



Truncated Sequential Non-Parametric Hypothesis Testing Based on Random Distortion Testing

Prashant Khanduri, Dominique Pastor, Vinod Kumar Sharma, Pramod Varshney

► To cite this version:

Prashant Khanduri, Dominique Pastor, Vinod Kumar Sharma, Pramod Varshney. Truncated Sequential Non-Parametric Hypothesis Testing Based on Random Distortion Testing. IEEE Transactions on Signal Processing, 2019, 67 (15), pp.4027-4042. 10.1109/TSP.2019.2923140 . hal-02173094

HAL Id: hal-02173094

<https://imt-atlantique.hal.science/hal-02173094>

Submitted on 4 Jul 2019

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Truncated Sequential Non-Parametric Hypothesis Testing Based on Random Distortion Testing

Prashant Khanduri, *Student Member, IEEE*, Dominique Pastor, *Member, IEEE*, Vinod Sharma, *Senior Member, IEEE* and Pramod K. Varshney, *Life Fellow, IEEE*

Abstract—In this work, we propose a new algorithm for sequential non-parametric hypothesis testing based on Random Distortion Testing (RDT). The data based approach is non-parametric in the sense that the underlying signal distributions under each hypothesis are assumed to be unknown. Our previously proposed non-truncated sequential algorithm, *SeqRDT*, was shown to achieve desired error probabilities under a few assumptions on the signal model. In this work, we show that the proposed truncated sequential algorithm, *T-SeqRDT*, requires even fewer assumptions on the signal model, while guaranteeing the error probabilities to be below pre-specified levels and at the same time makes a decision faster compared to its optimal fixed-sample-size (FSS) counterpart, *BlockRDT*. We derive bounds on the error probabilities and the average stopping times of the algorithm. Via numerical simulations, we compare the performance of *T-SeqRDT* to *SeqRDT*, *BlockRDT*, sequential probability ratio test (SPRT) and composite sequential probability ratio tests. We also show the robustness of the proposed approach compared to standard likelihood ratio based approaches.

Index Terms—Truncated sequential testing, non-parametric testing, robust hypothesis testing, random distortion testing (RDT), sequential probability ratio test (SPRT).

I. INTRODUCTION

In today's world, many applications are characterized by the availability of large amounts of complex-structured data. It is not always possible to fit the data to predefined models or distributions [2], [3]. Model dependent signal processing approaches are often susceptible to mismatches between the data and the assumed model. In cases where the data does not conform to the assumed model, providing sufficient performance guarantees becomes a challenging task [4], [5]. Therefore, it is important to devise methods that are model-independent, robust, provide sufficient performance guarantees for the task at hand and, at the same time, are simple to implement. In this work, we propose one such algorithm: *T-SeqRDT*, for binary hypothesis testing.

Standard binary hypothesis testing problems [6], based on a fixed number of samples, test the null (\mathcal{H}_0) versus the alternate

(\mathcal{H}_1) hypotheses. The decision is usually made through the Bayesian, minimax or Neyman-Pearson frameworks. Such tests are referred to as fixed-sample-size (FSS) tests. However, a majority of decision making problems are inherently sequential in nature, i.e., observations are collected sequentially and are processed one after the other [7]–[10]. In his seminal works [11], [12], Wald proposed his celebrated sequential procedure, namely, the sequential probability ratio test (SPRT) for testing two simple hypotheses. SPRT is optimal in the sense that it makes a decision faster on average, compared to all the procedures including FSS tests achieving the same probabilities of error. However, this optimality is lost in some cases when there is a mismatch between the assumed and true models for the underlying hypotheses to be tested [13]–[15], i.e., SPRTs can have larger stopping times on average compared to FSS tests that achieve the same error probabilities. To avoid these scenarios, a truncated version of SPRT was proposed in [15], where the truncation time was chosen based on the FSS test. However, the error probabilities achieved by truncated SPRTs are usually higher than those yielded by non-truncated SPRTs. In addition, larger truncation times are needed to guarantee error probabilities below predefined levels. The purpose of the present work is to propose a truncated sequential algorithm for non-parametric binary hypothesis testing.

