

A False Discovery Rate Oriented Approach to Parallel Sequential Change Detection Problems

Jie Chen, Wenyi Zhang , Senior Member, IEEE, and H. Vincent Poor , Fellow, IEEE

Abstract—The problem of sequentially detecting changes in parallel data streams is formulated and investigated. Each data stream may have its own change point at which the underlying probability distribution of its data changes, and the decision maker needs to declare, sequentially, which data streams have passed their change points. With a large number of parallel data streams, the error metric is the false discovery rate (FDR), which is the expected ratio of the number of falsely declared data streams to the total number of declared data streams. A data stream is falsely declared if the detected change point is ahead of its actual change point. Decision procedures that are guaranteed to control the FDR level are developed, and it is also shown that the average decision delays (ADDs) of these decision procedures do not grow with the number of data streams. Numerical simulations and case studies are conducted to corroborate the analytical results, and to illustrate the utility of the decision procedures.

Index Terms—Average decision delay, false discovery rate, large-scale inference, multiple change detection, multiple hypothesis testing.

I. INTRODUCTION

AS A fundamental breakthrough beyond the classical fixed-sample-size (FSS) binary hypothesis testing, sequential hypothesis testing, in which the decision maker is allowed to sequentially observe the data and make his/her decision at a randomly selected time, has been extensively studied since the seminal work by Wald [3]. For independent and identically distributed (i.i.d.) data, compared with the FSS Neyman-Pearson test [4], the sequential probability ratio test (SPRT) is superior in terms of the average sample size required for achieving the same error performance [5].

An important topic in the theory of sequential analysis is change detection. Suppose that the decision maker is monitoring

a stochastic system, in which at some unknown time there is an abrupt change of the underlying probability distribution of the system state. Change detection concerns about methods for sequentially detecting the occurrence of the change point, and the goal is to minimize the delay between the actual change point and the declared time, subject to certain constraint on the risk of false alarms [6], [7]. In the Bayesian formulation, a prior probability distribution of the change point is imposed and exploited when designing the decision procedure; see, e.g., [8]. In the non-Bayesian (or minimax) formulation, there is no prior knowledge about the change point and the design objective is to optimize the performance under the worst-case scenario; see, e.g., [9] where a decision procedure based on the cumulative sum (CUSUM) statistic [10] was shown to be optimal in an asymptotic sense, and [11] where the CUSUM decision procedure was further shown to be exactly optimal.

In basic formulations of change detection problems, the decision maker usually only needs to treat a single data stream. But in many modern applications, the rapid development of sensing technology allows for the generation of large-scale real-time streaming data [12], corresponding to a large number of parallel data streams. Detecting the change points in such parallel data streams is the situation considered in our work here.

With multiple data streams, a formulation different from ours is that an unknown subset of the data streams have a common change point. This setting is mainly motivated by surveillance applications, in which a system is monitored by multiple sensors and at an unknown time a disruption leads to a change in the observations of a subset of deployed sensors; see, e.g., [13]–[19]. In contrast, our formulation considers the scenario where each data stream has its own change point, and furthermore the data streams as well as their change points are mutually independent. This setting can be suitable for systems whose different components are essentially independent.

The formulation that a decision maker sequentially detects independent change points for parallel independent data streams may seem uninteresting at the first glance, since one may wonder that this simply boils down to a number of separate single-stream change detection procedures. Such a formulation, however, becomes interesting when the number of data streams becomes large, and the choice of error metric then turns out to be a key factor.

Let us start with the formulation of making K hypothesis tests given K observations, each corresponding to one of the tests. For any given decision procedure, Table I categorizes the outcomes of the K tests. A direct definition of the error metric is the familywise error rate (FWER) [20], [21], which is the probability of rejecting any null hypothesis, i.e., $\Pr\{V \geq 1\}$. A

Manuscript received July 15, 2019; revised January 21, 2020; accepted February 20, 2020. Date of publication March 2, 2020; date of current version March 20, 2020. The associate editor coordinating the review of this manuscript and approving it for publication was Prof. Remy Boyer. The work of Jie Chen and Wenyi Zhang was supported in part by the National Key Research and Development Program of China under Grant 2018YFA0701603, and in part by the National Natural Science Foundation of China under Grant 61722114. The work of H. Vincent Poor was supported by the U.S. National Science Foundation under Grant CCF-1908308. This article was presented in part at the 50th Annual Asilomar Conference on Signals, Systems, and Computers, Pacific Grove, CA, USA, 2016 [1], and in part at the 55th Annual Allerton Conference on Communications, Control, and Computing, Monticello, IL, USA, 2017 [2]. (Corresponding author: Wenyi Zhang.)

Jie Chen and Wenyi Zhang are with the CAS Key Laboratory of Wireless-Optical Communications, and the Department of Electronic Engineering and Information Science, University of Science and Technology of China, Hefei 230027, China (e-mail: cj12@mail.ustc.edu.cn; wenyizha@ustc.edu.cn).

H. Vincent Poor is with the Department of Electrical Engineering, Princeton University, Princeton, NJ 08544 USA (e-mail: poor@princeton.edu).

Digital Object Identifier 10.1109/TSP.2020.2977466

TABLE I
OUTCOMES OF K PARALLEL HYPOTHESIS TESTS

Hypothesis	Accepted	Rejected	Totals
Null true	L	V	K_0
Alternative true	P	S	K_1
Totals	W	R	K

simple decision procedure that guarantees the FWER no greater than α is Bonferroni's procedure, which essentially performs K separate hypothesis tests, each of which guarantees the probability of rejecting a null (i.e., Type I error probability) no greater than α/K . This requirement is very stringent for detecting weak signals when K is large, and decision procedures that control the FWER generally have very low detection power. This motivates the proposal of an alternative error metric called the false discovery rate (FDR) [22], [23], which is the expected proportion of falsely rejected hypotheses to all rejected hypotheses, i.e., $\text{FDR} = \mathbb{E}[\frac{V}{R \vee 1}]$.

Compared with FWER-oriented decision procedures, decision procedures that control the FDR are more powerful in identifying alternative hypotheses [12], at the cost of increased error rates. The most widely known FDR-oriented decision procedure is the Benjamini-Hochberg (BH) procedure [23]. These two error criteria have been widely and successfully used in applications with large-scale data streams or when many comparisons are needed, notably biological and medical signal processing, including high throughput gene and protein expression data, brain imaging, and clinical trials; see [12] for a variety of examples. There have been a few studies about the application of FDR and FWER in sequential testing of multiple hypotheses; see, e.g., [24] and [25]. Therein the focus is mainly about extending the metrics of FDR and FWER to sequential multiple hypotheses testing problems, rather than change detection. Some techniques in [24] have inspired our current work on sequential change detection of multiple data streams. As for the application to sequential change detection over multiple data streams, [26] considered a setting with multiple changes between two known distributions and extended the setting to multiple data streams. The work established an FDR control, but did not analyze average detection delay within multiple data streams.

Returning to our problem of sequential change detection over parallel data streams, a rejection corresponds to declaring a change point for a data stream, and a false rejection, i.e., a rejection of a null, corresponds to declaring a change point before the actual change point of the data stream. Thus the FWER is the probability of making an early change decision for at least one of the data streams. Similar to the multiple hypothesis testing scenario in the previous paragraph, to guarantee a prescribed level of the FWER, the change detection for each data stream is highly strict in declaring a change, in turn leading to excessively large decision delays when the number of data streams is large. To remedy this, we propose to use the FDR as the error metric, which quantifies the expected ratio between the number of false rejections and the total number of rejections. As will be shown in this work, FDR-oriented decision procedures substantially reduce the average decision delay (ADD) compared with FWER-oriented decision procedures. In fact, for FDR-oriented decision procedures, the ADD does not grow with the number of data streams, in sharp contrast with FWER-oriented decision

procedures for which the ADD grows logarithmically with the number of data streams.

The main contributions of this work are as follows.

- 1) We formulate the problem of change detection for parallel data streams, extend the FDR error metric to this problem, and develop the corresponding decision procedure, which is motivated by the classical BH procedure.
- 2) For the developed FDR-oriented decision procedure, we establish a theoretical guarantee for the FDR. Furthermore, we obtain the asymptotic behavior of the ADD as the number of data streams grows large, which does not grow with the number of data streams; in contrast, the ADD of FWER-oriented decision procedures grows logarithmically with the number of data streams.
- 3) To cope with the scenario where the post-change statistics are under multiple possible hypotheses, we extend the problem formulation of change detection to change detection and isolation, for parallel data streams, and develop the corresponding decision procedures and performance guarantees.
- 4) We test the developed decision procedures for simulated data sets as well as an application related to cognitive radios, to corroborate the analytical results, and to illustrate the utility of the decision procedures.

In our earlier work [27], we considered the non-Bayesian minimax problem formulation where no prior knowledge about change points is assumed, and developed a decision procedure which does not take into account truncation of data. In the current work, we consider the Bayesian problem formulation, and take into account the effect of truncation when analyzing the performance.

The remaining part of this paper is organized as follows. Section II describes the system model, formulates the multiple change detection problem, and introduces the key definitions. Section III presents the decision procedures that aim at controlling the FDR and the FWER, respectively. Section IV establishes the FDR guarantee and the asymptotic behavior of the ADD. Section V extends the problem formulation, decision procedures, as well as their analysis to the multiple change detection and isolation problem. Section VI presents the simulation results, and Section VII presents a case study where we apply the developed decision procedures to the multichannel dynamic spectrum access problem for cognitive radios. Finally Section VIII concludes this paper.

II. PROBLEM SETUP

Consider $K \geq 2$ parallel data streams:

$$X_1^{(k)}, X_2^{(k)}, \dots, X_{t^{(k)}-1}^{(k)}, X_{t^{(k)}}^{(k)}, \dots \text{ for } k \in [K] = \{1, \dots, K\}. \quad (1)$$

Denote the data at time epoch n by $\underline{X}_n^{[K]} = (X_n^{(1)}, X_n^{(2)}, \dots, X_n^{(K)})$. For the k -th data stream, there exists a time epoch $t^{(k)} \geq 1$ called its *change point*. In particular, we allow $t^{(k)} = \infty$; that is, the underlying statistics of a data stream may never change, and thus a data stream may not have any (finite) change point.

For a measurable space $(\Omega, \mathcal{F})$, consisting of a sample space Ω and a σ -field $\mathcal{F}$ of events, consider a family $\{P_\pi^{(k)}, k \in [K]\}$

of probability measures to describe the prior distribution of $t^{(k)}$ and the distribution of $\{X_n^{(k)}, k \in [K], n = 1, 2, \dots\}$. Under $P_\pi^{(k)}$, $t^{(k)}$ is a random variable with prior distribution $\pi_m = \Pr\{t^{(k)} = m\}$; conditioned upon $t^{(k)}$, $X_1^{(k)}, X_2^{(k)}, \dots, X_{t^{(k)}-1}^{(k)}$ are i.i.d. with a (pre-change) distribution h_0 , and $X_{t^{(k)}}^{(k)}, X_{t^{(k)}+1}^{(k)}, \dots$ are i.i.d. with another (post-change) distribution h_1 and independent of $X_1^{(k)}, X_2^{(k)}, \dots, X_{t^{(k)}-1}^{(k)}$. In Section V we will address detection and isolation problems where there are multiple possible post-change distributions. We assume mutual independence among the K data streams, and that h_0 and h_1 are probability densities with respect to some measure P on $(\Omega, \mathcal{F})$.

