \end{aligned} \tag{3}$$

where $\{\bar{\alpha}_0, \bar{\alpha}_1, \bar{n}_1, \cdots, \bar{n}_K\}$ are all given constants. We solve this problem in two steps. In Section III, we show that the optimal sequential detection rule for any $\boldsymbol{p}$ is an SPRT. Then, in Section IV, we characterize the optimal $\boldsymbol{p}$.

III. Optimal Sequential Test

Intuitively, SPRT should be the optimal sequential decision rule in problem (3) since the observations are conditionally i.i.d. given any realization of the sensor selection sequence S_1^{∞} and since the sensors are also selected in an i.i.d. fashion. In this section, we show that this intuition is indeed correct. We first prove that SPRT is the optimal sequential test for the problem specified in (3) but without the sensor usage constraints. Then, we use this result to prove the optimality of SPRT even in the presence of the sensor usage constraints.

The following result provides the generalization of the classical Wald-Wolfowitz inequality [3], [23] that formalizes the optimality of the SPRT without the sensor usage constraints.

Theorem 1 (Optimality of SPRT Without Sensor Usage Constraints): For any probability vector $\boldsymbol{p} \in \Delta_{K-1}$, let $(\hat{\varphi}, \hat{\delta})$ denote the SPRT(a, b) and (φ, δ) be any other sequential detection rule such that $\alpha_0(\varphi, \delta, \boldsymbol{p}) \leq \alpha_0(\hat{\varphi}, \hat{\delta}, \boldsymbol{p})$ and $\alpha_1(\varphi, \delta, \boldsymbol{p}) \leq \alpha_1(\hat{\varphi}, \hat{\delta}, \boldsymbol{p})$. Then, we have $\mathbb{E}_i[\sum_{j=1}^{N} C_j | \varphi, \boldsymbol{p}] \geq \mathbb{E}_i[\sum_{j=1}^{N} C_j | \hat{\varphi}, \boldsymbol{p}]$ for all $i \in \{0, 1\}$.

Proof: At every time j, the information obtained by the fusion center is the stochastic pair (S_j, X_j). Note that $\{(S_j, X_j)\}_{j=1}^{\infty}$ is an i.i.d. sequence of random pairs when either hypothesis is given. Thus, the log-likelihood ratio for the observation pair $\{(s_i, x_i)\}_{i=1}^{j}$ is given by

$$\log \frac{\prod_{i=1}^{j} p_{s_i} f_1^{(s_i)}(x_i)}{\prod_{i=1}^{j} p_{s_i} f_0^{(s_i)}(x_i)} = \sum_{i=1}^{j} \log \frac{f_1^{(s_i)}(x_i)}{f_0^{(s_i)}(x_i)}.$$

But this is identical to the expression in (1). Thus, with a given sensor selection probability vector $\boldsymbol{p}$, the sequential test with multiple sensors coincides with the problem framework of classical sequential hypothesis tests. The classical Wald-Wolfowitz inequality [3], [23] can thus be used to obtain the expression $\mathbb{E}_i[N|\varphi,\boldsymbol{p}] \geq \mathbb{E}_i[N|\hat{\varphi},\boldsymbol{p}]$ for all $i \in \{0,1\}$. Now note that C_1^∞ is a conditionally i.i.d. random sequence. Using Wald's identity (see Appendix A) yields

$$\mathbb{E}_i\left[\sum_{j=1}^{N} C_j|\varphi,\boldsymbol{p}\right] = \mathbb{E}[C_j|\boldsymbol{p}]\mathbb{E}_i[N|\varphi,\boldsymbol{p}] = (\boldsymbol{m}^T\boldsymbol{p})\mathbb{E}_i[N|\varphi,\boldsymbol{p}]. \quad (4)$$

Since $\boldsymbol{m}^T\boldsymbol{p}$ is irrespective of any stopping rule, the result follows immediately. ∎

Theorem 1 implies that the optimal stopping rule and terminal decision rule in problem (3) are given by an SPRT(a,b) with the thresholds a and b chosen such that the resulting reliabilities $\alpha_i(\varphi,\delta,\boldsymbol{p})$ equal the constraints $\bar{\alpha}_i, \forall i \in \{0,1\}$. Thus, in the sequel, we suppress the dependence of the reliabilities $\alpha_i(\varphi,\delta,\boldsymbol{p})$ on φ and δ and simply write $\alpha_i(\boldsymbol{p})$. As in the classical case, closed form expressions of a and b for reaching $\bar{\alpha}_i$ seem not to be tractable and we resort to Wald's approximations [3] (suitably generalized for multiple sensors) to set these thresholds. Denote by $Q_{s_1^n}$ the event that the test is terminated at step n and H_0 is accepted given that $S_1^n = s_1^n$; thus $Q_{s_1^n} \triangleq \{x_1^n : N = n, L(x_1^n, s_1^n) \leq a\}$. Since $(Q_{s_1^1}, Q_{s_1^2}, \cdots)$ is a sequence of disjoint events, we can write

$$\begin{aligned}
\alpha_1(\boldsymbol{p}) &= P_1\left(L\left(X_1^N, S_1^N\right) \leq a\right) \\
&= \sum_{n=1}^{\infty}\sum_{s_1^n\in\mathcal{S}^n} P\left(S_1^n = s_1^n\right) P_1\left(X_1^n \in Q_{s_1^n}|s_1^n\right) \\
&= \sum_{n=1}^{\infty}\sum_{s_1^n\in\mathcal{S}^n} P\left(S_1^n = s_1^n\right) \int\cdots\int_{Q_{s_1^n}} \prod_{i=1}^{n} f_1^{(s_i)}(x_i)\mathrm{d}x_i \\
&\leq \sum_{n=1}^{\infty}\sum_{s_1^n\in\mathcal{S}^n} P(S_1^n = s_1^n) \int\cdots\int_{Q_{s_1^n}} e^a \prod_{i=1}^{n} f_0^{(s_i)}(x_i)\mathrm{d}x_i \\
&= e^a P_0\left(L\left(X_1^N, S_1^N\right) \leq a\right) \\
&= e^a\left(1-\alpha_0(\boldsymbol{p})\right).
\end{aligned}$$

A similar argument yields $\alpha_0(\boldsymbol{p}) \leq e^{-b}(1-\alpha_1(\boldsymbol{p}))$. Thus, we have the inequalities

$$a \geq \log\frac{\alpha_1(\boldsymbol{p})}{1-\alpha_0(\boldsymbol{p})} \quad \text{and} \quad b \leq \log\frac{1-\alpha_1(\boldsymbol{p})}{\alpha_0(\boldsymbol{p})}. \quad (5)$$

As for classical SPRT, we assume that these inequalities hold with equality to obtain the thresholds a and b in terms of the desired conditional error probabilities. Thus, we set

$$a \approx \log\frac{\bar{\alpha_1}}{1-\bar{\alpha_0}} \quad \text{and} \quad b \approx \log\frac{1-\bar{\alpha_1}}{\bar{\alpha_0}}. \quad (6)$$

We now proceed to prove the optimality of SPRT in the presence of sensor usage constraints.

Theorem 2 (Optimality of SPRT With Sensor Usage Constraints): Assume that the optimization problem defined in (3) is feasible. Then, the optimal solution is the triplet $(\hat{\varphi},\hat{\delta},\boldsymbol{p})$ that yields the reliabilities $\bar{\alpha}_0$, $\bar{\alpha}_1$, where $(\hat{\varphi},\hat{\delta})$ is an SPRT(a,b), and $\boldsymbol{p}$ is an appropriately chosen probability vector.

Proof: We prove the theorem by contradiction. Assume that the triplet $(\varphi,\delta,\boldsymbol{p})$ is optimal in (3), in which the sequential detection rule (φ,δ) is not an SPRT.

First note that $(\hat{\varphi},\hat{\delta},\boldsymbol{p})$ is a feasible solution. The reliability constraints are satisfied by assumption. Since $(\varphi,\delta,\boldsymbol{p})$ is feasible, from Theorem 1 and (2), we obtain $\mathbb{E}[N_k|\hat{\varphi},\boldsymbol{p}] = p_k\mathbb{E}[N|\hat{\varphi},\boldsymbol{p}] \leq p_k\mathbb{E}[N|\varphi,\boldsymbol{p}] = \mathbb{E}[N_k|\varphi,\boldsymbol{p}] \leq \bar{n}_k$ for all $k \in \mathcal{S}$. Therefore, $(\hat{\varphi},\hat{\delta},\boldsymbol{p})$ is feasible.

Now, for any tests that satisfy $\alpha_0(\varphi,\delta,\boldsymbol{p}) \leq \bar{\alpha}_0 = \alpha_0(\hat{\varphi},\hat{\delta},\boldsymbol{p})$ and $\alpha_1(\varphi,\delta,\boldsymbol{p}) \leq \bar{\alpha}_1 = \alpha_1(\hat{\varphi},\hat{\delta},\boldsymbol{p})$, Theorem 1 yields $\mathbb{E}[\sum_{j=1}^{N} C_j|\varphi,\boldsymbol{p}] \geq \mathbb{E}[\sum_{j=1}^{N} C_j|\hat{\varphi},\boldsymbol{p}]$. This contradicts the assumption that $(\varphi,\delta,\boldsymbol{p})$ is optimal. Hence, our assumption must be wrong and the theorem follows. ∎

Remarks 1: It is worth noting that the proofs of Theorems 1 and 2 formalize the intuition that the optimal solution corresponds to SPRT since *both* the sensor selection process and the observations are conditionally i.i.d. Thus, for any given realization of the sensor selection sequence S_1^∞ (i.e. for any given $\boldsymbol{p}$), the optimal solution corresponds to SPRT, provided that $\boldsymbol{p}$ satisfies the constraints of the problem. If, for instance, the sensor selection process has memory or if the observations are not conditionally i.i.d., such optimality will not hold in general.

Theorem 2 shows that SPRT remains the optimal sequential decision rule even with multiple sensors with constraints on the usage of sensors. Consequently, the decision variables in (3) are reduced to $(a,b,\boldsymbol{p})$ where a and b are the thresholds of SPRT(a,b). Notice that (a,b) can be approximately solved for by using (6). Further, these approximations do not impact the choice of the sensor selection probability vector $\boldsymbol{p}$ or the sensor usage constraints. Hence, the problem (3) reduces to the design of the vector $\boldsymbol{p}$. This problem is considered in the next section.

IV. Optimal Sensor Selection Probability Vector

We now derive the optimal sensor selection probability vector $\boldsymbol{p}$, given that the sequential decision rule has been adopted to be an SPRT with thresholds chosen in the manner outlined in (6). We first formulate the sensor selection problem with sensor usage constraints as a sum-of-ratios Linear Fractional Programming (LFP) problem. Then we present computationally efficient algorithms to solve the problem.

