Cooperative Sensing via Sequential Detection

Qiyue Zou, *Member, IEEE*, Songfeng Zheng, *Member, IEEE*, and Ali H. Sayed, *Fellow, IEEE*

***Abstract*—Efficient and reliable spectrum sensing plays a critical role in cognitive radio networks. This paper presents a cooperative sequential detection scheme to reduce the average sensing time that is required to reach a detection decision. In the scheme, each cognitive radio computes the log-likelihood ratio for its every measurement, and the base station sequentially accumulates these log-likelihood statistics and determines whether to stop making measurement. The paper studies how to implement the scheme in a robust manner when the assumed signal models have unknown parameters, such as signal strength and noise variance. These ideas are illustrated through two examples in spectrum sensing. One assumes both the signal and noise are Gaussian distributed, while the other assumes the target signal is deterministic.**

***Index Terms*—Cognitive radio, composite hypothesis testing, cooperative sensing, sequential detection, spectrum sensing.**

I. Introduction

COGNITIVE radio has recently emerged as a useful technology to improve the efficiency of spectrum utilization [3], [4]. In the U.S., the spectrum is traditionally assigned by the Federal Communications Commission (FCC) to specific users or applications, and each user can only utilize its preassigned bandwidth for communication. This discipline causes some bandwidth to be overcrowded while some other bandwidth may be underutilized. Cognitive radio aims at providing a flexible way of spectrum management, permitting secondary users to temporally access spectrum that is not used by legacy users. In this regard, the FCC has taken a number of steps in the U.S. towards allowing low-power devices to operate in the broadcast TV bands that are not being used by TV channels [5]. The U.S. TV bands include the following portions of the VHF and UHF radio spectrum: 54–72, 76–88, 174–216, and 470–806 MHz. Each TV channel occupies a slot of 6-MHz bandwidth. If a TV frequency band is not used in a particular geographical region, it can be used by cognitive radios for transmission. To promote this development, IEEE has established the IEEE 802.22 Working Group to develop a standard for a cognitive radio-based device in TV bands [6].

A key challenge in the development of the IEEE 802.22 standard is that a cognitive radio should be able to reliably detect the presence of TV signals in a fading environment. Otherwise, the radio may use the frequency band that is occupied by a TV channel, and cause interference to the TV receivers nearby. Many sensing and detection schemes have been reported in the IEEE 802.22 community, e.g., [7]–[12]. These schemes can be classified into two categories: single-user sensing and cooperative sensing. Due to the large variation in the received signal strength that is caused by path loss and fading, single-user sensing has proven to be unreliable, which consequently triggered the FCC to require geolocation-based methods for identifying unused frequency bands [13], [14]. The geolocation approach is suitable for registered TV bands; however, its cost and operational overhead prevent its wide use in the opportunistic access to occasional "white spaces" in the spectrum. Cooperative sensing relies on multiple radios to detect the presence of primary users and provides a reliable solution for cognitive radio networks [10]–[12]. In this paper, we focus on how to achieve cooperative sensing in an efficient and robust manner.

The performance of spectrum sensing is usually measured by two key factors: probability of detection errors and sensing time. The traditional way to design a sensing strategy is based on the Neyman–Pearson criterion, and the resulting likelihood ratio test (LRT) fixes the number of required samples or the sensing time. In this framework, the probability of false alarm is required to be less than a predefined level α, and under this constraint, the probability of miss detection is optimized (minimized) by the proposed test [15]. In contrast to the Neyman–Pearson framework, another design methodology is to minimize the required sensing time, subject to a constraint on the detection errors [16]–[18]. The resulting test is called the sequential probability ratio test (SPRT) and was first developed in the seminal work by Wald [19]. A recent exposition about the theory behind the test can be found in [20]. Some recent papers have applied this technique to spectrum sensing for cognitive radio networks, e.g., [21] and [22]. In the scheme proposed in [21], the autocorrelation coefficient based log-likelihood ratios from different cognitive radios are combined in a sequential manner at the base station for quickly detecting the primary user. In [22], the sequential detection method is applied to the detection of cyclostationary features in the received signals. These techniques can reduce the sensing time and the amount of signal samples required in identifying the unused spectrum.

In this paper, we extend previous work on the sequential detection method for collaborative spectrum sensing. In the proposed framework, each cognitive radio computes the log-likelihood ratio for its every measurement, and the base station se-

Manuscript received February 27, 2010; accepted August 10, 2010. Date of publication August 26, 2010; date of current version November 17, 2010. The associate editor coordinating the review of this manuscript and approving it for publication was Prof. Huaiyu Dai. This work was supported in part by the National Science Foundation under Grants ECS-0601266, ECS-0725441, and CCF-0942936. Preliminary versions of this work appeared in the conference publications *Proceedings of the Tenth IEEE International Workshop on Signal Processing Advances in Wireless Communications* (SPAWC), June 2009, pp. 121–125, and the *Proceedings of the Fifteenth IEEE Workshop on Statistical Signal Processing*, August 2009, pp. 610–613.

Q. Zou was with the Electrical Engineering Department, University of California, Los Angeles, CA 90095 USA. He is now with the Marvell Semiconductor, Inc., Santa Clara, CA 95054 USA (e-mail: eqyzou@gmail.com).

S. Zheng is with the Department of Mathematics, Missouri State University, Springfield, MO 65897 USA (e-mail: songfengzheng@missouristate.edu).

A. H. Sayed is with the Electrical Engineering Department, University of California, Los Angeles, CA 90095 USA (e-mail: sayed@ee.ucla.edu).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TSP.2010.2070501

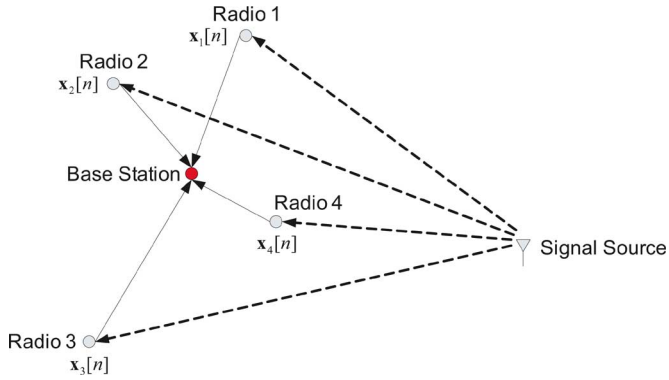

Fig. 1. Cognitive radio network for spectrum sensing.

quentially accumulates the log-likelihood statistics and determines whether to stop making new measurement. Due to uncertainties caused by fading and interference, we normally do not have exact information about some signal parameters, such as signal strength and noise variance. It is thus important to make the sequential detection algorithm sufficiently robust to the uncertainties in unknown parameters. Different from previous work which assumes complete knowledge about the distributions of the measurements, our work modifies the original SPRT in order to handle unknown parameters in the assumed signal models. In our proposed solution, unknown parameters are sequentially estimated by the maximum likelihood estimation, and the sequential detection algorithm is performed by using the estimated parameters. By doing so, the average sensing time depends on the signal conditions, rather than being fixed as in the Neyman–Pearson approach. With proper stopping conditions, the proposed scheme guarantees to achieve the desired sensing performance in terms of the probability of false alarm and miss detection. These ideas are illustrated through two spectrum sensing examples. One assumes both the signal and noise are Gaussian distributed, while the other assumes the target signal is deterministic.

Throughout this paper, we adopt the following definitions and notations. The network consists of M cognitive radios that are monitoring the frequency band of interest, as shown in Fig. 1. The two hypotheses corresponding to the signal-absent and signal-present events are defined as

$$\mathcal{H}_0: \text{target signal is absent}$$
$$\mathcal{H}_1: \text{target signal is present.}$$

The signal acquired by the mth $(m = 1, 2, \ldots, M)$ cognitive radio device is represented by

$$\mathcal{H}_0: \mathbf{x}_m[n] = \mathbf{s}_{0,m}[n],$$
$$\mathcal{H}_1: \mathbf{x}_m[n] = \mathbf{s}_{1,m}[n], \quad n = 1, 2, \ldots$$

where $\mathbf{s}_{1,m}[n]$ is the nth acquired signal sample when the target signal is present and $\mathbf{s}_{0,m}[n]$ is the nth acquired noise signal sample when the target signal is absent. The samples $\mathbf{x}_m[n]$ can be either a scalar or a vector, depending on the application of interest. Throughout the paper, we assume that the samples acquired by different radios are statistically independent, and that the samples acquired by the same radio are independent and identically distributed (i.i.d.).[1] Under $\mathcal{H}_0$ and $\mathcal{H}_1$, the distributions of the acquired signal at the mth radio are characterized by the probability density functions $p_{0,m}(\mathbf{x}_m[n])$ and $p_{1,m}(\mathbf{x}_m[n])$, respectively. The performance of detecting $\mathcal{H}_1$ against $\mathcal{H}_0$ is measured by the probability of false alarm and the probability of miss detection. The error of false alarm refers to the error of accepting $\mathcal{H}_1$ when $\mathcal{H}_0$ is true, while the error of miss detection is the error of accepting $\mathcal{H}_0$ when $\mathcal{H}_1$ is true. The probability of false alarm is represented by

$$P_{\text{FA}} = \Pr(\hat{\mathcal{H}} = \mathcal{H}_1 \,|\, \mathcal{H}_0)$$

and the probability of miss detection is represented by

$$P_{\text{MISS}} = \Pr(\hat{\mathcal{H}} = \mathcal{H}_0 \,|\, \mathcal{H}_1)$$

where $\hat{\mathcal{H}}$ represents the detector output.

The paper is organized as follows. Section II develops the sequential test for simple hypotheses and its application to cooperative sensing. Section III extends the discussion to composite hypothesis testing problems when the sensing models have unknown parameters and modeling uncertainties. The proposed scheme is evaluated in Section IV through computer simulations.

II. Sequential Sensing for Simple Hypotheses

To begin with, assume that the number N_{fix} of samples (acquired by each cognitive radio) is *fixed*. To detect $\mathcal{H}_0$ and $\mathcal{H}_1$, the likelihood ratio test (LRT) is performed according to

$$\begin{aligned} &\text{Accept } \mathcal{H}_1 \text{ if } \text{LLR} > \eta \\ &\text{Accept } \mathcal{H}_0 \text{ if } \text{LLR} \le \eta \end{aligned} \tag{1}$$

where the log-likelihood ratio (LLR) is computed by the base station as

$$\begin{aligned} \text{LLR} &= \ln\left(\prod_{n=1}^{N_{\text{fix}}}\prod_{m=1}^{M}\frac{p_{1,m}(\mathbf{x}_m[n])}{p_{0,m}(\mathbf{x}_m[n])}\right) \\ &= \sum_{n=1}^{N_{\text{fix}}}\sum_{m=1}^{M}\ln\left(\frac{p_{1,m}(\mathbf{x}_m[n])}{p_{0,m}(\mathbf{x}_m[n])}\right). \end{aligned}$$

The threshold value η and the sample size N_{fix} are selected such that the probability of false alarm and the probability of miss detection are bounded by some pre-assigned values $0 < \alpha < 1$ and $0 < \beta < 1$, respectively,[2] i.e.,

$$P_{\text{FA}} \le \alpha \quad \text{and} \quad P_{\text{MISS}} \le \beta. \tag{2}$$

To do so, the distributions of the test statistic, i.e., the LLR, under $\mathcal{H}_0$ and $\mathcal{H}_1$ need to be determined. The computation of

[1]The assumptions simplify the notation and derivations presented in the paper. They have extensions for many applications. For instance, in Example 2 further ahead, the samples acquired by different radios are statistically independent, given that the target signal and its amplitude are deterministic. Since the target signal in Example 2 is periodic, we define each sample as a signal vector over one period and the samples from the same radio can be regarded as i.i.d.

[2]The detector with a fixed sample size is designed according to the Neyman–Pearson criterion. The threshold η is determined by the probability of false alarm. That is, no matter what other conditions are, the threshold η is always set to ensure that P_{MISS} is minimized under the constraint $P_{\text{FA}} \le \alpha$. To ensure $P_{\text{MISS}} \le \beta$ by this design methodology, we need to choose appropriate N_{fix}.

the distributions is usually not easy and may involve complex numerical computations or simulations.

To reduce the number of required samples, instead of using a fixed sample size N_{fix}, we can implement the LRT for every acquired sample in a sequential manner motivated by Wald's work [19]. That is, for $N = 1, 2, \ldots$, we perform the following test:

$$\begin{aligned} &\text{Accept } \mathcal{H}_1 \text{ and terminate if } \text{LLR}_N \geq A \\ &\text{Accept } \mathcal{H}_0 \text{ and terminate if } \text{LLR}_N \leq B \\ &\text{Take one more sample to} \\ &\quad \text{repeat the test if } B < \text{LLR}_N < A \end{aligned} \tag{3}$$

where

$$\text{LLR}_N = \sum_{n=1}^{N} \sum_{m=1}^{M} \ln\left(\frac{p_{1,m}(\mathbf{x}_m[n])}{p_{0,m}(\mathbf{x}_m[n])}\right),$$

$A > 0$ and $B < 0$ are predetermined constants according to the sensing objective (2). In the context of cooperative sensing, each radio computes the log-likelihood ratio for its every acquired sample, and the base station sequentially accumulates the log-likelihood statistics and performs the above test, as described in Algorithm 1.

Algorithm 1: Cooperative Sequential Sensing for Simple Hypotheses

0: Set $N = 0$, and let $\text{LLR}_0 = 0$ at the base station.
1: **repeat**
2: $N = N + 1$.
3: The mth ($m = 1, 2, \ldots, M$) radio acquires sample $\mathbf{x}_m[N]$ and computes $\ln(p_{1,m}(\mathbf{x}_m[N])/p_{0,m}(\mathbf{x}_m[N]))$.
4: Each radio sends its $\ln(p_{1,m}(\mathbf{x}_m[N])/p_{0,m}(\mathbf{x}_m[N]))$ to the base station.
5: The base station updates the sequential log-likelihood ratio LLR_N according to

$$\text{LLR}_N = \text{LLR}_{N-1} + \sum_{m=1}^{M} \ln\left(\frac{p_{1,m}(\mathbf{x}_m[N])}{p_{0,m}(\mathbf{x}_m[N])}\right).$$

6: **until** $\text{LLR}_N \geq A$ or $\text{LLR}_N \leq B$.
7: If $\text{LLR}_N \geq A$, "$\mathcal{H}_1$: target signal is present" is claimed; if $\text{LLR}_N \leq B$, "$\mathcal{H}_0$: target signal is absent" is claimed.

Assume that the detection procedure terminates at $N = N_{\text{stop}}$. By the following lemma, the test stops at finite N_{stop} with probability one.

Lemma 2.1: (see [23, Lemma 1]): If the second moment of $\sum_{m=1}^{M} \ln(p_{1,m}(\mathbf{x}_m[n])/p_{0,m}(\mathbf{x}_m[n]))$ under $\mathcal{H}_i$ is not zero, then $\Pr(N_{\text{stop}} < \infty \,|\, \mathcal{H}_i) = 1$ for $i = 0, 1$.

To see how A and B are determined, we need to study how P_{FA} and P_{MISS} depend on A and B. Before proceeding, we present some regularity assumptions. Recall that the Kullback-Leibler (KL) distance $D(p(x) \parallel q(x))$ between the distributions $p(x)$ and $q(x)$ is defined as

$$D(p(x) \parallel q(x)) = \int p(x) \ln\left(\frac{p(x)}{q(x)}\right) dx.$$

Moreover, $D(p(x) \parallel q(x)) \geq 0$, where equality holds if, and only if, $p(x) = q(x)$. Notice that

$$\begin{aligned} &\mathbf{E}_{\mathcal{H}_0}\left\{\ln\left(\frac{p_{1,m}(\mathbf{x}_m[n])}{p_{0,m}(\mathbf{x}_m[n])}\right)\right\} \\ &\quad = -D(p_{0,m}(\mathbf{x}_m[n]) \parallel p_{1,m}(\mathbf{x}_m[n])) \\ &\mathbf{E}_{\mathcal{H}_1}\left\{\ln\left(\frac{p_{1,m}(\mathbf{x}_m[n])}{p_{0,m}(\mathbf{x}_m[n])}\right)\right\} \\ &\quad = D(p_{1,m}(\mathbf{x}_m[n]) \parallel p_{0,m}(\mathbf{x}_m[n])). \end{aligned}$$

Throughout the paper, we assume that

1a) $0 < \sum_{m=1}^{M} D(p_{0,m}(\mathbf{x}_m[n]) \parallel p_{1,m}(\mathbf{x}_m[n])) < \infty$;
1b) $0 < \sum_{m=1}^{M} D(p_{1,m}(\mathbf{x}_m[n]) \parallel p_{0,m}(\mathbf{x}_m[n])) < \infty$.

The ">0" condition ensures that the two hypotheses are distinguishable based on the underlying distributions, which also implies the condition required by Lemma 2.1; the "$<\infty$" condition ensures that the distributions are well behaved for the subsequent derivation.

At $N = N_{\text{stop}}$, we have

$$\text{LLR}_{N_{\text{stop}}} \approx A \quad \text{or} \quad \text{LLR}_{N_{\text{stop}}} \approx B$$

provided that the change in LLR_N at each step is relatively small compared to the absolute values of A and B, which is true when α and β are sufficiently small.[3] It can then be shown that (see Appendix A for the derivation[4])

$$P_{\text{FA}} = \Pr(\text{LLR}_{N_{\text{stop}}} \geq A \,|\, \mathcal{H}_0) \approx \frac{1 - e^{B}}{e^{A} - e^{B}} \tag{4}$$

$$P_{\text{MISS}} = \Pr(\text{LLR}_{N_{\text{stop}}} \leq B \,|\, \mathcal{H}_1) \approx \frac{e^{-A} - 1}{e^{-A} - e^{-B}}. \tag{5}$$

To find appropriate A and B, we set (4) and (5) to be equal to α and β, respectively, and result

$$A \approx \ln\left(\frac{1-\beta}{\alpha}\right), \quad B \approx \ln\left(\frac{\beta}{1-\alpha}\right). \tag{6}$$

Obviously, if α and β are sufficiently small, we have $A > 0$ and $B < 0$. We also see from (6) that A and B do not depend on specific distributions and are convenient to compute. Since α and β are normally much smaller than 1, we let

$$A = -\ln \alpha > 0 \quad \text{and} \quad B = \ln \beta < 0. \tag{7}$$

Although the stopping boundary (7) is obtained through approximation, we can prove that with (7), the sensing objective (2) is exactly achievable by the test, as shown in the following lemma.