Define the Kullback-Leibler (KL) divergence as

$$I = \int h_1(x) \log \left(\frac{h_1(x)}{h_0(x)} \right) P(dx), \quad (2)$$

the mean change point as

$$\bar{t} = \mathbb{E} [t^{(k)} | t^{(k)} < \infty], \quad (3)$$

which is independent of the index k since we have assumed that all streams have the same change point prior distribution (cf. Remark 4 in Section IV), and a parameter d as

$$d = - \lim_{m \rightarrow \infty} \frac{\log \Pr \{t^{(k)} \geq m + 1\}}{m}, \quad d \geq 0, \quad (4)$$

in which for prior distributions with exponential tails $d > 0$, and for prior distributions with heavy tails $d = 0$.

Suppose that a statistician sequentially observes the K parallel data streams and makes decision regarding whether changes have occurred for these data streams, up to some deadline $\bar{N}$. To be concrete, the statistician keeps observing $\underline{X}_n^{[K]}$ starting from $n = 1$, sequentially, until at a certain time epoch T_1 , declaring that a change has occurred for a certain data stream indexed by D_1 . Subsequently, the statistician keeps observing $\underline{X}_n^{[K] \setminus \{D_1\}}$ starting from $n = T_1 + 1$, with the D_1 -th data stream excluded in the observation, until at a certain time epoch T_2 , declaring that a change has occurred for a certain data stream indexed by D_2 ... Such a procedure is executed until either changes have been declared for all the K data streams, or the time epoch n has reached the deadline $\bar{N}$ and the statistician declares no change for the remain data streams.

We can formalize the above decision procedure as follows.

Definition: A decision procedure is a multiple stopping rule $\mathcal{R}$ consisting of a sequence of stopping times and decisions, as

$$\mathcal{R} = (T_1, D_1, T_2, D_2, \dots, T_K, D_K), \quad (5)$$

where $1 \leq T_1 \leq T_2 \leq \dots \leq T_K \leq \bar{N}$, and $\{D_k, k \in [K]\}$ constitute a permutation of $[K]$. Regarding the stopping times and decisions, $\{T_1 = n_1, D_1 = d_1\}$ depends upon $\underline{X}_1^{[K]}, \underline{X}_2^{[K]}, \dots, \underline{X}_{n_1}^{[K]}$, and, in a recursive way, for general $k \in [K]$, $\{T_k = n_k, D_k = d_k\}$ depends upon $\underline{X}_n^{[K]}, n = 1, 2, \dots, T_1; \underline{X}_n^{[K] \setminus \{D_1\}}, n = T_1 + 1, \dots, T_2; \dots; \underline{X}_n^{[K] \setminus \{D_1, \dots, D_{k-1}\}}, n = T_{k-1} + 1, \dots, n_k$.

We can also define the stopping time for every data stream.

Definition: For $k \in [K]$, denote by $T^{(k)}$ the stopping time for the k -th data stream, so that $T^{(k)} = T_l$ for some $l \in [K]$ such that $D_l = k$.

We then define the FDR and the FWER of a decision procedure $\mathcal{R}$ as follows.

Definition: Denote by $\bar{R}$ the number of change points declared by $\mathcal{R}$, i.e., the number of elements in $\{T^{(k)}, k \in [K]\}$ satisfying $T^{(k)} < \bar{N}$, and by $\bar{V}$ the number of falsely declared change points, i.e., the number of elements in $\{T^{(k)}, k \in [K]\}$ satisfying $T^{(k)} < \bar{N}$ and $T^{(k)} < t^{(k)}$. The FDR of $\mathcal{R}$ is

$$\text{FDR} = \mathbb{E}_\pi \left[\frac{\bar{V}}{\bar{R} \vee 1} \right], \quad (6)$$

where $a \vee b$ represents the maximum between a and b , and $\mathbb{E}_\pi$ represents the expectation with respect to $\{P_\pi^{(k)}, k \in [K]\}$. The FWER of $\mathcal{R}$ is

$$\text{FWER} = \Pr \{ \bar{V} \geq 1 \}. \quad (7)$$

Besides the FDR and the FWER, we are also interested in the ADD of a decision procedure. This is the average number of additional samples before declaring a change point after that change point, normalized by the number of data streams with a finite change point. We denote by K_∞ the number of data streams without a finite change point, which is a binomial random variable $\text{Bin}(K, \pi_\infty)$, with probability mass function $P(K_\infty = k_\infty) = C_K^{k_\infty} \pi_\infty^{k_\infty} (1 - \pi_\infty)^{K - k_\infty}$. Formally, we have the following definition.

Definition: For a decision procedure, the ADD is defined as

$$\text{ADD} = \sum_{k_\infty=0}^K P(K_\infty = k_\infty) \text{ADD}_{k_\infty}, \quad (8)$$

where ADD_{k_∞} is the conditional ADD upon $K_\infty = k_\infty$, which can be written as

$$\text{ADD}_{k_\infty} = \frac{1}{K - k_\infty} \sum_{k=1}^{K-k_\infty} \mathbb{E}_\pi [T^{(k)} - t^{(k)} | t^{(k)} \leq T^{(k)} < \bar{N}]. \quad (9)$$

III. DECISION PROCEDURES

For the k -th data stream, $k \in [K]$, we define a sequence of statistics

$$G_n^{(k)} = \mathbb{I}_{n+1}^{-1} \sum_{m=1}^n \pi_m \prod_{i=m}^n L_i^{(k)}, \quad n = 1, 2, \dots, \quad (10)$$

where $L_i^{(k)} = h_1(X_i^{(k)})/h_0(X_i^{(k)})$, and $\mathbb{I}_{n+1} = \Pr\{t^{(k)} \geq n + 1\}$. So $G_n^{(k)}$ is the likelihood ratio between the hypotheses that the change occurs at $t^{(k)} \leq n$ and that the change occurs at $t^{(k)} > n$. Setting $G_0^{(k)} = 0$, we have a recursive relationship for $G_n^{(k)}$ as

$$G_n^{(k)} = \mathbb{I}_{n+1}^{-1} \left(G_{n-1}^{(k)} \mathbb{I}_n + \pi_n \right) L_n^{(k)}, \quad (11)$$

which can be applied for practical implementation.

First we describe a decision procedure called MD-FDR that aims at controlling the FDR. Fix a parameter $\alpha \in (0, 1)$, and set an array of thresholds as

$$Q_s = \frac{K}{s\alpha} - 1, \quad Q_K \leq Q_{K-1} \leq \dots \leq Q_1. \quad (12)$$

For convenience of description, we divide the operation of a decision procedure into multiple stages. At the beginning, the time epoch $n = 1$, the decision procedure starts with the first

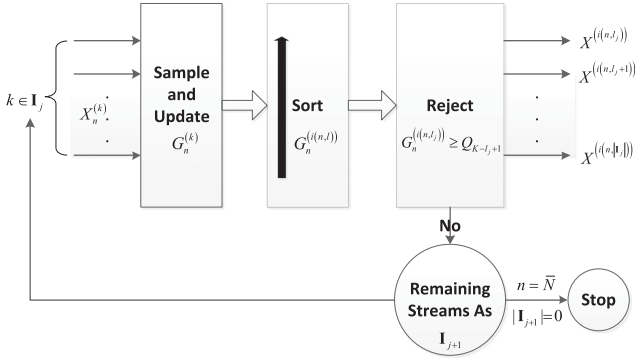


Fig. 1. Flowchart of the j -th stage of MD-FDR.

Algorithm 1: MD-FDR and MD-Hochberg.

For MD-FDR, set thresholds as (12); for MD-Hochberg, set thresholds as (15).

- 1) Initially set $q = 1$, $I_q = [K]$, $n = 1$, and $n_0 = 0$.
- 2) Sample the active data streams $\{X_n^{(k)}\}_{k \in I_q, n > n_{q-1}}$, and update the statistics $\{G_n^{(k)}\}_{k \in I_q, n > n_{q-1}}$ using (11).
- 3) Sort the statistics in ascending order as $\{G_n^{(i(n,l))}\}$, where $i(n, l)$ denotes the index of the l -th ordered statistic at time epoch n .
- 4) Repeat 2)–3) with n increased by 1 each time, until n equals

$$n_q = \bar{N} \wedge \min \{n > n_{q-1} | G_n^{(i(n,l))} \geq Q_{K-l+1}, \exists l \in [I_q]\}. \quad (13)$$

- a) If $n_q < \bar{N}$, stop the q -th stage and declare change points at n_q for the following data streams:

$$X_{n_q}^{(i(n_q, l_q))}, X_{n_q}^{(i(n_q, l_q+1))}, \dots, X_{n_q}^{(i(n_q, |I_q|))},$$

where

$$l_q = \min \{l \in [I_q] | G_n^{(i(n,l))} \geq Q_{K-l+1}\}. \quad (14)$$

Update I_{q+1} to exclude the indices of these declared data streams. If $|I_{q+1}| = 0$, then stop; otherwise, set $q = q + 1$, and go to 2) to begin the next stage.

- b) Otherwise, $n_q = \bar{N}$, declare that all the active data streams in I_q have no change points, and stop.
-

stage, and whenever at least a stopping time is reached and an associated decision for a data stream is declared, the current stage ends and a new stage begins from the next time epoch.

We use I_q to denote the indices of active data streams (i.e., the data streams that have not been stopped) at the beginning of the q -th stage, and use n_q to denote the time epoch at the end of the q -th stage. Initially we set $I_1 = [K]$ and $n_0 = 0$. We describe the decision procedure MD-FDR in Algorithm 1, which is also illustrated in Fig. 1.

For comparison, we also consider two decision procedures that control the FWER. Fix $\alpha \in (0, 1)$. The decision procedure based on Bonferroni's procedure, denoted as MD-Bonferroni,

has a single threshold $Q = K/\alpha - 1$, and is described in Algorithm 2. The decision procedure based on Hochberg's procedure [29], denoted as MD-Hochberg, is similar with MD-FDR, but has a different array of thresholds as

$$Q_s = \frac{K - s + 1}{\alpha} - 1, \quad Q_K \leq Q_{K-1} \leq \dots \leq Q_1. \quad (15)$$

The detailed description of MD-Hochberg is in Algorithm 1.

Algorithm 2: MD-Bonferroni.

- 1) Initially set $q = 1$, $I_q = [K]$, $n = 1$, and $n_0 = 0$.
- 2) Sample the active data streams $\{X_n^{(k)}\}_{k \in I_q, n > n_{q-1}}$, and update the statistics $\{G_n^{(k)}\}_{k \in I_q, n > n_{q-1}}$ using (11).
- 3) Repeat 2) with n increased by 1 each time, until n equals

$$n_q = \bar{N} \wedge \min \{n > n_{q-1} | G_n^{(k)} \geq Q, \exists k \in I_q\}. \quad (16)$$

- a) If $n_q < \bar{N}$, stop the q -th stage and declare change points at n_q for data streams that satisfy (16). Update I_{q+1} to exclude the indices of these declared data streams. If $|I_{q+1}| = 0$, then stop; otherwise, set $q = q + 1$, and go to 2) to begin the next stage.
 - b) Otherwise, $n_q = \bar{N}$, declare that all the active data streams in I_q have no change points, and stop.
-

IV. ANALYSIS OF DECISION PROCEDURES

In this section, we analyze the statistical properties of the proposed decision procedure MD-FDR. On one hand, we show that MD-FDR controls the FDR. On the other hand, we compare the ADDs achieved by MD-FDR and by FWER-oriented decision procedures.