A. Simplification of the Optimization Problem

We begin by reformulating the optimization problem by using the specific structure of the probability vector in our problem. First, we approximate $\mathbb{E}[N]$ following the arguments used by Wald in [3] for classical SPRT. Specifically, we use the approximation

$$\mathbb{E}[N] \approx \frac{(1-\pi_1)D_0}{-\boldsymbol{d}_0^T\boldsymbol{p}} + \frac{\pi_1 D_1}{\boldsymbol{d}_1^T\boldsymbol{p}}, \quad (7)$$

where $D_0 = (1-\bar{\alpha}_0)a + \bar{\alpha}_0 b$, $D_1 = \bar{\alpha}_1 a + (1-\bar{\alpha}_1)b$, $\boldsymbol{d}_i = [d_i^{(1)},\cdots,d_i^{(K)}]^T\ \forall i \in \{0,1\}$, $d_0^{(k)} = D(f_0^{(k)}\|f_1^{(k)})$ and $d_1^{(k)} = D(f_1^{(k)}\|f_0^{(k)})$. Note that $D_0 < 0$, $D_1 > 0$, $\boldsymbol{d}_0 > 0$ and $\boldsymbol{d}_1 > 0$. A detailed derivation of (7) is provided in Appendix A. Using (7) and the identity $\mathbb{E}[N_k] = p_k\mathbb{E}[N]$ and $\mathbb{E}[\sum_{j=1}^{N} C_j] =$

$(\boldsymbol{m}^T\boldsymbol{p})\mathbb{E}[N]$ as noted in (2) and (4), the problem in (3) is reformulated as

$$\min_{\boldsymbol{p}\in\Delta_{K-1}} \left\{ (\boldsymbol{m}^T\boldsymbol{p}) \times \left(\frac{(1-\pi_1)D_0}{-\boldsymbol{d}_0^T\boldsymbol{p}} + \frac{\pi_1 D_1}{\boldsymbol{d}_1^T\boldsymbol{p}} \right) \right\}$$
$$\text{subject to } p_k \left(\frac{(1-\pi_1)D_0}{-\boldsymbol{d}_0^T\boldsymbol{p}} + \frac{\pi_1 D_1}{\boldsymbol{d}_1^T\boldsymbol{p}} \right) \le \bar{n}_k, \ \ \forall k \in \mathcal{S}. \quad (8)$$

This optimization problem is a sum-of-ratios LFP [24], which is known to be, in general, computationally hard to solve. However, in our case, we can further simplify the problem.

Remark 2: We would like to emphasize that the optimization problem in (8) is an relaxation of our original problem (3) that is obtained using Wald's approximations. Thus, the solution of problem (8) may not be the optimal solution for (3). Nevertheless, Wald's approximations are used frequently in sequential detection literature since they are reasonably accurate for small error probabilities and make the implementation SPRT practically feasible.

Define the *weighted Kullback-Leibler divergence* of sensor k when H_0 is true as

$$e_0^{(k)} \triangleq \frac{-d_0^{(k)}}{(1-\pi_1)D_0 m_k},$$

and when H_1 is true as

$$e_1^{(k)} \triangleq \frac{d_1^{(k)}}{\pi_1 D_1 m_k}.$$

Denote $\boldsymbol{e}_i = [e_i^{(1)}, \cdots, e_i^{(K)}]^T > 0$, $i \in \{0,1\}$. Further, define two sensors j and k to be *equivalent*, if they have the same weighted KLD, i.e., $e_0^{(j)} = e_0^{(k)}$ and $e_1^{(j)} = e_1^{(k)}$. Let $q_k = \xi p_k m_k$ where ξ is the normalization factor such that $\boldsymbol{q} \in \Delta_{K-1}$. Simple algebraic manipulation then yields the equivalence between the optimization problem (8) and the following sum-of-ratios LFP with a constant numerator objective function,

$$\min_{\boldsymbol{q}\in\Delta_{K-1}} \left\{ \frac{1}{\boldsymbol{e}_0^T\boldsymbol{q}} + \frac{1}{\boldsymbol{e}_1^T\boldsymbol{q}} \right\}$$
$$\text{subject to } \frac{q_k}{\boldsymbol{e}_0^T\boldsymbol{q}} + \frac{q_k}{\boldsymbol{e}_1^T\boldsymbol{q}} \le \bar{n}_k m_k, \quad \forall k \in \mathcal{S}. \quad (9)$$

For ease of notation, we define the following quantities to represent the objective function and the constraints in (9):

$$g(\boldsymbol{q}) \triangleq \frac{1}{\boldsymbol{e}_0^T\boldsymbol{q}} + \frac{1}{\boldsymbol{e}_1^T\boldsymbol{q}}, \quad h_k(\boldsymbol{q}) \triangleq \frac{q_k}{\boldsymbol{e}_0^T\boldsymbol{q}} + \frac{q_k}{\boldsymbol{e}_1^T\boldsymbol{q}}, \quad \bar{n}_k' \triangleq \bar{n}_k m_k.$$

We may further reduce the number of decision variables in the optimization problem (9) if the condition in Lemma 1 is satisfied.

Lemma 1: Consider the problem (9) with the additional assumption that j sensors are equivalent. The optimal solution can then be found by solving an optimization problem of the same form as (9) but with $K - j + 1$ variables.

Proof: See Appendix B. ∎

From now on we consider the problem (9) and assume that the simplification in Lemma 1 has already been carried out. In other words, we assume that no two sensors are equivalent in the sense defined above.

B. A Special Case: Orderable Sensors

We first consider a special case in which an algorithm to solve (9) can be obtained that has complexity only linear in the number of sensors. This case is a generalization of the case of symmetric sensors that we presented in [1].

Specifically, in this section, we focus on the case when the sensors are *orderable* in the sense that we can order the indices of sensors such that

$$e_0^{(1)} \ge e_0^{(2)} \ge \cdots \ge e_0^{(K)} \quad \text{and} \quad e_1^{(1)} \ge e_1^{(2)} \ge \cdots \ge e_1^{(K)}. \quad (10)$$

Several hypothesis testing problems satisfy this assumption. We present two such examples.

Example 1: (Amplitude Detection): Consider the problem of detecting the presence of a constant signal embedded in additive white Gaussian noise. This is a classical binary hypothesis testing problem of deciding between the two hypotheses:

$$H_0 : f_0^{(k)} = \mathcal{N}\left(0, \sigma_k^2\right) \text{ vs.} H_1 : f_1^{(k)} = \mathcal{N}\left(\mu_k, \sigma_k^2\right),$$

where $\mathcal{N}(\mu, \sigma^2)$ denotes the Gaussian pdf with mean μ and variance σ^2. In this case, $d_1^{(k)} = d_0^{(k)} = \frac{\mu_k^2}{2\sigma_k^2}$. Clearly, for $i, j \in \mathcal{S}$, $e_0^{(i)} > e_0^{(j)}$ if and only if $e_1^{(i)} > e_1^{(j)}$.

Example 2: (Energy Detection): Consider the problem of energy detection with uniform observation costs. Let

$$H_0 : f_0^{(k)} = \mathcal{N}\left(0, \sigma_k^2\right) \text{ vs.} H_1 : f_1^{(k)} = \mathcal{N}\left(0, \sigma_k^2 + \nu_k^2\right),$$

and $\boldsymbol{m} = [1, 1, \cdots, 1]^T$. We obtain $d_0^{(k)} = \frac{1}{2}(\log(1+\gamma_k) - \frac{\gamma_k}{1+\gamma_k})$ and $d_1^{(k)} = \frac{1}{2}(\gamma_k - \log(1+\gamma_k))$ where $\gamma_k = \frac{\nu_k^2}{\sigma_k^2}$. Since $d_0^{(k)}$ and $d_1^{(k)}$ are both strictly increasing functions of γ_k on $\gamma_k > 0$, the sensors are orderable. Notice that this case is not symmetric as defined in [1].

We now develop an algorithm to solve (9) for the case when the sensors are orderable. The algorithm is based on the heuristic that to minimize the objective function in (9), we should select the sensors with "larger" weighted KLD with probability as high as possible, while satisfying the sensor usage constraints. We introduce the following terms. We say that a sensor is *active* if it is selected with probability greater than zero, and a sensor k is *fully used* if its sensor usage constraint is reached, i.e., $\mathbb{E}[N_k] = \bar{n}_k$. This heuristic can be justified by using the following theorem.

Theorem 3: Assume that (10) is satisfied and no two sensors are equivalent. A probability vector $\boldsymbol{q}^* \in \Delta_{K-1}$ is the optimal solution to the sum-of-ratios LFP in (9) if and only if all the following conditions are satisfied.

1) $\boldsymbol{q}^*$ is feasible.
2) $\boldsymbol{q}^* = [q_1^*, \cdots, q_k^*, 0, \cdots, 0]^T \in \Delta_{K-1}$ where $k \in \mathcal{S}$ and $h_j(\boldsymbol{q}^*) = \bar{n}_j'$ holds for all $j \in \{1, \cdots, k-1\}$.
3) There exists no $i < k$ such that a feasible $\boldsymbol{q} = [q_1, \cdots, q_i, 0, \cdots, 0]^T \in \Delta_{K-1}$ and $h_j(\boldsymbol{q}) = \bar{n}_j'$ holds for all $j \in \{1, \cdots, i-1\}$.

Proof: See Appendix C. ∎

Theorem 3 implies that a greedy algorithm to solve the sum-of-ratios LFP in (9) will result in the optimal solution. We seek a feasible $\boldsymbol{q}$ and the smallest set of active sensors such that all the sensors in the set are fully used, except for the sensor with the least weighted KLD among all sensors in the set. The greedy

algorithm for solving (9) can be conceptually described by the pseudocode in Algorithm 1.[1]

Algorithm 1: To Find the Optimal $\boldsymbol{q}$ for the Orderable Case

1: Order the indices of the sensors such that (10) is satisfied;
2: $\boldsymbol{q} \leftarrow [1, 0, \cdots, 0]^T$;
3: **if** $\boldsymbol{q} \in \mathcal{C}$ **then**
4: **return** $\boldsymbol{q}$;
5: **else**
6: **for** $k = 2 \rightarrow K$ **do**
7: Solve $\boldsymbol{q} \leftarrow [q_1, \cdots, q_k, 0, \cdots, 0]^T$ from the k equations: $h_j(\boldsymbol{q}) = \bar{n}'_j, \forall j \in \{1, \cdots, k-1\}$, and $\sum_{j=1}^{k} q_j = 1$;
8: **if** $\boldsymbol{q} \in \mathcal{C}$ **then**
9: **return** $\boldsymbol{q}$;
10: **end if**
11: **end for**
12: **end if**

While Algorithm 1 is efficient in the number of iterations, it still requires the solution of a $k \times k$ system of equations at each iteration k in the "for" loop. We can further reduce the computational complexity of the algorithm to $O(K)$ by implementing Algorithm 2 which is derived as follows. At the k-th iteration of Algorithm 1, notice that the first $k-1$ sensors are fully used, i.e., $h_j(\boldsymbol{q}) = \bar{n}'_j$ for all $j \in \{1, 2, \cdots, k-1\}$. These equality constraints yield $q_j \sum_{i=1}^{k-1} \bar{n}'_i = \bar{n}'_j \sum_{i=1}^{k-1} q_i$ for all $j \in \{1, 2, \cdots, k-1\}$. Letting $\rho = \frac{q_k}{\sum_{j=1}^{k-1} q_j}$, we can sum up all the equality constraints for the fully used sensors and obtain

$$\begin{aligned}\sum_{j=1}^{k-1} h_j(\boldsymbol{q}) &= \frac{\sum_{j=1}^{k-1} q_j}{\sum_{j=1}^{k} e_0^{(j)} q_j} + \frac{\sum_{j=1}^{k-1} q_j}{\sum_{j=1}^{k} e_1^{(j)} q_j} \\ &= \frac{\sum_{j=1}^{k-1} \bar{n}'_j}{\epsilon_0 + e_0^{(k)} \rho \sum_{j=1}^{k-1} \bar{n}'_j} + \frac{\sum_{j=1}^{k-1} \bar{n}'_j}{\epsilon_1 + e_1^{(k)} \rho \sum_{j=1}^{k-1} \bar{n}'_j} \\ &= \sum_{j=1}^{k-1} \bar{n}'_j, \end{aligned} \tag{11}$$

where $\epsilon_i = \sum_{j=1}^{k-1} \bar{n}'_j e_i^{(j)}$, $i \in \{0, 1\}$. Using (11), ρ can be solved from the quadratic equation

$$\frac{1}{\epsilon_0 + e_0^{(k)} \rho \sum_{j=1}^{k-1} \bar{n}'_j} + \frac{1}{\epsilon_1 + e_1^{(k)} \rho \sum_{j=1}^{k-1} \bar{n}'_j} = 1. \tag{12}$$