A key motivation in our approach is that SPRT loses its optimality properties if the hypotheses to be tested are composite [5], [12], [16]–[18]. For composite binary hypothesis testing problems, variants of SPRT have been developed. Of particular interest are invariant SPRT (ISPRT), weighted SPRT (WSPRT) and generalized SPRT (GSPRT) [16]. ISPRT relies on the principle of invariance [17], [19] to reduce the composite hypothesis to a simple one, which then makes it possible to apply Wald's SPRT [12]. However, this reduction imposes strong restrictions on the hypotheses to be tested [16], [20]. On the other hand, WSPRT assigns a suitable weight function to the unknown parameters [16], although it is not always possible to upper bound the probabilities of error and find an appropriate weight function, even in asymptotic regimes. In contrast, GSPRT approximates the likelihood ratio by replacing the unknown parameters in the likelihood by their maximum likelihood (ML) estimates [16], [21], [22]. Various versions of GSPRT have been proposed in the literature with different thresholds [23]–[25] and most of the literature is focused on the design of one-sided tests for testing single parameter families of distributions. Moreover, most algorithms are developed for exponential families of distributions and guarantees are asymptotic, which do not

A preliminary version of this work was presented at Allerton, 2018 [1]. The work of P. Khanduri and P. K. Varshney was supported by AFOSR Grant FA9550-16-1-0077. The work of D. Pastor was supported by Region Bretagne (France) and the European Regional Development Fund (ERDF).

P. Khanduri and P. K. Varshney are with Department of Electrical Engineering and Computer Science, Syracuse University, Syracuse, NY 13244 (email: pkhanduri@syr.edu; varshney@syr.edu).

D. Pastor is with IMT Atlantique, Lab-STICC, UBL, 29238 Brest, France (email: dominique.pastor@imt-atlantique.fr).

V. Sharma is with the Department of Electrical Communication Engineering, Indian Institute of Science, Bangalore 560012, India (e-mail: vinod@ece.iisc.ernet.in).

upper bound the probabilities of error [16], [23]–[25]. Importantly, GSPRT based approaches have heavy computational complexity even for simplest of models and, therefore, are difficult to implement online [26]. In contrast, in this work, we consider a two-sided binary hypothesis testing problem and propose a non-parametric truncated sequential testing algorithm, namely *T-SeqRDT*, to solve the hypothesis testing problem. *T-SeqRDT* has the following properties:

- The upper bounds on the probabilities of false alarm (PFA) and missed detection (PMD) are guaranteed to stay below pre-specified levels even in non-asymptotic regimes, which is naturally of practical interest. Moreover, the algorithm is guaranteed to be faster compared to its optimal FSS version, *BlockRDT*, to achieve the same performance.
- The underlying signal distribution under each hypothesis is assumed to be unknown, and importantly, the results derived do not rely on independence (or i.i.d) assumptions on the observations either. This makes the algorithm robust to mismatches in the distributions of the signals, compared to likelihood ratio based approaches.
- The algorithm is simple in structure with low computational complexity and, therefore, is easy to implement online.

It must be noted that non-parametric sequential hypothesis testing approaches have been considered in the past, with limited to no success, as guaranteeing both PFA and PMD below certain pre-specified levels may not be feasible for such non-parametric sequential testing problems [5], [17], [18]. The approaches proposed in [5] are based on approximating the likelihood ratio by employing estimates of the unknown parameters to be tested. These approaches impose restrictive assumptions on these estimates to guarantee robustness and asymptotic optimality when there is a mismatch between the assumed and the true distribution. This is of limited use in practical problems, which are non-asymptotic in nature.

From a general point of view, all the hypothesis testing procedures discussed above are based on likelihood theory. As such, they all assume prior knowledge of the distribution of the observations under each hypothesis, perhaps up to a few unknown parameters, to derive or approximate the likelihood ratios. Likelihood ratio based procedures are limited, because, usually, precise distribution models for the data in play are not available. Moreover, likelihood ratio based tests are not robust to model mismatches [5], [16], [20], [27]. In addition, as described earlier, many approaches in sequential hypothesis testing make stationarity as well as i.i.d. assumptions on the observed process under each hypothesis [16], [23]. Such assumptions are questionable in practice and emphasize the need for devising testing approaches that assume little knowledge of the underlying signals to be tested. In this respect, the truncated non-parametric sequential hypothesis test, *T-SeqRDT*, proposed in this paper and featuring the properties discussed earlier, results from an alternative sequential binary hypothesis testing formulation aimed at avoiding the aforementioned limitations.

To begin with, assume that Y is a one-dimensional observation, with probability distribution parameterized by ξ . Consider a two-sided hypothesis testing problem $\mathcal{H}_0 : \xi = \xi_0$

vs $\mathcal{H}_1 : \xi \neq \xi_0$. In practice, testing the signal for a precise value of ξ_0 might be too stringent due to measurement errors, environmental fluctuations other than noise and other factors [27]. Therefore, it is reasonable to allow for some fluctuations around ξ_0 and design the null hypothesis \mathcal{H}_0 to test for the signal in the neighborhood of ξ_0 . In this respect, we assume that Y is a corrupted observation of the signal to be tested, Ξ , and that Ξ is a random distorted version of ξ_0 with unknown distribution. The hypothesis testing problem then becomes:

$$\mathcal{H}_0 : |\Xi - \xi_0| \leq \tau \text{ vs } \mathcal{H}_1 : |\Xi - \xi_0| > \tau \quad (1)$$

where $\tau \in [0, \infty)$ represents the distortion. Note that, as discussed earlier, designing an asymptotically optimal GSPRT might not be feasible. However, one can formulate a WSPRT by assuming Ξ to be deterministic and observations to be Gaussian [12, Chapter 4], as discussed later in Section VI. Problem (1) was first considered in the form of random distortion testing (RDT) in [27], where the signal of interest, Ξ , with an unknown distribution, was embedded in i.i.d. Gaussian noise. The authors showed that the optimal tests (under certain criteria) were simple in design and, at the same time, independent of the signal distributions, thereby did not need the computation of likelihood ratios. To enhance detection performance with multiple samples, the authors extended the RDT formulation to FSS tests, *BlockRDT*, in [28]. The authors generalized the RDT formulation by replacing the signal Ξ , in (1), by its empirical mean over time. Although the detection performance improved with the number of samples, the designer had control only over PFA and no control over PMD. To overcome this limitation, the authors in [29], [30], extended the RDT formulation to sequential hypothesis testing and proposed *SeqRDT*. *SeqRDT* guarantees both PFA and PMD to be below certain levels, while making a decision faster on average. This was achieved by introducing a few key assumptions on the signal along with a buffer. Again, just like standard likelihood ratio based approaches, *SeqRDT* is not guaranteed to always make a decision faster compared to *BlockRDT* [29], [30]. To avoid such scenarios, we propose a truncated version of *SeqRDT*, *T-SeqRDT*, in this paper. We show that, the proposed algorithm requires fewer assumptions on the signal compared to *SeqRDT*, and at the same time always makes a decision faster compared to its FSS counterpart, *BlockRDT*. Similar to truncated SPRT [15], PFA and PMD of *T-SeqRDT* can be guaranteed to stay below pre-specified levels by designing larger truncation times. All this can be achieved with *T-SeqRDT* without the need for a buffer. We provide an analysis of the algorithm and show that it can be an alternative to the likelihood ratio based approaches as discussed earlier, when the signal distributions in play are unknown. The proposed approach is completely data dependent rather than being model dependent. The proposed algorithm is robust to model mismatches and is relatively simple to implement. In addition, the algorithm is able to provide sufficient performance guarantees with very few assumptions on the signal to be tested. Preliminary results were presented in [1] without proofs, for a fixed choice of thresholds and without control over the stopping time of the algorithm. In contrast, in the present work, we give the algorithm designer

freedom to choose the thresholds according to the desired performance guarantees. We provide complete analysis of the stopping time and probabilities of error behavior and their trade-offs with the designed thresholds. Moreover, we provide a method to choose the thresholds such that the stopping time is minimized. Summarizing, the main contributions of this work are as follows:

- We extend the RDT formulation for sequential non-parametric hypothesis testing and introduce a new truncated sequential algorithm, T-SeqRDT, to solve the binary hypothesis testing problem.
- We derive bounds on PFA and PMD of T-SeqRDT and show that the designed thresholds can guarantee pre-specified PFA and PMD.
- We analyze the average stopping time of T-SeqRDT and provide insights into the trade-off between the average stopping time and the error probabilities of T-SeqRDT.

In Section II, we formulate the problem. In Section III, we present the FSS algorithm, BlockRDT. In Section IV, we propose the truncated sequential algorithm, T-SeqRDT. In Section V, we derive the bounds on the error probabilities and stopping time of T-SeqRDT. In Section VI, we perform simulations to gain further insights into the behavior of T-SeqRDT. Finally, in Section VII, we conclude the paper.

Notation: All the random variables are defined on the same probability space $(\Omega, \mathcal{F}, \mathbb{P})$. $\mathcal{M}(\Omega, \mathbb{R})$ denotes the set of all real random variables defined on (Ω, \mathcal{F}) . Given $U \in \mathcal{M}(\Omega, \mathbb{R})$: $\mathbb{P}_U(B) = \mathbb{P}[U \in B]$ with $[U \in B] = \{\omega \in \Omega : U(\omega) \in B\}$ when B is a Borel set of \mathbb{R} . A domain \mathcal{B} of U is any Borel set B of \mathbb{R} such that $\mathbb{P}_U(B) = 1$. Given $\xi \in \mathbb{R}$ and $\sigma \in [0, \infty)$, $Z \sim \mathcal{N}(\xi, \sigma^2)$ implies Z is Gaussian distributed with mean ξ and variance σ^2 . The Generalized Marcum Function [31] with order $1/2$ is denoted by [32, Eq. (19) and Remark V.3],

$$Q_{\frac{1}{2}}(|\xi|, \eta) = \mathbb{P}[|Z| > \eta], \quad (2)$$

for $Z \sim \mathcal{N}(\xi, 1)$. For any $(a, b) \in [0, \infty) \times [0, \infty)$,

$$Q_{\frac{1}{2}}(a, b) = 1 - \Phi(b - a) + \Phi(-b - a) \quad (3)$$

where Φ is the cumulative distribution function (cdf) of Z . We remind the reader that [31, Theorem 1]:

Lemma 1.1 (Behavior of the Marcum function): Whatever its order, the Generalized Marcum function — and thus $Q_{\frac{1}{2}}$ — increases with its first argument and decreases with its second.

