Abstract—This paper considers cooperative spectrum sensing algorithms for Cognitive Radios which focus on reducing the number of samples to make a reliable detection. We develop an energy efficient detector with low detection delay using decentralized sequential hypothesis testing. Our algorithm at the Cognitive Radios employs an asynchronous transmission scheme which takes into account the noise at the fusion center. We start with a distributed algorithm, DualSPRT, in which Cognitive Radios sequentially collect the observations, make local decisions using SPRT (Sequential Probability Ratio Test) and send them to the fusion center. The fusion center sequentially processes these received local decisions corrupted by noise, using an SPRT-like procedure to arrive at a final decision. We theoretically analyse its probability of error and average detection delay. We also asymptotically study its performance. Even though DualSPRT performs asymptotically well, a modification at the fusion node provides more control over the design of the algorithm parameters which then performs better at the usual operating probabilities of error in Cognitive Radio systems. We also analyse the modified algorithm theoretically. Later we modify these algorithms to handle uncertainties in SNR and fading.

Index Terms—Cooperative spectrum sensing, distributed detection, sequential detection.

I. INTRODUCTION

Presently there is a scarcity of spectrum due to the proliferation of wireless services. Cognitive Radios (CRs) are proposed as a solution to this problem. They access the spectrum licensed to existing communication services (primary users) opportunistically and dynamically without causing much interference to the primary users. This is made possible via spectrum sensing by the Cognitive Radios (secondary users), to gain knowledge about the spectrum usage by the primary devices. However due to the strict spectrum sensing requirements ([1]) and the various inherent wireless channel impairments spectrum sensing has become one of the main challenges faced by the Cognitive Radios.

Multipath fading, shadowing and hidden node problem cause serious problems in spectrum sensing. Cooperative (decentralized or distributed) spectrum sensing in which different cognitive radios interact with each other exploiting spatial diversity, ([1], [23]) is proposed as an answer to these problems. Also it reduces the probability of false alarm and the probability of miss-detection. Cooperative spectrum sensing can be either centralized or distributed ([1]). In the centralized algorithm a central unit gathers sensing data from the Cognitive Radios and identifies the spectrum usage ([17]). On the other hand, in the distributed case each secondary user (SU) collects observations, makes a local decision and sends to a fusion center (FC) to make the final decision. Centralized algorithms provide better performance but also have more communication overhead in transmitting all the data to the fusion node. In the distributed case, the information that is exchanged between the secondary users and the fusion node can be a soft decision (summary statistic) or a hard decision. Soft decisions can give better gains at the fusion center but also consume higher bandwidth at the control channels (used for sharing information among secondary users). However hard decisions provide as good a performance as soft decisions when the number of cooperative users increases ([17]).

Spectrum sensing problem can be formulated in different ways, two of them being Neyman-Pearson framework (fixed sample size detection) and sequential detection framework which reduces the number of samples taken for deciding if a primary is transmitting or not. Sequential framework enables decision more quickly than the fixed sample size counterpart ([22]). Also, there are two types of sequential detection: one can consider detecting when a primary turns ON (or OFF) (change detection, see [2], [13] and the references therein) or just testing the hypothesis whether the primary is ON or OFF ([21], [25], [27] and references therein). In [13], cooperative spectrum sensing under sequential change detection framework with no coordination between the secondary users is considered, and random broadcast policies and several improvements are proposed. In sequential hypothesis testing one considers the case where the status of the primary channel is known to change very slowly, e.g., detecting occupancy of a TV transmission. Usage of idle TV bands by the Cognitive network is being targeted as the first application for cognitive radio. In this setup (minimising the expected sensing time with constraints on probability of errors) Walds’ SPRT (Sequential Probability Ratio Test) provides the optimal performance for a single Cognitive Radio ([22]). But the optimal solutions for cooperative setup are not available ([24]).

In this paper, we consider sequential hypothesis testing in cooperative setup. We first propose a decentralized algorithm DualSPRT, in which the secondary users sequentially collect the observations, make local decisions using SPRT and send them to the fusion center. Then the fusion center sequentially processes these received local decisions corrupted by noise, using a new sequential test, to arrive at a final decision. Unlike some of the previous works on cooperative spectrum sensing using sequential testing (see [21], [26] and references therein) we analyse this algorithm theoretically also. Feedback
from the fusion node to the CRs can possibly improve the performance. However that also requires an extra signalling channel which may not be available and has its own cost. Thus in our framework we assume that there is no feedback from the fusion center to the CRs. Furthermore, we consider the receiver noise at the fusion node and use physical layer fusion to reduce the transmission time of the decisions by the local nodes to the fusion node.

In sequential decentralized detection framework, optimization needs to be performed jointly over sensors and fusion center policies as well as over time. Unfortunately, this problem is intractable for most of the sensor configurations \((14)\). Specifically there is no optimal solution available for sensor configurations with no feedback from fusion center and limited local memory, which is more relevant in practical situations. Recently [7] and [14] proposed asymptotically optimal (order 1 (Bayes) and order 2 respectively) decentralized sequential hypothesis tests for such systems with full local memory. But these models do not consider noise at the fusion center and assume a perfect communication channel between the CR nodes and the fusion center. Also, often asymptotically optimal tests do not perform well at a finite number of observations.

Noisy channels between local nodes and fusion center are considered in [26] in decentralized sequential detection framework. But optimality of the tests are not discussed and the paper is more focussed on finding the best signalling schemes at the local nodes with the assumption of parallel channels between local nodes and the fusion center. Also fusion center tests are based on the assumption of perfect knowledge of local node probability of false alarm and probability of misdetection.

We study asymptotic performance of DualSPRT, with fusion center noise. It can approach the optimal centralized sequential solution (in Bayes and frequentist sense), which does not consider noise at FC. We assume a MAC (Multiple Access Channel) as the reporting channel at the fusion center and the test is not based on the local node probability of error. Later we modify DualSPRT to improve its performance. The parameters of the modified algorithm are easier to fine tune also. Furthermore we introduce a new way of quantizing SPRT decisions of local nodes and extend this algorithm to cover SNR uncertainties and fading channels. We also study its performance theoretically. We have seen via simulations that our algorithm works better than the algorithm in \([14]\) and almost as well as the algorithm in [7] even when the fusion center noise is not considered and MAC layer transmission delays are ignored in [7] and \([14]\).

In addition, we generalize our algorithm to include uncertainty in the received Signal to Noise Ratio (SNR) at the CRs and fading channels between primary and CR. This requires a composite hypothesis testing extension to the decentralized sequential detection problem and is not considered in any of the above references. \([25]\), \([27]\) also proposed cooperative sequential algorithms for spectrum sensing, but neither of them deal with the fusion center noise and SNR uncertainty case.

This paper is organised as follows. Section [I] presents the model. Section [III] provides the DualSPRT algorithm. An approximate theoretical performance of the algorithm is also provided. Section [IV] studies the asymptotic performance of DualSPRT. In Section [V] we improve over DualSPRT. We compare the different versions so obtained and also compare them with existing asymptotically optimal decentralized sequential algorithms. Sections [VI] extends these algorithms to consider the effect of fading and SNR uncertainty. Section [VII] concludes the paper.

II. System Model

We consider a Cognitive Radio system with one primary transmitter and \(L\) secondary users. The \(L\) nodes sense the channel to detect the spectral holes. The decisions made by the secondary users are transmitted to a fusion node via a reporting MAC for it to make a final decision.

Let \(X_{k,l}\) be the observation made at secondary user \(l\) at time \(k\). The \(\{X_{k,l}, k \geq 1\}\) are independent and identically distributed (i.i.d.). It is assumed that the observations are independent across Cognitive Radios. Based on \(\{X_{n,l}, n \leq k\}\) the secondary user \(l\) transmits \(Y_{k,l}\) to the fusion center. It is assumed that the secondary nodes are synchronised so that the fusion node receives \(Y_k = \sum_{l=1}^L Y_{k,l} + Z_k\), where \(\{Z_k\}\) is i.i.d. receiver noise. The fusion center uses \(Y_k\) and makes a decision. The observations \(\{X_{k,l}\}\) depend on whether the primary is transmitting (Hypothesis \(H_1\)) or not (Hypothesis \(H_0\)) as

\[
\text{Under } H_0: \quad X_{k,l} = \zeta_{k,l}, \quad k = 1, 2, \ldots
\]

\[
\text{Under } H_1: \quad X_{k,l} = h_l S_k + \zeta_{k,l}, \quad k = 1, 2, \ldots
\]

where \(h_l\) is the channel gain of the \(l^{th}\) user, \(S_k\) is the primary signal and \(\zeta_{k,l}\) is the observation noise at the \(l^{th}\) user at time \(k\). We assume \(\{\zeta_{k,l}, k \geq 1\}\) are i.i.d. Let \(N\) be the time to decide on the hypothesis by the fusion node. We assume that \(N\) is much less than the coherence time of the channel so that the slow fading assumption is valid. This means that \(h_l\) is random but remains constant during the spectrum sensing duration.

The general problem is to develop a distributed algorithm in the above setup which solves the problem:

\[
\min E_{DD} \triangleq E_l[N]
\]

subject to \(P_1(\text{Reject } H_1) \leq \alpha_1 \& P_0(\text{Reject } H_0) \leq \alpha_0\), \((1)\) where \(P_i\) is the probability measure and \(E_l\) the expectation when \(H_i\) is the true hypothesis, \(i \in \{0, 1\}\), and \(0 \leq \alpha_0, \alpha_1 \leq 1\). We will separately consider \(E_1[N]\) and \(E_0[N]\). It is well known that for a single node case \((L = 1)\) Wald’s SPRT performs optimally in terms of reducing \(E_1[N]\) and \(E_0[N]\) for given probability of errors. Motivated by the optimality of SPRT for a single node (and DualCUSUM in \([2]\)) we propose using DualSPRT in the next section and study its performance.