In general, the quadratic (12) has two roots. That the roots must be of opposite signs can be seen from that the fact that if we set $\rho = 0$, then $(\epsilon_0)^{-1} + (\epsilon_1)^{-1} > 1$ since the algorithm is not terminated at the previous iteration. We pick the positive root as the solution for ρ, since, by definition, ρ is non-negative. To determine the feasibility, we note that since $\boldsymbol{q} \in \Delta_{K-1}$ and ρ is solved based on the fact that the first $k-1$ sensors are fully used, the only thing we have to check is whether the condition $h_k(\boldsymbol{q}) \leq \bar{n}'_k$ is satisfied. We can express $h_k(\boldsymbol{q})$ as

$$\begin{aligned} h_k(\boldsymbol{q}) &= \frac{q_k}{\boldsymbol{e}_0^T \boldsymbol{q}} + \frac{q_k}{\boldsymbol{e}_1^T \boldsymbol{q}} = \frac{q_k}{\sum_{j=1}^{k-1} q_j} \left(\frac{\sum_{j=1}^{k-1} q_j}{\boldsymbol{e}_0^T \boldsymbol{q}} + \frac{\sum_{j=1}^{k-1} q_j}{\boldsymbol{e}_1^T \boldsymbol{q}} \right) \\ &= \rho \sum_{j=1}^{k-1} h_j(\boldsymbol{q}) = \rho \sum_{j=1}^{k-1} \bar{n}'_j. \end{aligned} \tag{13}$$

Hence, the feasibility of $\boldsymbol{q}$ can be determined by checking whether the condition $\rho \sum_{j=1}^{k-1} \bar{n}'_j \leq \bar{n}'_k$ is satisfied. Based on Theorem 3, Lemma 3, (12) and (13), we may thus solve for the optimal solution of (9) for the orderable case by implementing Algorithm 2. If the indices have been pre-ordered, the computational complexity of Algorithm 2 is linear in the number of sensors.

Algorithm 2: To Find the Optimal $\boldsymbol{q}$ for the Orderable Case

1: Order the indices of the sensors such that (10) is satisfied;
2: $\Sigma_{\bar{n}'} \leftarrow 0$, $\epsilon_0 \leftarrow 0$, $\epsilon_1 \leftarrow 0$;
3: **if** $(e_0^{(1)})^{-1} + (e_1^{(1)})^{-1} \leq \bar{n}'_1$ **then**
4: $\boldsymbol{q} \leftarrow [1, 0, \cdots, 0]^T$;
5: **return** $\boldsymbol{q}$;
6: **else**
7: **for** $k = 2 \rightarrow K$ **do**
8: $\epsilon_i \leftarrow \epsilon_i + e_i^{(k-1)} \bar{n}'_{k-1}, \forall i \in \{0, 1\}$;
9: $\Sigma_{\bar{n}'} \leftarrow \Sigma_{\bar{n}'} + \bar{n}'_{k-1}$;
10: $\rho \leftarrow$ The positive root of (12);
11: **if** $\rho \Sigma_{\bar{n}'} \leq \bar{n}'_k$ **then**
12: $\boldsymbol{q} \leftarrow [\frac{\bar{n}'_1}{\Sigma_{\bar{n}'}(1+\rho)}, \cdots, \frac{\bar{n}'_{k-1}}{\Sigma_{\bar{n}'}(1+\rho)}, \frac{\rho}{1+\rho}, 0, \cdots, 0]^T$;
13: **return** $\boldsymbol{q}$;
14: **end if**
15: **end for**
16: **end if**

C. The General Case

If the sensors are not "orderable" as discussed in the previous section, then Algorithm 1 is no longer optimal. One reason is that Algorithm 1 is based on the fact that at most one active sensor is not fully used in the orderable case. A counter-example in the general case can be found by considering the situation of unconstrained sensor usage, i.e., $\mathcal{C} = \Delta_{K-1}$. If there are no constraints on the sensor usage, then our problem set-up reduces to the one studied in [18] which showed that the optimal probability vector may be such that two active sensors are not fully used. Even though the feasible region of (9) is convex (see Appendix D for a proof), an arbitrary sum-of-ratios LFP even over a convex set may be NP-complete [20]. Rather than the computationally demanding algorithms available in the literature for this class of problems (see, e.g., [20], [25], [26]), we propose a computationally efficient sub-optimal algorithm that uses the specific structure of our problem. The sub-optimal algorithm is a generalization of the greedy algorithm proposed in Section IV-B using the following fact.

[1] In the algorithm, the feasible region of (9) is denoted by $\mathcal{C}$.

Algorithm 3: To Find a Solution to (9)

1: Order the indices of the groups such that $e_{G_1} \geq e_{G_2} \geq \cdots$;
2: $\mathcal{A} \leftarrow \emptyset$, $\mathcal{F} \leftarrow \emptyset$;
3: **for** $i = 1 \rightarrow |\mathcal{G}|$ **do**
4: **if** $G_i \subset \mathcal{A}$ **or** $\mathcal{F} \cap G_i \neq \emptyset$ **then**
5: **continue**;
6: **end if**
7: $\mathcal{A} \leftarrow \mathcal{A} \cup G_i$;
8: **if** $|\mathcal{A} \setminus \mathcal{F}| < 3$ **then**
9: $\boldsymbol{q}^* \leftarrow \arg\min_{\boldsymbol{q}} \{(\boldsymbol{e}_0^T \boldsymbol{q})^{-1} + (\boldsymbol{e}_1^T \boldsymbol{q})^{-1}\}$; s.t. $\boldsymbol{q} \in \Delta_{K-1}$, $h_j(\boldsymbol{q}) = \bar{n}'_j, \forall j \in \mathcal{F}$ $q_k = 0, \forall k \notin \mathcal{A}$
10: **else**
11: $\boldsymbol{q}^* \leftarrow \arg\min_{\boldsymbol{q}} \{(\boldsymbol{e}_0^T \boldsymbol{q})^{-1} + (\boldsymbol{e}_1^T \boldsymbol{q})^{-1}\}$; s.t. $\boldsymbol{q} \in \Delta_{K-1}$, $h_j(\boldsymbol{q}) = \bar{n}'_j, \forall j \in \mathcal{F}$ $q_k = 0, \forall k \notin \mathcal{A}$ $q_l = 0$ **or** $h_l(\boldsymbol{q}) = \bar{n}'_l$, for some $l \in \mathcal{A} \setminus \mathcal{F}$
12: **end if**
13: **if** $\boldsymbol{q}^* \in \mathcal{C}$ **then**
14: **return** $\boldsymbol{q}^*$;
15: **else**
16: $\mathcal{F} \leftarrow \mathcal{F} \cup \{k \in \mathcal{A} \setminus \mathcal{F} : h_j(\boldsymbol{q}^*) \geq \bar{n}'_j\}$;
17: **if** $|\mathcal{A} \setminus \mathcal{F}| > 1$ **then**
18: **jump** to line 9;
19: **end if**
20: **end if**
21: **end for**

Theorem 4: Assume $\mathcal{C} \neq \emptyset$. There exists an optimal solution $\boldsymbol{q}^*$ to (9) such that at most two active sensors are not fully used.

Proof: See Appendix E. ∎

Theorem 4 is a generalization of [18, Theorem 4] that proved a similar result for the case when there are no constraints on the sensor usage. Lemmas 3 and Theorem 4 imply that if we knew which sensors were active and which ones were fully used, then the optimal probability vector can be obtained by solving an optimization problem with three variables, irrespective of the total number of sensors. The source of computational complexity of the problem lies in identifying which sensors should be active, and which (one or two) active sensors are not fully used. In the orderable case this identification could be done by considering all the sensors sequentially. In the general case, two sensors may be simultaneously active and the sequential procedure breaks down.

To obtain a sub-optimal algorithm, we begin by defining the *efficiency of the sensor pair* $\{j, k\}$ to be

$$e_{\{j,k\}} \triangleq \left(\min_{q \in [0,1]} \frac{1}{q e_0^{(j)} + (1-q) e_0^{(k)}} + \frac{1}{q e_1^{(j)} + (1-q) e_1^{(k)}} \right)^{-1} \tag{14}$$

and the *efficiency of the single sensor* k as

$$e_k \triangleq \left(\frac{1}{e_0^{(k)}} + \frac{1}{e_1^{(k)}} \right)^{-1}. \tag{15}$$

Further, we say that a pair $\{j, k\}$ is *effective* if the minimum in (14) is achieved by a value $q \in (0, 1)$. The sub-optimal algorithm presented in Algorithm 3 is similar to Algorithm 1 considered for the case of orderable sensors. However, in this case, since we need to consider the possibility that two sensors may be simultaneously active and not fully used, we define the set of sensor groups $\mathcal{G}$ to be the collection of all effective pairs and all single sensors. Then, we calculate the efficiency of each member of this set by using either (14) or (15). Similar to Algorithm 1, we order the groups in the deceasing order of their efficiency, denoted by $G_1, G_2, \cdots$, and greedily activate sensor groups in this order till a feasible solution is obtained. In every iteration, we can define $\mathcal{A}$ and $\mathcal{F} \subset \mathcal{A}$ to be the set of available sensors and the set of fully used sensors, respectively. The relative complement of $\mathcal{F}$ in $\mathcal{A}$, denoted by $\mathcal{A} \setminus \mathcal{F}$, is the set of the sensors for which the usage probabilities need to be obtained. The algorithm has the following three main parts:

1) Initialization: This step is carried out once. For the first iteration, we let $\mathcal{A} = G_1$ and $\mathcal{F} = \emptyset$. We begin by obtaining the optimal sensor selection vector by temporarily ignoring the sensor usage constraints (in which case we can use the results of [18]). If there is only one sensor available, i.e., $|\mathcal{A}| = 1$, then the trivial sensor selection probability vector is optimal. If $|\mathcal{A}| = 2$, the optimal sensor selection probability vector $\boldsymbol{q}^*$ can be obtained by using the Karush-Kuhn-Tucker (KKT) conditions. We then evaluate whether this solution violates the sensor usage constraints. If it does not, then the solution is feasible for the original optimization problem as well and hence optimal. Otherwise, some other sensor(s) must be activated. Mimicking Algorithm 1, we force the sensors in $\mathcal{A}$ that reach or violate their sensor usage constraints to be fully used, i.e., set $\mathcal{F} = \{k \in \mathcal{A} : h_k(\boldsymbol{q}^*) \geq \bar{n}'_k\}$. Then we add one more group into the available set.