A. False Discovery Rate

The following result, Theorem 1, states that the proposed decision procedure MD-FDR controls FDR under our problem formulation.

Theorem 1: Suppose that $0 < I < \infty$, and $\bar{t} < \infty$. For any given $0 < \alpha < 1$, the decision procedure MD-FDR in Section III satisfies

$$\text{FDR} \leq \alpha + \frac{K\bar{t} + (I + d)^{-1} (K \log K - K \log \alpha) + o(K)}{\bar{N}}, \quad (17)$$

where $o(K)/K \rightarrow 0$ as $K \rightarrow \infty$. Hence if $\bar{N}$ as a function of K scales to satisfy $(K \log K)/\bar{N} \rightarrow 0$ as $K \rightarrow \infty$, we have

$$\text{FDR} \leq \alpha + o(1), \quad (18)$$

where $o(1) \rightarrow 0$ as $K \rightarrow \infty$.

Proof: We prove Theorem 1 in two parts. In the first part, we consider a decision procedure similar to MD-FDR but without the deadline $\bar{N}$, for which we prove that such a decision

procedure controls the FDR at the level α . In the second part, we impose the deadline $\bar{N}$ and estimate the extra cost upon the FDR.

Part 1:

As we remove the deadline $\bar{N}$ in MD-FDR (i.e., setting $\bar{N} = \infty$), we adjust the definition of the FDR as

$$\text{FDR}_\infty = \mathbb{E}_\pi \left[\frac{V}{R \vee 1} \right], \quad (19)$$

where R is the number of change points declared, i.e., the number of elements of $[K]$ such that $T^{(k)} < \infty$, and V is the number of falsely declared change points, i.e. the number of elements of $[K]$ such that $T^{(k)} < t^{(k)}$.

Let K_1 be the number of data streams satisfying $t^{(k)} < \infty$. For $k \in [K]$ and $s \in [K]$ define events

$$W_{k,s} = \{G_n^{(k)} \geq Q_s, \exists n < t^{(k)}\}. \quad (20)$$

For these events, we have

$$P_\pi(W_{k,s}) = P_\pi(G_n^{(k)} \geq Q_s, \exists n < t^{(k)}) \leq \frac{s\alpha}{K}. \quad (21)$$

In order to prove (21), consider for each data stream, say, the k th one, a stopping time with respect to the observations $X_1^{(k)}, X_2^{(k)}, \dots, X_{t^{(k)}-1}^{(k)}, X_{t^{(k)}}^{(k)}, \dots$ as follows:

$$T^{(k)} = \min \left\{ n \geq 1 \mid G_n^{(k)} \geq A \right\}, \quad A > 0. \quad (22)$$

The following result is standard and can be found in several references, e.g., [28, Sec. 4].

Lemma 2: For any $A > 0$, $P_\pi(T^{(k)} < t^{(k)}) \leq 1/(A + 1)$.

Hence, by setting $A = Q_s = K/(s\alpha) - 1$ when defining $T^{(k)}$, we have

$$P_\pi(W_{k,s}) = P_\pi(T^{(k)} < t^{(k)}) \leq \frac{1}{Q_s + 1} = \frac{s\alpha}{K}, \quad (23)$$

thus proving (21).

For $u \in \{0, \dots, K_1\}$ and $v \in \{1, \dots, K - u\}$, define

$$\Omega_v = \{\omega \subseteq \{1, \dots, K - u\} : |\omega| = v\}$$

$$V_{v,u}^\omega = \{v \text{ data streams with indices in } \omega \text{ stopped early,}$$

$$\text{i.e., } T^{(k)} < t^{(k)}, k \in \omega;$$

$$\text{and } u \text{ data streams correctly stopped}\}$$

$$V_{v,u} = \bigcup_{\omega \in \Omega_v} V_{v,u}^\omega.$$

Note that the sets $\{V_{v,u}^\omega\}$ in the union $V_{v,u}$ are mutually disjoint.

We claim that

$$P_\pi(W_{k,v+u} \cap V_{v,u}^\omega) \geq 1\{k \in \omega\} P_\pi(V_{v,u}^\omega). \quad (24)$$

The following proof of (24) is modified from an argument in [24]. If $k \notin \omega$, this inequality obviously holds. If $k \in \omega$, what we need to show is that $V_{v,u}^\omega \subseteq W_{k,v+u}$. For any outcome in $V_{v,u}^\omega$, the k -th data stream is stopped early, at some stage q . By (14) in the decision procedure MD-FDR, $k = i(n_q, l)$ for some $l \geq l_q$, so

$$G_{n_q}^{(k)} \geq G_{n_q}^{(i(n_q, l_q))} \geq Q_{K-l_q+1}. \quad (25)$$

Since $K - l_q + 1$ is the number of stopped data streams until stage q , this value is no greater than the total number of stopped data streams in $V_{v,u}^\omega$, $v + u$. Hence

$$G_{n_q}^{(k)} \geq Q_{K-l_q+1} \geq Q_{v+u}, \quad (26)$$

by noting that the thresholds $\{Q_s\}$ are decreasing with s . This thus shows that $V_{v,u}^\omega \subseteq W_{k,v+u}$ for any $k \in \omega$.

With (24) established, we now follow the argument of [30], with a few modifications suitable for our problem formulation:

$$\sum_{k=1}^{K-u} P_\pi(W_{k,v+u} \cap V_{v,u}) = \sum_{k=1}^{K-u} \sum_{\omega \in \Omega_v} P_\pi(W_{k,v+u} \cap V_{v,u}^\omega) \quad (27)$$

$$\geq \sum_{k=1}^{K-u} \sum_{\omega \in \Omega_v} 1\{k \in \omega\} P_\pi(V_{v,u}^\omega) \quad (28)$$

$$= \sum_{\omega \in \Omega_v} \sum_{k=1}^{K-u} 1\{k \in \omega\} P_\pi(V_{v,u}^\omega)$$

$$= \sum_{\omega \in \Omega_v} |\omega| P_\pi(V_{v,u}^\omega) = v P_\pi(V_{v,u}). \quad (29)$$

Here, (27) is because that the sets $\{V_{v,u}^\omega\}$ in the union $V_{v,u}$ are mutually disjoint, (28) is due to (24), and (29) is due to the definition of Ω_v .

Using (29), we have that conditioned upon K_1 , the corresponding (conditional) FDR $_\infty$ is

$$\begin{aligned} K_1\text{-FDR}_\infty &= \sum_{u=0}^{K_1} \sum_{v=1}^{K-u} \frac{v}{v+u} P_\pi(V_{v,u}) \\ &\leq \sum_{u=0}^{K_1} \sum_{v=1}^{K-u} \frac{v}{v+u} \left(\frac{1}{v} \sum_{k=1}^{K-u} P_\pi(W_{k,v+u} \cap V_{v,u}) \right) \\ &= \sum_{u=0}^{K_1} \sum_{v=1}^{K-u} \frac{1}{v+u} \sum_{k=1}^{K-u} P_\pi(W_{k,v+u} \cap V_{v,u}). \end{aligned} \quad (30)$$

Define $U_{v,u,k}$ to be the event that, conditioned upon the k th data stream being stopped early, some other $v - 1$ data streams are stopped early and some u data streams correctly stopped. So $W_{k,v+u} \cap V_{v,u} = W_{k,v+u} \cap U_{v,u,k}$. Define $U_{s,k} = \bigcup_{v+u=s} U_{v,u,k}$. For any k , $U_{1,k}, \dots, U_{K,k}$ partition the sample space. Then starting from (30), we have

$$\begin{aligned} K_1\text{-FDR}_\infty &\leq \sum_{u=0}^{K_1} \sum_{v=1}^{K-u} \frac{1}{v+u} \sum_{k=1}^{K-u} P_\pi(W_{k,v+u} \cap U_{v,u,k}) \\ &= \sum_{k=1, u \geq 0}^{K-u} \sum_{s=v+u=1}^K \frac{1}{s} P_\pi(W_{k,s} \cap U_{s,k}). \end{aligned} \quad (31)$$

Since the data streams are mutually independent, the event of the k th data stream being stopped early and the event of $U_{s,k}$ are

independent, and we have

$$\begin{aligned}
K_1\text{-FDR}_\infty &\leq \sum_{k=1, u \geq 0}^{K-u} \sum_{s=v+u=1}^K \frac{1}{s} P_\pi(W_{k,s}) P_\pi(U_{s,k}) \\
&\leq \sum_{k=1, u \geq 0}^{K-u} \sum_{s=1}^K \frac{1}{s} \left(\frac{s\alpha}{K} \right) P_\pi(U_{s,k}) \quad (32) \\
&\leq \frac{\alpha}{K} \sum_{k=1}^K \sum_{s=1}^K P_\pi(U_{s,k}) \\
&= \frac{\alpha}{K} \sum_{k=1}^K 1 = \alpha, \quad (33)
\end{aligned}$$

where (32) is due to (23). Because the bound (33) holds for any K_1 , we have $\text{FDR}_\infty \leq \alpha$.

Part 2:

Now we turn to the decision procedure MD-FDR with a finite $\bar{N}$ and analyze the relationship between FDR_∞ in (19) and FDR in (6). With $\bar{N}$ imposed, the number of change points declared, $\bar{R}$, is clearly no greater than R . Denote $R = \bar{R} + \Delta$, where Δ is the size of $\{T^{(k)} | \bar{N} \leq T^{(k)} < \infty\}$, which can be further divided into two parts. The first part has a size of $\Delta_1 = |\{T^{(k)} | \bar{N} \leq T^{(k)} < t^{(k)}\}|$, which is the reduction of the number of false alarm events, and the second part has a size of $\Delta_2 = |\{T^{(k)} | t^{(k)} \leq T^{(k)} < \infty, T^{(k)} \geq \bar{N}\}|$. We have

$$\begin{aligned}
\text{FDR} &= \mathbb{E}_\pi \left[\frac{\bar{V}}{\bar{R} \vee 1} \right] \\
&= \mathbb{E}_\pi \left[\frac{\bar{V}}{\bar{R} \vee 1} \middle| \Delta_2 = 0 \right] P_\pi(\Delta_2 = 0) \\
&\quad + \mathbb{E}_\pi \left[\frac{\bar{V}}{\bar{R} \vee 1} \middle| \Delta_2 \geq 1 \right] P_\pi(\Delta_2 \geq 1) \\
&\leq \mathbb{E}_\pi \left[\frac{\bar{V} + \Delta_1}{(\bar{R} + \Delta_1) \vee 1} \middle| \Delta_2 = 0 \right] P_\pi(\Delta_2 = 0) + P_\pi(\Delta_2 \geq 1) \\
&= \mathbb{E}_\pi \left[\frac{V}{R \vee 1} \middle| \Delta_2 = 0 \right] P_\pi(\Delta_2 = 0) + P_\pi(\Delta_2 \geq 1) \\
&\leq \mathbb{E}_\pi \left[\frac{V}{R \vee 1} \right] + P_\pi(\Delta_2 \geq 1) \\
&= \text{FDR}_\infty + P_\pi(\Delta_2 \geq 1) \leq \alpha + P_{\pi/\infty}(\Delta_2 \geq 1), \quad (34)
\end{aligned}$$

where $P_{\pi/\infty}(\cdot)$ means that this is the conditional probability $P_\pi(\cdot | t^{(k)} < \infty)$, and the last inequality is due to that Δ_2 satisfies $t^{(k)} \leq T^{(k)} < \infty$.