Lemma 2.2: If the stopping condition for the sequential test is set according to (7), then

$$P_{\text{FA}} \leq \alpha \quad \text{and} \quad P_{\text{MISS}} \leq \beta.$$

Proof: See Appendix B. ∎

[3]It is seen in (6) and (7) that the absolute values of A and B can be sufficiently large when α and β are sufficiently small.

[4]Appendices A and B are immediate consequences of some well-known facts in the sequential detection literature. They are given here to ensure readers who are new to this area can have good understanding of the theory that is frequently used in the subsequent sections.

In the test, the number of samples required to reach a decision in Algorithm 1 is a random variable. By using Wald's equation,[5] the average number of required samples under $\mathcal{H}_0$ and $\mathcal{H}_1$ is given by (see Appendix C for the derivation)

$$\mathbf{E}_{\mathcal{H}_0}\{N_{\text{stop}}\} \approx \frac{-(A-B-Ae^B+Be^A)}{(e^A-e^B)\left[\sum_{m=1}^{M} D\big(p_{0,m}(\mathbf{x}_m[n]) \,||\, p_{1,m}(\mathbf{x}_m[n])\big)\right]} \tag{8}$$

and

$$\mathbf{E}_{\mathcal{H}_1}\{N_{\text{stop}}\} \approx \frac{A-B-Ae^{-B}+Be^{-A}}{(e^{-A}-e^{-B})\left[\sum_{m=1}^{M} D(p_{1,m}(\mathbf{x}_m[n]) \,||\, p_{0,m}(\mathbf{x}_m[n]))\right]}. \tag{9}$$

To summarize the results, we have the following theorem.

Theorem 2.1: For the sequential sensing procedure defined in (3), if the thresholds A and B are given by (7), then $P_{\text{FA}} \le \alpha$ and $P_{\text{MISS}} \le \beta$. If α and β are sufficiently small, P_{FA} and P_{MISS} are approximated by expressions (4) and (5), respectively, and the average number of required samples under $\mathcal{H}_0$ and $\mathcal{H}_1$ is approximated by expressions (8) and (9), respectively.

Expressions (8) and (9) show that the average required sensing time depends on the KL distance provided by each sensing radio. The larger the KL distance, the less the required sensing time is, and this is consistent with intuition. In order to save processing power and communication bandwidth, we may need to select a subset of available radios for spectrum sensing. Based on (8) and (9), we can choose radios with larger KL distance in order to minimize

$$\max\left\{\mathbf{E}_{\mathcal{H}_0}\{N_{\text{stop}}\},\ \mathbf{E}_{\mathcal{H}_1}\{N_{\text{stop}}\}\right\}.$$

In the following, we illustrate the sequential sensing technique by examples.

Example 1—Detecting a Gaussian Random Signal with Known Variance: In the first example, the signal samples are scalars, i.e., let $\mathbf{x}_m[n] = X_m[n]$. The acquired signals under $\mathcal{H}_0$ and $\mathcal{H}_1$ are assumed to be i.i.d. Gaussian with mean zero and variances $\sigma_{0,m}^2$ and $\sigma_{1,m}^2$ ($\sigma_{1,m}^2 > \sigma_{0,m}^2$), respectively,[6] i.e.,

$$\mathcal{H}_i:\ X_m[n] \sim \frac{1}{\sqrt{2\pi\sigma_{i,m}^2}} e^{-\frac{(X_m[n])^2}{2\sigma_{i,m}^2}}, \quad i=0,1.$$

[5]Let $X_1, X_2, \ldots$ be any sequence of i.i.d. random variables with partial sums $S_n = X_1 + X_2 + \ldots + X_n$. Let T be any stopping time with respect to X_n $(n = 1, 2, \ldots)$. Wald's equation states that $\mathbf{E}\{S_T\} = \mathbf{E}\{X_1\}\mathbf{E}\{T\}$, provided that $\mathbf{E}\{X_1\} < \infty$ and $\mathbf{E}\{T\} < \infty$.

[6]This assumption is based on the statistical model used in the energy or power detectors, where primary users are detected based on the received signal power level.

For the test with a fixed sample size N_{fix}, the LRT given by (1) turns out to be an energy detector, i.e.,

$$\text{Accept } \mathcal{H}_1 \text{ if } \sum_{n=1}^{N_{\text{fix}}}\sum_{m=1}^{M}\left[\frac{1}{2}\left(\frac{1}{\sigma_{0,m}^2}-\frac{1}{\sigma_{1,m}^2}\right)(X_m[n])^2 + \frac{1}{2}\ln\left(\frac{\sigma_{0,m}^2}{\sigma_{1,m}^2}\right)\right] > \eta;$$

$$\text{Accept } \mathcal{H}_0 \text{ if } \sum_{n=1}^{N_{\text{fix}}}\sum_{m=1}^{M}\left[\frac{1}{2}\left(\frac{1}{\sigma_{0,m}^2}-\frac{1}{\sigma_{1,m}^2}\right)(X_m[n])^2 + \frac{1}{2}\ln\left(\frac{\sigma_{0,m}^2}{\sigma_{1,m}^2}\right)\right] \le \eta.$$

To find the value of η, we treat the summation $\sum_{n=1}^{N_{\text{fix}}}(X_m[n])^2$ approximately as Gaussian distributed and obtain an expression for the minimum N_{fix} to achieve (2) [see (10) at the bottom of the page and Appendix D for the derivation], where $Q(\cdot)$ is the Q-function of the standard normal distribution.[7] The associated η is given by

$$\eta \approx \frac{N_{\text{fix}}}{2}\sum_{m=1}^{M}\left(1-\frac{\sigma_{0,m}^2}{\sigma_{1,m}^2}\right) + \frac{N_{\text{fix}}}{2}\sum_{m=1}^{M}\ln\left(\frac{\sigma_{0,m}^2}{\sigma_{1,m}^2}\right) + \sqrt{\frac{N_{\text{fix}}}{2}\sum_{m=1}^{M}\left(1-\frac{\sigma_{0,m}^2}{\sigma_{1,m}^2}\right)^2}\,Q^{-1}(\alpha). \tag{11}$$

For the sequential test given in (3), we have

$$\text{LLR}_N = \sum_{n=1}^{N}\sum_{m=1}^{M}\left[\frac{1}{2}\left(\frac{1}{\sigma_{0,m}^2}-\frac{1}{\sigma_{1,m}^2}\right)(X_m[n])^2 + \frac{1}{2}\ln\left(\frac{\sigma_{0,m}^2}{\sigma_{1,m}^2}\right)\right].$$

The average number of required samples under $\mathcal{H}_0$ and $\mathcal{H}_1$ is given by (8) and (9), where

$$D(p_{0,m}(\mathbf{x}_m[n]) \,||\, p_{1,m}(\mathbf{x}_m[n])) = \frac{1}{2\rho_m} - \frac{1}{2} + \frac{1}{2}\ln\rho_m,$$

$$D(p_{1,m}(\mathbf{x}_m[n]) \,||\, p_{0,m}(\mathbf{x}_m[n])) = \frac{1}{2}\rho_m - \frac{1}{2} - \frac{1}{2}\ln\rho_m,$$

with $\rho_m = \sigma_{1,m}^2/\sigma_{0,m}^2 > 1$. The above two KL distance terms are monotonically increasing functions of ρ_m. So we prefer sensing radios with large ρ_m, which is consistent with common sense.

Example 2—Detecting a Known Periodic Signal in Gaussian Noise: In the second example, we detect a known *periodic*

[7]$Q(x)$ is defined as the probability that a standard normal random variable (zero mean, unit variance) exceeds x, i.e., $Q(x) = \int_x^\infty 1/\sqrt{2\pi}e^{-y^2/2}dy$.

$$N_{\text{fix}} \approx 2\left(\frac{\sqrt{\sum_{m=1}^{M}\left(1-\frac{\sigma_{0,m}^2}{\sigma_{1,m}^2}\right)^2}\,Q^{-1}(\alpha) + \sqrt{\sum_{m=1}^{M}\left(\frac{\sigma_{1,m}^2}{\sigma_{0,m}^2}-1\right)^2}\,Q^{-1}(\beta)}{\sum_{m=1}^{M}\frac{\sigma_{1,m}^2}{\sigma_{0,m}^2} + \sum_{m=1}^{M}\frac{\sigma_{0,m}^2}{\sigma_{1,m}^2} - 2M}\right)^2, \tag{10}$$

signal in white Gaussian noise.[8] Under $\mathcal{H}_0$ and $\mathcal{H}_1$, the acquired signal at each sensing radio is represented by

$$\mathcal{H}_0: X_m[n'] = W_m[n'],$$
$$\mathcal{H}_1: X_m[n'] = a_m s[n'] + W_m[n'], \quad n' = 1, 2, \ldots$$

where $s[n']$ is the periodic target signal, a_m is the amplitude of the received target signal, and $W_m[n']$ is the additive white Gaussian noise with mean zero and variance σ_m^2. Assume that $s[n']$ and a_m are known to the receivers, and the signals acquired by different radios are statistically independent. Since $s[n']$ is periodic, we define each $\mathbf{x}_m[n]$ as a signal vector over one period, i.e.,

$$\mathbf{x}_m[n] = [\, X_m[(n-1)K+1] \quad X_m[(n-1)K+2] \ldots X_m[nK]]$$

where K is the period of $s[n']$. We can see that the $\mathbf{x}_m[n]$ from the same radio are i.i.d. distributed. The LRT with fixed sample size is given by

$$\text{Accept } \mathcal{H}_1 \text{ if } \sum_{n=1}^{N_{\text{fix}}} \sum_{k=1}^{K} \sum_{m=1}^{M} \left[\frac{a_m}{\sigma_m^2} s[k] X_m[(n-1)K+k] - \frac{a_m^2}{2\sigma_m^2} (s[k])^2 \right] > \eta$$

$$\text{Accept } \mathcal{H}_0 \text{ if } \sum_{n=1}^{N_{\text{fix}}} \sum_{k=1}^{K} \sum_{m=1}^{M} \left[\frac{a_m}{\sigma_m^2} s[k] X_m[(n-1)K+k] - \frac{a_m^2}{2\sigma_m^2} (s[k])^2 \right] \leq \eta.$$

The minimum N_{fix} required to achieve (2) is given by (see Appendix E for the derivation)

$$N_{\text{fix}} \approx \frac{\left[Q^{-1}(\alpha) + Q^{-1}(\beta)\right]^2}{\left(\sum_{m=1}^{M} \frac{a_m^2}{\sigma_m^2}\right)\left(\sum_{k=1}^{K} (s[k])^2\right)}. \tag{12}$$

The associated η is given by

$$\eta \approx \frac{1}{2}([Q^{-1}(\alpha)]^2 - [Q^{-1}(\beta)]^2). \tag{13}$$

For the sequential test (3), we have

$$\text{LLR}_N = \sum_{n=1}^{N} \sum_{k=1}^{K} \sum_{m=1}^{M} \left[\frac{a_m}{\sigma_m^2} s[k] X_m[(n-1)K+k] - \frac{a_m^2}{2\sigma_m^2} (s[k])^2 \right].$$

The average number of required samples under $\mathcal{H}_0$ and $\mathcal{H}_1$ is given by (8) and (9), where

$$D(p_{0,m}(\mathbf{x}_m[n]) \,||\, p_{1,m}(\mathbf{x}_m[n])) = D(p_{1,m}(\mathbf{x}_m[n]) \,||\, p_{0,m}(\mathbf{x}_m[n])) = \frac{a_m^2}{2\sigma_m^2} \left(\sum_{k=1}^{K} (s[k])^2 \right).$$

III. Sequential Sensing With Composite Hypotheses

In Section II, we explained how to perform sequential sensing in a network when the distributions $p_{0,m}(\cdot)$ and $p_{1,m}(\cdot)$ are *exactly* known. In practice, however, there normally exist unknown parameters in the assumed statistical models of signal and noise. In a radio propagation environment, these unknown parameters might be the signal strength, noise variance, etc. With imperfect knowledge about these parameters, non-exact distributions are used in the sensing algorithm and may cause performance degradation. In this section, we first analyze the performance degradation caused by the modeling errors for the sequential test proposed in Section II, and then study how to modify the test to deal with unknown model parameters.

A. Sequential Detection With Modeling Errors

Rather than the true distributions $p_{0,m}(\mathbf{x}_m[n])$ and $p_{1,m}(\mathbf{x}_m[n])$ underlying $\mathcal{H}_0$ and $\mathcal{H}_1$, we assume that non-exact distributions $q_{0,m}(\mathbf{x}_m[n])$ and $q_{1,m}(\mathbf{x}_m[n])$ are used in the sequential test (3), i.e., the sequential log-likelihood ratio is computed by

$$\text{LLR}'_N = \sum_{n=1}^{N} \sum_{m=1}^{M} \ln \left(\frac{q_{1,m}(\mathbf{x}_m[n])}{q_{0,m}(\mathbf{x}_m[n])} \right). \tag{14}$$

The assumed distributions $q_{0,m}(\mathbf{x}_m[n])$ and $q_{1,m}(\mathbf{x}_m[n])$ are usually close to their true counterpart $p_{0,m}(\mathbf{x}_m[n])$ and $p_{1,m}(\mathbf{x}_m[n])$. Recall that the following assumptions for the distributions $p_{0,m}(\mathbf{x}_m[n])$ and $p_{1,m}(\mathbf{x}_m[n])$ are made in the derivation of the properties of the sequential test:

- 1a) $\sum_{m=1}^{M} D(p_{0,m}(\mathbf{x}_m[n]) \,||\, p_{1,m}(\mathbf{x}_m[n]))$, i.e., the mean of $\sum_{m=1}^{M} \ln(p_{0,m}(\mathbf{x}_m[n])/p_{1,m}(\mathbf{x}_m[n]))$ under $\mathcal{H}_0$, is positive and finite;
- 1b) $\sum_{m=1}^{M} D(p_{1,m}(\mathbf{x}_m[n]) \,||\, p_{0,m}(\mathbf{x}_m[n]))$, i.e., the mean of $\sum_{m=1}^{M} \ln(p_{1,m}(\mathbf{x}_m[n])/p_{0,m}(\mathbf{x}_m[n]))$ under $\mathcal{H}_1$, is positive and finite.

We make similar assumptions for $q_{0,m}(\mathbf{x}_m[n])$ and $q_{1,m}(\mathbf{x}_m[n])$ as follows:

- 2a) the mean of $\sum_{m=1}^{M} \ln(q_{0,m}(\mathbf{x}_m[n])/q_{1,m}(\mathbf{x}_m[n]))$ under $\mathcal{H}_0$ is positive and finite;
- 2b) the mean of $\sum_{m=1}^{M} \ln(q_{1,m}(\mathbf{x}_m[n])/q_{0,m}(\mathbf{x}_m[n]))$ under $\mathcal{H}_1$ is positive and finite.

The assumptions ensure that the sequential test with the log-likelihood ratio given by (14) stops with probability one. Intuitively, 2a) implies that the mean of LLR'_N under $\mathcal{H}_0$ is negative and goes unbounded as N increases, while 2b) implies that the mean of LLR'_N under $\mathcal{H}_1$ is positive and goes unbounded as N increases. The following theorem summarizes the performance of the sequential sensing algorithm in the presence of modeling errors. Part of the results can also be found in [24].

Theorem 3.1: For the test defined in (3) with LLR_N being replaced by LLR'_N in (14), assume that 2a) and 2b) hold. Then,

1) $\Pr(N_{\text{stop}} < \infty \,|\, \mathcal{H}_i) = 1$ for $i = 0, 1$;
2) there exist $\lambda_0 > 0$ and $\lambda_1 < 0$ such that

$$\mathbf{E}_{\mathcal{H}_0} \left\{ e^{\lambda_0 \sum_{m=1}^{M} \ln\left(\frac{q_{1,m}(\mathbf{x}_m[n])}{q_{0,m}(\mathbf{x}_m[n])}\right)} \right\} = 1$$

and

$$\mathbf{E}_{\mathcal{H}_1} \left\{ e^{\lambda_1 \sum_{m=1}^{M} \ln\left(\frac{q_{1,m}(\mathbf{x}_m[n])}{q_{0,m}(\mathbf{x}_m[n])}\right)} \right\} = 1;$$

[8] Sinusoidal signal is often used in communication channels for synchronization. It can be utilized as features for spectrum sensing.

3) if the thresholds A and B are given by (7), then

$$P_{\rm FA} \le \alpha^{\lambda_0} \quad \text{and} \quad P_{\rm MISS} \le \beta^{-\lambda_1};$$

4) if α and β are sufficiently small, $P_{\rm FA}$ and $P_{\rm MISS}$ are approximated by

$$P_{\rm FA} \approx \frac{1-e^{\lambda_0 B}}{e^{\lambda_0 A}-e^{\lambda_0 B}} \tag{15}$$

$$P_{\rm MISS} \approx \frac{1-e^{\lambda_1 A}}{e^{\lambda_1 B}-e^{\lambda_1 A}} \tag{16}$$

and the average number of samples required by the sensing scheme is given by (17) and (18), shown at the bottom of the page.