2) Updating $\mathcal{A}$: This step is carried out once every iteration. Notice that $\mathcal{F}$ is non-empty as long as $\boldsymbol{q}^*$ is not returned in the initialization step. During the i-th iteration, we make the sensors in G_i available if possible by updating $\mathcal{A}$ to $\mathcal{A} \cup G_i$. There are two cases in which this is not possible. First, it may be possible that $G_i \subset \mathcal{A}$ already. Secondly, it is also possible that $\mathcal{F} \cap G_i \neq \emptyset$ and $G_i \not\subset \mathcal{A}$ (i.e., if one sensor in G_i is not in $\mathcal{A}$, and the other sensor in G_i has been fully used). In both these cases, we skip this step and begin the next iteration. In the first case, it is because the sensors are already active. In the second case, the intuition is that the usage of the fully used sensor in G_i may have been occupied by another effective sensor pair more efficient than G_i.

3) Calculating $\boldsymbol{q}^*$ and updating $\mathcal{F}$: This step is carried out once every iteration if the set $\mathcal{A}$ has been updated during that iteration. Thus, this step is carried out if the set $\mathcal{A} \setminus \mathcal{F}$ is non-empty. To reduce the computational complexity of the algorithm, we will constrain that the condition $|\mathcal{A} \setminus \mathcal{F}| \leq 3$ is met in every iteration. Suppose that $|\mathcal{A} \setminus \mathcal{F}| < 3$. We ignore the sensor usage constraints and solve for $\boldsymbol{q}^*$ using Lemma 3 and the KKT conditions given that the sensors in $\mathcal{A}$ are active and the sensors in $\mathcal{F}$ are fully used, as shown in line 9. Suppose that $|\mathcal{A} \setminus \mathcal{F}| = 3$. We impose that one of the sensors in $\mathcal{A} \setminus \mathcal{F}$ be either inactive or fully used and solve for the optimal sensor selection probability vector by using KKT conditions (see line 11). For either

case, if the solution obtained in line 9 or line 11 is feasible with the sensor usage constraints, we return $\boldsymbol{q}^*$ as the solution. Otherwise, we set the sensors that are meeting their sensor usage constraints with equality, or that are violating their constraints, to be fully used (see line 16). Notice that after updating $\mathcal{F}$ in line 16, it is possible to obtain $|\mathcal{A} \setminus \mathcal{F}| = 2$, which may result $|\mathcal{A} \setminus \mathcal{F}| > 3$ after updating $\mathcal{A}$ in the next iteration. To solve this issue, we jump to the procedure stated in line 9 (see line 18) to guarantee the condition $|\mathcal{A} \setminus \mathcal{F}| \leq 3$ in every iteration.

Remark 3: The maximum number of iterations of Algorithm 3 is $|\mathcal{G}| \leq \binom{K}{2} + K$. In addition, at most K iterations are required to solve $\boldsymbol{q}^*$ using the KKT conditions. Hence the complexity of Algorithm 3 is $O(K^2)$ provided that the groups have been pre-ordered by their efficiencies.

Remark 4: In general, Algorithm 3 returns a sub-optimal sensor selection probability vector. However, for two special cases, the algorithm is optimal. If the sensors are orderable, $\mathcal{G}$ only consists of single sensor groups. In this case, Algorithm 1 coincides with Algorithm 3, which means that Algorithm 3 is optimal. Similarly, for the case when there are no sensor usage constraints (i.e., $\mathcal{C} = \Delta_{K-1}$), Algorithm 3 returns the optimal sensor selection probability vector in the first iteration.

V. An Upper Bound for the Exact Sensor Usage

The off-line sensor selection strategy as considered above is within the framework of the two approximations outlined in (6) and (7). These approximations are generalizations of similar relations identified by Wald for the classical sequential decision problem. In particular, the approximations of a, b as outlined in (6) are used to guarantee that the reliability constraints are never violated, since (6) is obtained from the lower and upper bounds of a and b, respectively. However, the approximation of $\mathbb{E}[N]$ as outlined in (7) does not guarantee that the realized value of $\mathbb{E}[N_k]$ satisfies the usage constraints imposed for the k-th sensor exactly. We now provide an upper bound for the realized value of $\mathbb{E}[N_k]$ which can be used to set a safety margin so that the sensor usage constraint is always satisfied.

Define the random variables

$$\Lambda_j \triangleq \log \frac{f_1^{(S_j)}(X_j)}{f_0^{(S_j)}(X_j)} \quad \text{and} \quad L_j \triangleq \sum_{i=1}^{j} \Lambda_i. \tag{16}$$

We begin by noting two consequence of Wald's identity (stated for completeness in Appendix A):

- $\mathbb{E}_i[L_N] = \mathbb{E}_i[N]\mathbb{E}_i[\Lambda_j]$ for any arbitrary time j (notice that Λ_j's form an i.i.d. sequence conditioned on H_i).
- For any arbitrary time j, the identities $\mathbb{E}_0[\Lambda_j] = -\boldsymbol{d}_0^T\boldsymbol{p}$ and $\mathbb{E}_1[\Lambda_j] = \boldsymbol{d}_1^T\boldsymbol{p}$ hold.

Combining these statements, we see that

$$\mathbb{E}_0[N] = \frac{\mathbb{E}_0[L_N]}{-\boldsymbol{d}_0^T\boldsymbol{p}} \quad \text{and} \quad \mathbb{E}_1[N] = \frac{\mathbb{E}_1[L_N]}{\boldsymbol{d}_1^T\boldsymbol{p}}. \tag{17}$$

Using (17) and Bayes' law, we can write

$$\mathbb{E}[N] = (1-\pi_1)\frac{\mathbb{E}_0[L_N]}{-\boldsymbol{d}_0^T\boldsymbol{p}} + \pi_1\frac{\mathbb{E}_1[L_N]}{\boldsymbol{d}_1^T\boldsymbol{p}}. \tag{18}$$

Now we lower bound $\mathbb{E}_0[L_N]$ and upper bound $\mathbb{E}_1[L_N]$ on the right hand side. Specifically, we use the facts that $\mathbb{E}_i[L_N|L_N \geq b] \geq b > 0$ and $\alpha_0 \leq \bar{\alpha}_0$ to obtain

$$\begin{aligned}\mathbb{E}_0[L_N] &= (1-\alpha_0)\mathbb{E}_0[L_N|L_N \leq a] + \alpha_0\mathbb{E}_0[L_N|L_N \geq b] \\ &\geq (1-\bar{\alpha}_0)\mathbb{E}_0[L_N|L_N \leq a] \\ &\geq \mathbb{E}_0[L_N|L_N \leq a]. \end{aligned} \tag{19}$$

Consider the term $\mathbb{E}_0[L_N|L_N \leq a]$ in (19). We can use Baye's law to write

$$\begin{aligned}\mathbb{E}_0[L_N|L_N \leq a] &= \mathbb{E}_0\left[L_{N-1} + \Lambda_N | \Lambda_N \leq a - L_{N-1}\right] \\ &= \sum_{n=1}^{\infty} P_0(N=n)\mathbb{E}_0 \\ &\qquad \left[L_{n-1} + \Lambda_n | \Lambda_n \leq a - L_{n-1}, N = n\right].\end{aligned}$$

Since N is the stopping time, the random variable L_{N-1} must belong to the open interval (a, b) with probability one. Thus, we can further bound

$$\begin{aligned}\mathbb{E}_0[L_N|L_N \leq a] \geq \sum_{n=1}^{\infty} P_0(N=n) \\ \times \inf_{l_{n-1}\in(a,b)} \mathbb{E}_0\left[l_{n-1} + \Lambda_n | \Lambda_n \leq a - l_{n-1}, N = n\right].\end{aligned}$$

We let $l_{n-1} = a - l$ and use the definition of the stopping time for SPRT. The right hand side can be rewritten to obtain

$$\begin{aligned}\mathbb{E}_0[L_N|L_N \leq a] \geq \sum_{n=1}^{\infty} P_0(N=n) \\ \times \inf_{l\in(a-b,0)} \left\{a - l + \mathbb{E}_0\left[\Lambda_n | \Lambda_n \leq l, \bigcap_{j=1}^{n-1}\{L_j \in (a,b)\}\right]\right\}.\end{aligned}$$

Because Λ_1^{∞} is an conditional i.i.d. sequence of random variables given either hypothesis, Λ_n must be independent of any event $\{L_j \in (a,b)\}$ for $1 \leq j \leq n-1$. Thus, we obtain

$$\begin{aligned}\mathbb{E}_0[L_N|L_N \leq a] &\geq \sum_{n=1}^{\infty} P_0(N=n) \\ &\qquad \times \inf_{l\in(a-b,0)} \left\{a - l + \mathbb{E}_0\left[\Lambda_n | \Lambda_n \leq l\right]\right\} \\ &= a + \inf_{l\in(a-b,0)} -l + \mathbb{E}_0[\Lambda_n|\Lambda_n < l]. \end{aligned} \tag{20}$$

Using (19) and (20) provides us the required lower bound

$$\mathbb{E}_0[L_N] \geq a + \inf_{l\in(a-b,0)} -l + \mathbb{E}_0[\Lambda_n|\Lambda_n < l]. \tag{21}$$

A similar procedure yields the upper bound

$$\begin{aligned}\mathbb{E}_1[L_N] &\leq \mathbb{E}_1[L_N|L_N \geq b] \\ &\leq b + \sup_{l\in(0,b-a)} -l + \mathbb{E}_1[\Lambda_n|\Lambda_n > l]. \end{aligned} \tag{22}$$

Using (18), (21), and (22) yields

$$\begin{aligned}\mathbb{E}[N] \leq (1-\pi_1)\frac{a + \inf_{l\in(a-b,0)} -l + \mathbb{E}_0[\Lambda_n|\Lambda_n < l]}{-\boldsymbol{d}_0^T\boldsymbol{p}} \\ + \pi_1\frac{b + \sup_{l\in(0,b-a)} -l + \mathbb{E}_1[\Lambda_n|\Lambda_n > l]}{\boldsymbol{d}_1^T\boldsymbol{p}}.\end{aligned} \tag{23}$$

Finally, plugging (23) into the identity $\mathbb{E}[N_k] = p_k \mathbb{E}[N]$ in (2) yields an upper bound for $\mathbb{E}[N_k]$. A numerical comparison of the upper bound with the realized value of $\mathbb{E}[N_k]$ is presented in Section VI. To demonstrate how the bound in (23) can be calculated, consider the detection problem considered in Example 1.