Then we choose $\bar{N}$ to control $P_{\pi/\infty}(\Delta_2 \geq 1)$. Let us write

$$\begin{aligned}
P_{\pi/\infty}(\Delta_2 \geq 1) &= 1 - P_{\pi/\infty}(\Delta_2 = 0) \\
&= 1 - [1 - P_{\pi/\infty}(T^{(k)} \geq \bar{N} | T^{(k)} \geq t^{(k)})]^K. \quad (35)
\end{aligned}$$

Following the argument in [28, Thm. 4], when $0 < I < \infty$ holds, for every $k \in [K]$, we have

$$\mathbb{E}_{\pi/\infty} [T^{(k)} - t^{(k)} | T^{(k)} \geq t^{(k)}] \leq \frac{1}{I+d} \log \frac{K}{\alpha} [1 + o(1)], \quad (36)$$

where $o(1) \rightarrow 0$ as $K \rightarrow \infty$. Applying Markov's inequality, we have

$$\begin{aligned}
P_{\pi/\infty}(T^{(k)} \geq \bar{N} | T^{(k)} \geq t^{(k)}) &\leq \frac{\mathbb{E}_{\pi/\infty} [T^{(k)} | T^{(k)} \geq t^{(k)}]}{\bar{N}} \\
&\leq \frac{\bar{t} + (I+d)^{-1} \log \frac{K}{\alpha} [1 + o(1)]}{\bar{N}}. \quad (37)
\end{aligned}$$

Back to (35) and (34), we have

$$\begin{aligned}
\text{FDR} &\leq \alpha + 1 - \left(1 - \frac{\bar{t} + (I+d)^{-1} \log \frac{K}{\alpha} [1 + o(1)]}{\bar{N}} \right)^K \\
&\leq \alpha + \frac{K\bar{t} + (I+d)^{-1} (K \log K - K \log \alpha + o(K))}{\bar{N}}. \quad (38)
\end{aligned}$$

This completes the proof of Theorem 1. $\blacksquare$

In our problem formulation, we explicitly incorporate the deadline $\bar{N}$ and establish the FDR control with the regularity conditions on the prior distribution of change point and the KL divergence between the pre and post change distributions. If we remove the deadline, the FDR is controlled without requiring such conditions, as already proved in the first part of Theorem 1. Also we consider the simple case where all the parallel data streams are mutually independent. Following the approach in [30], it is possible to provide some degree of FDR control even when the data streams are correlated, e.g., satisfying the so-called positive regression dependence on a subset. A systematic investigation along that direction is an interesting topic for future research.

B. Average Decision Delay

For the three decision procedures considered in Section III, we have the following result about the asymptotic behaviors of their ADDs.

Theorem 3: For the three decision procedures such that the FDR (for MD-FDR) and the FWER (for MD-Bonferroni and MD-Hochberg) are controlled by $\alpha > 0$, if $\pi_\infty > 0$ holds, as K grows without bound, and $\bar{N}$ grows with K , we have:

$$\text{MD-FDR : } \text{ADD} = O(1); \quad (39)$$

$$\text{MD-Bonferroni : } \text{ADD} \sim \frac{\log K}{I+d}; \quad (40)$$

$$\text{MD-Hochberg : } \text{ADD} \sim \frac{\log K}{I+d}. \quad (41)$$

Proof: First, consider MD-FDR which controls the FDR. According to [28, Thm. 4], if the k th data stream is stopped by threshold $Q_s = K/(s\alpha) - 1$, we have

$$\mathbb{E}_\pi [T^{(k)} - t^{(k)} | T^{(k)} \geq t^{(k)}] = \frac{1}{I+d} \log \frac{K}{s\alpha} [1 + o(1)], \quad (42)$$

where $o(1) \rightarrow 0$ as $K \rightarrow \infty$. To have the FDR controlled by α , as in the proof of Theorem 1, we make $\bar{N}$ as a function of K scale to satisfy $(K \log K) / \bar{N} \rightarrow 0$ as $K \rightarrow \infty$ and thus (37) goes to 0. Then, we have

$$\begin{aligned} \mathbb{E}_\pi [T^{(k)} - t^{(k)} | t^{(k)} \leq T^{(k)} < \bar{N}] \\ = \mathbb{E}_\pi [T^{(k)} - t^{(k)} | T^{(k)} \geq t^{(k)}] (1 + o(1)), \text{ as } K \rightarrow \infty. \end{aligned} \quad (43)$$

According to the definition of the ADD, we can derive its lower bound as follows:

$$\begin{aligned} \text{ADD} &= \sum_{k_\infty=0}^K P(K_\infty = k_\infty) \frac{1}{K - k_\infty} \sum_{k=1}^{K-k_\infty} \\ &\quad \times \mathbb{E}_\pi [T^{(k)} - t^{(k)} | T^{(k)} \geq t^{(k)}] \\ &\geq \sum_{k_\infty=0}^K P(K_\infty = k_\infty) \frac{(I+d)^{-1}}{(K - k_\infty)} \\ &\quad \times \sum_{k=1}^{K-k_\infty} \log \frac{K}{(K - k_\infty)\alpha} + o(1) \\ &= \frac{1}{I+d} \left[\log \frac{K}{\alpha} \right. \\ &\quad \left. - \sum_{k_\infty=0}^K P(K_\infty = k_\infty) \log(K - k_\infty) \right] + o(1) \\ &= \frac{1}{I+d} \left[\log \frac{K}{\alpha} - \log(K(1 - \pi_\infty)) \right. \\ &\quad \left. + \frac{\pi_\infty}{2K(1 - \pi_\infty)} \right] + o(1) \\ &= \frac{1}{I+d} |\log \alpha(1 - \pi_\infty)| (1 + o(1)), \end{aligned} \quad (44)$$

where the inequality in the second line is due to (42) and that for a fixed k_∞ , Q_{K-k_∞} is the smallest threshold.

On the other hand, we can derive its upper bound as follows:

$$\begin{aligned} \text{ADD} &= \sum_{k_\infty=0}^K P(K_\infty = k_\infty) \frac{1}{K - k_\infty} \sum_{k=1}^{K-k_\infty} \\ &\quad \times \mathbb{E}_\pi [T^{(k)} - t^{(k)} | T^{(k)} \geq t^{(k)}] \\ &\leq \sum_{k_\infty=0}^K P(K_\infty = k_\infty) \frac{(I+d)^{-1}}{(K - k_\infty)} \\ &\quad \times \sum_{k=1}^{K-k_\infty} \log \frac{K}{k\alpha} + o(1) \end{aligned}$$

$$\begin{aligned} &= \frac{1}{I+d} \left[\log \frac{K}{\alpha} - \sum_{k_\infty=0}^K P(K_\infty = k_\infty) \right. \\ &\quad \left. \times \frac{\log(K - k_\infty)!}{K - k_\infty} \right] + o(1) \\ &= \frac{1}{I+d} (|\log \alpha(1 - \pi_\infty)| + 1) (1 + o(1)), \end{aligned} \quad (45)$$

where the inequality in the second line is due to (42) and that the worst case of the thresholds occurs when the thresholds are ascending and different, and the equality in the last line is due to Stirling's approximation. Combining the lower and the upper bounds we have $\text{ADD} = O(1)$ which does not grow with K .

Then, consider MD-Bonferroni and MD-Hochberg which control the FWER. We have for MD-Bonferroni,

$$\begin{aligned} \text{ADD} &= \sum_{k_\infty=0}^K P(K_\infty = k_\infty) \frac{1}{K - k_\infty} \sum_{k=1}^{K-k_\infty} \\ &\quad \times \mathbb{E}_\pi [T^{(k)} - t^{(k)} | T^{(k)} \geq t^{(k)}] \\ &= \sum_{k_\infty=0}^K P(K_\infty = k_\infty) \frac{(I+d)^{-1}}{(K - k_\infty)} \\ &\quad \times \sum_{k=1}^{K-k_\infty} \log \frac{K}{\alpha} + o(1) \\ &= \frac{1}{I+d} (\log K + |\log \alpha|) (1 + o(1)), \end{aligned} \quad (46)$$

where the equality in the second line is due to that MD-Bonferroni uses a single threshold $Q = K/\alpha$; and for MD-Hochberg,

$$\begin{aligned} \text{ADD} &= \sum_{k_\infty=0}^K P(K_\infty = k_\infty) \frac{1}{K - k_\infty} \sum_{k=1}^{K-k_\infty} \\ &\quad \times \mathbb{E}_\pi [T^{(k)} - t^{(k)} | T^{(k)} \geq t^{(k)}] \\ &\geq \sum_{k_\infty=0}^K P(K_\infty = k_\infty) \frac{(I+d)^{-1}}{(K - k_\infty)} \\ &\quad \times \sum_{k=1}^{K-k_\infty} \log \frac{k_\infty + 1}{\alpha} + o(1) \\ &= \frac{1}{I+d} \left[\sum_{k_\infty=0}^K P(K_\infty = k_\infty) \log(k_\infty + 1) - \log \alpha \right] + o(1) \\ &= \frac{1}{I+d} \left[\log(K\pi_\infty + 1) - \frac{1 - \pi_\infty}{2K\pi_\infty} - \log \alpha \right] + o(1) \\ &= \frac{1}{I+d} [\log(K\pi_\infty + 1) + |\log \alpha|] (1 + o(1)), \end{aligned} \quad (47)$$

where the inequality in the second line is similar to that in (44) and due to the thresholds in MD-Hochberg. ■

From Theorem 3, we see that the ADD of MD-FDR does not increase with the number of data streams, and thus in this sense MD-FDR is a scalable decision procedure suitable for large scale problems. In contrast, the ADDs of MD-Bonferroni and MD-Hochberg scale at a logarithmic speed with the number of data streams. Such a discrepancy is due to that for procedures controlling the FDR, we allow a more relaxed decision criterion compared with procedures controlling the FWER.

Remark 4: Our assumption that all the data streams have the same prior distribution of their change points is mainly for convenience of analysis, but it can be relaxed. When different data streams have their own prior distributions, i.e. $\pi_m = \pi_m^{(k)}$, our results about FDR control and ADD asymptotic behavior can also be established, following the proofs of Theorems 1 and 3, and the general conclusion in Theorem 3 that under FDR control the ADD does not grow without bound with K still holds true. Similarly, the assumption that all the data streams have the same pre/post-change distributions can be relaxed as well.

V. EXTENSION TO DETECTION AND ISOLATION PROBLEMS

In this section, we generalize the problem formulation of change detection to change detection and isolation for parallel data streams. The change detection and isolation problem is about situations where after the change point, there are multiple distinct post-change hypotheses, only one of which is true, and the goal is to detect the change as soon as possible and identify the correct post-change distribution after the change occurs, subject to certain constraints on the false alarms and the false isolations. It is of importance for many applications, including fault diagnosis in dynamical systems and industrial processes, environment surveillance and monitoring, and target identification in radar and sonar signal processing; see, e.g., [7].