Proof: See Appendix F. ∎

It can be seen that in the absence of modeling errors, i.e., when

$$q_{i,m}(\mathbf{x}_m[n]) = p_{i,m}(\mathbf{x}_m[n]), \quad i=0,1,$$

we have $\lambda_0 = 1$ and $\lambda_1 = -1$, which is identical to the case we discussed in Section II. If $\lambda_0 \neq 1$, $P_{\rm FA}$ can be greater or less than the desired value α, if the stopping thresholds A and B are set according to (7), and a similar remark holds for λ_1. In many circumstances, modeling uncertainties can be represented by unknown parameters, for which the sequential test can be modified to incorporate sequential parameter estimation.

B. Sequential Sensing With Unknown Parameters

Assume that unknown parameters $\boldsymbol{\theta}_{0,m} \in \Theta_{0,m}$ and $\boldsymbol{\theta}_{1,m} \in \Theta_{1,m}$ are deterministic for $m = 1,2,\ldots,M$. The signal received by the mth $(m = 1,2,\ldots,M)$ sensing radio is distributed according to

$$\mathcal{H}_0 : \mathbf{x}_m[n] \sim p_{0,m}(\mathbf{x}_m[n];\boldsymbol{\theta}_{0,m})$$
$$\mathcal{H}_1 : \mathbf{x}_m[n] \sim p_{1,m}(\mathbf{x}_m[n];\boldsymbol{\theta}_{1,m}).$$

If the distributions $p_{0,m}(\cdot)$ and $p_{1,m}(\cdot)$ belong to the same family of distribution, we assume that the parameter spaces $\Theta_{0,m}$ and $\Theta_{1,m}$ are disjoint, i.e., $\Theta_{0,m} \cap \Theta_{1,m} = \emptyset$.[9] For a fixed sample size $N_{\rm fix}$, the log-likelihood ratio is given by

$$\text{LLR}'' = \sum_{n=1}^{N_{\rm fix}} \sum_{m=1}^{M} \ln\left(\frac{p_{1,m}(\mathbf{x}_m[n];\boldsymbol{\theta}_{1,m})}{p_{0,m}(\mathbf{x}_m[n];\boldsymbol{\theta}_{0,m})}\right). \tag{19}$$

Since $\boldsymbol{\theta}_{i,m}$ $(i = 0,1,\ m = 1,2,\ldots,M)$ are unknown, we exploit the generalized likelihood ratio test by replacing $\boldsymbol{\theta}_{i,m}$ in (19) with their maximum likelihood estimates. In other words, this test uses the so-called generalized log-likelihood ratio (GLLR), which is defined as

$$\text{GLLR} = \sum_{n=1}^{N_{\rm fix}} \sum_{m=1}^{M} \ln\left(\frac{p_{1,m}(\mathbf{x}_m[n];\hat{\boldsymbol{\theta}}_{1,m})}{p_{0,m}(\mathbf{x}_m[n];\hat{\boldsymbol{\theta}}_{0,m})}\right)$$

where $\hat{\boldsymbol{\theta}}_{1,m}$ and $\hat{\boldsymbol{\theta}}_{0,m}$ are the maximum likelihood estimates of $\boldsymbol{\theta}_{1,m}$ and $\boldsymbol{\theta}_{0,m}$, i.e.,

$$\hat{\boldsymbol{\theta}}_{0,m} = \arg\max_{\boldsymbol{\theta}_{0,m}\in\Theta_{0,m}} \sum_{n=1}^{N_{\rm fix}} \ln p_{0,m}(\mathbf{x}_m[n];\boldsymbol{\theta}_{0,m})$$
$$\hat{\boldsymbol{\theta}}_{1,m} = \arg\max_{\boldsymbol{\theta}_{1,m}\in\Theta_{1,m}} \sum_{n=1}^{N_{\rm fix}} \ln p_{1,m}(\mathbf{x}_m[n];\boldsymbol{\theta}_{1,m}).$$

The sequential test based on the above GLLR is defined as follows. Each time when the radios acquire new samples $\mathbf{x}_m[N]$ $(m = 1,2,\ldots,M)$, we update the parameter estimates according to[10]

$$\hat{\boldsymbol{\theta}}_{0,m}^{(N)} = \arg\max_{\boldsymbol{\theta}_{0,m}\in\Theta_{0,m}} \sum_{n=1}^{N} \ln p_{0,m}(\mathbf{x}_m[n];\boldsymbol{\theta}_{0,m}) \tag{20}$$

$$\hat{\boldsymbol{\theta}}_{1,m}^{(N)} = \arg\max_{\boldsymbol{\theta}_{1,m}\in\Theta_{1,m}} \sum_{n=1}^{N} \ln p_{1,m}(\mathbf{x}_m[n];\boldsymbol{\theta}_{1,m}) \tag{21}$$

and perform the following sequential test:

$$\begin{aligned}&\text{Accept } \mathcal{H}_1 \text{ and terminate if } \text{GLLR}_N \ge A_N\\ &\text{Accept } \mathcal{H}_0 \text{ and terminate if } \text{GLLR}_N \le B_N\\ &\text{Take one more sample to}\\ &\quad\text{repeat the test if } B_N < \text{GLLR}_N < A_N\end{aligned} \tag{22}$$

where GLLR_N is the generalized log-likelihood ratio based on the parameter estimates obtained from the N samples, i.e.,

$$\text{GLLR}_N = \sum_{n=1}^{N} \sum_{m=1}^{M} \ln\left(\frac{p_{1,m}\left(\mathbf{x}_m[n];\hat{\boldsymbol{\theta}}_{1,m}^{(N)}\right)}{p_{0,m}\left(\mathbf{x}_m[n];\hat{\boldsymbol{\theta}}_{0,m}^{(N)}\right)}\right),$$

A_N and B_N are predefined thresholds for stopping the test. Compared to the test defined in (3), where the thresholds A and B are constant, A_N and B_N are proposed to be functions of

[9]Throughout the paper, it is assumed that the parameter spaces $\Theta_{0,m}$ and $\Theta_{1,m}$ are known to the sensing algorithm.

[10]In the proposed sequential sensing algorithm, we use (20) and (21) to estimate the unknown parameters for every acquired sample, which is computationally expensive if the stopping condition requires a large amount of samples. As such, if they can be implemented by recursive algorithms, i.e., simply updating previous estimate at new samples, it will definitely help to reduce the computational cost. For instance, in Examples 3 and 4, (20) and (21) will be implemented recursively.

$$\mathbf{E}_{\mathcal{H}_0}\{N_{\rm stop}\} \approx \frac{-\left(A-B-Ae^{\lambda_0 B}+Be^{\lambda_0 A}\right)}{\left(e^{\lambda_0 A}-e^{\lambda_0 B}\right)\left[\sum_{m=1}^{M} D\left(p_{0,m}(\mathbf{x}_m[n])\,\|\,q_{1,m}(\mathbf{x}_m[n])\right)-\sum_{m=1}^{M} D\left(p_{0,m}(\mathbf{x}_m[n])\,\|\,q_{0,m}(\mathbf{x}_m[n])\right)\right]} \tag{17}$$

$$\mathbf{E}_{\mathcal{H}_1}\{N_{\rm stop}\} \approx \frac{A-B-Ae^{\lambda_1 B}+Be^{\lambda_1 A}}{\left(e^{\lambda_1 A}-e^{\lambda_1 B}\right)\left[\sum_{m=1}^{M} D\left(p_{1,m}(\mathbf{x}_m[n])\,\|\,q_{0,m}(\mathbf{x}_m[n])\right)-\sum_{m=1}^{M} D\left(p_{1,m}(\mathbf{x}_m[n])\,\|\,q_{1,m}(\mathbf{x}_m[n])\right)\right]}. \tag{18}$$

N. This is because the estimates $\hat{\boldsymbol{\theta}}_{1,m}^{(N)}$ and $\hat{\boldsymbol{\theta}}_{0,m}^{(N)}$ introduce estimation errors into the test, which will degrade the performance if A_N and B_N are set according to (6) or (7) as constants. To compensate for this effect, we choose appropriate functions A_N and B_N of N as the stopping thresholds. Intuitively, the average estimation error tends to decrease with the increase of N, and hence A_N and B_N should be functions of N to reflect this trend. When N is small, A_N and B_N should be set to tolerate large estimation errors.

The cooperative sensing algorithm is summarized in Algorithm 2. For the newly acquired samples $\mathbf{x}_m[N]$ $(m = 1, 2, \ldots, M)$ at time N, the radios update the parameter estimates and the base station recomputes the sequential GLLR_N. Then the GLLR_N is compared with A_N and B_N to decide whether to terminate or continue the test.

Algorithm 2: Cooperative Sequential Sensing for Composite Hypotheses

0: Set $N = 0$.
1: **repeat**
2: $N = N + 1$.
3: The mth $(m = 1, 2, \ldots, M)$ radio acquires sample $\mathbf{x}_m[N]$ and estimates the unknown parameters $\boldsymbol{\theta}_{0,m}$ and $\boldsymbol{\theta}_{1,m}$ by using (20) and (21). Since it is unknown that which hypothesis is true, expression (20) assumes that $\mathcal{H}_0$ is true and obtains $\hat{\boldsymbol{\theta}}_{0,m}^{(N)}$, and expression (21) assumes that $\mathcal{H}_1$ is true and obtains $\hat{\boldsymbol{\theta}}_{1,m}^{(N)}$.
4: Each radio computes its own generalized sequential log-likelihood ratio $\mathrm{GLLR}_{N,m}$ by

$$\mathrm{GLLR}_{N,m} = \sum_{n=1}^{N} \ln \left(\frac{p_{1,m}\left(\mathbf{x}_m[n]; \hat{\boldsymbol{\theta}}_{1,m}^{(N)}\right)}{p_{0,m}\left(\mathbf{x}_m[n]; \hat{\boldsymbol{\theta}}_{0,m}^{(N)}\right)} \right)$$

and sends the result to the base station.
5: The base station computes the generalized sequential log-likelihood ratio GLLR_N according to

$$\mathrm{GLLR}_N = \sum_{m=1}^{M} \mathrm{GLLR}_{N,m}.$$

6: **until** $\mathrm{GLLR}_N \geq A_N$ or $\mathrm{GLLR}_N \leq B_N$.
7: If $\mathrm{GLLR}_N \geq A_N$, "$\mathcal{H}_1$: target signal is present" is claimed; if $\mathrm{GLLR}_N \leq B_N$, "$\mathcal{H}_0$: target signal is absent" is claimed.

We now explore how to choose appropriate A_N and B_N for achieving the desired sensing objective. Let

$$\overline{\mathrm{GLLR}}_N^{(0)} = \sum_{n=1}^{N} \sum_{m=1}^{M} \ln \left(\frac{p_{1,m}(\mathbf{x}_m[n]; \tilde{\boldsymbol{\theta}}_{1,m})}{p_{0,m}\left(\mathbf{x}_m[n]; \boldsymbol{\theta}_{0,m}\right)} \right)$$

$$\overline{\mathrm{GLLR}}_N^{(1)} = \sum_{n=1}^{N} \sum_{m=1}^{M} \ln \left(\frac{p_{1,m}\left(\mathbf{x}_m[n]; \boldsymbol{\theta}_{1,m}\right)}{p_{0,m}(\mathbf{x}_m[n]; \tilde{\boldsymbol{\theta}}_{0,m})} \right)$$

where $\tilde{\boldsymbol{\theta}}_{1,m}$ and $\tilde{\boldsymbol{\theta}}_{0,m}$ are defined as[11]

$$\tilde{\boldsymbol{\theta}}_{1,m} = \underset{\bar{\boldsymbol{\theta}}_{1,m} \in \Theta_{1,m}}{\arg\min}\ D(p_{0,m}(\mathbf{x}_m[n]; \boldsymbol{\theta}_{0,m}) \,\|\, p_{1,m}(\mathbf{x}_m[n]; \bar{\boldsymbol{\theta}}_{1,m}))$$

$$\tilde{\boldsymbol{\theta}}_{0,m} = \underset{\bar{\boldsymbol{\theta}}_{0,m} \in \Theta_{0,m}}{\arg\min}\ D(p_{1,m}(\mathbf{x}_m[n]; \boldsymbol{\theta}_{1,m}) \,\|\, p_{0,m}(\mathbf{x}_m[n]; \bar{\boldsymbol{\theta}}_{0,m})).$$

Note that $\overline{\mathrm{GLLR}}_N^{(0)}$ is given by GLLR_N with $\hat{\boldsymbol{\theta}}_{0,m}^{(N)}$ and $\hat{\boldsymbol{\theta}}_{1,m}^{(N)}$ changed to $\boldsymbol{\theta}_{0,m}$ and $\tilde{\boldsymbol{\theta}}_{1,m}$ respectively, and $\overline{\mathrm{GLLR}}_N^{(1)}$ is given by GLLR_N with $\hat{\boldsymbol{\theta}}_{0,m}^{(N)}$ and $\hat{\boldsymbol{\theta}}_{1,m}^{(N)}$ changed to $\tilde{\boldsymbol{\theta}}_{0,m}$ and $\boldsymbol{\theta}_{1,m}$ respectively. It is easy to see that

$$\overline{\mathrm{GLLR}}_N^{(0)} = \mathrm{GLLR}_N + \Delta\mathrm{GLLR}_N^{(0)} \tag{23}$$

$$\overline{\mathrm{GLLR}}_N^{(1)} = \mathrm{GLLR}_N + \Delta\mathrm{GLLR}_N^{(1)} \tag{24}$$

where

$$\Delta\mathrm{GLLR}_N^{(0)} = \sum_{n=1}^{N} \sum_{m=1}^{M} \ln \left(\frac{p_{1,m}(\mathbf{x}_m[n]; \tilde{\boldsymbol{\theta}}_{1,m})}{p_{1,m}\left(\mathbf{x}_m[n]; \hat{\boldsymbol{\theta}}_{1,m}^{(N)}\right)} \right) + \sum_{n=1}^{N} \sum_{m=1}^{M} \ln \left(\frac{p_{0,m}\left(\mathbf{x}_m[n]; \hat{\boldsymbol{\theta}}_{0,m}^{(N)}\right)}{p_{0,m}(\mathbf{x}_m[n]; \boldsymbol{\theta}_{0,m})} \right)$$

$$\Delta\mathrm{GLLR}_N^{(1)} = \sum_{n=1}^{N} \sum_{m=1}^{M} \ln \left(\frac{p_{1,m}(\mathbf{x}_m[n]; \boldsymbol{\theta}_{1,m})}{p_{1,m}\left(\mathbf{x}_m[n]; \hat{\boldsymbol{\theta}}_{1,m}^{(N)}\right)} \right) + \sum_{n=1}^{N} \sum_{m=1}^{M} \ln \left(\frac{p_{0,m}\left(\mathbf{x}_m[n]; \hat{\boldsymbol{\theta}}_{0,m}^{(N)}\right)}{p_{0,m}(\mathbf{x}_m[n]; \tilde{\boldsymbol{\theta}}_{0,m})} \right).$$

In the above two expressions, we will see later that under some mild conditions, $\Delta\mathrm{GLLR}_N^{(0)}$ and $\Delta\mathrm{GLLR}_N^{(1)}$ can be asymptotically bounded in some sense. For example, the maximum likelihood estimation is in general asymptotically consistent and efficient, implying the following:

1) under $\mathcal{H}_0$, $\hat{\boldsymbol{\theta}}_{0,m}^{(N)}$ converges to $\boldsymbol{\theta}_{0,m}$ and $\hat{\boldsymbol{\theta}}_{1,m}^{(N)}$ converges to $\tilde{\boldsymbol{\theta}}_{1,m}$ in probability;
2) under $\mathcal{H}_1$, $\hat{\boldsymbol{\theta}}_{0,m}^{(N)}$ converges to $\tilde{\boldsymbol{\theta}}_{0,m}$ and $\hat{\boldsymbol{\theta}}_{1,m}^{(N)}$ converges to $\boldsymbol{\theta}_{1,m}$ in probability.

Therefore, we can view GLLR_N as the estimate of $\overline{\mathrm{GLLR}}_N^{(0)}$ under $\mathcal{H}_0$ with error term $\Delta\mathrm{GLLR}_N^{(0)}$ and as the estimate of $\overline{\mathrm{GLLR}}_N^{(1)}$ under $\mathcal{H}_1$ with error term $\Delta\mathrm{GLLR}_N^{(1)}$. Given the following regularity conditions[12]:

3a) $\Pr(N_{\mathrm{stop}} < \infty \,|\, \mathcal{H}_i) = 1$ and $\mathbf{E}_{\mathcal{H}_i}\{N_{\mathrm{stop}}\} < \infty$ for $i = 0, 1$;

[11]Throughout the paper, we assume that there exists only one $(\bar{\boldsymbol{\theta}}_{0,m}, \bar{\boldsymbol{\theta}}_{1,m})$ for each true parameter pair $(\boldsymbol{\theta}_{0,m}, \boldsymbol{\theta}_{1,m})$.

[12]Unlike in previous discussion, assumptions 3b) and 3c) do not imply assumption 3a), because the stopping thresholds A_N and B_N are in general not constant. However, it will be shown in later discussion that with mild regularity conditions, asymptotically A_N and B_N can be set as constants, for which 3a) is true, if provided that 3b) and 3c) are true.