Example 3: Consider the hypothesis testing problem in Example 1. For this problem, $\Lambda_n \sim \mathcal{N}(-d_0^{(k)}, 2d_0^{(k)})$ when H_0 is true and $S_n = k$, and $\Lambda_n \sim \mathcal{N}(d_1^{(k)}, 2d_1^{(k)})$ when H_1 is true and $S_n = k$. We first evaluate $\mathbb{E}_1[\Lambda_n | \Lambda_n > l]$ in (22),

$$\mathbb{E}_1[\Lambda_n|\Lambda_n > l] = \sum_{k=1}^{K} p_k \frac{\int_l^\infty \frac{x}{\sqrt{4\pi d_1^{(k)}}} \exp\left(\frac{-(x-d_1^{(k)})^2}{4d_1^{(k)}}\right) \mathrm{d}x}{\int_l^\infty \frac{1}{\sqrt{4\pi d_1^{(k)}}} \exp\left(\frac{-(x-d_1^{(k)})^2}{4d_1^{(k)}}\right) \mathrm{d}x}$$

$$= \sum_{k=1}^{K} p_k \left(d_1^{(k)} + \frac{\sqrt{d_1^{(k)}}}{\sqrt{\pi} Q\left(\frac{l-d_1^{(k)}}{\sqrt{2d_1^{(k)}}}\right)} e^{\frac{-(l-d_1^{(k)})^2}{4d_1^{(k)}}} \right), \quad (24)$$

where $Q(x) = \int_x^\infty \frac{1}{\sqrt{2\pi}} \exp(-\frac{1}{2}t^2)\mathrm{d}t$. Note that $-l + \mathbb{E}_1[\Lambda_n|\Lambda_n > l]$ is a monotonically decreasing function of l on $l \geq 0$. Thus the supremum occurs at $l = 0$, which yields

$$\mathbb{E}_1[N] \leq 1 + \frac{b + \sum_{k=1}^{K} \frac{p_k}{1 - Q\left(\sqrt{\frac{1}{2}d_1^{(k)}}\right)} \sqrt{\frac{d_1^{(k)}}{\pi}} e^{-\frac{d_1^{(k)}}{4}}}{\boldsymbol{d}_1^T \boldsymbol{p}}. \quad (25)$$

By symmetry, the bound of $\mathbb{E}_0[N]$ can be obtained via replacing b by $-a$ in (25).

Now we present a method that uses the upper bound to prevent the violation of the sensor usage constraints due to the error of Wald's approximations. We first solve for the sensor selection probability vector as presented in Section IV. Denote by $\Delta\bar{n}_k \geq 0$ the discrepancy between $\mathbb{E}[N_k]$ obtained by Wald's approximations and the upper bound of $\mathbb{E}[N_k]$. We can make the sensor usage constraints more conservative by imposing that the sensor usage is bounded by the term $\bar{n}_k - \Delta\bar{n}_k$ for every $k \in \mathcal{S}$. With the tightened constraints, the sensor selection probability vector and its corresponding upper bounds of the sensor usage can be obtained. We then repeat this procedure that makes the constraints progressively more conservative until the upper bounds of the sensor usage do not violate the original constraints.

VI. Numerical Examples

A. Orderable Case

To illustrate our algorithm, we consider a cooperative spectrum sensing problem in which the goal of a secondary user is to detect whether or not the spectrum is occupied by a primary user. The secondary user can obtain observations from $(K = 8)$ neighboring devices (i.e., sensors) to determine the availability of the spectrum. Let the spectrum sensing problem be modeled as an amplitude detection problem as presented in Example 1 with $\pi_1 = 0.2$. Let the signal-to-noise ratio (SNR) of the observation of sensor k be defined as $\mathrm{SNR}_k = \frac{\mu_k^2}{\sigma_k^2}$ be the k-th entry of the set {3.5, 3,2.5,2,1.5,1,0.5,0} dB, respectively. The different SNRs can arise from the variation in sensing times used by the sensors. An observation cost is incurred for every measurement that models the time and energy consumed for processing the observation taken by the sensor. In addition, sensor usage constraints arise from to the battery life of the sensors. We let the observation cost for the k-th sensor be $m_k = 1 + \sqrt{\mathrm{SNR}_k}$ where SNR_k is on a linear scale and the sensor usage constraint for the k-th sensor, denoted by $\bar{n}_k$ be the k-th element of the set {6,8,5,4,8,4,8,6}. We assume that the secondary user requires the reliability of the test to be $\alpha_0 \leq 10^{-9}$, $\alpha_1 \leq 10^{-10}$.

TABLE I
Numerical Results of SPRT With Each Sensor Being Selected With an Equal Probability

$\boldsymbol{p}$	$[0.125, 0.125, \cdots, 0.125]^T$
$\mathbb{E}[N]$	29.14 (Wald's apx.), 30.39 (sim.), 30.85 (upper bound)
$\mathbb{E}[N_k]$	$\{3.64, 3.64, \cdots, 3.64\}$ (Wald's apx.)
	$\{3.80, 3.80, \cdots, 3.80\}$ (sim.)
	$\{3.85, 3.85, \cdots, 3.85\}$ (upper bound)
$\mathbb{E}[\sum_{j=1}^N C_j]$	65.09 (Wald's apx.), 67.88 (sim.), 68.92 (upper bound)

TABLE II
Numerical Results of SPRT With Sensor Selection Governed by Algorithm 2

$\boldsymbol{p}$	$[0.257, 0.343, 0.214, 0.171, 0.015, 0, 0, 0]^T$
$\mathbb{E}[N]$	23.36 (Wald's apx.), 24.48 (sim.), 29.45 (upper bound)
$\mathbb{E}[N_k]$	$\{6, 8, 5, 4, 0.36, 0, 0, 0\}$ (Wald's apx.)
	$\{6.29, 8.39, 5.24, 4.19, 0.38, 0, 0, 0\}$ (sim.)
	$\{6.41, 8.56, 5.34, 4.27, 0.37, 0, 0, 0\}$ (upper bound)
$\mathbb{E}[\sum_{j=1}^N C_j]$	55.76 (Wald's apx.), 58.45 (sim.), 59.57 (upper bound)

We implement SPRT with two different off-line sensor selection schemes, one specified by the sensor selection probability vector obtained from Algorithm 2, and the other by a strategy in which every sensor is chosen with an equal probability (i.e., an "equally likely selection strategy"). Notice that the equally likely scheme may not be feasible for arbitrary sensor usage constraints. The comparisons between the two schemes are summarized in Tables I and II. Numerical results are obtained using both Monte Carlo simulations and Wald's approximations. It can be seen that the average total observation cost is reduced by around 14% by adopting the optimal sensor selection probability vector over the equally likely sensor selection strategy. Nevertheless, there is a slight discrepancy between the results obtained by Monte Carlo simulations and those obtained using Wald's approximations. This discrepancy is due to the error introduced by Wald's approximations as discussed in Section V, which may cause the sensor usage constraints to be violated. If we tighten the sensor usage constraints to $\bar{n}_k - \Delta\bar{n}_k = \{5.59, 7.44, 4.66, 3.73, 8, 4, 8, 6\}$, the upper bounds for $\mathbb{E}[N_k]$ are shown in Table III. We can see that the resulting upper bounds for $\mathbb{E}[N_k]$ in Table III all satisfy the original sensor usage constraints. Compared with Table II, there is a slight performance loss to $\mathbb{E}[\sum_{j=1}^N C_j]$, if we tighten the sensor usage constraints. However, the sensor usage constraints are not violated.

B. General Problem

If the sensors are not orderable, we may employ Algorithm 3 to obtain a sub-optimal solution. Note that by Lemma 1, the

TABLE III
NUMERICAL RESULTS OF SPRT WITH OPTIMAL SENSOR SELECTION AND TIGHTENED SENSOR USAGE CONSTRAINTS

$\boldsymbol{p}$	$[0.235, 0.314, 0.196, 0.157, 0.098, 0, 0, 0]^T$
$\mathbb{E}[N]$	23.94 (Wald's apx.), 25.07 (sim.), 25.55 (upper bound)
$\mathbb{E}[N_k]$	$\{5.62, 7.50, 4.69, 2.34, 0, 0, 0, 0\}$ (Wald's apx.)
	$\{5.59, 7.44, 4.66, 3.73, 2.51, 0, 0, 0\}$ (sim.)
	$\{5.97, 7.95, 4.97, 3.98, 2.68, 0, 0, 0\}$ (upper bound)
$\mathbb{E}[\sum_{j=1}^{N} C_j]$	56.66 (Wald's apx.), 59.40 (sim.), 60.54 (upper bound)

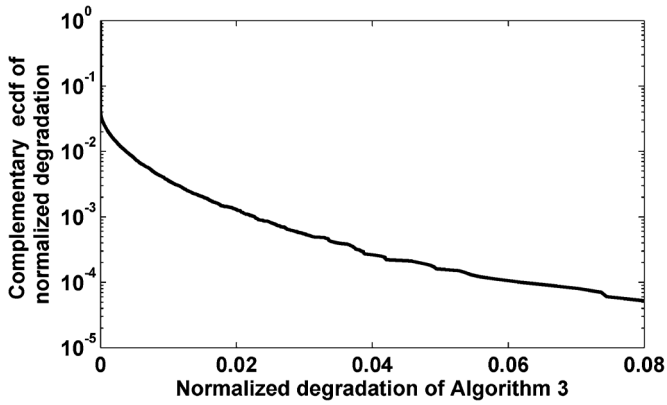

Fig. 2. The complementary empirical cumulative distribution function of normalized degradation D_{norm} using Algorithm 3.

problem (8) can be fully described by $\boldsymbol{e}_0, \boldsymbol{e}_1$ and $\bar{n}'_k$. Therefore, we perform a numerical experiment in which we generate values of $\boldsymbol{e}_0, \boldsymbol{e}_1$ and $\bar{n}'_k$ randomly. We choose $K = 10$ and perform 10^5 experiments in which $e_i^{(k)} \sim U(0,1)$ i.i.d. and $\bar{n}'_k \sim U(0, 0.2)$ i.i.d. where $U(x, y)$ denotes the uniform distribution between x and y. Only those parameter values are retained for which the feasible region is non-empty. For every such problem setting, we compute the performance C_{sub} obtained using Algorithm 3 and the performance C_{opt} obtained using the optimal sensor selection probability vector computed using a brute-force search over the probability simplex Δ_{K-1}. The normalized degradation of the performance using Algorithm 3 can then be defined as

$$D_{norm} = \frac{C_{sub} - C_{opt}}{C_{opt}}.$$

Fig. 2 plots the empirical complementary cumulative distribution function (cdf) of D_{norm}. We can see from the plot that Algorithm 3 renders near-optimal solutions in most cases. For instance, the proportion of the cases for which the normalized degradation has value larger than 0.02 (or 2%) is only about 10^{-3}. This shows that Algorithm 3 represents a good trade-off between computational efficiency and optimality.

VII. CONCLUDING REMARKS

In this paper, we have considered the problem of off-line sensor selection strategy for binary sequential hypothesis testing with multiple sensors and sensor usage constraints. By extending the Wald-Wolfowitz inequality, we showed that SPRT is the optimal sequential detection rule. We also formulated the sensor scheduling problem as a sum-of-ratios LFP. In particular, we showed that a greedy algorithm for solving the sensor selection probability vector is optimal for an important class of problems. For the general case, we proposed a computationally efficient algorithm that provides nearly optimal solutions in numerical simulations. In addition, an upper bound of the sensor usage is derived that enable us to set a safety margin for the sensor usage constraints.

This work can be extended in many directions. An efficient algorithm that is capable of determining the optimal sets of active sensors and fully used sensors for general cases is not addressed in this paper. In addition, the similar sensor selection problem in the on-line fashion with sensor usage constraints remains to be further investigated.

APPENDIX A
A DERIVATION OF THE APPROXIMATION (7)

We first introduce Wald's identity.