A. Model and Problem Formulation

Consider a family $\{P_{\pi,j}^{(k)}, k \in [K], j \in [J]\}$ of probability measures on $(\Omega, \mathcal{F})$ that describes both the prior distribution of $t^{(k)}$ and the distribution of $\{X_n^{(k)}, k \in [K], n = 1, 2, \dots\}$. Under $P_{\pi,j}^{(k)}$, $X_1^{(k)}, X_2^{(k)}, \dots, X_{t^{(k)}-1}^{(k)}$ are i.i.d. with a (pre-change) distribution h_0 , and $X_{t^{(k)}}^{(k)}, X_{t^{(k)}+1}^{(k)}, \dots$ are i.i.d. with another (post-change) distribution $h_{j^{(k)}}$, $j^{(k)} \in [J]$ and further independent of $X_1^{(k)}, X_2^{(k)}, \dots, X_{t^{(k)}-1}^{(k)}$. For simplicity, we write $P_{\pi,j}^{(k)}$ for $P_{\pi,j^{(k)}}^{(k)}$. We also impose a prior probability distribution $p(j)$ on the change type j , and that h_0 and h_j are probability densities with respect to some measure P on $(\Omega, \mathcal{F})$.

Now the decision maker needs to decide, for each data stream, both its change point and its change type (i.e., the post-change distribution). So we extend the definition of a decision procedure in Section II as follows.

Definition: A decision procedure for detection and isolation is a multiple stopping rule $\mathcal{R}$ consisting of a sequence of stopping times and decisions, as

$$\mathcal{R} = (T_1, D_1, \hat{J}_1, T_2, D_2, \hat{J}_2, \dots, T_K, D_K, \hat{J}_K), \quad (48)$$

where $\{T_k, k \in [K]\}$ and $\{D_k, k \in [K]\}$ are the same as those in (5) for a decision procedure for the basic problem formulation,

and $\hat{J}_k$ denotes the decided change type for the D_k -th data stream, $k \in [K]$.

We also define the stopping time and the decided change type for every data stream.

Definition: For $k \in [K]$, denote by $T^{(k)}$ the stopping time for the k -th data stream, the same as that in Section II, and by $\hat{J}^{(k)}$ the decided change type for the k -th data stream.

We then extend the definition of the FDR to incorporate the effect of deciding change types incorrectly.

Definition: Denote by $\bar{N}$ is the number of change points declared by $\mathcal{R}$, i.e., the number of elements in $\{T^{(k)}, k \in [K]\}$ satisfying $T^{(k)} < \bar{N}$, and by $\bar{V}$ the number of falsely declared change points or falsely classified change types, i.e., the number of elements in $\{T^{(k)}, k \in [K]\}$ satisfying $\{T^{(k)} < t^{(k)}, T^{(k)} < \bar{N}\} \cup \{t^{(k)} \leq T^{(k)} < \bar{N}, \hat{J}^{(k)} \neq j^{(k)}\}$. The FDR of $\mathcal{R}$ is

$$\text{FDR} = \mathbb{E}_{\pi,p} \left[\frac{\bar{V}}{\bar{N} \vee 1} \right], \quad (49)$$

where $\mathbb{E}_{\pi,p}$ represents the expectation with respect to $\{P_{\pi,j}^{(k)}, p(j), k \in [K], j \in [J]\}$.

B. Decision Procedure

For single-stream Bayesian change detection and isolation problems, thus far there has not been an optimal Bayesian solution satisfying the constraints on false alarm and false isolation probabilities for general prior distributions of the change point, except an asymptotically optimal Bayesian solution for the geometric prior in [31]. Most non-Bayesian solutions are proposed to satisfy constraints on average run length type of false alarm criteria, which are not suitable in the FDR-control problem formulation with multiple data streams; see, e.g. [32]–[34]. To construct a decision procedure controlling the FDR, we consider a non-Bayesian approach in [35], which aims at having a control of the weighted sum of false alarm and false isolation probabilities, and possesses asymptotic optimality with some conditions. For each data stream $k \in [K]$, we define a sequence of statistics according to [35, Thm. 6],

$$S_n^{(k)} = \max_{1 \leq j \leq J} \left\{ \max_{n-m_\alpha \leq l \leq n} \sum_{s=l}^n Z_s^{(k)}(j, 0) - \left(\max_{1 \leq g \leq J, g \neq j} \max_{n-m_\alpha \leq l \leq n} \sum_{s=l}^n Z_s^{(k)}(g, 0) \right)^+ \right\}, \quad (50)$$

in which m_α is a window size we choose and

$$Z_s^{(k)}(j, 0) = \log \left(h_j(X_s^{(k)}) / h_0(X_s^{(k)}) \right).$$

Let

$$I_{jg} = \int \log \left(\frac{h_j(x)}{h_g(x)} \right) h_j(x) P(dx).$$

In the sequel, we assume that

$$I^* = \min_{1 \leq j \leq J} \min_{0 \leq g \leq J, g \neq j} I_{jg} > 0. \quad (51)$$

Furthermore, we assume that the distribution of the change point satisfies

$$\log \left(\sum_{t=1}^{\infty} t \pi_t \right) = o \left(\left| \log \frac{\alpha}{K} \right| \right), \quad (52)$$

as $K \rightarrow \infty$.

In the following, we fix a parameter $\alpha \in (0, 1)$. According to [35], we choose m_α to satisfy

$$\liminf m_\alpha / \left| \log \frac{\alpha}{K} \right| > 1 / \min_{1 \leq j \leq J} \min_{0 \leq g \leq J, g \neq j} I_{jg},$$

$$\log m_\alpha = o \left(\left| \log \frac{\alpha}{K} \right| \right) \text{ as } K \rightarrow \infty. \quad (53)$$

We set an array of threshold values

$$Q_K \leq Q_{K-1} \leq \dots \leq Q_1, \quad (54)$$

according to

$$[2(J-1)m_\alpha + 2J]m_\alpha e^{-Q_s} = \frac{s}{K}\alpha, \quad (55)$$

which means that for threshold value Q_s , the weighted sum of the false alarm and false isolation probabilities does not exceed $\frac{s}{K}\alpha$ as shown in (59).

We present the decision procedure, called MDI-FDR, in Algorithm 3.

For the stopping times $\{T^{(k)}\}$, $k \in [K]$, obtained from MDI-FDR, we have

$$\sum_{j^{(k)}=1}^J \sum_{t^{(k)}=1}^{\infty} p(j^{(k)}) \pi_{t^{(k)}} P_{t^{(k)},j} \left(\{^{(k)} \leq T^{(k)} < \infty, \hat{j}^{(k)} \neq j^{(k)} \right) + \sum_{t^{(k)}=1}^{\infty} \pi_{t^{(k)}} P_{t^{(k)}} \left(T^{(k)} < t^{(k)} \right) \leq \left(\frac{s}{K} \right) \alpha, \forall s \in [K], \quad (56)$$

where the subscript $t^{(k)}$ in $P_{t^{(k)},j}$ and $P_{t^{(k)}}$ indicates that they are probability measures under a given change point $t^{(k)}$.

To see that the thresholds $\{Q_s\}$ satisfy (56), note that according to [35, Thm. 7], we have

$$\sup_{t \geq 1} P_{t^{(k)}} \left(t \leq T^{(k)} < t + m_\alpha \right) \leq \frac{2Jm_\alpha}{e^{Q_s}}, \quad (57)$$

$$\sup_{1 \leq t \leq r} P_{t^{(k)},j} \left(r \leq T^{(k)} < r + m_\alpha, \hat{j}^{(k)} \neq j^{(k)} \right) \leq \frac{2(J-1)m_\alpha}{e^{Q_s}}, \quad (58)$$

where $(T^{(k)}, \hat{j}^{(k)})$ is obtained by the decision procedure MDI-FDR. Then we have

$$\sum_{j^{(k)}=1}^J \sum_{t^{(k)}=1}^{\infty} p(j^{(k)}) \pi_{t^{(k)}} P_{t^{(k)},j} (t^{(k)} \leq T^{(k)} < \infty, \hat{j}^{(k)} \neq j^{(k)}) + \sum_{t^{(k)}=1}^{\infty} \pi_{t^{(k)}} P_{t^{(k)}} (T^{(k)} < t^{(k)}) \leq 2(J-1)m_\alpha^2 e^{-Q_s} + 2Jm_\alpha e^{-Q_s} \left\{ 1 + m_\alpha^{-1} \sum_{t=1}^{\infty} t \pi_t \right\} \leq [2(J-1)m_\alpha + 2J]m_\alpha e^{-Q_s}, \quad (59)$$

Algorithm 3: MDI-FDR.

- 1) Initially set $q = 1$, $I_q = [K]$, $n = 1$, and $n_0 = 0$.
- 2) Sample the active data streams $\{X_n^{(k)}\}_{k \in I_q, n > n_{q-1}}$, and update the statistics $\{S_n^{(k)}\}_{k \in I_q, n > n_{q-1}}$ using (50).
- 3) Sort the statistics in ascending order as $\{S_n^{(i(n,l))}\}$, where $i(n, l)$ denotes the index of the l -th ordered statistic at time epoch n .
- 4) Repeat 2)–3) with n increased by 1 each time, until n equals

$$n_q = \bar{N} \wedge \min \{n > n_{q-1} \mid S_n^{(i(n,l))} \geq Q_{K-l+1}, \exists l \in [I_q]\}. \quad (60)$$

- a) If $n_q < \bar{N}$, stop the q -th stage and declare change points at n_q for the following data streams:

$$X_{n_q}^{(i(n_q, l_q))}, X_{n_q}^{(i(n_q, l_q+1))}, \dots, X_{n_q}^{(i(n_q, |I_q|))},$$

where

$$l_q = \min \{l \in [I_q] \mid S_n^{(i(n,l))} \geq Q_{K-l+1}\}. \quad (61)$$

Meanwhile, declare change types as

$$\hat{j}^{(k)} = \arg \max_{1 \leq j \leq J} \left\{ \max_{n_q - m_\alpha \leq l \leq n_q} \sum_{s=l}^{n_q} Z_s(j, 0) \right\}, \quad (62)$$

for these declared data streams above. Update I_{q+1} to exclude the indices of these declared data streams. If $|I_{q+1}| = 0$, then stop; otherwise, set $q = q + 1$, and go to 2) to begin the next stage.

- b) Otherwise, $n_q = \bar{N}$, declare that all the active data streams in I_q have no change points, and stop.
-

where the first inequality follows from an argument similar to the proof of [35, Thm. 7], and the second inequality is from (52) and (53).

The following result provides a bound on the FDR of MDI-FDR.

Theorem 5: For a given $0 < \alpha < 1$, the decision procedure MDI-FDR satisfies

$$\text{FDR} \leq \alpha + \frac{K\bar{t} + (K \log K - K \log \alpha) / I^* + o(K)}{\bar{N}}, \quad (63)$$

where $o(K)/K \rightarrow 0$ as $K \rightarrow \infty$. Hence if we scale $\bar{N}$ with K such that $(K \log K) / \bar{N} \rightarrow 0$ as $K \rightarrow \infty$, we have

$$\text{FDR} \leq \alpha + o(1), \quad (64)$$

in which $o(1) \rightarrow 0$ as $K \rightarrow \infty$.

Proof: The proof is similar to that of Theorem 1, and thus we only indicate the difference between the proofs. Instead of (20), we define $W_{k,s}$ as

$$W_{k,s} = \{S_n^{(k)} \geq Q_s, \exists n < t^{(k)}, \text{ or } \hat{j}^{(k)} \neq j^{(k)}, \exists n \geq t^{(k)}\}. \quad (65)$$

So according to (56), we have $P_{\pi,p}(W_{k,s}) \leq s\alpha/K$, where the subscript p represents the prior distribution of the change type.