3b) the mean of

$$\sum_{m=1}^{M} \ln(p_{0,m}(\mathbf{x}_m[n];\boldsymbol{\theta}_{0,m})/p_{1,m}(\mathbf{x}_m[n];\tilde{\boldsymbol{\theta}}_{1,m}))$$

under $\mathcal{H}_0$ is positive and finite;

3c) the mean of

$$\sum_{m=1}^{M} \ln(p_{1,m}(\mathbf{x}_m[n];\boldsymbol{\theta}_{1,m})/p_{0,m}(\mathbf{x}_m[n];\tilde{\boldsymbol{\theta}}_{0,m}))$$

under $\mathcal{H}_1$ is positive and finite

the following lemma implies that if the test stops with a negligible probability (close to zero) for the event $\{\mathrm{GLLR}_{N_{\mathrm{stop}}} \geq A_{N_{\mathrm{stop}}}, \overline{\mathrm{GLLR}}^{(0)}_{N_{\mathrm{stop}}} < A\}$ under $\mathcal{H}_0$ and the event $\{\mathrm{GLLR}_{N_{\mathrm{stop}}} \leq B_{N_{\mathrm{stop}}}, \overline{\mathrm{GLLR}}^{(1)}_{N_{\mathrm{stop}}} > B\}$ under $\mathcal{H}_1$, the sensing objective (2) can be achieved.

Lemma 3.1: Assume that the sequential test (22) stops at N_{stop}. If the stopping thresholds A_N and B_N are set such that for any $0 < \epsilon < 1$,

$$\Pr\left(\mathrm{GLLR}_{N_{\mathrm{stop}}} \geq A_{N_{\mathrm{stop}}}, \overline{\mathrm{GLLR}}^{(0)}_{N_{\mathrm{stop}}} < A \middle| \mathcal{H}_0\right) \leq \epsilon \quad (25)$$

$$\Pr\left(\mathrm{GLLR}_{N_{\mathrm{stop}}} \leq B_{N_{\mathrm{stop}}}, \overline{\mathrm{GLLR}}^{(1)}_{N_{\mathrm{stop}}} > B \middle| \mathcal{H}_1\right) \leq \epsilon \quad (26)$$

where A and B are given by (7), then

$$P_{\mathrm{FA}} \leq \alpha + \epsilon \quad \text{and} \quad P_{\mathrm{MISS}} \leq \beta + \epsilon.$$

Proof: See Appendix G. ∎

By expressions (23) and (24), if we can bound the term $\Delta\mathrm{GLLR}^{(i)}_N$ under $\mathcal{H}_i$ for $i = 0, 1$ properly, we may find appropriate A_N and B_N to meet the conditions (25) and (26) required by Lemma 3.1. The following theorem states that this is feasible if the second moment of $\Delta\mathrm{GLLR}^{(i)}_N$ is finite under $\mathcal{H}_i$ for $i = 0, 1$.

Theorem 3.2: For the sensing algorithm defined in (22), assume that the second moment of $\Delta\mathrm{GLLR}^{(0)}_N$ under $\mathcal{H}_0$ and the second moment of $\Delta\mathrm{GLLR}^{(1)}_N$ under $\mathcal{H}_1$ are finite for all $N = 1, 2, \ldots$. Let

$$\gamma^{(i)}_N = \mathbf{E}_{\mathcal{H}_i}\left\{\left(\Delta\mathrm{GLLR}^{(i)}_N\right)^2\right\}, \quad i = 0, 1.$$

For any $0 < \epsilon < 1$, if the stopping thresholds are set as

$$A_N = A + \sqrt{\frac{\gamma^{(0)}_N}{\epsilon}} \quad \text{and} \quad B_N = B - \sqrt{\frac{\gamma^{(1)}_N}{\epsilon}}$$

where A and B are given by (7), then

$$P_{\mathrm{FA}} \leq \alpha + \epsilon \quad \text{and} \quad P_{\mathrm{MISS}} \leq \beta + \epsilon.$$

Proof: See Appendix H. ∎

By the above theorem, if we can obtain an upper bound for the second moment of $\Delta\mathrm{GLLR}^{(i)}_N$ under $\mathcal{H}_i$ for $i = 0, 1$, then we have a systematic way to set the stopping conditions to achieve the desired sensing performance.[13]

[13]The A_N and B_N suggested by Theorem 3.2 are usually too conservative and require a large number of samples to meet the objective. In practice, we can sometimes get tighter stopping conditions.

C. Asymptotic Upper Bound for the Second Moment of $\Delta\mathrm{GLLR}^{(i)}_N$

Theorem 3.2 requires an upper bound for the second moment of $\Delta\mathrm{GLLR}^{(i)}_N$ under $\mathcal{H}_i$ for $i = 0, 1$ and every N, which is normally too complicated to compute. In the following, we consider obtaining such a bound in the *asymptotic* sense, i.e., when N is large. This is motivated by two reasons. First, it is easier to compute such an asymptotic bound since the maximum likelihood estimator is asymptotically consistent and efficient; second, when α and β are small, there is a high chance that the test stops at large N, and the asymptotic bound provides a good approximation to actual bound in this case. To begin with, we note that under some mild regularity conditions, the maximum likelihood estimator is unbiased, asymptotically consistent and efficient [25], implying that (see Appendix I for a brief explanation):

4a) under $\mathcal{H}_0$, as $N \to \infty$, $\hat{\boldsymbol{\theta}}^{(N)}_{0,m}$ converges to $\boldsymbol{\theta}_{0,m}$ and $\hat{\boldsymbol{\theta}}^{(N)}_{1,m}$ converges to $\tilde{\boldsymbol{\theta}}_{1,m}$ in probability, and

$$\mathbf{E}_{\mathcal{H}_0}\left\{\left\|\hat{\boldsymbol{\theta}}^{(N)}_{0,m} - \boldsymbol{\theta}_{0,m}\right\|^2\right\} = O\left(\frac{1}{N}\right)$$

$$\mathbf{E}_{\mathcal{H}_0}\left\{\left\|\hat{\boldsymbol{\theta}}^{(N)}_{1,m} - \tilde{\boldsymbol{\theta}}_{1,m}\right\|^2\right\} = O\left(\frac{1}{N}\right);$$

4b) under $\mathcal{H}_1$, as $N \to \infty$, $\hat{\boldsymbol{\theta}}^{(N)}_{0,m}$ converges to $\tilde{\boldsymbol{\theta}}_{0,m}$ and $\hat{\boldsymbol{\theta}}^{(N)}_{1,m}$ converges to $\boldsymbol{\theta}_{1,m}$ in probability, and

$$\mathbf{E}_{\mathcal{H}_1}\left\{\left\|\hat{\boldsymbol{\theta}}^{(N)}_{0,m} - \tilde{\boldsymbol{\theta}}_{0,m}\right\|^2\right\} = O\left(\frac{1}{N}\right)$$

$$\mathbf{E}_{\mathcal{H}_1}\left\{\left\|\hat{\boldsymbol{\theta}}^{(N)}_{1,m} - \boldsymbol{\theta}_{1,m}\right\|^2\right\} = O\left(\frac{1}{N}\right).$$

$O(\cdot)$ is the Big-O notation.[14]

Note that $\Delta\mathrm{GLLR}^{(i)}_N$ $(i = 0, 1)$ are given by a summation over terms of the following form:

$$\pm \ln\left(\frac{p\left(\mathbf{x}[n];\hat{\boldsymbol{\theta}}^{(N)}\right)}{p(\mathbf{x}[n];\boldsymbol{\theta})}\right)$$

where $\hat{\boldsymbol{\theta}}^{(N)} \to \boldsymbol{\theta}$ in probability and $\mathbf{E}\{\|\hat{\boldsymbol{\theta}}^{(N)} - \tilde{\boldsymbol{\theta}}\|^2\} = O(1/N)$ as $N \to \infty$. It can be shown by using the Taylor series expansion that (see Appendix J for the derivation):

$$\mu^{(N)} = \mathbf{E}\left\{\ln\left(\frac{p(\mathbf{x}[n];\hat{\boldsymbol{\theta}}^{(N)})}{p\left(\mathbf{x}[n];\boldsymbol{\theta}\right)}\right)\right\} = O\left(\frac{1}{N}\right) \quad (27)$$

$$\lambda^{(N)} = \mathbf{E}\left\{\left(\ln\left(\frac{p(\mathbf{x}[n];\hat{\boldsymbol{\theta}}^{(N)})}{p\left(\mathbf{x}[n];\boldsymbol{\theta}\right)}\right)\right)^2\right\} = O\left(\frac{1}{N}\right). \quad (28)$$

[14]That is, $f(x) = O(g(x))$ as $x \to 0$ iff there exist positive numbers δ and T such that $|f(x)| \leq T|g(x)|$ for all $|x| < \delta$.

Since $\mathbf{x}[n]$ $(n = 1, 2, \ldots)$ are i.i.d., we approximately treat $\ln(p(\mathbf{x}[n]; \hat{\boldsymbol{\theta}}^{(N)})/p(\mathbf{x}[n]; \boldsymbol{\theta}))$ $(n = 1, 2, \ldots)$ as independent. Thus,

$$\begin{aligned}
&\mathbf{E}\left\{\left(\sum_{n=1}^{N} \ln\left(\frac{p(\mathbf{x}[n]; \hat{\boldsymbol{\theta}}^{(N)})}{p(\mathbf{x}[n]; \boldsymbol{\theta})}\right)\right)^2\right\} \\
&\quad \approx \left(\sum_{n=1}^{N} \mu^{(N)}\right)^2 + \sum_{n=1}^{N}\left[\lambda^{(N)} - \left(\mu^{(N)}\right)^2\right] \\
&\quad = N^2\left(\mu^{(N)}\right)^2 + N\left[\lambda^{(N)} - \left(\mu^{(N)}\right)^2\right] \\
&\quad = O(1).
\end{aligned}$$

This eventually implies

$$\mathbf{E}_{\mathcal{H}_i}\left\{\left(\Delta\mathrm{GLLR}_N^{(i)}\right)^2\right\} = O(1), \quad i = 0, 1.$$

That is, $\mathbf{E}\{(\Delta\mathrm{GLLR}_N^{(i)})^2\}$ $(i = 0, 1)$ are asymptotically bounded by a constant. This enables us to use Theorem 3.2 as a systematic approach to achieve the sensing objective. That is, we can set the stopping thresholds as

$$A_N = A + C_A \quad \text{and} \quad B_N = B - C_B \tag{29}$$

for some positive constants C_A and C_B. In this case, since A_N and B_N are finite constants and $\mathbf{E}\{(\Delta\mathrm{GLLR}_N^{(i)})^2\}$ $(i = 0, 1)$ are asymptotically bounded, the test must stop with probability 1 and assumption 3a) holds.

D. Complexity of the Proposed Algorithm

As a drawback, Algorithm 2 has higher complexity than the methods with fixed sample size. Assume that the method with fixed sample size needs N_{fix} samples to achieve the sensing objective. It then has computational complexity $O(MN_{\mathrm{fix}})$, since the parameter estimates $\hat{\boldsymbol{\theta}}_{0,m}$, $\hat{\boldsymbol{\theta}}_{1,m}$, and the GLLR are only computed once and the computational cost is normally proportional to the total number of samples MN_{fix}. To achieve the same sensing objective, Algorithm 2 has complexity $O(MN_{\mathrm{fix}}^2)$, because $\hat{\boldsymbol{\theta}}_{0,m}^{(N)}$, $\hat{\boldsymbol{\theta}}_{1,m}^{(N)}$, and the GLLR_N need to be computed for every $N$and the average number of samples is normally proportional to N_{fix}.

In Algorithm 2, the computation is carried out by all the radios, and each radio has the complexity $O(N_{\mathrm{fix}}^2)$. In practice, alternatively we can perform all the computation at the base station that usually has more computational power. In certain situations where the complexity of Algorithm 2 prevents it from being practical, e.g., in low-power sensor networks, our next goal is to seek suboptimal (or approximate) algorithms with less complexity.

In the following, we use two spectrum sensing examples to illustrate the idea.

Example 3—Detecting a Gaussian Random Signal with Unknown Variances: In this example, the signal model is identical to that in Example 1, except that $\sigma_{1,m}^2$ and $\sigma_{0,m}^2$ are unknown to the receivers. Assume that the signal and noise variances have different range:

$$\begin{aligned}
\mathcal{H}_0 &: \nu_{0,m}^{\mathrm{lower}} \le \sigma_{0,m}^2 \le \nu_{0,m}^{\mathrm{upper}} \\
\mathcal{H}_1 &: \nu_{1,m}^{\mathrm{lower}} \le \sigma_{1,m}^2 \le \nu_{1,m}^{\mathrm{upper}}
\end{aligned}$$

where $\nu_{0,m}^{\mathrm{lower}}$ and $\nu_{0,m}^{\mathrm{upper}}$ are the lower and upper bounds for $\sigma_{0,m}^2$, $\nu_{1,m}^{\mathrm{lower}}$ and $\nu_{1,m}^{\mathrm{upper}}$ are the lower and upper bounds for $\sigma_{1,m}^2$, and $\nu_{0,m}^{\mathrm{upper}} \le \nu_{1,m}^{\mathrm{lower}}$. The generalized likelihood ratio test with a fixed sample size is given by

$$\begin{aligned}
\text{Accept } \mathcal{H}_1 \text{ if } &\sum_{n=1}^{N_{\mathrm{fix}}}\sum_{m=1}^{M}\left[\frac{1}{2}\left(\frac{1}{\hat{\sigma}_{0,m}^2} - \frac{1}{\hat{\sigma}_{1,m}^2}\right)(X_m[n])^2\right. \\
&\left. + \frac{1}{2}\ln\left(\frac{\hat{\sigma}_{0,m}^2}{\hat{\sigma}_{1,m}^2}\right)\right] > \eta \\
\text{Accept } \mathcal{H}_0 \text{ if } &\sum_{n=1}^{N_{\mathrm{fix}}}\sum_{m=1}^{M}\left[\frac{1}{2}\left(\frac{1}{\hat{\sigma}_{0,m}^2} - \frac{1}{\hat{\sigma}_{1,m}^2}\right)(X_m[n])^2\right. \\
&\left. + \frac{1}{2}\ln\left(\frac{\hat{\sigma}_{0,m}^2}{\hat{\sigma}_{1,m}^2}\right)\right] \le \eta
\end{aligned} \tag{30}$$

where the maximum likelihood estimates of $\sigma_{0,m}^2$ and $\sigma_{1,m}^2$ are given by

$$\begin{aligned}
\hat{\sigma}_{0,m}^2 &= \max\left\{\min\left\{\frac{1}{N_{\mathrm{fix}}}\sum_{n=1}^{N_{\mathrm{fix}}}(X_m[n])^2, \nu_{0,m}^{\mathrm{upper}}\right\}, \nu_{0,m}^{\mathrm{lower}}\right\} \\
\hat{\sigma}_{1,m}^2 &= \max\left\{\min\left\{\frac{1}{N_{\mathrm{fix}}}\sum_{n=1}^{N_{\mathrm{fix}}}(X_m[n])^2, \nu_{1,m}^{\mathrm{upper}}\right\}, \nu_{1,m}^{\mathrm{lower}}\right\}.
\end{aligned}$$

As illustrated by Example 1 in Section II, if we have perfect knowledge of $\sigma_{0,m}^2$ and $\sigma_{1,m}^2$, the minimum number N_{fix} of required samples and the associated η are given by expressions (10) and (11). However, since $\sigma_{0,m}^2$ and $\sigma_{1,m}^2$ are unknown in this example, the optimal N_{fix} and η should be valid for all possible signal and noise variances, and can be found from the worst-case scenario, i.e., when $\sigma_{0,m}^2 = \nu_{0,m}^{\mathrm{upper}}$ and $\sigma_{1,m}^2 = \nu_{1,m}^{\mathrm{lower}}$. In other words, the optimal N_{fix} is the minimum sample size for achieving $P_{\mathrm{FA}} \le \alpha$ and $P_{\mathrm{MISS}} \le \beta$ when $\sigma_{0,m}^2 = \nu_{0,m}^{\mathrm{upper}}$ and $\sigma_{1,m}^2 = \nu_{1,m}^{\mathrm{lower}}$ for all $m = 1, 2, \ldots, M$.

The proposed sequential scheme is a natural solution to this type of problem because the required sensing time depends on the signals and the associated parameter estimates. In the test (22), we have

$$\begin{aligned}
\mathrm{GLLR}_N = \sum_{n=1}^{N}\sum_{m=1}^{M}&\left[\frac{1}{2}\left(\frac{1}{(\hat{\sigma}_{0,m}^{(N)})^2} - \frac{1}{(\hat{\sigma}_{1,m}^{(N)})^2}\right)(X_m[n])^2\right. \\
&\left. + \frac{1}{2}\ln\left(\frac{(\hat{\sigma}_{0,m}^{(N)})^2}{(\hat{\sigma}_{1,m}^{(N)})^2}\right)\right]
\end{aligned}$$

where

$$\begin{aligned}
\left(\hat{\sigma}_{0,m}^{(N)}\right)^2 &= \max\left\{\min\left\{\frac{1}{N}\sum_{n=1}^{N}(X_m[n])^2, \nu_{0,m}^{\mathrm{upper}}\right\}, \nu_{0,m}^{\mathrm{lower}}\right\} \\
(\hat{\sigma}_{1,m}^{(N)})^2 &= \max\left\{\min\left\{\frac{1}{N}\sum_{n=1}^{N}(X_m[n])^2, \nu_{1,m}^{\mathrm{upper}}\right\}, \nu_{1,m}^{\mathrm{lower}}\right\}.
\end{aligned}$$

In the above, the sum $\sum_{n=1}^{N}(X_m[n])^2$ can be updated recursively at each new sample to save the complexity. Let $\theta_{i,m} = \sigma_{i,m}^2$ $(i=0,1)$, $\tilde{\theta}_{1,m} = \nu_{1,m}^{\text{lower}}$, and $\tilde{\theta}_{0,m} = \nu_{0,m}^{\text{upper}}$. It is easy to verify that under $\mathcal{H}_0$,

$$\hat{\theta}_{0,m}^{(N)} \to \theta_{0,m} \quad \text{and} \quad \hat{\theta}_{1,m}^{(N)} \to \tilde{\theta}_{1,m}$$

in probability, and

$$\mathbf{E}_{\mathcal{H}_0}\left\{\left(\hat{\theta}_{0,m}^{(N)} - \theta_{0,m}\right)^2\right\} = \frac{2}{N}\theta_{0,m}^2$$
$$\mathbf{E}_{\mathcal{H}_0}\left\{\left(\hat{\theta}_{1,m}^{(N)} - \tilde{\theta}_{1,m}\right)^2\right\} \le \frac{2}{N}\theta_{0,m}^2.$$

Similarly, under $\mathcal{H}_1$ we have

$$\hat{\theta}_{0,m}^{(N)} \to \tilde{\theta}_{0,m} \quad \text{and} \quad \hat{\theta}_{1,m}^{(N)} \to \theta_{1,m}$$

in probability, and

$$\mathbf{E}_{\mathcal{H}_1}\left\{\left(\hat{\theta}_{0,m}^{(N)} - \tilde{\theta}_{0,m}\right)^2\right\} \le \frac{2}{N}\theta_{1,m}^2$$
$$\mathbf{E}_{\mathcal{H}_1}\left\{\left(\hat{\theta}_{1,m}^{(N)} \to \theta_{1,m}\right)^2\right\} = \frac{2}{N}\theta_{1,m}^2.$$

Following from previous discussion, we set the stopping conditions as (29), where A and B are given by (7) and C_A, C_B are some positive constants. Theorem 3.2 advises on how to select proper C_A and C_B; however, the C_A and C_B obtained from Theorem 3.2 are normally too large and require too many samples. In our simulations, we obtain the values for C_A and C_B through computer experiments.