Lemma 2 (Wald's Identity [27]): Consider an i.i.d. sequence of random variables Y_1^∞ where $\mathbb{E}[|Y_j|] < \infty$ and the random variable Y_j is determined by X_1^j and S_1^j. Then we have

$$\mathbb{E}\left[\sum_{j=1}^{N} Y_j\right] = \mathbb{E}[Y_j]\mathbb{E}[N]. \tag{26}$$

□

Using the definition in (16), note that Λ_1^∞ is a conditional i.i.d. random sequence given either hypothesis. Using Wald's identity, we have $\mathbb{E}_i[L_N] = \mathbb{E}_i[N]\mathbb{E}_i[\Lambda_j]$ where $\mathbb{E}_0[\Lambda_j] = -\boldsymbol{d}_0^T\boldsymbol{p}$ and $\mathbb{E}_1[\Lambda_j] = \boldsymbol{d}_1^T\boldsymbol{p}$. By Bayes rule, we have

$$\mathbb{E}_0[L_N] = (1-\alpha_0)\mathbb{E}_0\left[L_N | L_N \le a\right] + \alpha_0\mathbb{E}_0\left[L_N | L_N \ge b\right].$$

We may take $\alpha_0 \approx \bar{\alpha}_0$ and approximate that L_N stays exactly on either a or b with probability one, which yields $\mathbb{E}_0[L_N | L_N \le a] \approx a$ and $\mathbb{E}_0[L_N | L_N \ge b] \approx b$. Then we have $\mathbb{E}[L_N|H_0] \approx (1-\bar{\alpha}_0)a + \bar{\alpha}_0 b$. Similarly, $\mathbb{E}[L_N|H_1] \approx \bar{\alpha}_1 a + (1-\bar{\alpha}_1)b$. Notice that these approximations are reasonably accurate when α_0 and α_1 are small enough. Plugging these two approximations into $\mathbb{E}[N] = (1-\pi_1)\mathbb{E}_0[N] + \pi_1\mathbb{E}_1[N]$ gives (7).

APPENDIX B
PROOFS OF LEMMA 1

We assume without loss of generality that sensors 1, 2, $\cdots$, j are equivalent, so that $e_0^{(1)} = e_0^{(2)} = \cdots = e_0^{(j)}$ and $e_1^{(1)} = e_1^{(2)} = \cdots = e_1^{(j)}$. Consider the optimization problem

$$\begin{aligned} &\min_{\bar{\boldsymbol{q}} \in \Delta_{K-j}} g(\bar{\boldsymbol{q}}) \\ &\text{subject to } h_1(\bar{\boldsymbol{q}}) \le \sum_{i=1}^{j} \bar{n}'_i \\ &\qquad h_k(\bar{\boldsymbol{q}}) \le \bar{n}'_k, \quad \forall k = 2, 3, \cdots, K-j+1. \end{aligned} \tag{27}$$

While the new problem (27) depends only on $K-j+1$ variables, the feasible sets for (9) and (27) are equivalent. If $\bar{\boldsymbol{q}}$ is feasible in the new problem, we let $\boldsymbol{q} = [q_1, q_2, \cdots, q_K]^T$ where $q_k = \frac{\bar{n}'_k}{\sum_{i=1}^{j}\bar{n}'_i}\sum_{i=1}^{j}\bar{q}_i$ if $k \le j$ and $q_k = \bar{q}_k$ if $k > j$. Thus, $\boldsymbol{q}$ is feasible in the problem (9), and $g(\boldsymbol{q}) = g(\bar{\boldsymbol{q}})$. On the other hand, starting from a feasible solution $\boldsymbol{q}$ of the problem (9), we

can generate $\bar{\boldsymbol{q}} = [\bar{q}_1, \bar{q}_2, \cdots, \bar{q}_{K-j+1}]^T$ where $\bar{q}_1 = \sum_{i=1}^{j} q_i$ and $\bar{q}_k = q_k$ if $k > 1$. Hence, the lemma follows.

APPENDIX C
PROOF OF THEOREM 3

We begin by presenting a lemma that can reduce the number of decision variables in the problem (9) if some of the sensors are known to be fully used.

Lemma 3: Suppose that j sensors are known to be fully used in the problem (9). Then, the optimization problem (9) is equivalent to an optimization problem with $K - j + 1$ variables.

Proof: We assume without loss of generality that the sensors $1, 2, \cdots, j$ are the fully used sensors. Thus, for all $k \in \{1, \cdots, j\}$, we have $h_k(\boldsymbol{q}) = \bar{n}'_k$ or, in turn,

$$\frac{\bar{n}'_1}{q_1} = \cdots = \frac{\bar{n}'_j}{q_j} = \frac{1}{\boldsymbol{e}_0^T \boldsymbol{q}} + \frac{1}{\boldsymbol{e}_1^T \boldsymbol{q}}.$$

This implies that the $j-1$ variables $q_2, \cdots, q_j$ can be treated as dummy variables, which are fully determined by q_1. In particular, we let

$$\tilde{\boldsymbol{q}} \triangleq \left[\sum_{k=1}^{j} q_1, q_{j+1}, q_{j+2}, \cdots, q_K\right]^T \in \mathbb{R}^{K-j+1},$$

$$\tilde{\boldsymbol{e}}_i \triangleq \left[\frac{\sum_{k=1}^{j} \bar{n}'_k e_i^{(k)}}{\sum_{k=1}^{j} \bar{n}'_k}, e_i^{(j+1)}, e_i^{(j+2)}, \cdots, e_i^{(K)}\right]^T \in \mathbb{R}^{K-j+1},$$

$$\tilde{\boldsymbol{n}}' \triangleq \left[\sum_{k=1}^{j} \bar{n}'_k, \bar{n}'_{j+1}, \bar{n}'_{j+2}, \cdots, \bar{n}'_K\right]^T \in \mathbb{R}^{K-j+1}.$$

Then we can rewrite the problem (9) as

$$\begin{aligned} &\max_{\tilde{\boldsymbol{q}} \in \Delta_{K-j}} \{\tilde{q}_1\} \\ &\text{subject to } \frac{\tilde{q}_1}{\tilde{\boldsymbol{e}}_0^T \tilde{\boldsymbol{q}}} + \frac{\tilde{q}_1}{\tilde{\boldsymbol{e}}_1^T \tilde{\boldsymbol{q}}} = \tilde{n}'_1 \\ &\qquad \tilde{q}_k \le \frac{\tilde{n}'_k}{\tilde{n}'_1} \tilde{q}_1, \quad \forall k \in \{2, \cdots, K-j+1\}, \end{aligned} \tag{28}$$

which completes the proof. ∎

We now prove the theorem. First, we present a necessary condition for a probability vector to be the optimal solution to (9) is that if sensor i is active, all sensors with larger weighted KLD than the weighted KLD of sensor i must be fully used. Second, we provide a sufficient condition of the optimal probability vector, that is, if we can find a feasible $\boldsymbol{q}$ with the least number of active sensors that satisfies the necessary condition, then it is optimal. The proofs of the two conditions are as follows.

(*Necessity*:) For the sum-of-ratios LFP in (9) for the case when the sensors are orderable. Suppose that a vector $\boldsymbol{q}$ with $q_i > 0$ is feasible in (9), and there exists a sensor $j < i$ such that $h_j(\boldsymbol{q}) < \bar{n}'_j$. The necessary condition can be proved by showing that the vector $\Delta\boldsymbol{q} \in \mathbb{R}^K$, where $\Delta q_i = -1$, $\Delta q_j = 1$ and zero elsewhere, is a feasible descent direction for the objective function at $\boldsymbol{q}$. Namely, there exists $\lambda > 0$ such that $\boldsymbol{q} + \lambda\Delta\boldsymbol{q}$ is feasible, and $g(\boldsymbol{q} + \lambda\Delta\boldsymbol{q}) < g(\boldsymbol{q})$. To prove this statement, note that $\frac{d}{d\lambda}\boldsymbol{e}_0^T(\boldsymbol{q} + \lambda\Delta\boldsymbol{q}) = e_0^{(j)} - e_0^{(i)} \ge 0$ and $\frac{d}{d\lambda}\boldsymbol{e}_1^T(\boldsymbol{q} + \lambda\Delta\boldsymbol{q}) = e_1^{(j)} - e_1^{(i)} \ge 0$. Since no two sensors are equivalent by assumption, it follows that $\Delta\boldsymbol{q}$ is a descent direction. Since h_k decreases along $\Delta\boldsymbol{q}$ at $\boldsymbol{q}$ for every $k \ne j$, and $\boldsymbol{q}$ does not satisfy the constraint $h_j(\boldsymbol{q}) \le \bar{n}'_j$ with equality, the direction is feasible as well.

(*Sufficiency*:) We prove the sufficient condition inductively. Notice that if the probability vector $[1, 0, \cdots, 0]^T$ is feasible, clearly, it is the trivial solution to the problem (9). If the vector $[1, 0, \cdots, 0]^T$ is not feasible, we consider a probability vector $\boldsymbol{q}$ such that sensor 1 is fully used and every sensor $j > 2$ is inactive (i.e., $q_j = 0$ for $j > 2$). We now show that, if $\boldsymbol{q}$ is feasible for (9), then $\boldsymbol{q}$ must be the optimal solution to (9). Using Lemma 3 and (28), it is enough to prove that decreasing q_j for every $j > 2$ gives a larger value of q_1 while sensor 1 remains fully used. Note that q_1 can be viewed as an implicit function of $q_3, \cdots, q_K$ through the equations $h_1(\boldsymbol{q}) = \bar{n}'_1$ and $q_2 = 1 - \sum_{k \ne 2} q_k$. The implicit derivative with respect to q_j for each $j > 2$ is given by

$$\begin{aligned} &\frac{\partial h_1(\boldsymbol{q})}{\partial q_j} \\ &= \sum_{i=0}^{1} \frac{(\partial_{q_j} q_1)\left(e_i^{(2)} + \sum_{k \ne 2}\left(e_i^{(k)} - e_i^{(2)}\right) q_k\right)}{\left(\boldsymbol{e}_i^T \boldsymbol{q}\right)^2} \\ &\quad + \frac{q_1\left(\left(e_i^{(1)} - e_i^{(2)}\right)\partial_{q_j} q_1 + e_i^{(j)} - e_i^{(2)}\right)}{\left(\boldsymbol{e}_i^T \boldsymbol{q}\right)^2} \\ &= \sum_{i=0}^{1} \frac{(\partial_{q_j} q_1)\left(e_i^{(2)} + \sum_{k=3}^{K}\left(e_i^{(k)} - e_i^{(2)}\right) q_k\right) - q_1\left(e_i^{(j)} - e_i^{(2)}\right)}{\left(\boldsymbol{e}_i^T \boldsymbol{q}\right)^2} \\ &= \frac{\partial \bar{n}'_1}{\partial q_j} = 0. \end{aligned} \tag{29}$$

Note that $e_0^{(j)} - e_0^{(2)} \le 0$, $e_1^{(j)} - e_1^{(2)} \le 0$ for all $j > 2$ and at least one of them is nonzero because of the assumption that no two sensors are equivalent. In addition, $e_i^{(2)} + \sum_{k=3}^{K}(e_i^{(k)} - e_i^{(2)})q_k = e_i^{(2)}(q_1 + q_2) + \sum_{k=3}^{K} e_i^{(k)} q_k > 0$. From (29), we obtain $\partial_{q_j} q_1 < 0$. Hence, the statement holds. On the other hand, if such a vector $\boldsymbol{q}$ is not feasible, then we must active more sensors, which implies that the first two sensors must be fully used. We can once again invoke Lemma 3 to obtain a problem of the form (28) with the number of unknown variables further reduced by one. Therefore, this procedure can be applied recursively until we obtain a feasible solution, which then must be optimal.