We use \(P_{MD}\) for \(P_1(\text{reject } H_1)\) and \(P_{FA}\) for \(P_0(\text{reject } H_0)\). In case of \(E_{DD}\), hypothesis under consideration can be understood from the context.

III. Decentralized Sequential Tests: DualSPRT

In this section we develop DualSPRT algorithm for decentralized sequential detection and also study its performance.
A. DualSPRT algorithm

To explain the setup and analysis we start with the simple case, where the channel gain, \( h_i = 1 \) for all \( i \)'s. We will consider fading in the next section. DualSPRT is as follows:

1) Secondary node \( l \), computes at step \( k \),
\[
W_{0,l} = 0, \\
W_{k,l} = W_{k-1,l} + \log \left[ f_{l,k}(X_{k,l}) \right] f_{0,l}(X_{k,l}) \right], \] for \( k \geq 1 \),
where \( f_{l,k} \) is the density of \( X_{k,l} \) under \( H_1 \) and \( f_{0,l} \) is the density of \( X_{k,l} \) under \( H_0 \) (w.r.t. a common distribution).

2) Secondary node \( l \) transmits a constant \( b_1 \) at time \( k \) if \( W_{k,l} \geq \gamma_{1,l} \) or transmits \( b_0 \) when \( W_{k,l} \leq -\gamma_{0,l} \). When \( W_{k,l} \) does not cross the interval \((-\gamma_{0,l}, \gamma_{1,l})\), node \( l \) does not transmit anything, i.e.,
\[
Y_{k,l} = b_1 I\{W_{k,l} \geq \gamma_{1,l}\} + b_0 I\{W_{k,l} \leq -\gamma_{0,l}\}
\]
where \( \gamma_{0,l}, \gamma_{1,l} > 0 \) and \( I\{A\} \) denotes the indicator function of set \( A \). Parameters \( b_1, b_0, \gamma_{1,l}, \gamma_{0,l} \) are chosen appropriately.

3) Physical layer fusion is used at the fusion Centre, i.e., \( Y_k = \sum_{l=1}^{K} Y_{k,l} + Z_k \), where \( \{Z_k\} \) is the i.i.d. noise at the fusion node.

4) Finally, fusion center calculates the log-likelihood ratio:
\[
F_k = F_{k-1} + \log \left[ g_{\mu_l}(Y_k) / g_{-\mu_0}(Y_k) \right], \] for \( k \geq 0, \mu_0 > 0, \mu_1 > 0, \]
where \( g_{-\mu_0} \) is the density of \( Z_k - \mu_0 \) and \( g_{\mu_l} \) is the density of \( Z_k + \mu_1, \mu_0 \) and \( \mu_1 \) being positive constants appropriately chosen.

5) The fusion center decides about the hypothesis at time \( N \) where
\[
N = \inf \{ k : F_k \geq \beta_1 \} \text{ or } F_k \leq -\beta_0, \]
and \( \beta_0, \beta_1 > 0 \). The decision at time \( N \) is \( H_1 \) if \( F_N \geq \beta_1 \), otherwise \( H_0 \).

Performance of this algorithm depends on \( (\gamma_{1,l}, \gamma_{0,l}, \beta_1, \beta_0, b_1, b_0, \mu_1, \mu_0) \). In particular these parameters should be chosen such that the overall probabilities of error are less than \( \alpha_1 \) and \( \alpha_0 \) respectively. Any prior information available about \( H_0 \) or \( H_1 \) can be used to decide constants (via, say, formulating this problem in the Bayesian framework; we will comment on this again). Also we choose these parameters such that the probability of false alarm/miss-detection, \( P_{FA}/P_{MD} \) at local nodes is higher than \( P_{PA}/P_{MD} \). A good set of parameters for given SNR values can be obtained from our analysis below.

Deciding at local nodes and transmitting decisions to the fusion node reduces the transmission rate and transmit energy used by the local nodes in communication with the fusion node. Also, physical layer fusion in Step 3 reduces transmission time, but requires synchronisation of different local nodes. If synchronisation is not possible, then some other MAC algorithm, e.g., TDMA can be used with channel coding. But this will incur extra delay.

Using sequential tests at SUs and at FC (without physical layer synchronization and fusion receiver noise) has been shown to perform well in \([27], [21]\). In the next subsection we analyse the performance under our setup.

B. Performance Analysis

We first provide the analysis for the mean detection delay \( E_{DD} \) and then for \( P_{FA} \).

KL-divergence of two probability distributions \( P \) and \( Q \) on the same measurable space \((\Omega, F)\) is defined as
\[
D(P||Q) = \int \log \frac{dP}{dQ} \, dP, \quad \text{if } P << Q, \\
\infty \text{ otherwise},
\]
where \( P << Q \) denotes that \( P \) is absolutely continuous w.r.t. \( Q \). More explicitly, at node \( l \), let
\[
\delta_{i,l} = E_i \left[ \log \frac{f_{l,i}(X_{i,l})}{f_{0,l}(X_{i,l})} \right], \quad \rho_{l,i}^2 = \text{Var}_{H_i} \left[ \log \frac{f_{l,i}(X_{i,l})}{f_{0,l}(X_{i,l})} \right].
\]
Then \( \delta_{i,l} = D(f_{l,i}||f_{0,i}) \) and \( \delta_{0,i} = -D(f_{0,i}||f_{l,i}) \). We will assume \( \delta_{i,l} \) finite throughout this paper. Sometimes we will also need \( \rho_{l,i}^2 < \infty \). When the true hypothesis is \( H_1 \), by Jensen’s Inequality, \( \delta_{i,l} > 0 \) and when it is \( H_0 \), \( \delta_{i,l} < 0 \).

At secondary node \( l \), SPRT sum \( \{W_{k,l}, k \geq 0\} \) is a random walk with drift given by \( \delta_{i,l} \) under the true hypothesis \( H_i \).

Let
\[
N_i = \inf \{ k : W_{k,l} \notin (-\gamma_{0,l}, \gamma_{1,l}) \}, \quad N_i^1 = \inf \{ k : W_{k,l} \geq \gamma_{1,l} \} \text{ and } N_i^0 = \inf \{ k : W_{k,l} \leq -\gamma_{0,l} \}.
\]
Then \( N_i = \min\{N_i^1, N_i^0\} \). Also let \( N^0 = \inf \{ k : F_k \leq -\beta_0 \} \) and \( N^1 = \inf \{ k : F_k \geq \beta_1 \} \). Then stopping time of DualSPRT, \( N = \min\{N^1, N^0\} \).

For simplicity in the rest of the section, we take \( \gamma_{1,l} = \gamma_1, \beta_1 = \beta, b_1 = -b_0 = b \) and \( \mu_1 = \mu_0 = \mu \). Of course the analysis will carry over for the general case.

For convenience we summarize the important notation used in this paper in Table I. Notation specific to some algorithms are also mentioned separately.

1) \( E_{DD} \) Analysis: At the fusion node \( F_k \) crosses \( \beta \) under \( H_1 \) when a sufficient number of local nodes transmit \( b_1 \). The dominant event occurs when the number of local nodes transmitting are such that the mean value of the increments of the sum \( F_k \) will just have turned positive. In the following we find the mean time to this event and then the time to cross \( \beta \) after this. The \( E_{DD} \) analysis is same under hypothesis \( H_0 \) and \( H_1 \). Hence we provide the analysis for \( H_1 \).

The following lemmas provide justification for considering only the events \( \{N_i^1\} \) and \( \{N_i^1\} \) for analysis of \( E_{DD} = E_i[N] \).

Lemma 1. For \( i = 0,1 \), \( P_i(N_i = N_i^1) \to 1 \) as \( \gamma \to \infty \) and \( P_i(N_i = N_i^1) \to 1 \) as \( \gamma \to \infty \) and \( \beta \to \infty \).

Proof: From random walk results ([9], Chapter IV) we know that if a random walk has negative drift then its maximum is finite with probability one. This implies that \( P_i(\gamma_i < \infty) \to 0 \) as \( \gamma \to \infty \) for \( i \neq j \) but \( P_i(\gamma_i < \infty) = 1 \) for any \( \gamma < \infty \). Thus \( P_i(N_i = N_i^1) \to 1 \) as \( \gamma \to \infty \). This also implies that as \( \gamma \to \infty \), the mean of increments of \( F_k \) is positive for \( H_1 \) and negative for \( H_0 \). Therefore, \( P_i(N_i = N_i^1) \to 1 \) as \( \gamma \to \infty \) and \( \beta \to \infty \).

Lemma 2. Under \( H_i, \ i = 0,1 \) and \( j \neq i \),
\[ (a) \ |N_i - N_i^1| \to 0 \text{ a.s. as } \gamma \to \infty \text{ and } \lim_{\gamma \to \infty} \frac{N_i^1}{\gamma} = \lim_{\gamma \to \infty} \frac{N_i^1}{\gamma} = \frac{1}{D(f_{i,i}||f_{j,i})} \text{ a.s. and in } L^1, \]
\[ (b) \ |N_i - N_i^1| \to 0 \text{ a.s. and } \lim_{\gamma \to \infty} \frac{N_i^1}{\gamma} = \lim_{\gamma \to \infty} \frac{N_i^1}{\gamma} = \text{a.s. and in } L^1, \]
as \( \gamma \to \infty \) and \( \beta \to \infty \).
Thus when $\gamma$ is large, we can approximate $E_t[N_t]$ by $\gamma D(f_{1,1||f_0,1})$. Also under $H_1$, by central limit theorem for the first passage time $N_t^1$ (Theorem 5.1, Chapter III in [9]),

$$N_t^1 \sim \mathcal{N}(\frac{\gamma}{\delta_{t,1}}, \frac{\rho_{1,1}^2 \gamma}{\delta_{t,1}^2}),$$

where $\mathcal{N}(a,b)$ denotes Gaussian distribution with mean $a$ and variance $b$. From Lemma 2 we can use this result for $N_t$ also. Similarly we can obtain the results under $H_0$ and at the fusion node. Let $\delta_{t,FC}^i$ be the mean of increments of the fusion center test sum $F_{k, i}$, under $H_1$, when $j$ local nodes are transmitting. Let $t_j$ be the point at which the mean of increments of $F_{k, i}$ changes from $\delta_{t,FC}^i$ to $\delta_{t,FC}^{i+1}$ and let $F_j = E[F_{t_j-1, i}]$, the mean value of $F_{k, i}$ just before transition epoch $t_j$. The following lemma holds.