Example 4—Detecting a Periodic Signal with Unknown Amplitude: The signal model in this example is identical to that in Example 2, except that the signal amplitude a_m are unknown to the receivers. Assume that

$$0 < a_m^{\text{lower}} \le a_m \le a_m^{\text{upper}}$$

where a_m^{lower} and a_m^{upper} are the lower and upper bounds for a_m. The generalized likelihood ratio test with a fixed sample size is given by

$$\text{Accept } \mathcal{H}_1 \text{ if } \sum_{n=1}^{N_{\text{fix}}}\sum_{k=1}^{K}\sum_{m=1}^{M}\left[\frac{\hat{a}_m}{\sigma_m^2}s[k]X_m[(n-1)K+k] - \frac{(\hat{a}_m)^2}{2\sigma_m^2}(s[k])^2\right] > \eta$$
$$\text{Accept } \mathcal{H}_0 \text{ if } \sum_{n=1}^{N_{\text{fix}}}\sum_{k=1}^{K}\sum_{m=1}^{M}\left[\frac{\hat{a}_m}{\sigma_m^2}s[k]X_m[(n-1)K+k] - \frac{(\hat{a}_m)^2}{2\sigma_m^2}(s[k])^2\right] \le \eta \qquad (31)$$

where

$$\hat{a}_m = \max\left\{\min\left\{\frac{\sum_{n=1}^{N_{\text{fix}}}\sum_{k=1}^{K}s[k]X_m[(n-1)K+k]}{N_{\text{fix}}\left(\sum_{k=1}^{K}(s[k])^2\right)}, a_m^{\text{upper}}\right\}, a_m^{\text{lower}}\right\}.$$

The optimal N_{fix} and its associated η are selected such that the resulting P_{FA} and P_{MISS} are no greater than α and β, respectively, if $a_m = a_m^{\text{lower}}$ for all $m = 1, 2, \ldots, M$.

In the proposed sequential test (22), we have

$$\text{GLLR}_N = \sum_{n=1}^{N}\sum_{k=1}^{K}\sum_{m=1}^{M}\left[\frac{\hat{a}_m^{(N)}}{\sigma_m^2}s[k]X_m[(n-1)K+k] - \frac{(\hat{a}_m^{(N)})^2}{2\sigma_m^2}(s[k])^2\right]$$

where

$$\hat{a}_m^{(N)} = \max\left\{\min\left\{\frac{\sum_{n=1}^{N}\sum_{k=1}^{K}s[k]X_m[(n-1)K+k]}{N\left(\sum_{k=1}^{K}(s[k])^2\right)}, a_m^{\text{upper}}\right\}, a_m^{\text{lower}}\right\}.$$

Similar to Example 3, we can calculate $\hat{a}_m^{(N)}$ by designing iterative procedure for saving the computational complexity. It is easy to check that under $\mathcal{H}_1$,

$$\hat{a}_m^{(N)} \to a_m, \quad \mathbf{E}_{\mathcal{H}_1}\left\{\left(\hat{a}_m^{(N)} - a_m\right)^2\right\} = \frac{\sigma_m^2}{N\left(\sum_{k=1}^{K}(s[k])^2\right)}$$

and under $\mathcal{H}_0$,

$$\hat{a}_m^{(N)} \to 0, \quad \mathbf{E}_{\mathcal{H}_0}\left\{\left(\hat{a}_m^{(N)}\right)^2\right\} \le \frac{\sigma_m^2}{N\left(\sum_{k=1}^{K}(s[k])^2\right)}.$$

The stopping conditions can then be set as (29). In the simulations, we obtain the values for C_A and C_B through computer experiments.

IV. Simulation Results

In this section, we simulate the proposed cooperative sequential sensing scheme and compare its performance with the conventional method with fixed sample size. The simulated network has four cognitive radios for spectrum sensing, i.e., $M = 4$. Four different sensing scenarios are considered, corresponding to the four examples presented in Sections II and III. In the first two examples, we assume that the exact signal and noise models are known to the radios, while in the last two examples, the assumed signal and noise models have unknown parameters.

In the first scenario, the target signal and observation noise are Gaussian distributed with zero mean. The received signal variances under hypotheses $\mathcal{H}_0$ and $\mathcal{H}_1$ are listed in Table I. The base station has full knowledge of the distributions of the signal and noise in terms of $\sigma_{1,m}^2$ and $\sigma_{0,m}^2$ $(m = 1, \ldots, 4)$. The sensing method with a fixed sample size uses the N_{fix} and η determined by expressions (10) and (11), while the sequential method uses the A and B determined by (7). In the simulations, the sensing objective is set according to (2) with $\alpha = \beta$. For different $\alpha = \beta$, Fig. 2 plots the average number of samples that is required to achieve $P_{\text{FA}} \le \alpha$ and $P_{\text{MISS}} \le \beta$. It is observed that the sequential method substantially reduces

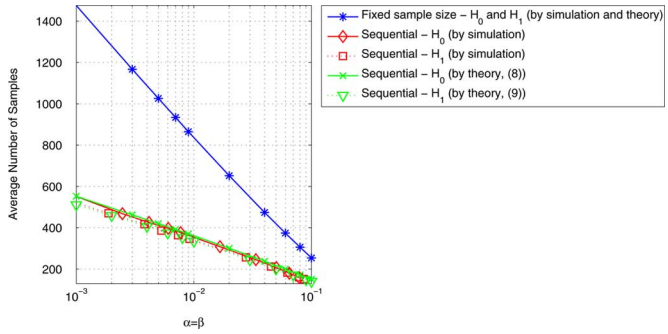

Fig. 2. Detection of zero-mean Gaussian signals with *known* variances. The signal variance $\sigma^2_{1,m}$ and noise variance $\sigma^2_{0,m}$ are listed in Table I. The sensing method with a fixed sample size uses the $N_{\rm fix}$ and η determined by (10) and (11). The parameters A and B in the sequential method are determined by (7).

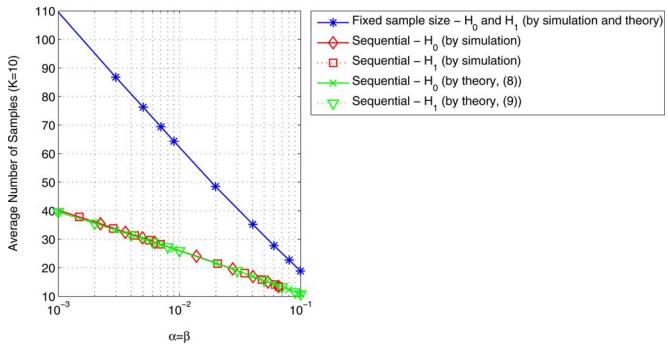

Fig. 3. Detection of a sinusoidal signal with *known* amplitude in white Gaussian noise. The signal amplitude a_m and noise variance σ^2_m are listed in Table II. The sensing method with a fixed sample size uses the $N_{\rm fix}$ and η determined by (12) and (13). The parameters A and B in the sequential method are determined by (7).

TABLE I
SIMULATED SIGNAL AND NOISE VARIANCES IN FIGS. 2 AND 4

	m=1	m=2	m=3	m=4
$\sigma^2_{0,m}$	1.00	0.90	0.80	0.70
$\sigma^2_{1,m}$	1.10	1.00	0.90	0.80
$\nu^{\rm lower}_{0,m}$	0.10	0.10	0.10	0.10
$\nu^{\rm upper}_{0,m}$	1.025	0.925	0.825	0.725
$\nu^{\rm lower}_{1,m}$	1.075	0.975	0.875	0.775
$\nu^{\rm upper}_{1,m}$	2.00	2.00	2.00	2.00

TABLE II
SIMULATED SINUSOIDAL AMPLITUDE AND NOISE VARIANCE IN FIGS. 3 AND 5

	m=1	m=2	m=3	m=4
a_m	0.10	0.12	0.14	0.16
σ^2_m	1.00	1.00	1.00	1.00
$a^{\rm lower}_m$	0.05	0.05	0.05	0.05
$a^{\rm upper}_m$	0.20	0.20	0.20	0.20

the sensing time. The second scenario considers the detection of a sinusoidal signal in white Gaussian noise. The normalized signal frequency is 0.1 and thus $K = 10$. The received signal amplitude a_m and noise variances σ^2_m are listed in Table II, and the simulation results are shown in Fig. 3. Note that each "sample" in this scenario is a "vectorized" sample, consisting of $K = 10$ actual samples in a full period.

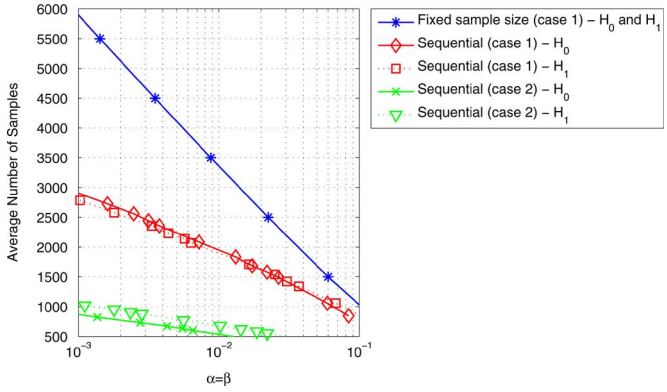

Fig. 4. Detection of zero-mean Gaussian signals with *unknown* variances. The simulated signal and noise variances are listed in Table I. In Case 1, the simulated signal and noise variances are $\nu^{\rm lower}_{1,m}$ and $\nu^{\rm upper}_{0,m}$, respectively; in Case 2, the simulated signal and noise variances are $\sigma^2_{1,m}$ and $\sigma^2_{0,m}$, respectively. The sensing method with a fixed sample size uses the $N_{\rm fix}$ and η determined from the worst case, i.e., Case 1. In the sequential method, A_N and B_N are determined by (29) with $C_A = 3.8$ and $C_B = 4.5$.

The third scenario assumes the same signal model as in the first scenario, but the base station has no exact information about the signal and noise variances. Prior knowledge about the range of the signal and noise variances is given in Table I, i.e., $\nu^{\rm lower}_{0,m} \le \sigma^2_{0,m} \le \nu^{\rm upper}_{0,m}$ and $\nu^{\rm lower}_{1,m} \le \sigma^2_{1,m} \le \nu^{\rm upper}_{1,m}$ for $m = 1, \ldots, 4$. The sensing method with a fixed sample size uses the test given by expression (30). Monte Carlo simulations are performed for various values of η to find the optimal η that requires minimum $N_{\rm fix}$ to achieve $P_{\rm FA} \le \alpha$ and $P_{\rm MISS} \le \beta$ ($\alpha = \beta$). In the sequential test, the stopping condition is given by (29), where $C_A = 3.8, C_B = 4.5$,[15] and A, B are set according to (7). Fig. 4 plots the average number of required samples versus $\alpha = \beta$ for the sequential method and the method with a fixed sample size. In Case 1, the actually simulated signal and noise variances are $\nu^{\rm lower}_{1,m}$ and $\nu^{\rm upper}_{0,m}$, respectively; in Case 2, the actually simulated signal and noise variances are $\sigma^2_{1,m}$ and $\sigma^2_{0,m}$, respectively. Case 1 represents the worst signal condition for spectrum sensing and normally requires more samples than Case 2. Here, the "worst case" refers to the signal condition that requires the maximum number of samples in order to achieve the sensing objectives for both $\mathcal{H}_0$ and $\mathcal{H}_1$, given that the detection thresholds or stopping rules have been optimized for each case. Obviously, Case 1 has the worst signal-to-noise ratio for detection. Since the method with a fixed sample size cannot distinguish between Case 1 and Case 2, the number of signal samples it requires should satisfy the requirement of Case 1, although it might be too many for Case 2. The proposed sequential method adjusts the number of samples adaptively for each case, and substantially saves the sensing time. The last scenario assumes the same signal model as in the second scenario, but the amplitude of the target sinusoidal signal is unknown. The lower and upper bounds of the received sinusoidal amplitude at each radio are given in Table II. The sensing method with a fixed sample size

[15]The expressions provided in Theorem 3.2 for determining C_A and C_B are usually computationally expensive, and the computed C_A and C_B are normally too large and require too many samples. In the simulations, we simply tried many different combinations for the values of C_A and C_B, and selected the one that meets the requirement, i.e., $P_{\rm FA} \le \alpha$ and $P_{\rm MISS} \le \beta$, with the smallest $\max\{\mathbf{E}_{\mathcal{H}_0}\{N_{\rm stop}\}, \mathbf{E}_{\mathcal{H}_1}\{N_{\rm stop}\}\}$ for the worst signal scenario.

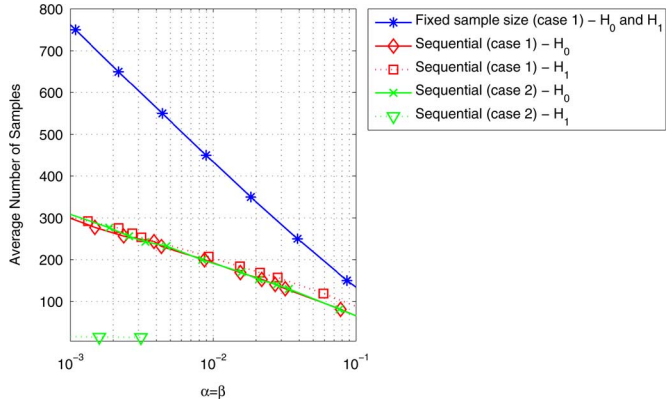

Fig. 5. Detection of a sinusoidal signal with *unknown* amplitude in white Gaussian noise. The signal amplitude and noise variance are listed in Table II. In Case 1, the simulated signal amplitude and noise variance are a_m^{lower} and σ_m^2, respectively; in Case 2, the simulated signal amplitude and noise variance are a_m and σ_m^2, respectively. The sensing method with a fixed sample size uses the N_{fix} and η determined from the worst case, i.e., Case 1. In the sequential method, A_N and B_N are determined by (29) with $C_A = 3.0$ and $C_B = 1.0$.

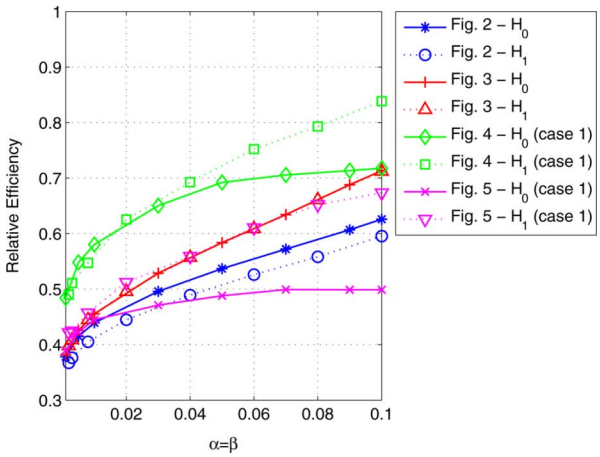

Fig. 6. Relative efficiency of the sequential algorithm for the simulation results presented in Figs. 2–5. Relative efficiency is defined as the ratio between the average number of samples required by the sequential algorithm and that required by the method with fixed sample size.

uses the test given by (31). In the sequential algorithm, we set $C_A = 3.0$ and $C_B = 1.0$ for the stopping condition given by (29). As shown in Fig. 5, the sequential method substantially reduces the average number of required signal samples.

Fig. 6 shows the relative efficiency of the sequential algorithm with respect to the method with fixed sample size. Relative efficiency is defined as the ratio between the average number of samples required by the sequential algorithm and that required by the method with fixed sample size. Fig. 6 plots the relative efficiency computed from the simulation results presented in Figs. 2–5. For the simulations with unknown signal parameters, only the worst-case relative efficiency, i.e., Case 1 in Figs. 4 and 5, is plotted. The plots show that the achieved saving in the average number of samples depends on α, β, and signal distributions.

In addition to the sequential algorithm presented in the paper, we can also perform sequential detection by the method in [1] and [24]. Instead of estimating the unknown model parameters, [1] and [24] always assume the signal model for the worst scenario, i.e., Case 1, and performs the sequential test by using the algorithm presented in Section II. This approach avoids estimating the unknown parameters and can have less complexity. However, if the signal is not the assumed case, the method in [1] and [24] may require a larger sample size and lose flexibility, while a major part (Section III) of this paper is to solve this problem.