APPENDIX D
THE CONVEXITY OF THE FEASIBLE REGION IN (9)

We need the following preliminary result.

Lemma 4: For $i \in \{0, 1\}$, let $G_i : [0, 1] \to \mathbb{R}$ be

$$G_i(x) = \frac{sx + t(1-x)}{u_i x + v_i(1-x)}$$

where $s, t, u_i, v_i > 0$. Then $G = G_0 + G_1$ is a quasiconvex function.

Proof: If $sv_0 = tu_0$ and $sv_1 = tu_1$, G can be written in the form of a linear fractional function of x, which is known to be quasiconvex. Otherwise, it is enough to show that G has no

local maximum on the open interval (0,1). The first and second order derivatives of G are given by

$$\begin{aligned} G'(x) &= G_0'(x) + G_1'(x) \\ &= \frac{sv_0 - tu_0}{(u_0x + v_0(1-x))^2} + \frac{sv_1 - tu_1}{(u_1x + v_1(1-x))^2} \\ G''(x) &= -2\frac{(sv_0 - tu_0)(u_0 - v_0)}{(u_0x + v_0(1-x))^3} - 2\frac{(sv_1 - tu_1)(u_1 - v_1)}{(u_1x + v_1(1-x))^3} \\ &= -\frac{2G_0'(x)(u_0 - v_0)}{u_0x + v_0(1-x)} - \frac{2G_1'(x)(u_1 - v_1)}{u_1x + v_1(1-x)}. \end{aligned}$$

Since G is smooth on (0,1), i.e., its derivatives of all orders exist, a local extreme point x^*, if it exists, on the interval (0,1) must satisfy $G'(x^*) = 0$ and $G''(x^*) \neq 0$. Moreover, the condition $G'(x^*) = 0$ gives $G_0'(x^*) = -G_1'(x^*)$. Consequently, $G''(x^*)$ can be rewritten as

$$\begin{aligned} G''(x^*) &= 2G_1'(x^*)\left(\frac{u_0 - v_0}{u_0x + v_0(1-x)} - \frac{u_1 - v_1}{u_1x + v_1(1-x)}\right) \\ &= 2G_1'(x^*)\frac{u_0v_1 - u_1v_0}{(u_0x + v_0(1-x))(u_1x + v_1(1-x))}. \end{aligned}$$

Consider the case of $G_1'(x^*) > 0$. Then we have $tu_0 > sv_0 > 0$ and $sv_1 > tu_1 > 0$, which yield $u_0v_1 - u_1v_0 > 0$. Likewise, if $G_1'(x^*) < 0$, then we obtain $u_0v_1 - u_1v_0 < 0$. Both cases imply $G''(x^*) > 0$. We can conclude that x^* must be a local minimum, not maximum. Hence, G is a quasiconvex function. ■

Now we show that the feasible region of (9), $\mathcal{C} = \{\boldsymbol{q} \in \Delta_{K-1} : h_k(\boldsymbol{q}) \leq \bar{n}_k', \forall k \in \mathcal{S}\}$, is a convex set. Suppose that $\boldsymbol{p}, \boldsymbol{q} \in \mathcal{C}$. Consider the vector $\boldsymbol{r} = x\boldsymbol{p} + (1-x)\boldsymbol{q}$ where $0 \leq x \leq 1$. Obviously, $\boldsymbol{r}$ is on the probability simplex Δ_{K-1}. Let $s = p_k$, $t = q_k$, $u_i = \boldsymbol{e}_i^T\boldsymbol{p}$, $v_i = \boldsymbol{e}_i^T\boldsymbol{q}$. Using the notation in Lemma 4, $G(0) = h_k(\boldsymbol{q}) \leq \bar{n}_k'$, $G(1) = h_k(\boldsymbol{p}) \leq \bar{n}_k'$, and $G(x) = h_k(\boldsymbol{r})$. From Lemma 4, $G(x)$ is quasiconvex on $x \in [0,1]$. Then we obtain $G(x) = G((1-x)\times 0 + x \times 1) \leq \max\{G(0), G(1)\} \leq \max\{\bar{n}_k', \bar{n}_k'\} = \bar{n}_k'$ which gives $\boldsymbol{r} \in \mathcal{C}$. Consequently, $\mathcal{C}$ is a convex set.

Appendix E
Proof of Theorem 4

Lemma 5: Given any $\boldsymbol{e}_0 > 0$, $\gamma > 0$, $\zeta\bar{n}_k' > 0$, $K > 2$ such that $\mathcal{Q} = \{\boldsymbol{q} \in \Delta_{K-1} : \boldsymbol{e}_0^T\boldsymbol{q} = \gamma, q_k \leq \zeta\bar{n}_k', \forall k \in \mathcal{S}\}$ is non-empty, there exists $\boldsymbol{q}' \in \mathcal{Q}$ such that at least $K-2$ elements of $\boldsymbol{q}'$ satisfy $q_k' = 0$ or $q_k' = \zeta\bar{n}_k'$.

Proof: We prove the lemma by induction. For $K = 3$, the statement can be illustrated geometrically. Let $\mathcal{Q}_1 = \{\boldsymbol{q} \in \Delta_2 : \boldsymbol{e}_0^T\boldsymbol{q} = \gamma\}$, $\mathcal{Q}_2 = \{\boldsymbol{q} \in \Delta_2 : q_k \leq \zeta\bar{n}_k', \forall k\}$. If $\mathcal{Q}_1$ consists of a single point or $\mathcal{Q}_1 = \Delta_2$, then the result is trivial. Otherwise, $\mathcal{Q}_1$ is a line segment in $\mathbb{R}^3$, and $\mathcal{Q}_2$ is a polygon in $\mathbb{R}^3$. Since $\mathcal{Q} = \mathcal{Q}_1 \cap \mathcal{Q}_2$ is non-empty, $\mathcal{Q}$ contains at least one point on an edge of the polygon $\mathcal{Q}_2$, which implies that at least one k satisfies $q_k = 0$ or $q_k = \zeta\bar{n}_k'$. Assume that the statement is true for all $K \leq n$. For $K = n+1$, let $\boldsymbol{q}$ be an arbitrary element in $\mathcal{Q}$. For notational simplicity, we assume that the first j entries of $\boldsymbol{q}$ satisfy neither $q_k = 0$ nor $q_k = \zeta\bar{n}_k'$ and the other entries satisfy one of these two equations, in which we only consider the non-trivial case of $j > 2$. If $j < K = n+1$, let $\tilde{\boldsymbol{q}} = \xi[q_1, \cdots, q_j]^T$, $\tilde{\boldsymbol{e}}_0 = [e_0^{(1)}, \cdots, e_0^{(j)}]^T$, $\tilde{\mathcal{Q}} = \{\tilde{\boldsymbol{q}} \in \Delta_{j-1} : \tilde{\boldsymbol{e}}_0^T\tilde{\boldsymbol{q}} = \xi(\gamma - \sum_{k=j+1}^K e_0^{(k)}q_k), \tilde{q}_k \leq \xi\zeta\bar{n}_k'\}$ where $\xi = (\sum_{k=1}^j q_k)^{-1}$. By assumption, the statement holds for all $K \leq n$. Since $\tilde{\boldsymbol{q}} \in \tilde{\mathcal{Q}}$, there exists $\tilde{\boldsymbol{q}}' \in \tilde{\mathcal{Q}}$ such that at least $j-2$ elements of $\tilde{\boldsymbol{q}}'$ satisfy $\tilde{q}_k' = 0$ or $\tilde{q}_k' = \xi\zeta\bar{n}_k'$. We can take $\boldsymbol{q}' = [\frac{\tilde{q}_1'}{\xi}, \cdots, \frac{\tilde{q}_j'}{\xi}, q_{j+1}, \cdots, q_K]^T$ and it obviously satisfies the desired condition. If $j = K$, i.e., none of q_k satisfies $q_k = 0$ or $q_k = \zeta\bar{n}_k'$, then we can repeat the above procedure twice and get a desired $\boldsymbol{q}'$. Thus the lemma follows. ■

Suppose that $\boldsymbol{q}^*$ is an optimal solution to (9). We may assume without loss of generality that all sensors are active by removing the elements of $\boldsymbol{q}^*$ and $\boldsymbol{e}_i$ corresponding to the inactive sensors. This assumption gives $\boldsymbol{q}^* > 0$.

If $\boldsymbol{q}^*$ lies in the interior of $\mathcal{C}$, then none of the sensors is fully used. For all k, we have

$$\begin{aligned} &\frac{\partial}{\partial q_k}\left(\frac{1}{\boldsymbol{e}_0^T\boldsymbol{q}} + \frac{1}{\boldsymbol{e}_1^T\boldsymbol{q}} + \lambda\left(1 - \sum_{l=1}^K q_l\right)\right)\Bigg|_{\boldsymbol{q}=\boldsymbol{q}^*} \\ &\qquad = -\frac{e_0^{(k)}}{\left(\boldsymbol{e}_0^T\boldsymbol{q}^*\right)^2} - \frac{e_1^{(k)}}{\left(\boldsymbol{e}_1^T\boldsymbol{q}^*\right)^2} - \lambda = 0 \end{aligned} \quad (30)$$

where λ is the Lagrange multiplier. Thus, the all-ones vector $\boldsymbol{1} \in \text{span}\{\boldsymbol{e}_0, \boldsymbol{e}_1\}$, and there exist $u, v > 0$ such that $\boldsymbol{e}_1 = u\boldsymbol{1} - v\boldsymbol{e}_0$. Then $\boldsymbol{e}_0^T\boldsymbol{q}^* = \gamma$ must satisfy the following equation

$$\min_{\boldsymbol{q} \in \Delta_{K-1}} \frac{1}{\boldsymbol{e}_0^T\boldsymbol{q}} + \frac{1}{u - v\boldsymbol{e}_0^T\boldsymbol{q}} = \frac{1}{\gamma} + \frac{1}{u - v\gamma}. \quad (31)$$

Let $\zeta = (\frac{1}{\gamma} + \frac{1}{u-v\gamma})^{-1}$. By Lemma 5, there exists $\boldsymbol{q}' \in \Delta_{K-1}$ such that $\boldsymbol{e}_0^T\boldsymbol{q}' = \gamma$ and at most two active sensors are not fully used. Namely, $\boldsymbol{q}'$ is an optimal solution as well. Therefore, the statement is true in this case.