**Lemma 3.** Under $H_i$, $i = 0, 1$, as $\gamma \to \infty$,

$$P_t(\text{Decision at time } t_k = H_i) \quad \text{and} \quad \tau_k$$

is the $k$th order statistics of $N_t^1, N_t^2, \ldots, N_t^L \to 1$.

**Proof:** From Lemma 1, $P_t(\text{Decision at time } t_k = H_i)$ and $\tau_k$ is the $k$th order statistics of $N_t^1, N_t^2, \ldots, N_t^L \geq P_t(N_t^1 < N_t^j, j \neq i, \quad 1 \leq l \leq \ldots, L) \to 1$, as $\gamma \to \infty$.

We use Lemma 1 and equation 5 in the following to obtain an approximation for $EDD$ when $\gamma$ and $\beta$ are large. Large $\gamma$ and $\beta$ are needed for small probability of error. Then we can assume that the local nodes are making correct decisions. Although $F_{k, i}$ is a random walk before $t_j$, it is not so between $t_j$ and $t_{j+1}$ for $j = 1, \ldots, L$. But we assume that in the following approximation.

Let

$$l^* = \min\{j : \delta_{t,FC}^j > 0 \text{ and } \beta - \frac{F_j}{\delta_{t,FC}^j} < E[t_{j+1}] - E[t_j]\},$$

$F_j$ can be iteratively calculated as

$$F_j = F_{j-1} + \delta_{t,FC}^j \left(E[t_j] - E[t_{j-1}]\right), \quad F_0 = 0.$$ (6)

Note that $\delta_{t,FC}^j$ ($0 \leq j \leq L$) can be found by assuming $E_t[Y_k]$ as $bj$ and $t_j$ as the $j$th order statistics of $\{N_t^1, 0 \leq l \leq L\}$. The Gaussian approximation [5] can be used to calculate the expected value of the order statistics using the method given in [3]. This implies that $E[t_j]$’s and hence $F_j$’s are available offline. By using these values $E_{DD}$ ($\approx E_1(1)$) can be approximated as

$$E_{DD} \approx E[t_{l^*}] + \frac{\beta - F_{l^*}}{\delta_{t,FC}^{l^*}},$$ (7)

where the first term on R.H.S. is the mean time till the mean of increments becomes positive at the fusion node while the second term indicates the mean time for $F_{k, i}$ to cross $\beta$ from $t_{l^*}$ onwards.

2) $P_{MD}/P_{FA}$ Analysis: We provide analysis under $H_1$. $P_{FA}$ analysis is same as that of $P_{MD}$ analysis with obvious changes. When the thresholds at local nodes are reasonably large, according to Lemma 3 with a large probability local nodes are making the right decisions and $t_k$ can be taken as the order statistics assuming that all local nodes make the right decisions. Then for missed detection the dominant event

<table>
<thead>
<tr>
<th>Notation</th>
<th>Meaning</th>
</tr>
</thead>
<tbody>
<tr>
<td>$L$</td>
<td>Number of CRs</td>
</tr>
<tr>
<td>$X_{k,l}$</td>
<td>Observation at CR $l$ at time $k$</td>
</tr>
<tr>
<td>$Y_{k,l}$</td>
<td>Transmitted value from CR $l$ to FC at time $k$.</td>
</tr>
<tr>
<td>$b_l$</td>
<td>Channel gain of the $l$th CR</td>
</tr>
<tr>
<td>$Z_k$</td>
<td>FC MAC noise at time $k$</td>
</tr>
<tr>
<td>$f_{i,1}$, $g_n$</td>
<td>PDF of $X_{1,1}$ under $H_1$, PDF of $Z_k + \mu$</td>
</tr>
<tr>
<td>$W_{k,l}$</td>
<td>Test statistic at CR $l$ at time $k$.</td>
</tr>
<tr>
<td>$F_k$</td>
<td>Test statistic at FC at time $k$ (1)</td>
</tr>
<tr>
<td>$F_k^a$, $F_k^b$</td>
<td>Test statistics at FC (2)</td>
</tr>
<tr>
<td>$\xi_k$</td>
<td>LLR at FC (1)</td>
</tr>
<tr>
<td>$\xi_k^+$</td>
<td>LLR when all CR’s transmit wrong decisions (1)</td>
</tr>
<tr>
<td>$\tau_k$</td>
<td>Worst case value of $E_t[\xi_k] = E_t[\xi_k^+]$ (1)</td>
</tr>
<tr>
<td>$F_{t,1}$, $F_{t,n}$</td>
<td>$\sum_{i=1}^{n} \xi_i^+ (\sum_{i=1}^{\gamma} \xi_i^+)$ (1)</td>
</tr>
<tr>
<td>$A^1, \Delta(A^1)$</td>
<td>[all CRs transmit $b_i$ under $H_j$}, $E_t[\xi_k A^1]$ (1)</td>
</tr>
<tr>
<td>$\gamma_l, \gamma_0, l$</td>
<td>Thresholds at CR $l$ (1, 2)</td>
</tr>
<tr>
<td>$g(l)$</td>
<td>Threshold at CR (3, 4)</td>
</tr>
<tr>
<td>$\beta_l, \beta_0$</td>
<td>Thresholds at FC</td>
</tr>
<tr>
<td>$\mu_l, \mu_0$</td>
<td>Design parameters in FC LLR</td>
</tr>
<tr>
<td>$b_l, b_0$</td>
<td>Transmitting values to the FC at CR $l$ (1, 3)</td>
</tr>
<tr>
<td>$b_l^+$</td>
<td>Transmitting values to the FC at CR $l$ (2, 4)</td>
</tr>
<tr>
<td>$N$</td>
<td>First time $F_k$ crosses $-\beta_0, \beta_1$ (1)</td>
</tr>
<tr>
<td>$N_{t, N^0}$</td>
<td>First time $F_k$ crosses $-\beta_0$ (1)</td>
</tr>
<tr>
<td>$N_{t, N^1, N^0}$</td>
<td>Corresponding values of $N, N^1, N^0$ at CR $l$ (1)</td>
</tr>
<tr>
<td>$N_{l, (g,c)}$</td>
<td>First time $W_{k,l}$ crosses $g(c)$ $H_1$ at time $l$, $n$</td>
</tr>
<tr>
<td>$\tau_0, \tau_0$</td>
<td>First time $F_k^0$ crosses $-\beta_0, F_k$ crosses $\beta_1$ (2)</td>
</tr>
<tr>
<td>$\delta_{t,FC}$</td>
<td>Mean and variance of LLR at CR $l$ under $H_1$</td>
</tr>
<tr>
<td>$\delta_{t,FC}^i$</td>
<td>Mean and variance of LLR at FC under $H_i$ when $j$ CRs transmit</td>
</tr>
<tr>
<td>$t_j$</td>
<td>Time epoch when $\delta_{t,FC}^i$ changes to $\delta_{t,FC}^{i+1}$ (1)</td>
</tr>
<tr>
<td>$T_j$</td>
<td>Time epoch when $\delta_{t,FC}^i$ changes to $\delta_{t,FC}^{i+1}$ (2)</td>
</tr>
<tr>
<td>$F_j$</td>
<td>$E[F_{t_j-1}]$</td>
</tr>
<tr>
<td>$D_{tot}, D_{tot}^1$</td>
<td>$\sum_{l=1}^{L} D_{f_0, f_1</td>
</tr>
<tr>
<td>$r_0, \mu$</td>
<td>$D_{f_0, f_1</td>
</tr>
<tr>
<td>$\tau(c)$</td>
<td>Last time RW with drift $\delta_0$, will be above $-\log c$</td>
</tr>
<tr>
<td>$R_i$</td>
<td>$\min_{1 \leq l \leq L} \tau(c)$</td>
</tr>
<tr>
<td>$G_c, G_1, g_1, g_1$</td>
<td>CDF of $</td>
</tr>
<tr>
<td>$\Lambda_c, \Lambda_a, \Lambda_a$</td>
<td>sup (d - $\log g_1$), sup (d - $\log g_1$)</td>
</tr>
<tr>
<td>$\alpha_1^2, \alpha_1^2$</td>
<td>ess sup $</td>
</tr>
<tr>
<td>$R_c(\delta)$</td>
<td>Bayes Risk of test $\delta$ with cost $c$</td>
</tr>
<tr>
<td>$\nu(\alpha)$</td>
<td>First time RW</td>
</tr>
</tbody>
</table>

| TABLE I: List of important notations. 1- DualSPRT, 2- SPRT-CSRT, 3- GLR-SPRT, 4- GLR-CSRT |
is $P_{1}(N^{0} < t_{1})$. Also for reasonable performance we should select thresholds such that $P_{1}(N^{1} < t_{1})$ is small. Then
\[ P_{MD} = P_{1}(N^{0} < N^{1}) \geq P_{1}(N^{0} < t_{1}, N^{1} > t_{1}) \]
\[ \approx P_{1}(N^{0} < t_{1}). \]  
(8)

Under the above conditions, this lower bound should give a good approximation. In the following, we get an approximation for this.