V. Conclusion

This paper presents a sequential method for spectrum sensing in a cognitive radio network. In contrast to the conventional Neyman-Pearson detector, the sequential method significantly reduces the average sensing time that is required to achieve the sensing objective. For signals received through wireless channels, we consider the case when the assumed signal and noise models have unknown parameters. The traditional sequential test is modified in order to detect signals with unknown parameters. The simulation results coincide with the theory and demonstrate the effectiveness of the proposed cooperative sensing scheme.

Appendix

A. Derivation of Expressions (4) and (5)

Under $\mathcal{H}_0$ and $\mathcal{H}_1$, the acquired samples are distributed according to $\prod_{n=1}^{N}\prod_{m=1}^{M} p_{0,m}(\mathbf{x}_m[n])$ and $\prod_{n=1}^{N}\prod_{m=1}^{M} p_{1,m}(\mathbf{x}_m[n])$, respectively. Let

$$Y[n] = \ln\left(\frac{\prod_{m=1}^{M} p_{1,m}(\mathbf{x}_m[n])}{\prod_{m=1}^{M} p_{0,m}(\mathbf{x}_m[n])}\right).$$

The log-likelihood ratio can be written as

$$\text{LLR}_N = \sum_{n=1}^{N} \ln\left(\frac{\prod_{m=1}^{M} p_{1,m}(\mathbf{x}_m[n])}{\prod_{m=1}^{M} p_{0,m}(\mathbf{x}_m[n])}\right) = \sum_{n=1}^{N} Y[n].$$

It is easy to see that

$$\mathbf{E}_{\mathcal{H}_0}\left\{e^{Y[n]}\right\} = \mathbf{E}_{\mathcal{H}_0}\left\{\frac{\prod_{m=1}^{M} p_{1,m}(\mathbf{x}_m[n])}{\prod_{m=1}^{M} p_{0,m}(\mathbf{x}_m[n])}\right\} = 1$$

which implies that e^{LLR_N} is a martingale under $\mathcal{H}_0$ for $N = 1, 2, \ldots$. Since

$$\mathbf{E}_{\mathcal{H}_0}\{Y[n]\} = -\sum_{m=1}^{M} D\left(p_{0,m}(\mathbf{x}_m[n]) \,\|\, p_{1,m}(\mathbf{x}_m[n])\right) < 0,$$

we have $\mathbf{E}_{\mathcal{H}_0}\{N_{\text{stop}}\} < \infty$.[16] By the optional stopping theorem,

$$\mathbf{E}_{\mathcal{H}_0}\{e^{\text{LLR}_{N_{\text{stop}}}}\} = 1. \tag{32}$$

Note that at N_{stop}, we have either $\text{LLR}_{N_{\text{stop}}} \approx A$ or $\text{LLR}_{N_{\text{stop}}} \approx B$. Thus,

$$\mathbf{E}_{\mathcal{H}_0}\{e^{\text{LLR}_{N_{\text{stop}}}}\} \approx P_{\text{FA}} e^A + (1 - P_{\text{FA}}) e^B. \tag{33}$$

Combining (32) and (33), we solve

$$P_{\text{FA}} e^A + (1 - P_{\text{FA}}) e^B \approx 1$$

[16] This is a direct application of Proposition 7.1.1 in [26].

and get (4). By further noting that $e^{-\mathrm{LLR}_N}$ is a Martingale under $\mathcal{H}_1$, we can get (5) through the same argument.

B. Proof of Lemma 2.2

At $N = N_{\mathrm{stop}}$,

$$\mathrm{LLR}_{N_{\mathrm{stop}}} \geq A > 0 \quad \text{or} \quad \mathrm{LLR}_{N_{\mathrm{stop}}} \leq B < 0.$$

Since e^{LLR_N} is a martingale under $\mathcal{H}_0$, we have

$$1 = \mathbf{E}_{\mathcal{H}_0}\{e^{\mathrm{LLR}_{N_{\mathrm{stop}}}}\} \geq P_{\mathrm{FA}} e^{A}$$

which leads to

$$P_{\mathrm{FA}} \leq e^{-A} = \alpha.$$

On the other hand,

$$1 = \mathbf{E}_{\mathcal{H}_1}\{e^{-\mathrm{LLR}_{N_{\mathrm{stop}}}}\} \geq P_{\mathrm{MISS}} e^{-B}$$

and hence

$$P_{\mathrm{MISS}} \leq e^{B} = \beta.$$

C. Derivation of Expressions (8) and (9)

By Wald's equation,

$$\begin{aligned}
&\mathbf{E}_{\mathcal{H}_0}\{\mathrm{LLR}_{N_{\mathrm{stop}}}\} \\
&\quad = \mathbf{E}_{\mathcal{H}_0}\left\{\sum_{n=1}^{N_{\mathrm{stop}}}\sum_{m=1}^{M} \ln\left(\frac{p_{1,m}(\mathbf{x}_m[n])}{p_{0,m}(\mathbf{x}_m[n])}\right)\right\} \\
&\quad = \mathbf{E}_{\mathcal{H}_0}\{N_{\mathrm{stop}}\}\mathbf{E}_{\mathcal{H}_0}\left\{\sum_{m=1}^{M} \ln\left(\frac{p_{1,m}(\mathbf{x}_m[n])}{p_{0,m}(\mathbf{x}_m[n])}\right)\right\} \\
&\quad = \mathbf{E}_{\mathcal{H}_0}\{N_{\mathrm{stop}}\}\left(-\sum_{m=1}^{M} D(p_{0,m}(\mathbf{x}_m[n]) \,||\, p_{1,m}(\mathbf{x}_m[n]))\right).
\end{aligned} \tag{34}$$

Recall that either $\mathrm{LLR}_{N_{\mathrm{stop}}} \approx A$ or $\mathrm{LLR}_{N_{\mathrm{stop}}} \approx B$, hence

$$\begin{aligned}
\mathbf{E}_{\mathcal{H}_0}\{\mathrm{LLR}_{N_{\mathrm{stop}}}\} &\approx P_{\mathrm{FA}} A + (1 - P_{\mathrm{FA}})B \\
&\approx \frac{A - B - Ae^{B} + Be^{A}}{e^{A} - e^{B}}.
\end{aligned} \tag{35}$$

Combining (34) and (35) leads to (8). Similarly, we can get (9).

D. Derivation of Expressions (10) and (11)

Under $\mathcal{H}_0$, according to the central limit theorem, when N_{fix} is large, the test statistic is approximately distributed as

$$\begin{aligned}
\mathcal{Z}_0 &= \sum_{n=1}^{N_{\mathrm{fix}}}\sum_{m=1}^{M}\left[\frac{1}{2}\left(\frac{1}{\sigma_{0,m}^2} - \frac{1}{\sigma_{1,m}^2}\right)(X_m[n])^2 + \frac{1}{2}\ln\left(\frac{\sigma_{0,m}^2}{\sigma_{1,m}^2}\right)\right] \\
&\sim \mathcal{N}\left(\frac{N_{\mathrm{fix}}}{2}\sum_{m=1}^{M}\left(1 - \frac{\sigma_{0,m}^2}{\sigma_{1,m}^2}\right) + \frac{N_{\mathrm{fix}}}{2}\sum_{m=1}^{M}\ln\left(\frac{\sigma_{0,m}^2}{\sigma_{1,m}^2}\right),\right. \\
&\qquad \left.\frac{N_{\mathrm{fix}}}{2}\sum_{m=1}^{M}\left(1 - \frac{\sigma_{0,m}^2}{\sigma_{1,m}^2}\right)^2\right).
\end{aligned}$$

Hence,

$$\begin{aligned}
P_{\mathrm{FA}} &= \Pr(\mathcal{Z}_0 > \eta) \\
&\approx Q\left(\frac{\eta - \frac{N_{\mathrm{fix}}}{2}\sum_{m=1}^{M}\left(1 - \frac{\sigma_{0,m}^2}{\sigma_{1,m}^2}\right) - \frac{N_{\mathrm{fix}}}{2}\sum_{m=1}^{M}\ln\left(\frac{\sigma_{0,m}^2}{\sigma_{1,m}^2}\right)}{\sqrt{\frac{N_{\mathrm{fix}}}{2}\sum_{m=1}^{M}\left(1 - \frac{\sigma_{0,m}^2}{\sigma_{1,m}^2}\right)^2}}\right)
\end{aligned}$$

which gives

$$\begin{aligned}
\eta &\approx \frac{N_{\mathrm{fix}}}{2}\sum_{m=1}^{M}\left(1 - \frac{\sigma_{0,m}^2}{\sigma_{1,m}^2}\right) + \frac{N_{\mathrm{fix}}}{2}\sum_{m=1}^{M}\ln\left(\frac{\sigma_{0,m}^2}{\sigma_{1,m}^2}\right) \\
&\quad + \sqrt{\frac{N_{\mathrm{fix}}}{2}\sum_{m=1}^{M}\left(1 - \frac{\sigma_{0,m}^2}{\sigma_{1,m}^2}\right)^2}\, Q^{-1}(P_{\mathrm{FA}}).
\end{aligned} \tag{36}$$

Under $\mathcal{H}_1$, when N_{fix} is large, the test statistic is approximately distributed as

$$\begin{aligned}
\mathcal{Z}_1 &= \sum_{n=1}^{N_{\mathrm{fix}}}\sum_{m=1}^{M}\left[\frac{1}{2}\left(\frac{1}{\sigma_{0,m}^2} - \frac{1}{\sigma_{1,m}^2}\right)(X_m[n])^2 + \frac{1}{2}\ln\left(\frac{\sigma_{0,m}^2}{\sigma_{1,m}^2}\right)\right] \\
&\sim \mathcal{N}\left(\frac{N_{\mathrm{fix}}}{2}\sum_{m=1}^{M}\left(\frac{\sigma_{1,m}^2}{\sigma_{0,m}^2} - 1\right) + \frac{N_{\mathrm{fix}}}{2}\sum_{m=1}^{M}\ln\left(\frac{\sigma_{0,m}^2}{\sigma_{1,m}^2}\right),\right. \\
&\qquad \left.\frac{N_{\mathrm{fix}}}{2}\sum_{m=1}^{M}\left(\frac{\sigma_{1,m}^2}{\sigma_{0,m}^2} - 1\right)^2\right)
\end{aligned}$$

which yields (37), shown at the bottom of the page. By substituting (36) into (37), we need to find N_{fix} such that [see the equation shown at the bottom of the next page], which finally yields (10) by letting $P_{\mathrm{FA}} = \alpha$ and $P_{\mathrm{MISS}} = \beta$. Expression (11) is obtained from (36) by letting $P_{\mathrm{FA}} = \alpha$.

$$P_{\mathrm{MISS}} = \Pr(\mathcal{Z}_1 \leq \eta) \approx Q\left(\frac{\frac{N_{\mathrm{fix}}}{2}\sum_{m=1}^{M}\left(\frac{\sigma_{1,m}^2}{\sigma_{0,m}^2} - 1\right) + \frac{N_{\mathrm{fix}}}{2}\sum_{m=1}^{M}\ln\left(\frac{\sigma_{0,m}^2}{\sigma_{1,m}^2}\right) - \eta}{\sqrt{\frac{N_{\mathrm{fix}}}{2}\sum_{m=1}^{M}\left(\frac{\sigma_{1,m}^2}{\sigma_{0,m}^2} - 1\right)^2}}\right). \tag{37}$$

E. Derivation of Expressions (12) and (13)

Under $\mathcal{H}_0$, the test statistic is distributed as

$$\sum_{n=1}^{N_{\text{fix}}}\sum_{k=1}^{K}\sum_{m=1}^{M}\left[\frac{a_m}{\sigma_m^2}s[k]X_m[(n-1)K+k]-\frac{a_m^2}{2\sigma_m^2}\left(s[k]\right)^2\right]$$
$$=\sum_{n=1}^{N_{\text{fix}}}\sum_{k=1}^{K}\sum_{m=1}^{M}\left[\frac{a_m}{\sigma_m^2}s[k]W_m[(n-1)K+k]-\frac{a_m^2}{2\sigma_m^2}\left(s[k]\right)^2\right]$$
$$\sim\mathcal{N}\left\{-\frac{N_{\text{fix}}}{2}\left(\sum_{m=1}^{M}\frac{a_m^2}{\sigma_m^2}\right)\left(\sum_{k=1}^{K}(s[k])^2\right),\right.$$
$$\left.N_{\text{fix}}\left(\sum_{m=1}^{M}\frac{a_m^2}{\sigma_m^2}\right)\left(\sum_{k=1}^{K}(s[k])^2\right)\right\}$$

and hence

$$P_{\text{FA}}=Q\left(\frac{\eta+\frac{N_{\text{fix}}}{2}\left(\sum_{m=1}^{M}\frac{a_m^2}{\sigma_m^2}\right)\left(\sum_{k=1}^{K}\left(s[k]\right)^2\right)}{\sqrt{N_{\text{fix}}\left(\sum_{m=1}^{M}\frac{a_m^2}{\sigma_m^2}\right)\left(\sum_{k=1}^{K}\left(s[k]\right)^2\right)}}\right). \quad (38)$$

Under $\mathcal{H}_1$, the test statistic is distributed as

$$\sum_{n=1}^{N_{\text{fix}}}\sum_{k=1}^{K}\sum_{m=1}^{M}\left[\frac{a_m}{\sigma_m^2}s[k]X_m[(n-1)K+k]-\frac{a_m^2}{2\sigma_m^2}\left(s[k]\right)^2\right]$$
$$=\sum_{n=1}^{N_{\text{fix}}}\sum_{k=1}^{K}\sum_{m=1}^{M}\left[\frac{a_m}{\sigma_m^2}s[k]\left(a_m s[k]+W_m[(n-1)K+k]\right)\right.$$
$$\left.-\frac{a_m^2}{2\sigma_m^2}\left(s[k]\right)^2\right]$$
$$\sim\mathcal{N}\left\{\frac{N_{\text{fix}}}{2}\left(\sum_{m=1}^{M}\frac{a_m^2}{\sigma_m^2}\right)\left(\sum_{k=1}^{K}(s[k])^2\right),\right.$$
$$\left.N_{\text{fix}}\left(\sum_{m=1}^{M}\frac{a_m^2}{\sigma_m^2}\right)\left(\sum_{k=1}^{K}(s[k])^2\right)\right\}$$

and hence

$$P_{\text{MISS}}=Q\left(\frac{\frac{N_{\text{fix}}}{2}\left(\sum_{m=1}^{M}\frac{a_m^2}{\sigma_m^2}\right)\left(\sum_{k=1}^{K}\left(s[k]\right)^2\right)-\eta}{\sqrt{N_{\text{fix}}\left(\sum_{m=1}^{M}\frac{a_m^2}{\sigma_m^2}\right)\left(\sum_{k=1}^{K}\left(s[k]\right)^2\right)}}\right). \quad (39)$$

To achieve $P_{\text{FA}}=\alpha$ and $P_{\text{MISS}}=\beta$, we solve (38) and (39) to get (12) and (13).

F. Proof of Theorem 3.1

1) It follows directly from the assumptions that $\Pr(N_{\text{stop}}<\infty\,|\,\mathcal{H}_i)=1$ [23, Lemma 1] and $\mathbf{E}_{\mathcal{H}_i}\{N_{\text{stop}}\}<\infty$ [26, Proposition 7.1.1] for $i=0,1$.

2) To show that there exists $\lambda_0>0$ such that

$$\mathbf{E}_{\mathcal{H}_0}\left\{e^{\lambda_0\sum_{m=1}^{M}\ln\left(\frac{q_{1,m}(\mathbf{x}_m[n])}{q_{0,m}(\mathbf{x}_m[n])}\right)}\right\}=1, \quad (40)$$

it is sufficient to prove that there exist $\kappa_1>0$ and $\kappa_2>0$ such that

$$\mathbf{E}_{\mathcal{H}_0}\left\{e^{\kappa_1\sum_{m=1}^{M}\ln\left(\frac{q_{1,m}(\mathbf{x}_m[n])}{q_{0,m}(\mathbf{x}_m[n])}\right)}\right\}<1$$

and

$$\mathbf{E}_{\mathcal{H}_0}\left\{e^{\kappa_2\sum_{m=1}^{M}\ln\left(\frac{q_{1,m}(\mathbf{x}_m[n])}{q_{0,m}(\mathbf{x}_m[n])}\right)}\right\}>1.$$

We note that as $\kappa_1\to 0^+$,

$$\mathbf{E}_{\mathcal{H}_0}\left\{e^{\kappa_1\sum_{m=1}^{M}\ln\left(\frac{q_{1,m}(\mathbf{x}_m[n])}{q_{0,m}(\mathbf{x}_m[n])}\right)}\right\}$$
$$=\mathbf{E}_{\mathcal{H}_0}\left\{1+\kappa_1\sum_{m=1}^{M}\ln\left(\frac{q_{1,m}(\mathbf{x}_m[n])}{q_{0,m}(\mathbf{x}_m[n])}\right)+O(\kappa_1^2)\right\}<1$$

as $\mathbf{E}_{\mathcal{H}_0}\{\sum_{m=1}^{M}\ln(q_{1,m}(\mathbf{x}_m[n])/q_{0,m}(\mathbf{x}_m[n]))\}<0$ by assumption 2a). Moreover, assumption 2a) implies

$$\Pr\left(\prod_{m=1}^{M}q_{0,m}(\mathbf{x}_m[n])\neq\prod_{m=1}^{M}q_{1,m}(\mathbf{x}_m[n])\,\middle|\,\mathcal{H}_0\right)>0,$$

and hence[17]

$$\Pr\left(\frac{\prod_{m=1}^{M}q_{1,m}(\mathbf{x}_m[n])}{\prod_{m=1}^{M}q_{0,m}(\mathbf{x}_m[n])}>1\,\middle|\,\mathcal{H}_0\right)>0.$$