On the other hand, suppose that an optimal solution $\boldsymbol{q}^*$ is given, in which some of the sensors are fully used. Likewise, we assume without loss of generality that all sensors are active, and only sensor 1 is fully used. Further, we let $\rho_k = \frac{q_k}{q_1} > 0$ since $q_1 > 0$. It can be shown that to maximize q_1 in (28) is equivalent to minimize $\sum_{l=2}^K \rho_l$. Since sensor 1 is the only fully used sensor, for all $k > 1$, we have,

$$\begin{aligned} &\frac{\partial}{\partial \rho_k}\left[\sum_{l=2}^K \rho_l + \lambda\left(\frac{1}{\boldsymbol{e}_0^T\boldsymbol{\rho}} + \frac{1}{\boldsymbol{e}_1^T\boldsymbol{\rho}} - n_1'\right)\right]\Bigg|_{\boldsymbol{\rho}=\boldsymbol{\rho}^*} \\ &\qquad = 1 - \lambda\left(\frac{e_0^{(k)}}{\left(\boldsymbol{e}_0^T\boldsymbol{\rho}^*\right)^2} + \frac{e_1^{(k)}}{\left(\boldsymbol{e}_1^T\boldsymbol{\rho}^*\right)^2}\right) = 0, \end{aligned} \quad (32)$$

where $\boldsymbol{\rho}^* = \boldsymbol{q}^*/q_1$. Similar to (30), the condition (32) implies that $e_1^{(k)} = u - ve_0^{(k)}$ for all $k > 1$ where $u, v > 0$, i.e., we have $\tilde{\boldsymbol{e}}_1 = u\boldsymbol{1} - v\tilde{\boldsymbol{e}}_0$ where $\tilde{\boldsymbol{e}}_i = [e_i^{(2)}, \cdots, e_i^{(K)}]^T \in \mathbb{R}^{K-1}$, $\forall i \in \{0,1\}$. Now we would like to find an optimal $\boldsymbol{q}' = [q_1', \cdots, q_K']^T$ such that at most two active sensors are not fully used. Let $q_1' = q_1^*$, $\tilde{\boldsymbol{q}}' = \frac{1}{1-q_1^*}[q_2', \cdots, q_K']^T \in \mathbb{R}^{K-1}$, $\tilde{\boldsymbol{q}}^* = \frac{1}{1-q_1^*}[q_2^*, \cdots, q_K^*]^T \in \mathbb{R}^{K-1}$. Let $\gamma = \tilde{\boldsymbol{e}}_0^T\tilde{\boldsymbol{q}}^*$, $\zeta = (\frac{1-q_1^*}{\boldsymbol{e}_0^T\boldsymbol{q}^*} + \frac{1-q_1^*}{\boldsymbol{e}_1^T\boldsymbol{q}^*})^{-1}$, $\mathcal{Q} = \{\tilde{\boldsymbol{q}} \in \Delta_{K-2} : \tilde{\boldsymbol{e}}_0^T\tilde{\boldsymbol{q}} = \gamma, \tilde{q}_k \leq \zeta\bar{n}_k'\}$. Note that if $\tilde{\boldsymbol{q}}' \in \mathcal{Q}$, it results that $\boldsymbol{q}'$ is optimal and sensor 1 is fully used. From Lemma 5, the desired $\boldsymbol{q}'$ exists. Hence the theorem follows.

REFERENCES

[1] C.-Z. Bai, V. Gupta, and Y.-F. Huang, "Sequential hypothesis testing with off-line randomized sensor selection strategy," in *Proc. IEEE Int. Conf. on Acoust., Speech, Signal Process. (ICASSP 2012)*, Kyoto, Japan, Mar. 2012.

[2] V. Katewa and V. Gupta, "On the optimality of sequential test with multiple sensors," in *Proc. Amer. Contr. Conf. (ACC 2012)*, Montréal, Canada, Jun. 2012.

[3] A. Wald, *Sequential Analysis*. New York, NY, USA: Wiley, 1947.

[4] S. Chaudhari, V. Koivunen, and H. V. Poor, "Autocorrelation-based decentralized sequential detection of OFDM signals in cognitive radios," *IEEE Trans. Signal Process.*, vol. 57, no. 7, pp. 2690–2700, Mar. 2009.

[5] A. Dejonghe *et al.*, "Versatile spectrum sensing on mobile devices?," in *Proc. IEEE Symp. New Front. Dyn. Spectrum*, Singapore, Apr. 2010, pp. 1–6.

[6] G. V. Moustakides Y. Yilmaz and X. Wang, "Cooperative sequential spectrum sensing based on level-triggered sampling," *IEEE Trans. Signal Process.*, vol. 60, no. 9, pp. 4509–4524, Sep. 2012.

[7] V. Gupta, T. H. Chung, B. Hassibi, and R. M. Murray, "On a stochastic sensor selection algorithm with applications in sensor scheduling and sensor coverage," *Automatica*, vol. 42, no. 2, pp. 251–260, Feb. 2006.

[8] Y. Mo, R. Ambrosino, and B. Sinopoli, "Sensor selection strategies for state estimation in energy constrained wireless sensor networks," *Automatica*, vol. 47, no. 7, pp. 1330–1338, Jul. 2011.

[9] S. Joshi and S. Boyd, "Sensor selection via convex optimization," *IEEE Trans. Signal Process.*, vol. 57, no. 2, pp. 451–462, Feb. 2009.

[10] D. Bajović, B. Sinopoli, and J. Xavier, "Sensor selection for event detection in wireless sensor networks," *IEEE Trans. Signal Process.*, vol. 59, no. 10, pp. 4938–4953, Oct. 2011.

[11] C. K. Sestok, M. R. Said, and A. V. Oppenheim, "Randomized data selection in detection with applications to distributed signal processing," *Proc. IEEE*, vol. 9, no. 8, pp. 1184–1198, Aug. 2003.

[12] Y. Wang, I. I. Hussein, and R. S. Erwin, "Risk-based sensor management for integrated detection and estimation," in *Proc. Amer. Contr. Conf. (ACC 2011)*, San Francisco, CA, USA, Jun. 2011, pp. 3633–3638.

[13] Y. Polyanskiy and S. Verdú, "Binary hypothesis testing with feedback (slides)," in *Proc. Inf. Theory Appl. Workshop (ITA)*, San Diego, CA, USA, Feb. 2011.

[14] M. Naghshvar and T. Javidi, "Active sequential hypothesis testing," *Ann. Statist.*, vol. 41, no. 6, Dec. 2013.

[15] C.-Z. Bai and V. Gupta, "An on-line sensor selection algorithm for SPRT with multiple sensors," in *Proc. IEEE Conf. Decision Contr. (CDC 2014)*, Los Angeles, CA, USA, Dec. 2014.

[16] S. Nitinawarat, G. K. Atia, and V. V. Veeravalli, "Controlled sensing for multihypothesis testing," *IEEE Trans. Autom. Control*, vol. 58, no. 10, pp. 2451–2464, Oct. 2013.

[17] V. Srivastava, K. Plarre, and F. Bullo, "Adaptive sensor selection in sequential hypothesis testing," in *Proc. IEEE Conf. on Decision and Contr. and Eur. Contr. Conf. (CDC-ECC)*, Orlando, FL, USA, Dec. 2011.

[18] V. Srivastava, K. Plarre, and F. Bullo, "Randomized sensor selection in sequential hypothesis testing," *IEEE Trans. Signal Process.*, vol. 59, no. 5, pp. 2342–2354, May 2011.

[19] C. W. Baum and V. V. Veeravalli, "A sequential procedure for multihypothesis testing," *IEEE Trans. Inf. Theory*, vol. 40, no. 6, pp. 1994–2007, Nov. 1994.

[20] R. W. Freund and F. Jarre, "Solving the sum-of-ratios problem by an interior-point method," *J. Global Optimiz.*, vol. 19, no. 1, pp. 83–102, Jan. 2001.

[21] T. Cover and J. Thomas, *Elements of Information Theory*, 2nd ed. Hoboken, NJ, USA: Wiley, 2006.

[22] T. S. Ferguson, *Mathematical Statistics: A Decision Theoretic Approach*. New York, NY, USA: Academic, 1967.

[23] A. Wald and J. Wolfowitz, "Optimum character of the sequential probability ratio test," *Ann. Math. Statist.*, vol. 19, no. 3, pp. 326–339, 1948.

[24] S. Schaible and J. Shi, "Fractional programming: The sum-of-ratios case," *Optimiz. Methods Software*, vol. 18, no. 2, pp. 219–229, 2003.

[25] H. P. Benson, "On the global optimization of sums of linear fractional functions over a convex set," *J. Optimiz. Theory Appl.*, vol. 121, no. 1, pp. 19–39, Apr. 2004.

[26] H. P. Benson, "Solving sum of ratios fractional programs via concave minimization," *J. Optimiz. Theory Appl.*, vol. 135, no. 1, pp. 1–28, Jan. 2007.

[27] D. Siegmund, *Sequential Analysis*. New York, NY, USA: Springer-Verlag, 1985.

Cheng-Zong Bai (S'12) received the B.S. and M.S. degrees in communication engineering from National Chiao Tung University, Hsinchu, Taiwan, in 2005, and the National Taiwan University, Taipei, Taiwan, in 2007, respectively. He received the M.S. and Ph.D. degrees in electrical engineering from the University of Notre Dame, IN, USA, in 2012 and 2015, respectively.

From 2009 to 2010, he was a Research Assistant with the Institute of Earth Sciences, Academia Sinica, Taipei. His research interests include statistical signal processing, cyberphysical security, and information theory.

Vaibhav Katewa received the B.Tech. degree from the Indian Institute of Technology, Kanpur, and the M.S. degree from the University of Notre Dame, Notre Dame, IN, USA, both in electrical engineering.

He is a Ph.D. candidate in the Department of Electrical Engineering, University of Notre Dame. His research interests include decentralized control and protocol design for networked control systems.

Vijay Gupta (S'00–M'06) received the B. Tech degree from the Indian Institute of Technology, Delhi, and the M.S. and Ph.D. degrees from the California Institute of Technology, Pasadena, USA, all in electrical engineering.

He is an Associate Professor in the Department of Electrical Engineering , University of Notre Dame, Notre Dame, IN, USA. He has served as a Research Associate with the Institute for Systems Research, University of Maryland, College Park, USA, and as a consultant to the Systems Group at the United Technology Research Center, Hartford, CT, USA. His research interests include various topics at the intersection of communication, computation, and control.

Dr. Gupta received the NSF Career award in 2009 and the Donald P. Eckman award from the American Automatic Control Council in 2013.

Yih-Fang Huang (S'80–M'82–SM'94–F'95) received the B.S.E.E. degree from National Taiwan University, the M.S.E.E. degree from the University of Notre Dame, Notre Dame, IN, USA, and the Ph.D. from Princeton University, Princeton, NJ, USA.

He is a Professor with the Department of Electrical Engineering and Senior Associate Dean of the College of Engineering at the University of Notre Dame. He served as Chair of the Electrical Engineering Department at the University of Notre Dame from 1998 to 2006. In spring 1993, he received the Toshiba Fellowship and was Toshiba Visiting Professor at Waseda University, Tokyo, Japan. From April to July 2007, he was a visiting professor at Munich University of Technology, Germany. His research interests focus on theory and applications of statistical signal detection and estimation, and adaptive signal processing.

Dr. Huang received the Golden Jubilee Medal of the IEEE Circuits and Systems Society in 1999, served as Vice President in 1997–1998, and was a Distinguished Lecturer for the same society in 2000–2001. At the University of Notre Dame, he received the Presidential Award in 2003, the Electrical Engineering department's Outstanding Teacher Award in 1994 and 2011, the Rev. Edmund P. Joyce, CSC Award for Excellence in Undergraduate Teaching in 2011, and the College of Engineering's Teacher of the Year Award in 2013. In fall 2007, he was awarded the Fulbright-Nokia scholarship for lectures/research at Helsinki University of Technology in Finland.