Let $\xi_{k} = \log \left[ g_{\mu_{k}}(Y_{k}) / g_{\mu_{0}}(Y_{k}) \right]$. Then $F_{k} = \xi_{1} + \xi_{2} + \ldots + \xi_{k}$ and if we assume that $\xi_{k}$ before $t_{1}$, has mean zero and has distribution symmetric about zero (e.g., $\sim N(0, \sigma^{2})$) then,
\[ P_{1}(\text{reject } H_{1} \text{ before } t_{1}) \]
\[ \approx \sum_{k=1}^{\infty} P_{1}\left\{ F_{k} < -\beta \cap \sum_{n=1}^{k-1} \{ F_{n} > -\beta \} \right\} t_{1} > k \right\} \right] P[t_{1} > k] \]
\[ = \sum_{k=1}^{\infty} \left( P_{1}\left[ \sum_{n=1}^{k-1} \{ F_{n} > -\beta \} \right] \right) (1 - \Phi_{t_{1}}(k)) \]
\[ (A) \sum_{k=1}^{\infty} \left( P_{1}[F_{k} < -\beta | F_{k-1} > -\beta] \right) \]
\[ P_{1}\left[ \inf_{1 \leq n \leq k-1} F_{n} > -\beta \right] \right] (1 - \Phi_{t_{1}}(k)) \]
\[ (B) \sum_{k=1}^{\infty} \left( \int_{c=0}^{\infty} P_{1}[\xi_{k} < -c | F_{k-1} > -\beta] dc \right) \]
\[ \left( 1 - 2P_{1}[F_{k-1} \leq -\beta] \right) (1 - \Phi_{t_{1}}(k)), \]
where $\Phi_{t_{1}}$ is the Cumulative Distribution Function of $t_{1}$. Since we are considering only $F_{k}, k \leq t_{1}$, we remove the dependencies on $t_{1}$. In the above equations (A) is because of the Markov property of the random walk and (B) is due to the following lemma. This lemma can be obtained from [4, p. 525].

**Lemma 4.** If $\xi_{1}$ has mean zero and distribution symmetric about zero,
\[ P\left[ \inf_{1 \leq n \leq k-1} F_{n} > -\theta \right] \geq 1 - 2P\left[ F_{k-1} \leq -\theta \right]. \]

Similarly we can write an upper bound by replacing $P[\sum_{n=1}^{k-1} \{ F_{n} > -\beta \}]$ with $P[F_{k-1} > -\theta]$. We can make the lower bound tighter if we do the same analysis for the random walk between $t_{1}$ and $t_{2}$ with appropriate changes and add to the above bounds.

3) **Example 1:** We apply the DualSPRT on the following example and compare the $E_{DD}$ and $P_{FA}$ via analysis provided above with the simulation results. We assume that $f_{0}$ and $f_{1}$ are Gaussian with different means. This model is relevant when the noise and interference are log-normally distributed (23), and when $X_{k,l}$ is the sum of energy of a large number of observations at the secondary nodes at a low SNR.

Parameters used for simulation are as follows: $L = 5$, $f_{0} \sim N(0, 1)$ and $f_{1} \sim N(1, 1)$. Also $f_{0} = f_{0,l}$ and $f_{1} = f_{1,l}$ for $1 \leq l \leq L$, and $b = 1$. We plot $P_{E}(=P_{FA}$ under $H_{0}$ and $P_{MD}$ under $H_{1})$ and $E_{DD}(=E_{1}[N]$ or $E_{0}[N])$ versus $\beta$ in Figure 1. Here $\gamma$, $\mu$ and $b$ are fixed for ease of calculation and they are chosen to provide good performance for the given

$P_{MD}/P_{FA}$. The figure also contains the results obtained via analysis. We see a good match in theory and simulations. For comparison, Figure 1 also contains asymptotic results which are presented in Section IV below.

Fig. 1: DualSPRT-Comparison between theory and simulation

The above example is for the case when $X_{k,l}$ have the same distribution for different $l$ under the hypothesis $H_{0}$ and $H_{1}$. However in practice the $X_{k,l}$ for different local nodes $l$ will often be different because their receiver noise can have different variances and/or the path losses from the primary transmitter to the secondary nodes can be different. An example is provided here to illustrate the application of the above analysis to such a scenario. Now the order statistics $t_{1}$, in (7) needs to be appropriately computed.

4) **Example 2:** There are five secondary nodes with primary to secondary channel gain being 0, −1.5, −2.5, −4, and −6 dB respectively (corresponding post change means are 1, 0.84, 0.75, 0.63, 0.5). $f_{0} \sim N(0, 1)$, $f_{0,l} = f_{0,l}$ for $1 \leq l \leq 5$. Figure 2 provides the $E_{DD}$ and $P_{FA}$ via analysis and simulations. We see a good match.

Unlike normal SPRT for i.i.d. observations where bounds are available on error probabilities based on the thresholds, here, in DualSPRT, the desired error probabilities are complicated functions of $\gamma_{0,l}, \gamma_{1,l}, \beta_{0}, \beta_{1}, b_{0}, b_{1}, \mu_{1}$ and $\mu_{0}$. From the analysis provided in this section, it is possible to provide beforehand, at least approximately, the set of values for thresholds to achieve desired error probabilities and these can be used to design the test.

IV. ASYMPTOTIC PROPERTIES OF DUALSPRT

In this section we prove asymptotic properties of DualSPRT. We use the following notation:
\[ D_{tot}^{0} = \sum_{l=1}^{L} D(f_{0,l}||f_{1,l}), \]
\[ D_{tot}^{1} = \sum_{l=1}^{L} D(f_{1,l}||f_{0,l}), \]
Let $\mathcal{A}^i$ be the event that all the secondary users transmit $b_i$ when the true hypothesis is $H_i$. Also let $\Delta(\mathcal{A}^i)$ be the mean of increments of $F_k$ when $\mathcal{A}^i$ happens, i.e., $\Delta(\mathcal{A}^i) = E_i (\log \frac{g_{i0}(Y_k)}{g_{i0}(Y_k)} | A^i)$. We use $\bar{\gamma}_i$ as the mean of the increments of $F_k$ when all the local nodes transmit wrong decisions under $H_i$.

In the rest of this section, local node thresholds are $\gamma_{0,i} = -r_i |\log c|, \gamma_{1,i} = \rho_i |\log c|$ and fusion center thresholds are $\beta_0 = -|\log c|, \beta_1 = |\log c|.

We will also need,

$$
\tau_1(c) \overset{\Delta}{=} \sup \{ n \geq 1 : W_{n,i} \geq -r_i |\log c| \},
\tau(c) \overset{\Delta}{=} \max_{1 \leq i \leq L} \tau_i(c).
$$

Let $F_k^n$ be another likelihood ratio sequence (at FC) with expected value of its components as $\bar{\gamma}_i$ under $H_i$, the worst case value of the mean of the increments of $F_k$. Let the increments of $F_k^n$ be $\xi_1^n, \ldots, \xi_L^n$ which are i.i.d. Under $H_0$, $\bar{\theta}_0 > 0$ and under $H_1$, $\bar{\theta}_1 < 0$.

**Theorem 1.** For all $l$ and for some $\alpha > 1$, let $E_i[|\xi_1^n|^{\alpha + 1}] < \infty$ and $E_i \left[ \left| \log \frac{f_{i0}(X_{1,l})}{f_{i0}(X_{1,l})} \right|^{\alpha + 1} \right] < \infty, i = 0, 1$. Then, under $H_i$,

$$
\lim_{c \to 0} \frac{N}{c |\log c|} \leq \frac{1}{D_{tot}^l} + M_i \text{ a.s. and in } L_1,
$$

where $M_i = C_i / \Delta(\mathcal{A}^i), C_0 = -\left[ 1 + \frac{E_0[|\xi_1^n|]}{D_{tot}^l} \right]$ and $C_1 = \left[ 1 + \frac{E_1[|\xi_1^n|]}{D_{tot}^l} \right].$

**Proof:** See Appendix A.

**Remark 1.** When $\alpha_i = \infty$ which is generally true, $A_i(\alpha_i) = \infty$ (19) and in Theorem 2a and 2b we need to consider only $\bar{R}_i < A_i(\alpha_i)$. Also $A_i$ can be computed from $A_i$ using Contraction principle in Large Deviation theory (6).

**Theorem 2.** Let $g_i(\lambda) < \infty$ in a neighbourhood of zero. Then,

(a) $\lim_{c \to 0} \frac{P_{FA}}{c} = 0$ if for some $0 < \eta < R_0, s_0(\eta) > 1$.

(b) $\lim_{c \to 0} \frac{P_{MD}}{c} = 0$ if for some $0 < \eta < R_1, s_1(\eta) > 1$.

**Proof:** See Appendix B.

**Remark 2.** In (2) Lemma 1-Appendix A, it is proved that log likelihood ratio converts a large class of distributions into light tailed distributions and then $g_i(\lambda)$ is finite in a neighbourhood of zero. For instance, consider a regularly varying distribution for $Z_k, P(Z_k > x) = t'(x)x^{-\alpha},$ where $t'(x)$ is a slowly varying function and $\alpha > 0$. Then, $\log g_{y_0}(x)/g_{y_0}(x) = (l(x - \mu_1)/(l(x + \mu_0))(x - \mu_1)^{-\alpha}(x + \mu_0)^{\alpha} \leq x^{\beta_1 + \beta_2}$ for large $x$, any $\beta_0 > 0$ and an appropriately chosen $\beta_2 > 0$. This proves the conditions for (2) Lemma 1 and hence exponential tail for $G_i(t)$ follows.

We compare the asymptotic results obtained in Theorems 1 and 2 with that of SPRT with all the data available at the local nodes centrally without noise. Let $N_o$ be the stopping time of such an SPRT. Then, from (3) Theorem 2.11.1 and 2.11.2,

$$
\lim_{c \to 0} \frac{E_i[N_o]}{|\log c|} = \frac{1}{D_{tot}^l}.
$$

Theorem 2 implies the asymptotics (12) on $P_{FA}$ and $P_{MD}$ for DualSPRT. Comparing Theorem 1 with (11), we see that the rates of convergence of DualSPRT are optimal. For the limits to equal, we need $M_0$ and $M_1$ to be zero. In Section IV-A we compute $M_0$ and $M_1$ for Gaussian fusion center noise.