[17]Here, we assume that the distributions $p_{0,m}(\mathbf{x}_m[n])$, $p_{1,m}(\mathbf{x}_m[n])$, $q_{0,m}(\mathbf{x}_m[n])$, and $q_{1,m}(\mathbf{x}_m[n])$ satisfy the following regularity condition. For each $m=1,2,\ldots,M$, there exists a set $\mathcal{D}_m$ such that i) $p_{0,m}(\mathbf{x}_m[n])\neq 0$ iff $\mathbf{x}_m[n]\in\mathcal{D}_m$; ii) $p_{1,m}(\mathbf{x}_m[n])\neq 0$ iff $\mathbf{x}_m[n]\in\mathcal{D}_m$; iii) $q_{0,m}(\mathbf{x}_m[n])\neq 0$ iff $\mathbf{x}_m[n]\in\mathcal{D}_m$; iv)$q_{1,m}(\mathbf{x}_m[n])\neq 0$ iff $\mathbf{x}_m[n]\in\mathcal{D}_m$. Intuitively, $\mathcal{D}_m$ is the set that consists of all "observable" measurements and the probability density functions have nonzero values on those measurements. Define the Cartesian product $\mathcal{D}$ as $\mathcal{D}=\mathcal{D}_1\times\mathcal{D}_2\times\ldots\times\mathcal{D}_M$, and let $\mathbf{x}[n]=(\mathbf{x}_1[n],\mathbf{x}_2[n],\ldots,\mathbf{x}_M[n])$. Note the following fact:

$$\int_{\mathcal{D}}\prod_{m=1}^{M}q_{0,m}(\mathbf{x}_m[n])d\mathbf{x}[n]=\int_{\mathcal{D}}\prod_{m=1}^{M}q_{1,m}(\mathbf{x}_m[n])\,d\mathbf{x}[n]=1.$$

Then,

$$\int_{\mathcal{D}}\mathbf{1}_{\left\{\mathbf{x}[n]:\prod_{m=1}^{M}q_{0,m}(\mathbf{x}_m[n])>\prod_{m=1}^{M}q_{1,m}(\mathbf{x}_m[n])\right\}}d\mathbf{x}[n]>0$$

iff

$$\int_{\mathcal{D}}\mathbf{1}_{\left\{\mathbf{x}[n]:\prod_{m=1}^{M}q_{0,m}(\mathbf{x}_m[n])<\prod_{m=1}^{M}q_{1,m}(\mathbf{x}_m[n])\right\}}d\mathbf{x}[n]>0$$

where $\mathbf{1}_{\{\cdot\}}$ is the indicator function.

$$P_{\text{MISS}}\approx Q\left(\frac{\frac{N_{\text{fix}}}{2}\sum_{m=1}^{M}\left(\frac{\sigma_{1,m}^2}{\sigma_{0,m}^2}-1\right)-\frac{N_{\text{fix}}}{2}\sum_{m=1}^{M}\left(1-\frac{\sigma_{0,m}^2}{\sigma_{1,m}^2}\right)-\sqrt{\frac{N_{\text{fix}}}{2}\sum_{m=1}^{M}\left(1-\frac{\sigma_{0,m}^2}{\sigma_{1,m}^2}\right)^2}Q^{-1}(P_{\text{FA}})}{\sqrt{\frac{N_{\text{fix}}}{2}\sum_{m=1}^{M}\left(\frac{\sigma_{1,m}^2}{\sigma_{0,m}^2}-1\right)^2}}\right)$$

This further implies that

$$\Pr\left(\sum_{m=1}^{M}\ln\left(\frac{q_{1,m}(\mathbf{x}_m[n])}{q_{0,m}(\mathbf{x}_m[n])}\right)>\rho\middle|\mathcal{H}_0\right)>0$$

for some $\rho>0$. We then have

$$\mathbf{E}_{\mathcal{H}_0}\left\{e^{\kappa_2\sum_{m=1}^{M}\ln\left(\frac{q_{1,m}(\mathbf{x}_m[n])}{q_{0,m}(\mathbf{x}_m[n])}\right)}\right\}\geq\Pr\left(\sum_{m=1}^{M}\ln\left(\frac{q_{1,m}(\mathbf{x}_m[n])}{q_{0,m}(\mathbf{x}_m[n])}\right)>\rho\middle|\mathcal{H}_0\right)e^{\kappa_2\rho}.$$

If κ_2 is sufficiently large,

$$\mathbf{E}_{\mathcal{H}_0}\left\{e^{\kappa_2\sum_{m=1}^{M}\ln\left(\frac{q_{1,m}(\mathbf{x}_m[n])}{q_{0,m}(\mathbf{x}_m[n])}\right)}\right\}>1.$$

Since $\mathbf{E}_{\mathcal{H}_0}\{e^{\lambda\sum_{m=1}^{M}\ln((q_{1,m}(\mathbf{x}_m[n]))/(q_{0,m}(\mathbf{x}_m[n])))}\}$ is a continuous function of λ, there must exist $\lambda_0>0$ between κ_1 and κ_2 such that (40) holds. It can be similarly shown that there exists $\lambda_1<0$ such that

$$\mathbf{E}_{\mathcal{H}_1}\left\{e^{\lambda_1\sum_{m=1}^{M}\ln\left(\frac{q_{1,m}(\mathbf{x}_m[n])}{q_{0,m}(\mathbf{x}_m[n])}\right)}\right\}=1.$$

3) Because of (40), $e^{\lambda_0\mathrm{LLR}'_N}$ is a martingale. Hence,

$$1=\mathbf{E}_{\mathcal{H}_0}\{e^{\lambda_0\mathrm{LLR}'_{N_{\mathrm{stop}}}}\}\geq P_{\mathrm{FA}}e^{\lambda_0 A}$$

which leads to $P_{\mathrm{FA}}\leq e^{-\lambda_0 A}=\alpha^{\lambda_0}$. Similarly, we can show that $P_{\mathrm{MISS}}\leq e^{-\lambda_1 B}=\beta^{-\lambda_1}$.

4) If α and β are sufficiently small, i.e., A and B are sufficiently large, we have either $\mathrm{LLR}'_{N_{\mathrm{stop}}}\approx A$ or $\mathrm{LLR}'_{N_{\mathrm{stop}}}\approx B$. Then,

$$1=\mathbf{E}_{\mathcal{H}_0}\{e^{\lambda_0\mathrm{LLR}'_{N_{\mathrm{stop}}}}\}\approx P_{\mathrm{FA}}e^{\lambda_0 A}+(1-P_{\mathrm{FA}})e^{\lambda_0 B}$$

from which we get (15). Similarly, by noting that $e^{\lambda_1\mathrm{LLR}'_N}$ is a martingale under $\mathcal{H}_1$, we get (16). To compute the average number of required samples under $\mathcal{H}_0$, we have

$$\begin{aligned}\mathbf{E}_{\mathcal{H}_0}\{\mathrm{LLR}'_{N_{\mathrm{stop}}}\}&\approx AP_{\mathrm{FA}}+B(1-P_{\mathrm{FA}})\\&\approx\frac{A-B-Ae^{\lambda_0 B}+Be^{\lambda_0 A}}{e^{\lambda_0 A}-e^{\lambda_0 B}}\end{aligned}$$

and

$$\begin{aligned}&\mathbf{E}_{\mathcal{H}_0}\left\{\sum_{m=1}^{M}\ln\left(\frac{q_{1,m}(\mathbf{x}_m[n])}{q_{0,m}(\mathbf{x}_m[n])}\right)\right\}\\&=\sum_{m=1}^{M}\mathbf{E}_{\mathcal{H}_0}\left\{\ln\left(\frac{q_{1,m}(\mathbf{x}_m[n])}{p_{0,m}(\mathbf{x}_m[n])}\right)\right\}+\sum_{m=1}^{M}\mathbf{E}_{\mathcal{H}_0}\left\{\ln\left(\frac{p_{0,m}(\mathbf{x}_m[n])}{q_{0,m}(\mathbf{x}_m[n])}\right)\right\}\\&=\sum_{m=1}^{M}\left[-D\left(p_{0,m}(\mathbf{x}_m[n])\,\|\,q_{1,m}(\mathbf{x}_m[n])\right)+D\left(p_{0,m}(\mathbf{x}_m[n])\,\|\,q_{0,m}(\mathbf{x}_m[n])\right)\right].\end{aligned}$$

By Wald's equation,

$$\mathbf{E}_{\mathcal{H}_0}\{N_{\mathrm{stop}}\}=\frac{\mathbf{E}_{\mathcal{H}_0}\{\mathrm{LLR}'_{N_{\mathrm{stop}}}\}}{\mathbf{E}_{\mathcal{H}_0}\left\{\sum_{m=1}^{M}\ln\left(\frac{q_{1,m}(\mathbf{x}_m[n])}{q_{0,m}(\mathbf{x}_m[n])}\right)\right\}}$$

which yields (17). Expression (18) can be obtained similarly.

G. Proof of Lemma 3.1

Under $\mathcal{H}_0$, it is easy to verify that $e^{\overline{\mathrm{GLLR}}_N^{(0)}}$ is a martingale. By the optional stopping theorem,

$$\mathbf{E}_{\mathcal{H}_0}\left\{e^{\overline{\mathrm{GLLR}}_{N_{\mathrm{stop}}}^{(0)}}\right\}=1.$$

Hence,

$$\Pr\left(\overline{\mathrm{GLLR}}_{N_{\mathrm{stop}}}^{(0)}\geq A\middle|\mathcal{H}_0\right)\leq e^{-A}=\alpha.$$

It then follows that

$$\begin{aligned}P_{\mathrm{FA}}&=\Pr\left(\mathrm{GLLR}_{N_{\mathrm{stop}}}\geq A_{N_{\mathrm{stop}}}|\mathcal{H}_0\right)\\&=\Pr\left(\mathrm{GLLR}_{N_{\mathrm{stop}}}\geq A_{N_{\mathrm{stop}}},\overline{\mathrm{GLLR}}_{N_{\mathrm{stop}}}^{(0)}\geq A\middle|\mathcal{H}_0\right)\\&\quad+\Pr\left(\mathrm{GLLR}_{N_{\mathrm{stop}}}\geq A_{N_{\mathrm{stop}}},\overline{\mathrm{GLLR}}_{N_{\mathrm{stop}}}^{(0)}<A\middle|\mathcal{H}_0\right)\\&\leq\Pr\left(\overline{\mathrm{GLLR}}_{N_{\mathrm{stop}}}^{(0)}\geq A\middle|\mathcal{H}_0\right)+\epsilon\\&\leq\alpha+\epsilon.\end{aligned}$$

Similarly, we can show that $P_{\mathrm{MISS}}\leq\beta+\epsilon$ under the given assumptions.

H. Proof of Theorem 3.2

By Chebyshev's inequality, we have

$$\Pr\left(\left|\mathrm{GLLR}_N-\overline{\mathrm{GLLR}}_N^{(i)}\right|\geq\sqrt{\frac{\gamma_N^{(i)}}{\epsilon}}\middle|\mathcal{H}_i\right)=\Pr\left(\left|\Delta\mathrm{GLLR}_N^{(i)}\right|\geq\sqrt{\frac{\gamma_N^{(i)}}{\epsilon}}\middle|\mathcal{H}_i\right)\leq\epsilon.$$

Let

$$A_N=A+\sqrt{\frac{\gamma_N^{(0)}}{\epsilon}}\quad\text{and}\quad B_N=B-\sqrt{\frac{\gamma_N^{(1)}}{\epsilon}}.$$

At $N=N_{\mathrm{stop}}$, we have

$$\mathrm{GLLR}_{N_{\mathrm{stop}}}\geq A_{N_{\mathrm{stop}}}\quad\text{or}\quad\mathrm{GLLR}_{N_{\mathrm{stop}}}\leq B_{N_{\mathrm{stop}}}.$$

It then follows that

$$\begin{aligned}&\Pr\left(\mathrm{GLLR}_{N_{\mathrm{stop}}}\geq A_{N_{\mathrm{stop}}},\overline{\mathrm{GLLR}}_{N_{\mathrm{stop}}}^{(0)}<A\middle|\mathcal{H}_0\right)\\&\leq\Pr\left(\left|\mathrm{GLLR}_{N_{\mathrm{stop}}}-\overline{\mathrm{GLLR}}_{N_{\mathrm{stop}}}^{(0)}\right|\geq A_{N_{\mathrm{stop}}}-A\middle|\mathcal{H}_0\right)\\&=\Pr\left(\left|\mathrm{GLLR}_{N_{\mathrm{stop}}}-\overline{\mathrm{GLLR}}_{N_{\mathrm{stop}}}^{(0)}\right|\geq\sqrt{\frac{\gamma_{N_{\mathrm{stop}}}^{(0)}}{\epsilon}}\middle|\mathcal{H}_0\right)\\&\leq\epsilon\end{aligned}$$

and

$$\begin{aligned}
&\Pr\left(\mathrm{GLLR}_{N_{\mathrm{stop}}} \le B_{N_{\mathrm{stop}}}, \overline{\mathrm{GLLR}}^{(1)}_{N_{\mathrm{stop}}} > B \middle| \mathcal{H}_1\right)\\
&\quad\le \Pr\left(\left|\mathrm{GLLR}_{N_{\mathrm{stop}}} - \overline{\mathrm{GLLR}}^{(1)}_{N_{\mathrm{stop}}}\right| \ge B - B_{N_{\mathrm{stop}}} \middle| \mathcal{H}_1\right)\\
&\quad= \Pr\left(\left|\mathrm{GLLR}_{N_{\mathrm{stop}}} - \overline{\mathrm{GLLR}}^{(1)}_{N_{\mathrm{stop}}}\right| \ge \sqrt{\frac{\gamma^{(1)}_{N_{\mathrm{stop}}}}{\epsilon}} \middle| \mathcal{H}_1\right)\\
&\quad\le \epsilon.
\end{aligned}$$

It then follows from Lemma 3.1 immediately that

$$P_{\mathrm{FA}} \le \alpha + \epsilon \quad \text{and} \quad P_{\mathrm{MISS}} \le \beta + \epsilon.$$

I. A Brief Justification of Assumptions 4a) and 4b)

In this section, we give a brief explanation on assumptions 4a) and 4b). We refer to [25] for rigorous arguments and statements. The maximum likelihood estimate of $\boldsymbol{\theta}_{0,m}$ and $\boldsymbol{\theta}_{1,m}$ are given by (20) and (21), respectively. Under hypothesis $\mathcal{H}_0$, we consider the following equivalent expressions for $\hat{\boldsymbol{\theta}}^{(N)}_{0,m}$ and $\hat{\boldsymbol{\theta}}^{(N)}_{1,m}$:

$$\hat{\boldsymbol{\theta}}^{(N)}_{0,m} = \underset{\bar{\boldsymbol{\theta}}_{0,m}\in\Theta_{0,m}}{\arg\max} \frac{1}{N}\sum_{n=1}^{N} \ln\left(\frac{p_{0,m}(\mathbf{x}_m[n];\bar{\boldsymbol{\theta}}_{0,m})}{p_{0,m}(\mathbf{x}_m[n];\boldsymbol{\theta}_{0,m})}\right) \tag{41}$$

$$\hat{\boldsymbol{\theta}}^{(N)}_{1,m} = \underset{\bar{\boldsymbol{\theta}}_{1,m}\in\Theta_{1,m}}{\arg\max} \frac{1}{N}\sum_{n=1}^{N} \ln\left(\frac{p_{1,m}(\mathbf{x}_m[n];\bar{\boldsymbol{\theta}}_{1,m})}{p_{1,m}(\mathbf{x}_m[n];\tilde{\boldsymbol{\theta}}_{1,m})}\right). \tag{42}$$