We can then follow the proof of FDR_∞ in Theorem 1. Upon imposing $\bar{N}$, we utilize the following result about the expected delay in detection and isolation problems:

Lemma 6: [35, Thm. 7] If $(T^{(k)}, \hat{J}^{(k)})$ satisfies (56), then for every $1 \leq j \leq J$,

$$\sum_{t^{(k)}=1}^{\infty} \pi_{t^{(k)}} \mathbb{E}_{t^{(k)},j} [T^{(k)} - t^{(k)} | T^{(k)} \geq t^{(k)}] \sim (1 + o(1)) |\log \alpha| / \min_{0 \leq g \leq J, g \neq j} I_{jg} \quad \text{as } \alpha \rightarrow 0. \quad (66)$$

Hence, we have for every $k \in [K]$,

$$\begin{aligned} & \mathbb{E}_{\pi,p} [T^{(k)} - t^{(k)} | T^{(k)} \geq t^{(k)}] \\ &= \sum_{j=1}^J \sum_{t^{(k)}=1}^{\infty} p(j^{(k)}) \pi_{t^{(k)},j} [T^{(k)} - t^{(k)} | T^{(k)} \geq t^{(k)}] \\ &\leq \frac{1}{I^*} \log \frac{K}{\alpha} (1 + o(1)), \end{aligned} \quad (67)$$

where $o(1) \rightarrow 0$ as $K \rightarrow \infty$ in our situation. The expectation $\mathbb{E}_{t^{(k)},j}$ is taken with respect to $P_{t^{(k)},j}$. Then we use (67) to replace (36) in the proof of Theorem 1, and the remaining part readily follows. ■

We also study the ADD performance of MDI-FDR, which is defined in the same way as that in Section II. Applying (67), we have the following result regarding the ADD.

Theorem 7: For the decision procedure MDI-FDR controlling the FDR by α , as K grows without bound, we have:

$$\text{ADD} \leq \frac{1}{I^*} (|\log \alpha(1 - \pi_\infty)| + 1) (1 + o(1)). \quad (68)$$

The proof is similar to that of Theorem 3 and is thus omitted here.

We can see that the statistics in MDI-FDR do not exploit the knowledge of the distribution of change types. So we also propose a heuristic statistic with a Bayesian flavor as:

$$S_n^{(k)} = \max_{1 \leq j \leq J} \left\{ \frac{G_n^{(k)}(j, 0)}{\left(\max_{1 \leq g \leq J, g \neq j} G_n^{(k)}(g, 0) \vee 1 \right)} \right\}, \quad (69)$$

where $G_n^{(k)}(j, 0)$ is a statistic similar to (10) as:

$$G_n^{(k)}(j, 0) = \mathbb{I}_{n+1}^{-1} \left(G_{n-1}^{(k)}(j, 0) \mathbb{I}_n + \pi_n \right) L_n^{(k)}(j, 0), \quad (70)$$

where $\mathbb{I}_{n+1} = \Pr\{t^{(k)} \geq n+1\}$ and $L_n^{(k)}(j, 0) = h_j(X_n^{(k)})/h_0(X_n^{(k)})$, $1 \leq j \leq J$. We use the thresholds $\{Q_s\}$ in (12). This thus leads to a heuristic decision procedure called MDI-Heuristic in Algorithm 4. Unfortunately we have not established FDR control for MDI-Heuristic, but our numerical experiments in Section VI-B suggest that this heuristic decision procedure performs well. A more thorough theoretic analysis of MDI-Heuristic is left for future research.

Algorithm 4: MDI-Heuristic.

- 1) Initially set $q = 1$, $I_q = [K]$, $n = 1$, and $n_0 = 0$.
- 2) Sample the active data streams $\{X_n^{(k)}\}_{k \in I_q, n > n_{q-1}}$, and update the statistics $\{S_n^{(k)}\}_{k \in I_q, n > n_{q-1}}$ using (69).
- 3) Sort the statistics in ascending order as $\{S_n^{(i(n,l))}\}$, where $i(n, l)$ denotes the index of the l -th ordered statistic at time epoch n .
- 4) Repeat 2)–3) with n increased by 1 each time, until n equals

$$n_q = \bar{N} \wedge \min \{n > n_{q-1} | S_n^{(i(n,l))} \geq Q_{K-l+1}, \exists l \in [I_q]\}. \quad (71)$$

- a) If $n_q < \bar{N}$, stop the q -th stage and declare change points at n_q for the following data streams:

$$X_{n_q}^{(i(n_q, l_q))}, X_{n_q}^{(i(n_q, l_q+1))}, \dots, X_{n_q}^{(i(n_q, |I_q|))},$$

where

$$l_q = \min \{l \in [I_q] | S_n^{(i(n,l))} \geq Q_{K-l+1}\}. \quad (72)$$

Meanwhile, declare change types as

$$\hat{J}^{(k)} = \arg \max_{1 \leq j \leq J} \left\{ \frac{G_n^{(k)}(j, 0)}{\left(\max_{1 \leq g \leq J, g \neq j} G_n^{(k)}(g, 0) \vee 1 \right)} \right\}, \quad (73)$$

for these declared data streams above. Update I_{q+1} to exclude the indices of these declared data streams. If $|I_{q+1}| = 0$, then stop; otherwise set $q = q + 1$, and go to 2) to begin the next stage.

- b) Otherwise, $n_q = \bar{N}$, declare that all the active data streams in I_q have no change points, and stop.
-

Finally, we also consider decision procedures that control the FWER, which is defined as $\Pr\{\bar{V} \geq 1\}$. Here we provide a decision procedure based on Bonferroni's procedure, denoted as MDI-Bonferroni, as described in Algorithm 5.

VI. SIMULATION RESULTS

In this section, we present the numerical results of Monte Carlo simulations designed to compare the proposed decision procedures controlling the FDR and the FWER. In order to make a fair comparison, we set the upper bounds of the FDR and the FWER the same, as $\alpha = 0.1$. The deadline $\bar{N}$ is 2000.

A. Multiple Change Detection

Regarding the prior probability distribution of change points, we set the probability of that there is no finite change point as $\pi_\infty = 0.2$, and set the distribution of finite change points as a (conditional) geometric distribution with parameter $p = 0.1$. Table II displays the estimated FDR/FWER for K i.i.d. data

TABLE II
ERROR PERFORMANCE FOR CHANGE DETECTION WITH I.I.D.
GAUSSIAN DATA STREAMS

K	Procedure	Estimated FDR/FWER
10	MD-FDR	0.0261
	MD-Bonferroni	0.0153
	MD-Hochberg	0.0275
100	MD-FDR	0.0230
	MD-Bonferroni	0.0129
	MD-Hochberg	0.0211
200	MD-FDR	0.0225
	MD-Bonferroni	0.0135
	MD-Hochberg	0.0238
500	MD-FDR	0.0221
	MD-Bonferroni	0.0127
	MD-Hochberg	0.0238
1000	MD-FDR	0.0222
	MD-Bonferroni	0.0116
	MD-Hochberg	0.0220

Algorithm 5: MDI-Bonferroni.

- 1) Initially set $q = 1$, $I_q = [K]$, $n = 1$, and $n_0 = 0$.
- 2) Sample the active data streams $\{X_n^{(k)}\}_{k \in I_q, n > n_{q-1}}$, and update the statistics $\{S_n^{(k)}\}_{k \in I_q, n > n_{q-1}}$ using (50).
- 3) Repeat 2) with n increased by 1 each time, until n equals

$$n_q = \bar{N} \wedge \min \{n > n_{q-1} \mid S_n^{(k)} \geq Q_1, \exists k \in I_q\}. \quad (74)$$

- a) If $n_q < \bar{N}$, stop the q -th stage and declare change points at n_q for data streams that satisfy (74). Meanwhile, declare change types as

$$\hat{J}^{(k)} = \arg \max_{1 \leq j \leq J} \left\{ \max_{n_q - m_q \leq l \leq n_q} \sum_{s=l}^{n_q} Z_s^{(k)}(j, 0) \right\}, \quad (75)$$

for these declared data streams above. Update I_{q+1} to exclude the indices of these declared data streams. If $|I_{q+1}| = 0$, then stop; otherwise, set $q = q + 1$, and go to 2) to begin the next stage.

- b) Otherwise, $n_q = \bar{N}$, declare that all the active data streams in I_q have no change points, and stop.

streams, with pre/post-change probability distributions:

$$h_0 \sim \mathcal{N}(0, 1) \quad \text{v.s.} \quad h_1 \sim \mathcal{N}(1, 1). \quad (76)$$

We see that all the decision procedures well satisfy the upper bound α .

Fig. 2 compares the ADD performance of the decision procedures. We see that the ADD is not visibly affected by the number of data streams K for MD-FDR, but tends to grow logarithmically with K for MD-Bonferroni and MD-Hochberg. This observation confirms the analytical result in Theorem 3, which indicates that FDR control decision procedures tend to be more scalable compared with FWER control decision procedures.

B. Multiple Change Detection and Isolation

We again consider the prior probability distribution of change points as a mixture of a singleton at infinity with probability

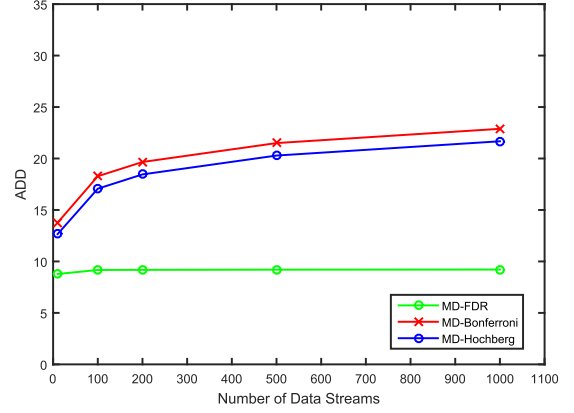


Fig. 2. ADD comparison of FDR control decision procedures and FWER control decision procedures.

TABLE III
PERFORMANCE FOR CHANGE DETECTION AND ISOLATION WITH I.I.D.
GAUSSIAN DATA STREAMS UNDER CONFIGURATION P1

$P1$				
K	Procedure	Estimated FDR/FWER	ADD	$\bar{R}$
10	MDI-FDR	0.0143	3.49	9.13
	MDI-Bonferroni	0.0154	3.98	9.02
100	MDI-FDR	0.0137	3.50	90.87
	MDI-Bonferroni	0.0137	4.68	89.65
500	MDI-FDR	0.0136	3.54	455.53
	MDI-Bonferroni	0.0151	5.29	449.50
1000	MDI-FDR	0.0138	3.54	912.00
	MDI-Bonferroni	0.0146	5.53	900.05

π_∞ and a geometric distribution with parameter $p = 0.05$. We further set the post-change type as a uniform distribution on $\{1, \dots, J\}$, and consider K i.i.d. data streams with pre/post-change probability distributions:

$$h_0 \sim \mathcal{N}(0, 1) \quad \text{v.s.} \quad h_j \sim \mathcal{N}(\mu_j, 1), \quad p(j) = 1/J.$$

We run simulations for four configurations as follows:

$$P1 : J = 2, \pi_\infty = 0.1, [\mu_1, \mu_2] = [2, 6],$$

$$P2 : J = 2, \pi_\infty = 0.1, [\mu_1, \mu_2] = [2, -2],$$

$$P3 : J = 2, \pi_\infty = 0.6, [\mu_1, \mu_2] = [2, 6],$$

$$P4 : J = 3, \pi_\infty = 0.1, [\mu_1, \mu_2, \mu_3] = [2, 6, 4].$$

Tables III through VI display the results, wherein we include $\bar{R}$, the average number of change points declared.