We can consider the asymptotic performance in the Bayesian framework also. Then the two hypotheses $H_0$ and $H_1$ are assumed to have known prior probabilities $\pi \text{ and } 1 - \pi$ respectively. A cost $c (\geq 0)$ is assigned to each time step taken for decision. Let $W_i > 0, i = 0, 1$ be the cost of falsely
rejecting $H_0$. Then Bayes risk of a test $\delta$ with stopping time $N$ is defined as,

$$R_r(\delta) = \pi[cE_0(N) + W_0P_0\{\text{reject } H_0\}] + (1 - \pi)[cE_1(N) + W_1P_1\{\text{reject } H_1\}]$$

(13)

Optimising (13) makes sense even when one does not have prior $\pi$ (i.e., within the frequentist framework) because then taking $\pi$ and $W_i$ appropriately, one can think of selecting a decision rule that asymptotically minimizes a weighted sum of $E[c \pi(N)]$ and $P_1\{\text{reject } H_1\}$, $i = 1, 2$.

Let $R_r(\delta_{\text{cent.}})$ and $R_r(\delta_{\text{DualSPRT}})$ be the Bayes’s Risk of the optimal centralized SPRT without considering fusion center noise and of DualSPRT respectively. Then, (14) p. 2076,

$$\lim_{c \rightarrow 0} \frac{R_r(\delta_{\text{cent.}})}{c \log |c|} = \left(\frac{\pi}{D_{\text{tot}}^0} + 1 - \pi \right) \frac{1}{D_{\text{tot}}^1}$$

From Theorem 1 and Theorem 2(a) and 2(b), using (13), for DualSPRT with fusion center noise,

$$\lim_{c \rightarrow 0} \frac{R_r(\delta_{\text{DualSPRT}})}{c \log |c|} = \left(\frac{\pi}{D_{\text{tot}}^0} + 1 - \pi \right) \frac{1}{D_{\text{tot}}^1} + C$$

where $C = M_0\pi + M_1(1 - \pi)$. The constant $C$ can be made arbitrarily small by making $M_0$ and $M_1$ small.

A. Example-Gaussian distribution

In the following we apply Theorems 1 and 2 when the fusion center noise is Gaussian $N(0, \sigma_F^2)$. We take $\mu_1 = \mu_0 = \mu > 0$ and $b_1 = -b_0 = b > 0$. For Theorem 1 $\Delta(\alpha) = -2\mu b/\sigma_F^2$ and $\Delta(\lambda) = 2\mu b/\sigma_F^2$. Therefore $M_0$ and $M_1$ in Theorem 1 $\rightarrow 0$ if $L \rightarrow \infty$ and/or $b \rightarrow \infty$. This also happens if $\sigma_F^2 \rightarrow 0$.

Using Remark 1 the condition in Theorem 2(a) is $\sigma_F^2 \eta/(4\mu^2 \sqrt{2\eta} - 2\mu b)$ for some $0 < \eta < R_0$ and that for Theorem 2(b) is $\sigma_F^2 \eta/(4\mu^2 \sqrt{2\eta} + 2\mu b)$ for some $0 < \eta < R_1$. Combining these two, it is sufficient to satisfy later condition with $0 < \eta < \min(R_0, R_1)$. For Gaussian input observations at the local nodes, assuming $f_{1,l} = f_1, f_{0,l} = f_0$ for $1 \leq l \leq L$, we get $\delta_{l,i} = \delta_i$ and $\rho_{l,i} = \rho_i, R_i = \frac{\delta_i^2}{2\rho_i^2}$. This specifies upper-bounds for the choice of $\mu, L$ and $b$.

V. IMPROVED DECENTRALIZED SEQUENTIAL TESTS:

SPRT-CSPRT

This section considers some improvements over DualSPRT. The improved algorithms are theoretically analysed and their performance is compared with existing decentralized schemes.

New Algorithms: SPRT-CSPRT and DualCSPRT

In DualSPRT presented in Section III-A, observations $\{Y_k\}$ to the fusion center are not always identically distributed. Till the first transmission from secondary nodes, these observations come from i.i.d. noise distribution, but not after that. Since the non-asymptotic optimality of SPRT is known for i.i.d. observations only (22), using SPRT at the fusion center is not optimal.

We improve DualSPRT with the following modifications. Steps (1)-(3) (corresponding to the algorithm run at the local nodes) are same as in DualSPRT. The steps (4) and (5) are replaced by:

4) Fusion center runs two algorithms:

$$F_k^1 = (F_{k-1}^1 + \log |g_{\mu_1}(Y_k)/g_{\mu_2}(Y_k)|)^+, \ F_{k}^0 = 0, \ (14)$$

$$F_k^0 = (F_{k-1}^0 + \log |g_{\rho_1}(Y_k)/g_{\rho_2}(Y_k)|), \ F_{k}^0 = 0, \ (15)$$

where $(x)^+ = \max(0, x), (x)^- = \min(0, x)$, $\mu_1$ and $\mu_0$ are positive constants, $g_{\mu}$ is the pdf of i.i.d. noise $\{Z_k\}$ at the fusion center and $g_\mu$ is the pdf of $\mu + Z_k$.

5) The fusion center decides about the hypothesis at time

$\inf\{k : F_k^1 \geq \beta_1 \text{ or } F_k^0 \leq -\beta_0\}$

and $\beta_0, \beta_1 > 0$. The decision is $H_1$ if $F_k^1 \geq \beta_1$ and $H_0$ if $F_k^0 \leq -\beta_0$.

The following discussion provides motivation for this test.

1) If the SPRT sum defined in (2) goes below zero it delays in crossing the positive threshold $\beta_1$. Hence if we keep SPRT sum at zero whenever it goes below zero, it reduces $E_{DD}$. This happens in CUSUM (15). Similarly one can use a CUSUM statistic under $H_0$ also. These ideas are captured in (14) and (15).

2) The proposed test is also capable of reducing false alarms caused by noise $Z_k$ before first transmission at $t_1$ from the local nodes. For $F_k^1$ and $F_k^0$ to move away from zero, the mean of increments should be positive and negative respectively. Let $\hat{\mu}_k = E[Y_k]$ at time $k$. Then,

$$E[\hat{\mu}_k] \left[ \log \frac{g_{\mu_1}(Y_k)}{g_{\mu_2}(Y_k)} \right] = D(g_{\hat{\mu}_k} || g_\mu) - D(g_{\hat{\mu}_k} || g_{\mu_1}). \ (16)$$

Hence before $t_1$, positive mean value of increments is not possible. After $t_1$ under $H_1$ (assuming the local nodes make correct decisions, the justification for which is provided in Section III), the mean of increments becomes more positive. Similarly for $F_k^0$. But in case of DualSPRT, SPRT sum at the fusion center has the increments given by $\log y_{\rho_1}(Y_k)/y_{\rho_2}(Y_k)$. This is difficult to keep zero only before $t_1$ and thus creates more errors due to noise $Z_k$.

3) Even though the problem under consideration is hypothesis testing, this is essentially a change detection problem at the fusion center. The observations at the fusion center have the distribution of noise before $t_1$ and after $t_1$ the mean changes. But in our scenario, this is a composite sequential change detection problem with the observations that are not i.i.d. and we look for change in both directions, it is difficult to use existing algorithms available for sequential change detection. Nevertheless our test (14)-(15) provides a guaranteed performance in this scenario.

We consider one more improvement. When a local Cognitive Radio SPRT sum crosses its threshold, it transmits $b_1/b_0$. This node transmits till the fusion center SPRT sum crosses the threshold. If it is not a false alarm, then its SPRT sum keeps on increasing (decreasing). But if it is a false alarm, then the sum will eventually move towards the other threshold. Hence instead of transmitting $b_1/b_0$ the Cognitive Radio can transmit a higher / lower value in an intelligent fashion. This should improve the performance. Thus we modify step (3) in
DualSPRT as,

\[ Y_{k,l} = \sum_{i=1}^{4} b_{i}^{l}\{W_{k,l} \in [\gamma_{1} + (i - 1)\Delta_{1}, \gamma_{1} + i\Delta_{1}]\} + \]

\[ b_{0}^{l}\{W_{k,l} \in [-\gamma_{0} - (i - 1)\Delta_{0}, -\gamma_{0} - i\Delta_{0}]\} \quad (17) \]

where \( \Delta_{1} \) and \( \Delta_{0} \) are the parameters to be tuned at the Cognitive Radio. \( 4\Delta_{1} \) and \( 4\Delta_{0} \) are taken as \( \infty \). The drift under \( H_{1} (H_{0}) \) is a good choice for \( \Delta_{1} (\Delta_{0}) \).

We call the algorithm with the above two modifications as SPRT-CSPRT (with ‘C’ as an indication about the motivation from CUSUM).

If we use CSPRT at both the secondary nodes and the fusion center with the proposed quantisation methodology (we call it DualCSPRT) it works better as we will show via simulations in Section V-A. In Section V-B we will theoretically analyse SPRT-CSPRT. As the performance of DualCSPRT (Figure 3a) is close to that of SPRT-CSPRT, we analyse only SPRT-CSPRT.

A. Performance Comparison

Throughout the rest of this section we use \( \gamma_{l,t} = \gamma_{0,l} = \gamma \), \( \beta_{1} = \beta_{0} = \beta \) and \( \mu_{1} = \mu_{0} = \mu \) for the simplicity of simulations and analysis.

We apply DualSPRT, SPRT-CSPRT and DualCSPRT on the following example and compare their \( E_{DD} \) for various values of \( P_{MD} \). We assume that the pre-change distribution \( f_{0} \) and the post change distribution \( f_{1} \) are Gaussian with different means.