Note that $\mathbf{x}_m[n]$ are distributed according to $p_{0,m}(\mathbf{x}_m[n];\boldsymbol{\theta}_{0,m})$ under $\mathcal{H}_0$. By the law of large numbers, as $N\to\infty$,

$$\begin{aligned}
&\frac{1}{N}\sum_{n=1}^{N} \ln\left(\frac{p_{0,m}(\mathbf{x}_m[n];\bar{\boldsymbol{\theta}}_{0,m})}{p_{0,m}(\mathbf{x}_m[n];\boldsymbol{\theta}_{0,m})}\right)\\
&\quad\to \int p_{0,m}(\mathbf{x}_m[n];\boldsymbol{\theta}_{0,m}) \ln\left(\frac{p_{0,m}(\mathbf{x}_m[n];\bar{\boldsymbol{\theta}}_{0,m})}{p_{0,m}(\mathbf{x}_m[n];\boldsymbol{\theta}_{0,m})}\right) d\mathbf{x}_m[n]\\
&\quad= -D(p_{0,m}(\mathbf{x}_m[n];\boldsymbol{\theta}_{0,m}) \,||\, p_{0,m}(\mathbf{x}_m[n];\bar{\boldsymbol{\theta}}_{0,m}))
\end{aligned} \tag{43}$$

in probability, and

$$\begin{aligned}
&\frac{1}{N}\sum_{n=1}^{N} \ln\left(\frac{p_{1,m}(\mathbf{x}_m[n];\bar{\boldsymbol{\theta}}_{1,m})}{p_{1,m}(\mathbf{x}_m[n];\tilde{\boldsymbol{\theta}}_{1,m})}\right)\\
&\quad\to \int p_{0,m}(\mathbf{x}_m[n];\boldsymbol{\theta}_{0,m}) \ln\left(\frac{p_{1,m}(\mathbf{x}_m[n];\bar{\boldsymbol{\theta}}_{1,m})}{p_{1,m}(\mathbf{x}_m[n];\tilde{\boldsymbol{\theta}}_{1,m})}\right) d\mathbf{x}_m[n]\\
&\quad= \int p_{0,m}(\mathbf{x}_m[n];\boldsymbol{\theta}_{0,m}) \ln\left(\frac{p_{0,m}(\mathbf{x}_m[n];\boldsymbol{\theta}_{0,m})}{p_{1,m}(\mathbf{x}_m[n];\tilde{\boldsymbol{\theta}}_{1,m})}\right) d\mathbf{x}_m[n]\\
&\qquad- \int p_{0,m}(\mathbf{x}_m[n];\boldsymbol{\theta}_{0,m}) \ln\left(\frac{p_{0,m}(\mathbf{x}_m[n];\boldsymbol{\theta}_{0,m})}{p_{1,m}(\mathbf{x}_m[n];\bar{\boldsymbol{\theta}}_{1,m})}\right) d\mathbf{x}_m[n]\\
&\quad= D(p_{0,m}(\mathbf{x}_m[n];\boldsymbol{\theta}_{0,m}) \,||\, p_{1,m}(\mathbf{x}_m[n];\tilde{\boldsymbol{\theta}}_{1,m}))\\
&\qquad- D(p_{0,m}(\mathbf{x}_m[n];\boldsymbol{\theta}_{0,m}) \,||\, p_{1,m}(\mathbf{x}_m[n];\bar{\boldsymbol{\theta}}_{1,m}))
\end{aligned} \tag{44}$$

in probability. Combining (41) with (43) and (42) with (44), we can in general justify that under $\mathcal{H}_0$,

$$\begin{aligned}
\hat{\boldsymbol{\theta}}^{(N)}_{0,m} &\to \underset{\bar{\boldsymbol{\theta}}_{0,m}\in\Theta_{0,m}}{\arg\min}\ D(p_{0,m}(\mathbf{x}_m[n];\boldsymbol{\theta}_{0,m}) \,||\, p_{0,m}(\mathbf{x}_m[n];\bar{\boldsymbol{\theta}}_{0,m}))\\
&= \boldsymbol{\theta}_{0,m}\\
\hat{\boldsymbol{\theta}}^{(N)}_{1,m} &\to \underset{\bar{\boldsymbol{\theta}}_{1,m}\in\Theta_{1,m}}{\arg\min}\ D(p_{0,m}(\mathbf{x}_m[n];\boldsymbol{\theta}_{0,m}) \,||\, p_{1,m}(\mathbf{x}_m[n];\bar{\boldsymbol{\theta}}_{1,m}))\\
&= \tilde{\boldsymbol{\theta}}_{1,m}
\end{aligned}$$

in probability. By similar argument, we can justify that under $\mathcal{H}_1$,

$$\begin{aligned}
\hat{\boldsymbol{\theta}}^{(N)}_{0,m} &\to \underset{\bar{\boldsymbol{\theta}}_{0,m}\in\Theta_{0,m}}{\arg\min}\ D(p_{1,m}(\mathbf{x}_m[n];\boldsymbol{\theta}_{1,m}) || p_{0,m}(\mathbf{x}_m[n];\bar{\boldsymbol{\theta}}_{0,m}))\\
&= \tilde{\boldsymbol{\theta}}_{0,m}\\
\hat{\boldsymbol{\theta}}^{(N)}_{1,m} &\to \underset{\bar{\boldsymbol{\theta}}_{1,m}\in\Theta_{1,m}}{\arg\min}\ D(p_{1,m}(\mathbf{x}_m[n];\boldsymbol{\theta}_{1,m}) || p_{1,m}(\mathbf{x}_m[n];\bar{\boldsymbol{\theta}}_{1,m}))\\
&= \boldsymbol{\theta}_{1,m}
\end{aligned}$$

in probability.

Under some mild regularity conditions, the maximum likelihood estimator is asymptotically efficient, i.e., under $\mathcal{H}_0$,

$$\sqrt{N}\left(\hat{\boldsymbol{\theta}}^{(N)}_{0,m} - \boldsymbol{\theta}_{0,m}\right) \to \mathcal{N}\left(0, \mathbf{I}^{-1}_{\boldsymbol{\theta}_{0,m}}\right)$$

where $\mathbf{I}_{\boldsymbol{\theta}_{0,m}}$ is the Fisher information matrix evaluated from $p_{0,m}(\mathbf{x}_m[n];\boldsymbol{\theta}_{0,m})$ and its elements at the k_1th row and k_2th column are given by

$$\mathbf{I}_{\boldsymbol{\theta}_{0,m}}(k_1,k_2) = -\mathbf{E}\left\{\frac{\partial^2 \ln p_{0,m}(\mathbf{x}_m[n];\boldsymbol{\theta}_{0,m})}{\partial\boldsymbol{\theta}_{0,m}(k_1)\partial\boldsymbol{\theta}_{0,m}(k_2)}\right\}$$

where $\boldsymbol{\theta}_{0,m}(k_1)$ and $\boldsymbol{\theta}_{0,m}(k_2)$ are the k_1th and k_2th elements of the parameter vector $\boldsymbol{\theta}_{0,m}$. Moreover, $\hat{\boldsymbol{\theta}}^{(N)}_{0,m}$ is unbiased. It then follows immediately that

$$\mathbf{E}_{\mathcal{H}_0}\left\{\hat{\boldsymbol{\theta}}^{(N)}_{0,m} - \boldsymbol{\theta}_{0,m}\right\} = 0 = O\left(\frac{1}{N}\right)$$

and

$$\mathbf{E}_{\mathcal{H}_0}\left\{\left\|\hat{\boldsymbol{\theta}}^{(N)}_{0,m} - \boldsymbol{\theta}_{0,m}\right\|^2\right\} = O\left(\frac{1}{N}\right).$$

For $\hat{\boldsymbol{\theta}}^{(N)}_{1,m}$, normally we also have[18]

$$\begin{aligned}
\mathbf{E}_{\mathcal{H}_0}\left\{\hat{\boldsymbol{\theta}}^{(N)}_{1,m} - \tilde{\boldsymbol{\theta}}_{1,m}\right\} &= O\left(\frac{1}{N}\right)\\
\mathbf{E}_{\mathcal{H}_0}\left\{\left\|\hat{\boldsymbol{\theta}}^{(N)}_{1,m} - \tilde{\boldsymbol{\theta}}_{1,m}\right\|^2\right\} &= O\left(\frac{1}{N}\right).
\end{aligned}$$

Similarly, under $\mathcal{H}_1$,

$$\begin{aligned}
\mathbf{E}_{\mathcal{H}_1}\left\{\hat{\boldsymbol{\theta}}^{(N)}_{0,m} - \tilde{\boldsymbol{\theta}}_{0,m}\right\} &= O\left(\frac{1}{N}\right)\\
\mathbf{E}_{\mathcal{H}_1}\left\{\left\|\hat{\boldsymbol{\theta}}^{(N)}_{0,m} - \tilde{\boldsymbol{\theta}}_{0,m}\right\|^2\right\} &= O\left(\frac{1}{N}\right)
\end{aligned}$$

[18]This is true for Examples 3 and 4.

and

$$\mathbf{E}_{\mathcal{H}_1}\left\{\hat{\boldsymbol{\theta}}_{1,m}^{(N)} - \boldsymbol{\theta}_{1,m}\right\} = O\left(\frac{1}{N}\right)$$
$$\mathbf{E}_{\mathcal{H}_1}\left\{\left\|\hat{\boldsymbol{\theta}}_{1,m}^{(N)} - \boldsymbol{\theta}_{1,m}\right\|^2\right\} = O\left(\frac{1}{N}\right).$$

J. Derivation of Expressions (27) and (28)

Let $\Delta\boldsymbol{\theta}^{(N)} = \boldsymbol{\theta}^{(N)} - \boldsymbol{\theta}$. By the Taylor series expansion,

$$\begin{aligned}\ln&\left(\frac{p(\mathbf{x}[n];\hat{\boldsymbol{\theta}}^{(N)})}{p(\mathbf{x}[n];\boldsymbol{\theta})}\right)\\ &= \frac{1}{p(\mathbf{x}[n];\boldsymbol{\theta})}(\nabla p)^T\Delta\boldsymbol{\theta}^{(N)} + \frac{1}{2p(\mathbf{x}[n];\boldsymbol{\theta})}\left(\Delta\boldsymbol{\theta}^{(N)}\right)^T\\ &\quad\times\left[\nabla^2 p - \frac{1}{p(\mathbf{x}[n];\boldsymbol{\theta})}\nabla p(\nabla p)^T\right]\\ &\quad\cdot\Delta\boldsymbol{\theta}^{(N)} + O\left(\left\|\Delta\boldsymbol{\theta}^{(N)}\right\|^3\right)\end{aligned} \tag{45}$$

where ∇p is the gradient vector of $p(\mathbf{x}[n];\boldsymbol{\theta})$ with respect to $\boldsymbol{\theta}$, $(\cdot)^T$ denotes the matrix transpose, and $\nabla^2 p$ is the Hessian matrix of $p(\mathbf{x}[n];\boldsymbol{\theta})$ with respect to $\boldsymbol{\theta}$. In the expression, $\Delta\boldsymbol{\theta}^{(N)}$ depends on all the N currently received samples, while the other terms only depend on the sample $\mathbf{x}[n]$. The mean of (45) is given by

$$\begin{aligned}\mu^{(N)} &= \mathbf{E}\left\{\frac{1}{p(\mathbf{x}[n];\boldsymbol{\theta})}(\nabla p)^T\Delta\boldsymbol{\theta}^{(N)} + O\left(\left\|\Delta\boldsymbol{\theta}^{(N)}\right\|^2\right)\right\}\\ &= O\left(\frac{1}{N}\right)\end{aligned}$$

and the second moment of (45) is given by

$$\begin{aligned}\lambda^{(N)} &= \mathbf{E}\left\{\frac{1}{p^2(\mathbf{x}[n];\boldsymbol{\theta})}(\nabla p)^T\Delta\boldsymbol{\theta}^{(N)}\right.\\ &\qquad\left.\times\left(\Delta\boldsymbol{\theta}^{(N)}\right)^T\nabla p + O\left(\left\|\Delta\boldsymbol{\theta}^{(N)}\right\|^3\right)\right\}\\ &= O\left(\frac{1}{N}\right).\end{aligned}$$

ACKNOWLEDGMENT

The authors would like to thank the anonymous reviewers for their constructive comments, which helped improve the presentation in the manuscript.

REFERENCES

[1] Q. Zou, S. Zheng, and A. H. Sayed, "Cooperative spectrum sensing via sequential detection for cognitive radio networks," in *Proc. 10th IEEE Int. Workshop on Signal Processing Advances in Wireless Communications (SPAWC)*, Jun. 2009, pp. 121–125.

[2] Q. Zou, S. Zheng, and A. H. Sayed, "Cooperative spectrum sensing via coherence detection," in *Proc. 15th IEEE Workshop on Statistical Signal Processing*, Aug. 2009, pp. 610–613.

[3] J. Mitola, III and G. Q. Maguire, "Cognitive radio: Making software radios more personal," *IEEE Personal Commun.*, vol. 6, no. 4, pp. 13–18, 1999.

[4] S. Haykin, "Cognitive radio: Brain-empowered wireless communications," *IEEE J. Sel. Areas Commun.*, vol. 23, no. 2, pp. 201–220, Feb. 2005.

[5] "First report and order and further notice of proposed rulemaking: In the matter of unlicensed operation in the TV broadcast bands," Federal Communications Commission, Oct. 2006.

[6] IEEE 802.22 Working Group on Wireless Regional Area Networks (WRANs) [Online]. Available: http://www.ieee802.org/22

[7] Z. Tian and G. B. Giannakis, "A wavelet approach to wideband spectrum sensing for cognitive radios," in *Proc. Int. Conf. Cognitive Radio Oriented Wireless Networks Communications (CROWNCOM)*, Jun. 2006, pp. 1–5.

[8] D. Cabric, A. Tkachenko, and R. W. Brodersen, "Spectrum sensing measurements of pilot, energy, and collaborative detection," in *Proc. MILCOM*, Oct. 2006, pp. 1–7.

[9] Z. Quan, S. Cui, H. V. Poor, and A. H. Sayed, "Collaborative wideband sensing for cognitive radios," *IEEE Signal Process. Mag.*, vol. 25, no. 6, pp. 60–73, Nov. 2008.

[10] M. Matsui, H. Shiba, K. Akabane, and K. Uehara, "A novel cooperative sensing technique for cognitive radio," in *Proc. 18th IEEE Int. Symp. Personal, Indoor, Mobile Radio Communications (PIMRC)*, Sep. 2007, pp. 1–5.

[11] J. Unnikrishnan and V. V. Veeravalli, "Cooperative sensing for primary detection in cognitive radio," *IEEE J. Sel. Topics Signal Process.*, vol. 2, no. 1, pp. 18–27, Feb. 2008.

[12] Z. Quan, S. Cui, and A. H. Sayed, "Optimal linear cooperation for spectrum sensing in cognitive radio networks," *IEEE J. Sel. Topics Signal Process.*, vol. 2, no. 1, pp. 28–40, Feb. 2008.

[13] R. Tandra and A. Sahai, "SNR walls for signal detection," *IEEE J. Sel. Topics Signal Process.*, vol. 2, no. 1, pp. 4–17, Feb. 2008.

[14] "Second report and order and memorandum opinion and order: In the matter of unlicensed operation in the TV broadcast bands," Federal Communications Commission, Nov. 2008.

[15] S. M. Kay, *Fundamentals of Statistical Signal Processing: Detection Theory*. Englewood Cliffs, NJ: Prentice-Hall, 1998, vol. 2.

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

[17] S. Tantaratana, "Some recent results on sequential detection," in *Advances in Statistical Signal Processing*, H. V. Poor and J. B. Thomas, Eds. Greenwich: JAI Press, 1993, vol. II, ch. 8.

[18] S. Tantaratana, "Relative efficiency of the sequential probability ratio test in signal detection," *IEEE Trans. Inf. Theory*, vol. 24, no. 1, pp. 22–31, Jan. 1978.

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

[20] H. V. Poor, *Quickest Detection*. New York: Cambridge Univ. Press, 2009.

[21] 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, Jul. 2009.

[22] K. W. Choi, W. S. Jeon, and D. G. Jeong, "Sequential detection of cyclostationary signal for cognitive radio systems," *IEEE Trans. Wireless Commun.*, vol. 8, no. 9, pp. 4480–4485, Sep. 2009.

[23] A. Wald, "On cumulative sums of random variables," *Ann. Math. Statist.*, vol. 15, no. 3, pp. 283–296, Sep. 1944.

[24] P. X. Quang, "Robust sequential testing," *Ann. Statist.*, vol. 13, no. 2, pp. 638–649, 1985.

[25] A. W. van der Vaart, *Asymptotic Statistics*. New York: Cambridge Univ. Press, 1998.

[26] S. M. Ross, *Stochastic Processes*, 2nd ed. New York: Wiley, 1996.

Qiyue Zou (S'06–M'10) received the B.Eng. and M.Eng. degrees in electrical and electronic engineering from Nanyang Technological University (NTU), Singapore, in 2001 and 2004, respectively, and the M.A. degree in mathematics and the Ph.D. degree in electrical engineering from the University of California, Los Angeles (UCLA), in 2008.

From February 2001 to September 2004, he was with the Centre for Signal Processing of NTU, working on statistical and array signal processing. From August 2008 to April 2009, he was with the WiLinx Corporation, San Diego, CA, working on ultrawideband communications. Since April 2009, he has been with the Marvell Semiconductor, Inc., Santa Clara, CA, working on signal processing for data storage IC. His research interests include signal processing and digital communications.

Dr. Zou received the Young Author Best Paper Award in 2007 from the IEEE Signal Processing Society.

Songfeng Zheng (S'06–M'10) received the B.S. degree in electrical engineering, and the M.S. degree in computer science, from Xi'an Jiao Tong University, China, in 2000 and 2003, respectively, and the Ph.D. degree in statistics from the University of California, Los Angeles (UCLA), in 2008.

Since 2008, he has been an Assistant Professor with the Department of Mathematics, Missouri State University. His research interests include statistical learning, statistical computation, and image analysis.

Ali H. Sayed (F'01) received the Ph.D. degree from Stanford University, Stanford, CA, in 1992.

He is Professor of Electrical Engineering at the University of California, Los Angeles (UCLA), where he leads the Adaptive Systems Laboratory. He has published widely, in the areas of statistical signal processing, estimation theory, adaptive filtering, signal processing for communications, adaptive networks, and bio-inspired cognition. He is coauthor of the textbook *Linear Estimation* (Englewood Cliffs, NJ: Prentice-Hall, 2000), of the research monograph *Indefinite Quadratic Estimation and Control* (Philadelphia, PA: SIAM, 1999), and coeditor of *Fast Algorithms for Matrices with Structure* (Philadelphia, PA: SIAM, 1999). He is also the author of the textbooks *Fundamentals of Adaptive Filtering* (Hoboken, NJ: Wiley, 2003), and *Adaptive Filters* (Hoboken, NJ: Wiley, 2008). He has contributed several encyclopedia and handbook articles.

Dr. Sayed is a Fellow of IEEE for his contributions to adaptive filtering and estimation algorithms. He has served on the Editorial Boards of the *IEEE Signal Processing Magazine*, the *European Signal Processing Journal*, the *International Journal on Adaptive Control and Signal Processing*, and the *SIAM Journal on Matrix Analysis and Applications.* He has also served as the Editor-in-Chief of the IEEE TRANSACTIONS ON SIGNAL PROCESSING from 2003 to 2005, and the EURASIP *Journal on Advances in Signal Processing* from 2006 to 2007. He has served on the Publications (2003–2005), Awards (2005), and Conference (2007–present) Boards of the IEEE Signal Processing Society. He also served on the Board of Governors of the IEEE Signal Processing Society from 2007 to 2008 and is currently the Vice-President of Publications of the same Society. His work has received several recognitions, including the 1996 IEEE Donald G. Fink Award, the 2002 Best Paper Award from the IEEE Signal Processing Society, the 2003 Kuwait Prize in Basic Sciences, the 2005 Terman Award, and the 2005 Young Author Best Paper Award from the IEEE Signal Processing Society. He has served as a 2005 Distinguished Lecturer of the IEEE Signal Processing Society and as General Chairman of the IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP) 2008.