Comparing Tables III and IV, we see that as the gap between μ_1 and μ_2 decreases, the FDR/FWER and the ADD increase, reflecting that the task of isolation of post-change type becomes challenging. Comparing Tables III and V, we see that a larger probability of infinite change point considerably worsens the FDR/FWER, whereas still controlled by the prescribed upper bound. From Table VI, we see that an increase in the number of post-change types decreases the FDR/FWER, but at the cost of increasing the ADD. This is because in the decision procedure the thresholds increase with J and the error probability is with respect to the worst case. This suggests that when J is large, it may be possible to set smaller thresholds to control the error rates at a similar level but with reduced decision delays.

TABLE IV
PERFORMANCE FOR CHANGE DETECTION AND ISOLATION WITH I.I.D.
GAUSSIAN DATA STREAMS UNDER CONFIGURATION P2

<i>P2</i>				
<i>K</i>	Procedure	Estimated FDR/FWER	ADD	$\bar{R}$
10	MDI-FDR	0.0306	4.96	9.28
	MDI-Bonferroni	0.0379	5.69	9.04
100	MDI-FDR	0.0319	5.06	92.38
	MDI-Bonferroni	0.0392	6.89	89.68
500	MDI-FDR	0.0297	5.07	462.91
	MDI-Bonferroni	0.0343	7.77	449.52
1000	MDI-FDR	0.0300	5.07	926.56
	MDI-Bonferroni	0.0320	8.17	900.07

TABLE V
PERFORMANCE FOR CHANGE DETECTION AND ISOLATION WITH I.I.D.
GAUSSIAN DATA STREAMS UNDER CONFIGURATION P3

<i>P3</i>				
<i>K</i>	Procedure	Estimated FDR/FWER	ADD	$\bar{R}$
10	MDI-FDR	0.0835	3.59	4.39
	MDI-Bonferroni	0.0917	3.88	4.09
100	MDI-FDR	0.0790	3.80	43.81
	MDI-Bonferroni	0.0787	4.77	40.43
500	MDI-FDR	0.0788	3.79	217.31
	MDI-Bonferroni	0.0760	5.27	200.27
1000	MDI-FDR	0.0790	3.80	435.56
	MDI-Bonferroni	0.0800	5.54	401.40

TABLE VI
PERFORMANCE FOR CHANGE DETECTION AND ISOLATION WITH I.I.D.
GAUSSIAN DATA STREAMS UNDER CONFIGURATION P4

<i>P4</i>				
<i>K</i>	Procedure	Estimated FDR/FWER	ADD	$\bar{R}$
10	MDI-FDR	0.0031	7.21	9.04
	MDI-Bonferroni	0.0028	8.13	9.02
100	MDI-FDR	0.0028	7.31	90.44
	MDI-Bonferroni	0.0035	9.70	90.20
500	MDI-FDR	0.0029	7.32	451.75
	MDI-Bonferroni	0.0025	10.97	450.52
1000	MDI-FDR	0.0029	7.33	903.22
	MDI-Bonferroni	0.0027	11.62	900.79

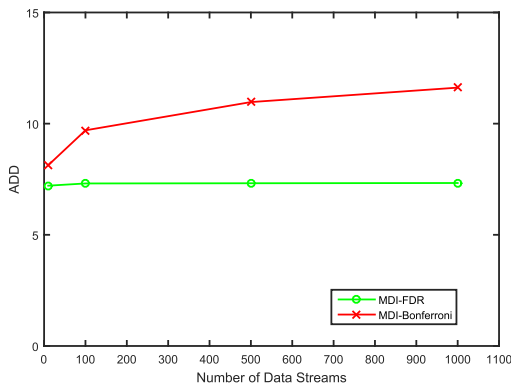


Fig. 3. ADD of MDI-FDR and MDI-Bonferroni under configuration P4.

In terms of the ADD, MDI-FDR outperforms MDI-Bonferroni. The performance shown in Fig. 3 exhibits a similar trend as that in Fig. 2: the ADD in MDI-FDR is not visibly affected by K , but tends to grow logarithmically with K in MDI-Bonferroni.

TABLE VII
COMPARISON OF MDI-FDR AND MDI-Heuristic FOR CHANGE DETECTION AND ISOLATION WITH I.I.D. GAUSSIAN DATA STREAMS

<i>P4</i>			
<i>K</i>	Procedure	Estimated FDR	ADD
10	MDI-FDR	0.0236	5.97
	MDI-Heuristic	0.0310	3.52
100	MDI-FDR	0.0217	6.06
	MDI-Heuristic	0.0265	3.61
500	MDI-FDR	0.0224	6.07
	MDI-Heuristic	0.0266	3.60
1000	MDI-FDR	0.0220	6.07
	MDI-Heuristic	0.0266	3.60

We also conduct simulations for the heuristic decision procedure MDI-Heuristic, and compare its performance with MDI-FDR in Table VII. We see that both decision procedures control the FDR at a similar level, and that MDI-Heuristic achieves lower ADD compared with MDI-FDR, which is consistent with our intuition that MDI-Heuristic makes better use of the prior information.

VII. A CASE STUDY: MULTICHANNEL SPECTRUM SENSING IN COGNITIVE RADIOS WITH KNOWN PARAMETERS

We conduct a case study in which we apply the proposed decision procedures to the multichannel dynamic spectrum access problem for cognitive radios.

Cognitive radio has been a potential solution to the problem of spectrum effectiveness in wireless communications [36][37]. The technology relies on detection of idle resources in the licensed spectrums to the primary users and dynamic utilization of these spectrum resources by the cognitive users. To make this technology feasible, the cognitive users should vacate the spectrum bands as soon as the primary users start to transmit signals so as to avoid undesirable interference. A small detection delay will maximally make use of vacant spectrum bands, and also minimize interference to the primary network after it resumes transmissions. Thus, the detection delay is a performance metric in spectrum sensing as important as the false alarm rate, especially in bursty applications. The resumption of the primary transmission will result in a change in the probability distribution of the observed signal at the cognitive user. Therefore, sequential change detection can be an appropriate method for spectrum sensing in cognitive radios [38][39].

Now we consider a multichannel dynamic spectrum access model for cognitive radios, as shown in Fig. 4, where each channel corresponds to a data stream, and the beginning of a primary user's transmission corresponds to a change. For simplicity, we only consider the single cognitive user mode for each channel. There is a fusion center that can process the information from all cognitive users without loss.

We assume there are K frequency channels allocated to K primary users without dependence. We make each cognitive user monitor the corresponding spectrum band and take signal samples sequentially. Let $X_n^{(k)}$ denote the received signal at the k -th cognitive user at time slot n . When the primary user is not transmitting, $X_n^{(k)} = W_n^{(k)}$, where $W_n^{(k)}$ is assumed to be circularly symmetric complex Gaussian with mean zero and variance σ^2 . When the primary user is active and transmitting,

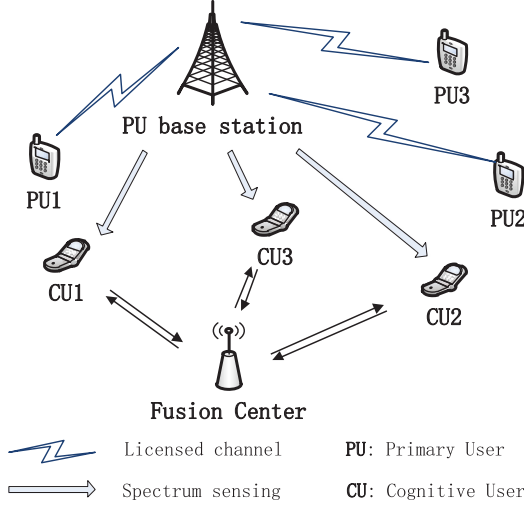


Fig. 4. Illustration of a multichannel spectrum access network.

$X_n^{(k)} = Z_n^{(k)} + W_n^{(k)}$, where $Z_n^{(k)}$ is the faded received primary signal at the k -th cognitive user. We assume a time-invariant channel gain $h^{(k)}$ of the k -th frequency channel, and $Z_n^{(k)} = h^{(k)} S_n^{(k)}$ to be zero-mean circularly symmetric complex Gaussian with variance $P^{(k)}$. The primary user is initially inactive, and at an unknown time slot $t^{(k)}$ it becomes active for the k -th spectrum channel. Thus there is a change in the distribution of data samples from $CN(0, \sigma^2)$ to $CN(0, \sigma^2 + P^{(k)})$ at the unknown time slot $t^{(k)}$ for the k -th data stream. Here, the post-change distributions are different for each data stream, which is reasonable because the channel gains cannot be the same among all spectrum channels. As for the distribution of change points, in wireless communications, the arrival time of the primary user's traffic can be simply assumed to follow a geometric distribution with parameter p . In practice, because the communication signal is a complex number, we usually take the absolute square of the received signal, i.e., $|X_n^{(k)}|^2$ as the input of the corresponding detection procedures. The change in the distribution of $|X_n^{(k)}|^2$ should be from an exponential distribution $\text{Exp}(\sigma^2)$ to another one $\text{Exp}(\sigma^2 + P^{(k)})$.

Thus our proposed decision procedure MD-FDR can be suitable for this problem, even though the post-change distributions are different for each data stream, which is mentioned before in Section IV. It is natural to interpret the FDR as the severity of the "wasted" spectrum (i.e., those unused channels mistakenly detected as occupied by primary users), and the ADD as the amount of mean interference duration induced by the cognitive users. We can apply the decision procedure MD-FDR to this multichannel dynamic spectrum sensing problem to detect the retransmitting of primary users as quickly as possible to give primary users more protection while controlling the waste of spectrum resources.

We present simulation results in Fig. 5 to compare the performance of the proposed decision procedures controlling the FDR and the FWER with different numbers of spectrum channels. The upper bounds of the FDR and the FWER are set to be the same, as $\alpha = 0.05, 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8$, respectively. The parameter of the prior distribution of change points is set as $p = 0.05$. As for the variances of pre/post-change probability

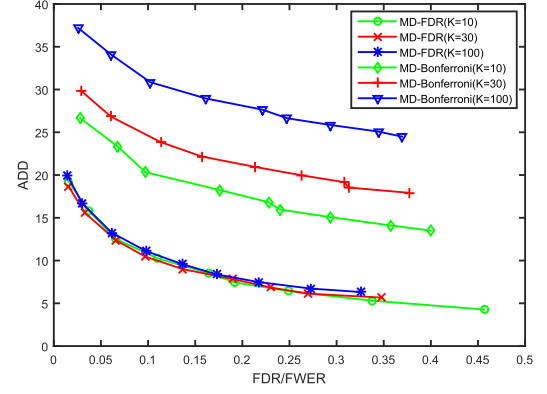


Fig. 5. ADD versus FDR/FWER for different number of frequency channels.

distributions, we set $\sigma^2 = 2$ and $P^{(k)}$ is drawn from $[1, 2]$ uniformly for different K . We can see that the ADD decreases with the increase of FDR/FWER for all decision procedures, and the FDR control decision procedures actually have a smaller mean inference duration compared with the FWER control decision procedures. The ADD of FWER control decision procedures has a larger increase for large K compared with that of FDR control decision procedures. This is due to the fact that, for large K the FWER metric will lead to a more stringent constraint on the false alarm rate and then make the ADD increase, which is consistent with the results in Section IV-B and Section VI-A. The ADD of MD-FDR varies only slightly with K because there are more different post-change probability distributions with the increase of K and the ADD is associated with different KL divergences.