For simulations we have used the following parameters. There are 5 nodes (\( L = 5 \)) and \( f_{0,l} \sim \mathcal{N}(0,1) \), for \( 1 \leq l \leq L \). Primary to secondary channel gains are 0, \(-1.5, -2.5, -4 \) and \(-6dB \) respectively (the corresponding post change means of Gaussian distribution with variance 1 are 0, 0.84, 0.75, 0.63 and 0.5). We assume \( Z_{k} \sim \mathcal{N}(0,5) \) and the mean of increments of DualSPRT and SPRT-CSPRT at the fusion center is taken as \( 2\mu Y_{k} \), with \( \mu \) being 1. We also take \( D_{0} = D_{1} = 0 \), \( \{b_{1}, b_{1}, b_{1}, b_{1}\} = \{1, 2, 3, 4\} \), \( \{b_{0}, b_{0}, b_{0}, b_{0}\} = \{-1, -2, -3, -4\} \) and \( b_{1} = -b_{0} = 1 \) (for DualSPRT). Parameters \( \gamma \) and \( \beta \) are chosen from a range of values to achieve a particular \( P_{FA} \). Figure 3a provides the \( E_{DD} \) and \( P_{MD} \) via simulations. We see a significant improvement in \( E_{DD} \) compared to DualSPRT. The difference increases as \( P_{MD} \) decreases. The performance under \( H_{0} \) is similar.

Performance comparisons with the asymptotically optimal decentralized sequential algorithms which do not consider fusion center noise (DSPRT [7], Mei’s SPRT [14]) are given in Figure 3b. Note that DualSPRT and SPRT-CSPRT include fusion center noise. We take \( f_{0,l} \sim \mathcal{N}(0,1), f_{1,l} \sim \mathcal{N}(1,1) \) for \( 1 \leq l \leq L \) and \( Z_{k} \sim \mathcal{N}(0,1) \). We find that the performance of SPRT-CSPRT is close to that of DSPRT (which is second order asymptotically optimal) and better than Mei’s SPRT (which is first order asymptotically optimal). Similar comparisons were obtained with other data sets.

B. Performance Analysis of SPRT-CSPRT

\( E_{DD} \) and \( P_{MD}/P_{FA} \) analysis is same under \( H_{1} \) and \( H_{0} \). Hence we provide analysis under \( H_{1} \) only.

1) \( P_{MD} \) Analysis: Between each change of mean of increments (which occurs due to the change in number of Cognitive Radios transmitting to the fusion node and due to the change in the value transmitted according to the quantisation rule (17)) at the fusion center, under \( H_{1} \), (14) has a positive drift and behaves approximately like a normal random walk. Under \( H_{1} \) (15) also has a positive drift, but due to the min in its expression it will stay around zero and as the event of crossing negative threshold is rare (15) becomes a reflected random walk between each drift change. Similarly under \( H_{0} \), (14) and (15) become reflected random walk and normal random walk respectively. The false alarm occurs when the reflected random walk crosses its threshold.

Under \( H_{1} \), let \( \tau_{\beta} \triangleq \inf\{k \geq 1 : F^{k}_{1} \leq -\beta\} \) and \( T_{\beta} \triangleq \inf\{k \geq 1 : F^{k}_{1} \geq \beta\} \).

Following the argument in Section III-B2 for \( P_{MD} \), we get,

\[ P_{1}(\text{reject } H_{1}) = P_{1}(\tau_{\beta} < T_{\beta}) \approx P_{1}(\tau_{\beta} < t_{1}). \]

\[ P_{1}(\tau_{\beta} < t_{1}) = \sum_{k=1}^{\infty} P_{1}(\tau_{\beta} \leq k, k < t_{1}) = \sum_{k=1}^{\infty} P_{1}(\tau_{\beta} \leq k | k < t_{1}) P_{1}(t_{1} > k). \quad (18) \]

In the following we compute \( P_{1}(\tau_{\beta} > x | \tau_{\beta} < t_{1}) \) and \( P_{1}(t_{1} > k) \). It is shown in (18) that,

\[ \lim_{\beta \to \infty} P_{1}(\tau_{\beta} > x | \tau_{\beta} < t_{1}) = \exp(-\lambda_{\beta}x), \quad x > 0, \quad (19) \]

where \( \lambda_{\beta} \) is obtained by finding solution to an integral equation obtained via renewal arguments (19). Let \( L(s) \) be the mean of \( \tau_{\beta} \) with \( F^{0}_{1} = s \) and \( S_{k} = \log \frac{g_{Y_{k}}(Y_{k})}{g_{Y_{k}}(Y_{k})} \). Note that \( \{S_{k}, k < t_{1}\} \) are i.i.d. From the renewal arguments,
by conditioning on $S_1 = z$,
\[ L(s) = P(S_1 > s)(L(0) + 1) \]
\[ + \int_{-\beta - s}^{-s} (L(s + z) + dF_S(z)) dz + F_S(-\beta - s), \]
where $F_S$ is the distribution of $S_k$ before the first transmission from the local nodes. This is a Fredholm integral equation of the second kind (5[20]). Existence of a unique solution for it is shown in [2]. By solving these equations numerically, we get $\lambda_\beta = 1/L(0)$.

From the central limit theorem approximation given in Section II-B, we can find the distribution of $t_1$. Thus [18] provides,
\[ P_i(\text{False alarm before } t_i) \approx \sum_{k=1}^{\infty} (1 - e^{-\lambda_i k}) L(1 - \Phi_N(k)), \]
where $\Phi_N$ is the Cumulative Distribution Function of $N_i$, obtained from the Gaussian approximation.

2) $E_{DD}$ Analysis: In this section we compute $E_{DD}$ theoretically. Recall that $t_i$ also approximates the first time at which $i$ local nodes are transmitting. Mean of $t_i$ can be computed from the method explained in [3], for finding $k^{th}$ central moment of non i.i.d. $t_i$ order statistics.

Between $t_i$ and $t_{i+1}$ the mean of the increments at the fusion center is not necessarily constant because there are four thresholds (each corresponds to different quantizations) at the secondary node. The transmitted value changes after crossing each threshold, $b_1^i \rightarrow b_2^i \rightarrow b^3_i$. Let $t_i^j, 1 \leq j \leq 3$ be the time points at which a node changes the transmitting values from $b_j^i$ to $b_{j+1}^i$ between $t_i$ and $t_{i+1}$. We assume that with a high probability the secondary node with the lowest first passage time mean will transmit first, the node with the second lowest mean will transmit second and so on. This is justified by the fact that the distribution of the first passage time of $\gamma > 0$ by a random walk with drift $\delta > 0$ and variance $\sigma^2$ is $\mathcal{N}(\frac{\gamma}{\delta}, \frac{\sigma^2}{\delta^2})$. Thus if $\delta$ is large, the mean $\gamma/\delta$ is small and the variance $\sigma^2/\gamma/\delta^3$ is much smaller. In the following we will make computations under these approximations. The time difference between $t_i^{th}$ and $t_{i+1}^{th}$ transmission can be calculated if we take the second assumption ($= \Delta_1/\delta_1$). We know $E[t_i]$ for every $i$ from an argument given earlier. Suppose $i^{th}$ node transmits at $t_i^{th}$ instant and if $E[t_i] + \Delta_1/\delta_1 < E[t_{i+1}]$ then $E[t_i] = E[t_i] + \Delta_1/\delta_1$. Similarly if $E[t_i] + \Delta_1/\delta_1 < E[t_{i+1}]$ then $E[t_i^2] = E[t_i]^2 + \Delta_1/\delta_1$ and so on. Let us represent the sequence $t = \{t_1, t_1^2, t_2, t_2^2, t_3, t_3^2, \ldots \}$ (entry only for existing ones by the above criteria) by $T = \{T_1, T_2, T_3, \ldots \}$.

Let $\delta_{FC}^k$ be the mean of the increments at the fusion center between $T_k$ and $T_{k+1}$, under $H_i$. Thus $T_k$’s are the transition epochs at which the mean of the increments of fusion center changes from $\delta_{FC}^{k-1}$ to $\delta_{FC}^k$. Also let $F_{FC} = E[F_{FC} - 1]$ be the mean value of $F_{FC}$ just before the transition epoch $T_k$. With the assumption of the very low $P_{fa}$ at the local nodes and from the knowledge of the sequence $t$ we can easily calculate $\delta_{FC}^k$ for each $T_k$. Similarly $F_{k+1} = F_{k} + \delta_{FC}^k(E[T_{k+1} - E[T_k])$. Then,
\[ E_{DD} \approx E[T_{i+}] + \frac{\beta - F_{i+}}{\delta_{FC}^i}, \tag{20} \]
where
\[ l^* = \min\{j : \delta_{FC}^j > 0 \text{ and } \frac{\beta - F_{j+}}{\delta_{FC}^j} < E[T_{j+1} - E[T_j]] \} \]

The above approximation of $E_{DD}$ is based on Central Limit Theorem and Law of Large Numbers and hence is valid for any distributions with finite second moments.

Figure 4 provides the comparison between simulation and analysis. We used the same setup as in Section V-A (with $Z_k \sim \mathcal{N}(0, 1)$). We see a reasonable approximation.

---

VI. UNKNOWN RECEIVED SNRs AND FADING

This section considers the extensions of DualSPRT and SPRT-CSPT to take care of the SNR uncertainty and the slow fading between the primary user and a Cognitive Radio. Since the transmissions from CR to FC are in CR network, we assume reporting channel to FC as AWGN only. This assumption is commonly made (11, 23).

A. Different and unknown SNRs

We consider the case where the received signal power from the PU to a CR node is fixed but not known to the local Cognitive Radio nodes. This can happen if the transmit power of the primary is not known and/or there is unknown shadowing. Now we limit ourselves to the energy detector where the observations $X_{k,l}$ are average energy of $M$ samples received by the $i^{th}$ Cognitive Radio node. Then for somewhat large $M$, the distributions of $X_{k,l}$ under $H_0$ and $H_1$ can be approximated by Gaussian distributions: $f_{0,l} \sim \mathcal{N}(\sigma_1^2, 2\sigma_1^2/M)$ and $f_{1,l} \sim \mathcal{N}(p_1 + \sigma_1^2, 2(p_1 + \sigma_1^2)^2/M)$, where $P_1$ is the
received power and $\sigma_l^2$ is the noise variance at the $l^{th}$ CR node. Under low SNR conditions $(P_l + \sigma_l^2)^2 \approx \sigma_l^2$ and hence $X_{k,l}$ are Gaussian distributed with mean change under $H_0$ and $H_1$. Now taking $X_{k,l} - \sigma_l^2$ as the data for the detection algorithm at the $l^{th}$ node, since $P_l$ is unknown we can formulate this problem as a sequential hypothesis testing problem with
\[ H_0 : \theta = 0; \quad H_1 : \theta \geq \theta_1, \]
where $\theta$ is $P_l$ under $H_1$ and $\theta_1$ is appropriately chosen.

The problem
\[ H_0 : \theta \leq \theta_0; \quad H_1 : \theta \geq \theta_1, \]
subject to
\[ P_0\{\text{reject} \ H_0\} \leq \alpha, \quad \text{for } \theta \leq \theta_0, \]
\[ P_\theta\{\text{reject} \ H_1\} \leq \beta, \quad \text{for } \theta \geq \theta_1, \]
for exponential family of distributions is well studied in [12]. The following algorithm of Lai [12] is asymptotically Bayes optimal and hence we use it at the local nodes instead of SPRT. Let $\theta \in A = [a_1, a_2]$. Define
\[ W_{n,l} = \max\left\{ \sum_{k=1}^{n} \log \frac{f_{\theta_0}(X_k)}{f_{\theta_1}(X_k)}, \sum_{k=1}^{n} \log \frac{f_{\theta_1}(X_k)}{f_{\theta_0}(X_k)} \right\}, \]
\[ N_l(g, c) = \inf\{n : W_{n,l} \geq g(n(c))\}, \]
where $g()$ is a time varying threshold and $c > 0$ is a design parameter. The function $g$ satisfies $g(t) \approx \log(1/t) t \rightarrow 0$ and is the boundary of an associated optimal stopping problem for the Wiener process [12]. $\hat{\theta}_l$ is the Maximum-Likelihood estimate of $\theta$ bounded by $a_1$ and $a_2$. For Gaussian $f_0$ and $f_1$, $\hat{\theta}_l = \max\{a_1, \min[S_n/n(a_2)]\}$. At time $N_l(g, c)$ decide upon $H_0$ or $H_1$ according as $\hat{\theta}_l \in \{0\}$ or $\hat{\theta}_l \in \{1\}$, i.e., $\theta^*$. $\theta^*$ is obtained by solving $D(f_0||f_0) = D(f_0||f_1)$. $\theta^*$ is a time varying threshold and $c > 0$ is a design parameter.

For our case where $H_0 : \theta = 0$, unlike in [12] where $H_0 : \theta \leq 0, E_0[N_l(g, c)]$ largely depends upon the value $\theta_1$. As $\theta_1$ increases, $E_0[N_l(g, c)]$ decreases and $E_1[N_l(g, c)]$ increases. If $P_l \in [P_D, \bar{P}]$ for all $l$ then a good choice of $\theta_1$, is $(P_D - \bar{P})/2$.

1) GLR-SPRT: First we modify DualSPRT. In the distributed setup with the received power at the local nodes unknown, the local nodes will use the Lai’s algorithm mentioned above while the fusion node runs the SPRT. All other details remain same. We call this algorithm GLR-SPRT.

2) GLR-CSPRT: This is a modified version of SPRT-CSPRT. Here, we modify GLR-SPRT to GLR-CSPRT with appropriate change in quantisation and using CSPRT at the fusion center instead of SPRT. The quantisation [17] is changed in the following way: if $\hat{\theta}_l \geq \theta^*$, let $I_1 = \{g(kc), g(kc + 3\Delta)\}$, $I_2 = \{g(kc + 3\Delta), g(kc + 2\Delta), g(kc + \Delta), \infty\}$, $Y_{k,l} = b_{l,k}$ if $W_{k,l} \in I_2$ for some $n$. If $\hat{\theta}_l \leq \theta^*$ we will transmit from $\{b_{l,0}, b_{l,1}, b_{l,2}, b_{l,3}\}$ under the same conditions. Here, $\Delta$ is a tuning parameter and $0 \leq 3\Delta \leq 1$.

The performance comparison of GLR-SPRT and GLR-CSPRT for the example in Section VI-A with $Z_k \sim N(0,1)$ is given in Figure [5]. Here $\Delta = 0.25$. As the performance under $H_1$ and $H_0$ are different, we give the values under both. We can see that GLR-SPRT is always inferior to GLR-CSPRT. For $E_{DD}$ under $H_1$, interestingly GLR-CSPRT has lesser values than that of SPRT-CSPRT for $P_{FA} = 0.02$ (note that SPRT-CSPRT has complete knowledge of the SNRs), while under $H_0$ it has higher values than SPRT-CSPRT.

Fig. 5: Comparison among SPRT-CSPRT, GLR-SPRT and GLR-CSPRT for different SNR’s between the Primary and the secondary users

B. Channel with Fading

In this section we consider the system where the channels from the primary transmitter to the secondary nodes have fading ($h_l \neq 1$). We assume slow fading, i.e., the channel coherence time is longer than the hypothesis testing time.

When the fading gain $h_l$ is known to the $l^{th}$ secondary node then this case can be considered as the different SNR case as in the example given in Section III-B4. Thus we consider the case where the channel gain $h_l$ is not known to the $l^{th}$ node.

We consider the energy detector setup of Section VI-A. However, now $P_l$, the received signal power at the local node $l$ is random. If the fading is Rayleigh distributed then $P_l$ has exponential distribution. The hypothesis testing problem becomes
\[ H_0 : f_{0,l} \sim N(0, \sigma^2); \quad H_1 : f_{1,l} \sim N(\theta, \sigma^2) \]
where $\theta$ is random with exponential distribution and $\sigma^2$ is the variance of noise. We will assume that $\sigma^2$ is known at the nodes.

We are not aware of this problem being handled via sequential hypothesis testing before. However we use Lai’s algorithm in Section VI-A where we take $\theta_1$ to be the median of the distribution of $\theta$, i.e., $P(\theta \geq \theta_1) = 1/2$. This seems a good choice for $\theta_1$ as a compromise between $E_0[N]$ and $E_1[N]$.

1) GLR-SPRT: First we apply the technique on GLR-SPRT. We use an example where $\sigma^2 = 1, \theta = \exp(1), \text{Var}(Z_k) = 1$, and $L = 5$. The performance of this algorithm is compared with that of DualSPRT (with perfect channel state information) in Figure [5]. We observe that under $H_1$, for high $P_{MD}$ this algorithm works better than DualSPRT with channel state information, but as $P_{MD}$ decreases DualSPRT becomes better and the difference increases. For $H_0$, GLR-SPRT is always worse and the difference is almost constant.
2) **GLR-CSPRT**: Figure 6 also provides comparison of DualSPRT, GLR-SPRT and GLR-CSPRT. Notice that the comment given for \( E_{DD} \) for Figure 5a is also valid here.

![Comparison among DualSPRT, GLR-SPRT and GLR-CSPRT with slow fading between the primary and the secondary users](image)

**VII. CONCLUSIONS**

This paper presents fast algorithms for cooperative spectrum sensing satisfying reliability constraints. We have presented and analysed DualSPRT, a decentralized sequential hypothesis test. Simulation results corroborate the theoretical study of DualSPRT. Asymptotic properties of DualSPRT are also explored and its performance can approach asymptotically Bayes optimal tests. Improvement over DualSPRT using CUSUM statistics for the fusion center test leads to another algorithm in which the selection of parameters is easy to choose apart from performance enhancement. We also provide approximate theoretical analysis of the algorithm. Numerical experiments show that this algorithm performs as well as an asymptotic order-2 optimal algorithm without fusion center noise, proposed in literature. We further extend our algorithms to cover the case of unknown SNR and channel fading and obtain satisfactory performance compared to perfect channel state information case.

**APPENDIX A**

**PROOF OF THEOREM [1]**

We will prove the theorem under \( H_0 \). The proof under \( H_1 \) will follow in the same way.