Note that we only consider the simplest situation in which the parameters of pre/post-change distributions are known. In practice, because of the unknown primary signal statistics and channel gains, it is difficult to make the parameters of the post-change distributions precise. Thus, we may consider that $P^{(k)}$ belongs to a certain set of possible values, which is similar to the model in Section V.

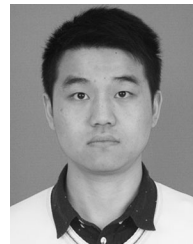
VIII. CONCLUSION

To cope with the emerging situation of analyzing large-scale real-time streaming data, in this work, we have examined the problem of sequentially detecting changes in parallel data streams, and brought the error metric of FDR into the problem formulation. Our proposed FDR-oriented decision procedure is a sequential variant of the classical BH procedure, where the p -value is replaced with a sequential detection statistic. Thanks to the characteristics of FDR-oriented decision procedures, when it comes to change detection problems, the average decision delay is significantly reduced compared with conventional FWER-oriented decision procedures. Our theoretical findings are also corroborated by numerical experiments and case studies related to cognitive radios.

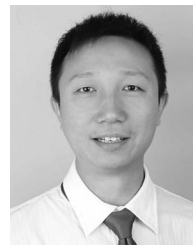
Note that we have not established the optimality of our proposed decision procedure for the problem of sequentially detecting changes in multiple data streams. It appears to be difficult to establish a lower bound for the ADD of any decision rule in the class of rules with a given FDR upper bound. This is an important yet unsolved issue for future research.

REFERENCES

- [1] J. Chen, W. Zhang, and H. V. Poor, "On parallel sequential change detection controlling false discovery rate," in *Proc. 50th Annu. Asilomar Conf. Signals, Syst. Comput.*, Nov. 2016, pp. 107–111.
- [2] J. Chen, W. Zhang, and H. V. Poor, "An FDR-oriented approach to multiple sequential fault detection and isolation," in *Proc. 55th Annu. Allerton Conf. Commun., Control, Comput.*, Oct. 2017, pp. 629–633.
- [3] A. Wald, *Sequential Analysis*. New York, NY, USA: Wiley, 1947.
- [4] J. Neyman and E. S. Pearson, "On the problem of the most efficient tests of statistical hypotheses," *Philos. Trans. Roy. Soc. London A.*, vol. 231, no. 694–706, pp. 289–337, Jan. 1933.
- [5] A. Wald and J. Wolfowitz, "Optimum character of the sequential probability ratio test," *Ann. Math. Statist.*, vol. 19, no. 3, pp. 326–339, 1948.
- [6] H. V. Poor and O. Hadjiladis, *Quickest Detection*. Cambridge, UK: Cambridge Univ. Press, 2009.
- [7] A. G. Tartakovsky, I. V. Nikiforov, and M. Basseville, *Sequential Analysis: Hypothesis Testing and Changepoint Detection*. Boca Raton, FL, USA: CRC Press, 2014.
- [8] A. N. Shiryaev, "On optimum methods in quickest detection problems," *Theory Probability Appl.*, vol. 8, pp. 22–46, 1963.
- [9] G. Lorden, "Procedures for reacting to a change in distribution," *Ann. Math. Statist.*, vol. 42, pp. 1897–1908, 1971.
- [10] E. S. Page, "Continuous inspection schemes," *Biometrika*, vol. 41, pp. 100–115, 1954.
- [11] G. V. Moustakides, "Optimal stopping times for detecting changes in distributions," *Ann. Statist.*, vol. 14, pp. 1379–1387, 1986.
- [12] B. Efron, *Large-Scale Inference: Empirical Bayes Methods for Estimation, Testing, and Prediction*. Cambridge, UK: Cambridge Univ. Press, 2010.
- [13] Y. Mei, "Efficient scalable schemes for monitoring a large number of data streams," *Biometrika*, vol. 97, no. 2, pp. 419–433, 2010.
- [14] N. R. Zhang and D. Siegmund, "Model selection for high-dimensional, multi-sequence change-point problems," *Statistica Sinica*, vol. 22, no. 4, pp. 1507–1538, 2012.
- [15] Y. Xie and D. Siegmund, "Sequential multi-sensor change-point detection," *Ann. Statist.*, vol. 41, no. 2, pp. 670–692, 2013.
- [16] T. Banerjee and V. V. Veeravalli, "Data-efficient quickest change detection in sensor networks," *IEEE Trans. Signal Process.*, vol. 63, no. 14, pp. 3727–3735, Jul. 2015.
- [17] G. Fellouris and G. Sokolov, "Second-order asymptotic optimality in multisensor sequential change detection," *IEEE Trans. Inf. Theory*, vol. 62, no. 6, pp. 3662–3675, Jun. 2016.
- [18] S. Banerjee and G. Fellouris, "Scalable multichannel joint sequential change detection and isolation," in *Proc. IEEE Int. Symp. Inf. Theory*, 2017, pp. 3095–3099.
- [19] H. P. Chan, "Optimal sequential detection in multi-stream data," *Ann. Statist.*, vol. 45, no. 6, pp. 2736–2763, 2017.
- [20] S. Holm, "A simple sequentially rejective multiple test procedure," *Scandinavian J. Statist.*, vol. 6, no. 2, pp. 65–70, 1979.
- [21] G. Hommel, "A stagewise rejective multiple test procedure based on a modified Bonferroni test," *Biometrika*, vol. 75, pp. 383–386, 1988.
- [22] G. Eklund and P. Seeger, "Masssignifikansanalys," *Statistisk Tidskrift*, vol. 5, pp. 355–365, 1965.
- [23] Y. Benjamini and Y. Hochberg, "Controlling the false discovery rate: A practical and powerful approach to multiple testing," *J. Roy. Statist. Soc. Ser. B*, vol. 57, pp. 289–300, 1995.
- [24] J. Bartroff and J. Song, "Sequential tests of multiple hypotheses controlling false discovery and nondiscovery rates," 2019. *ArXiv:1311.3350*.
- [25] J. Bartroff and J. Song, "Sequential tests of multiple hypotheses controlling type I and II familywise error rates," *J. Statistical Planning Inference*, vol. 153, pp. 100–114, 2014.
- [26] A. Gandy and F. D. Lau, "Non-restarting cumulative sum charts and control of the false discovery rate," *Biometrika*, vol. 100, no. 1, pp. 261–268, 2012.
- [27] J. Chen, W. Zhang, and H. V. Poor, "Non-Bayesian multiple change-point detection controlling false discovery rate," in *Proc. IEEE Int. Symp. Inf. Theory*, 2016, pp. 31–35.
- [28] A. G. Tartakovsky and V. V. Veeravalli, "General asymptotic Bayesian theory of quickest change detection," *Theory Probability Appl.*, vol. 49, no. 3, pp. 458–497, 2005.
- [29] Y. Hochberg, "A sharper Bonferroni procedure for multiple tests of significance," *Biometrika*, vol. 75, pp. 800–802, 1988.
- [30] Y. Benjamini and D. Yekutieli, "The control of the false discovery rate in multiple testing under dependency," *Ann. Statist.*, vol. 29, pp. 1165–1188, 2001.
- [31] S. Dayanik, W. B. Powell, and K. Yamazaki, "Asymptotically optimal Bayesian sequential change detection and identification rules," *Ann. Oper. Res.*, vol. 208, pp. 337–370, 2013.
- [32] I. V. Nikiforov, "A generalized change detection problem," *IEEE Trans. Inf. Theory*, vol. 41, no. 1, pp. 171–187, Jan. 1995.
- [33] I. V. Nikiforov, "A simple recursive algorithm for diagnosis of abrupt changes in random signals," *IEEE Trans. Inf. Theory*, vol. 46, no. 7, pp. 2740–2746, Nov. 2000.
- [34] A. G. Tartakovsky, "Multidecision quickest change-point detection: Previous achievements and open problems," *Sequential Anal.*, vol. 27, pp. 201–231, 2008.
- [35] T. L. Lai, "Sequential multiple hypothesis testing and efficient fault detection-isolation in stochastic systems," *IEEE Trans. Inf. Theory*, vol. 46, no. 2, pp. 595–608, Mar. 2000.
- [36] J. Mitola and G. Q. Maguire, "Cognitive radios: Making software radios more personal," *IEEE Pers. Commun.*, vol. 6, no. 4, pp. 13–18, Aug. 1999.
- [37] S. Haykin, "Cognitive radio: Brain-empowered wireless communications," *IEEE J. Sel. Areas Commun.*, vol. 23, no. 2, pp. 201–220, Feb. 2005.
- [38] L. Lai, Y. Fan, and H. V. Poor, "Quickest detection in cognitive radio: A sequential change detection framework," in *Proc. IEEE Global Commun. Conf.*, 2008, pp. 1–5.
- [39] H. Li, C. Li, and H. Dai, "Quickest spectrum sensing in cognitive radio," in *Proc. Conf. Inf. Sci. Syst.*, Princeton, NJ, USA, 2008, pp. 203–208.



Jie Chen received the B.E. degree in communication engineering in 2013 from the University of Science and Technology of China, Hefei, China, where he is currently working toward the Ph.D. degree in information and communication engineering. His research interests include sequential change detection and isolation for large-scale data streams.



Wenyi Zhang (Senior Member, IEEE) received the bachelor's degree in automation from Tsinghua University, Beijing, China, in 2001, and the master's and Ph.D. degrees in electrical engineering from the University of Notre Dame, Notre Dame, IN, USA, in 2003 and 2006, respectively. He is currently a Professor with the Department of Electronic Engineering and Information Science, University of Science and Technology of China, Hefei, China. He was with the Communication Science Institute, University of Southern California, as a Postdoctoral Research Associate, and with Qualcomm, Inc., Corporate Research and Development. His research interests include wireless communications and networking, information theory, and statistical signal processing. He was an Editor for the IEEE COMMUNICATIONS LETTERS, and is currently an Editor for the IEEE TRANSACTIONS ON WIRELESS COMMUNICATIONS.



H. Vincent Poor (Fellow, IEEE) received the Ph.D. degree in EECS from Princeton University, Princeton, NJ, USA, in 1977. From 1977 to 1990, he was with the faculty of the University of Illinois at Urbana-Champaign. Since 1990, he has been on the faculty at Princeton, where he is the Michael Henry Strater University Professor of Electrical Engineering. From 2006 to 2016, he was the Dean of Princeton's School of Engineering and Applied Science. He has also held visiting appointments at several other institutions, including most recently at Berkeley and Cambridge.

His research interests are in the areas of information theory, signal processing and machine learning, and their applications in wireless networks, energy systems, and related fields. Among his publications in these areas is the forthcoming book *Advanced Data Analytics for Power Systems* (Cambridge University Press, 2020). Dr. Poor is a member of the National Academy of Engineering and the National Academy of Sciences, and is a Foreign Member of the Chinese Academy of Sciences, the Royal Society, and other national and international academies. He received the Technical Achievement and Society Awards of the IEEE Signal Processing Society in 2007 and 2011, respectively. Recent recognition of his work includes the 2017 IEEE Alexander Graham Bell Medal, the 2019 ASEE Benjamin Garver Lamme Award, a D.Sc. *honoris causa* from Syracuse University, awarded in 2017, and a D.Eng. *honoris causa* from the University of Waterloo, awarded in 2019.