Let \( \nu(a) \) be the stopping time when a random walk starting at zero and formed by the sequence \( \{ \log \frac{g_{n+1}(x_k)}{g_n(x_k)} + (\Delta(A^0) - E_0[\log \frac{g_{n+1}(x_k)}{g_n(x_k)}]) \}, k \geq \tau(c) + 1 \} \) (with \(-\)ve drift under \( H_0 \)) crosses \( a \). Then,

\[
N \leq N^0 \leq \tau(c) + \nu(-|\log c| - F_{\tau(c)+1}^c).
\]

Therefore,

\[
\frac{N}{|\log c|} \leq \frac{\tau(c)}{|\log c|} + \frac{\nu(-|\log c| - F_{\tau(c)+1}^c)}{|\log c|}.
\]

We consider the first term on the R.H.S. of (24). From [9] Remark 4.4, p. 90 as \( c \to 0, \tau(c) \to \infty \) a.s. and

\[
\lim_{c \to 0} \frac{\tau(c)}{|\log c|} = \frac{1}{\delta_{0,l}} = \frac{1}{D_{tot}^0} \text{ a.s. Therefore,}
\]

\[
\frac{\tau(c)}{|\log c|} \to \max_{l} \left\{ \frac{-r_l}{\delta_{0,l}} \right\} = \frac{1}{D_{tot}^0} \text{ a.s.}
\]

Furthermore, from [11] proof of Theorem 1 (i) \( \Rightarrow \) (ii) p. 871, it can be seen that \{\( \tau(c)/|\log c| \)\} is uniformly integrable for each \( l \). Therefore, \{\( \tau(c)/|\log c| \)\} is also uniformly integrable and hence,

\[
E_0[\tau(c)] \to \frac{1}{D_{tot}^0}.
\]

The second term in R.H.S. of (24),

\[
\nu(-|\log c| - F_{\tau(c)+1}^c) \leq \nu(-|\log c|) + \nu(-F_{\tau(c)+1}^c).
\]

We know, from [9] Chapter III, as \( c \to 0 \),

\[
\nu(-|\log c|) \to -\frac{1}{\Delta(A^0)} \text{ a.s. and in } L_1.
\]

Next consider \( \nu(-F_{\tau(c)+1}^c) \). Let \( \hat{F}_k^c \) be a random walk formed from \( |\xi_k^c| \). It can be shown that \( \hat{F}_k^c \) stochastically dominates \( F_k \) and thus we can make \( \hat{F}_k^c \geq F_k \) a.s. for all \( k \geq 0 \). Then,

\[
\frac{\nu(-F_{\tau(c)+1}^c)}{|\log c|} \leq \nu(-\hat{F}_{\tau(c)+1}^c).
\]

Also,

\[
\frac{\hat{F}_{\tau(c)+1}^c}{|\log c|} = \frac{\hat{F}_{\tau(c)+1}^c}{\tau(c) + 1} \to E[|\xi_1^c|/D_{tot}^0] \text{ a.s.}
\]

Thus,

\[
\nu(-F_{\tau(c)+1}^c) \to -1 \frac{E[|\xi_1^c|]}{\Delta(A^0)} D_{tot}^0 \text{ a.s.}
\]

From (24), (25), (27), (28) and (29),

\[
\lim_{c \to 0} \frac{N}{|\log c|} \leq \frac{1}{D_{tot}^0} - \frac{1}{\Delta(A^0)} E_0[|\xi_1^c|/D_{tot}^0] \text{ a.s.}
\]

Now we show \( L_1 \) convergence. For \( \alpha > 1 \),

\[
E_0[\nu(-\hat{F}_{\tau(c)+1}^c)^{\alpha}] \to \frac{1}{|\log c|^{\alpha}} \int_0^{\log c} E_0[\nu(-x)^{\alpha}] dP_{\hat{F}_{\tau(c)+1}^c}(x)
\]
Therefore, \( \nu(\cdot) \) is uniformly integrable and hence, from (29),
\[
\lim_{c \to 0} E_0[\nu(-\tilde{F}_c^{*+})/|\log c|] = -\frac{1}{\Delta(A^c)} \frac{E[|\xi_0^*|]}{D_0^{\text{tot}}}. \]

This, with (24), (26), (27) and (28), implies that (since \( \epsilon \) can be taken arbitrarily small),
\[
\lim_{c \to 0} E_0[N] \leq \frac{1}{D_0^{\text{tot}}} + M_0,
\]
where \( M_0 = -\frac{1}{\Delta(A^c)} \left[ 1 + E_0[|\xi_0^*|]/D_0^{\text{tot}} \right]. \)

Similarly we can prove \( \lim_{c \to 0} E_1[N] \leq \frac{1}{D_0^{\text{tot}}} + M_1 \), where
\[
M_1 = \frac{1}{\Delta(A^c)} \left[ 1 + E_0[|\xi_1^*|]/D_0^{\text{tot}} \right]. \]

\section*{Appendix B

\textsc{Proof of Theorem 2}}

We prove the result for \( P_{FA} \). For \( P_{MD} \) it can be proved in the same way.

Probability of False Alarm can be written as,
\[
P_0(\text{Reject } H_0) = P_0[\text{FA before } \tau(c)] + P_0[\text{FA after } \tau(c)]. \tag{33}
\]

Consider the first term in the R.H.S. of (33) It can be shown that \( F^*_c \) stochastically dominates \( F_c \) under \( H_0 \). Thus we can construct \( \{ F^*_k \} \) such that \( F^*_k \geq F_k \) a.s. for all \( k \geq 0 \) and hence
\[
P_0[\text{FA before } \tau(c)] \leq P_0[\sup_{0 \leq k \leq \tau(c)} F^*_k \geq |\log c|] = P_0\left( \sum_{k=0}^\infty |\xi_k^*| \geq |\log c| \right). \tag{34}
\]

From [10] Theorem 1.3 \( E_0[e^{\eta\tau(c)}] < \infty \), for \( 0 < \eta < R_0^l \) and \( R_0^l = -\log \inf_{t \geq 0} E_0[e^{\tau^*_l(t)}] \). Combining this fact with \( \tau(c) < \sum_{l=1}^L \tau_l(c) \) and the fact that \( \tau_l(c) \) is independent of each other (see [9]), yields \( E_0[e^{\eta\tau(c)}] \leq E_0[e^{\sum_{l=1}^L \eta \tau_l(c)}] < \infty \), for \( 0 < \eta < R_0 \) = min \( R_0^l \). Therefore, from Markov inequality, with \( k_1 = E_0[e^{\eta\tau(c)}] \),
\[
P_0[\text{FA before } \tau(c)] \leq k_1 e^{\eta \left( -s_0(\eta)/\log c \right)} \tag{35}
\]

Let \( \tilde{F}_c^* = \sum_{k=1}^n |\xi_k^*| \). Then, with (35), the expected value of \( |\xi_k^*| \) being positive and with exponential tail assumption of \( G_0(t) \), from [5] Theorem 1, Remark 1, (34) is,
\[
P_0[F^*_c > \log c] \leq k_2 \exp(-s_0(\eta)/\log c), \tag{36}
\]
for any \( 0 < \eta < R_0 \), \( k_2 \) is a constant and \( s_0(\eta) \) is defined in [10]. Therefore,
\[
P_0[FA before \tau(c)] \leq k_2 \frac{e^{s_0(\eta)}}{c} \to 0, \tag{37}
\]
if \( s_0(\eta) > 1 \) for some \( \eta \).

Now we consider the second term in (33),
\[
P_0[FA after \tau(c)] = P_0[FA after \tau(c); A^0] + P_0[FA after \tau(c); (A^0)^c]
\]

Since events \{ FA after \( \tau(c) \) \} and \( (A^0)^c \) are mutually exclusive, the second term in the above expression is zero. Now consider \( P_0[FA after \tau(c); A^0] \). For \( 0 < r < 1 \),
\[
P_0[FA after \tau(c); A^0] \leq P_0[\text{Random walk with drift } \Delta(A^0) \text{ and initial value } F_{\tau(c)+1} \text{ crosses } |\log c|] \leq P_0[\text{Random walk with drift } \Delta(A^0) \text{ and } F_{\tau(c)+1} \leq r |\log c| \text{ crosses } |\log c|] + P_0[\text{Random walk with drift } \Delta(A^0) \text{ and } F_{\tau(c)+1} > r |\log c| \text{ crosses } |\log c|] \leq P_0[\text{Random walk with drift } \Delta(A^0) \text{ and } F_{\tau(c)+1} \leq r |\log c| \text{ crosses } |\log c|] + P_0[\text{Random walk with drift } \Delta(A^0) \text{ and } F_{\tau(c)+1} > r |\log c| \text{ crosses } |\log c|]. \tag{38}
\]
Considering the first expression, we have
\[ (P_0|\text{Random walk with drift } \Delta(A^0) \leq r| \log c| \text{ crosses } | \log c|) / c \leq (P_0|\text{Random walk with drift } \Delta(A^0) \leq r| \log c| \text{ crosses } | \log c|) / c \]
\[ \overset{(A) \exp(-(1-r)|\log c|s')}{\leq} \frac{\exp((-1-r)s')}{c} \to 0, \quad (39) \]
iff \((1-r)s' > 1\). Here (A) follows from \[16\] p. 78-79.

We choose \(s'<1\) and \(0 < r < 1\) to satisfy \((1-r)s' > 1\).

Consider the second term in (38). Using the stochastic dominance of \(\{F_k\}\) by \(\{F_k^*\}\),
\[ P_0[F_{\tau(c)+1} > r| \log c|] \leq P_0[F_{\tau^*(c)+1} > r| \log c|]. \]
We have \(P[\tau(c) + 1 > t] = P[\tau(c) > t - 1] \leq k_1 \exp(-\eta t)\),
where \(k_1 = \exp(\eta\sigma^*c)\). Therefore, following (36),
\[ P_0[F_{\tau(c)+1} > r| \log c|] \leq k_2 \exp(s_0(\eta)) / c \to 0, \quad \text{if } r > s_0(\eta) > 1, \]
as in (37). Then \(1 / s_0(\eta) < r \leq 1 / s'\).

REFERENCES


\[1\] For a random walk \(W_n = \sum_{i=1}^{n} X_i\), with stopping times \(T_a = \inf\{n \geq 1 : W_n \leq a\}\), \(T_b = \inf\{n \geq 1 : W_n \leq b\}\) and \(T_{a,b} = \min(T_a, T_b)\), \(a < 0 < b\), let \(s'\) be the non-zero solution to \(M(s') = 1\), where \(M\) denotes the M.G.F. of \(X_i\). Then, \(s' < 0\) if \(E[X_i] > 0\), and \(s' > 0\) if \(E[X_i] < 0\) and \(E[\exp(s'W_{T_{a,b}})] = 1\) (\[16\] p. 78-79). Then it can be shown that \(P(W_{T_a}) \leq \exp(-s'a)\) when \(E[X_i] > 0\) and \(P(W_{T_b}) \leq \exp(-s'b)\) when \(E[X_i] < 0\).