Given $\gamma \in (0, 1)$ and $\rho \in [0, \infty)$, $\lambda_\gamma(\rho)$ is defined as the unique solution in x to $Q_{\frac{1}{2}}(\rho, x) = \gamma$ [27, Lemma 2, statement (i)], so that:

$$Q_{\frac{1}{2}}(\rho, \lambda_\gamma(\rho)) = \gamma. \quad (4)$$

The set of all sequences defined on \mathbb{N} (resp. $[1, N] = \{1, 2, \dots, N\}$) and valued in $\mathcal{M}(\Omega, \mathbb{R})$ is denoted by $\mathcal{M}(\Omega, \mathbb{R})^{\mathbb{N}}$ (resp. $\mathcal{M}(\Omega, \mathbb{R})^{[1, N]}$). Given U in $\mathcal{M}(\Omega, \mathbb{R})^{\mathbb{N}}$ (resp. $U \in \mathcal{M}(\Omega, \mathbb{R})^{[1, N]}$), the realization of U at $n \in \mathbb{N}$ (resp. $n \in [1, N]$) is called a sample of U and denoted by U_n . Each U_n is an element of $\mathcal{M}(\Omega, \mathbb{R})$. Given $N \in \mathbb{N}$, the sample mean of U over the N samples U_1, \dots, U_N is $\langle U \rangle_N = \frac{1}{N} \sum_{n=1}^N U_n$. The minimum of two real numbers a_1 and a_2 is denoted by $a_1 \wedge a_2$ and $\bigwedge_{i=1}^n a_i$ denotes the minimum of n real numbers a_1, a_2, \dots, a_n .

II. PROBLEM STATEMENT

Let $\Xi = (\Xi_n)_{n \in \mathbb{N}}$ be an element of $\mathcal{M}(\Omega, \mathbb{R})^{\mathbb{N}}$. This discrete time random process models the random mixture of a distorted signal of interest and possible interferences. Standard hypothesis testing approaches assume that the random process under null hypothesis (\mathcal{H}_0) is generated from an unknown underlying joint distribution \mathcal{P}_0 , i.e., $\Xi = (\Xi_n)_{n \in \mathbb{N}} \sim \mathcal{P}_0$, and under alternate hypothesis (\mathcal{H}_1), by any underlying, yet unknown, arbitrary joint distribution, other than \mathcal{P}_0 , i.e., $\Xi = (\Xi_n)_{n \in \mathbb{N}} \sim \mathcal{P}_0$. No assumption is made on the stationarity or the distribution of $\Xi = (\Xi_n)_{n \in \mathbb{N}}$. In this respect, the samples Ξ_n are not necessarily i.i.d. As pointed out in [29], [30], this problem is difficult to tackle as very little or no knowledge of the underlying signal distributions is assumed under both hypotheses; thereby, likelihood ratio based tests (SPRT or GSPRT) are not suitable for such problems. As an alternative to likelihood ratio based approaches, we propose tests based on RDT [27], where we associate a non-parametric distance related criterion with each hypothesis which is independent of the distributions of the actual hypotheses. This non-parametric criterion serves as a surrogate to the actual hypotheses to be tested. Next, we present the model in more detail.

We assume that Ξ is observed in additive and independent Gaussian noise $X = (X_n)_{n \in \mathbb{N}}$. The observation process is $Y = (Y_n)_{n \in \mathbb{N}}$ such that $Y_n = \Xi_n + X_n$ for all $n \in \mathbb{N}$, and we write $Y = \Xi + X$. In our formulation, Ξ models the distortion around a fixed known and deterministic model ξ_0 . We, however, expect that, for N large enough, the empirical mean $\langle \Xi \rangle_N$ remains close to ξ_0 under \mathcal{H}_0 and drifts significantly away from ξ_0 under \mathcal{H}_1 . Using standard terminology in statistical inference, we say that this problem is the testing of the null hypothesis — a random event, actually — $\mathcal{H}_0 : |\langle \Xi \rangle_N - \xi_0| \leq \tau$ against the alternate hypothesis (event) $\mathcal{H}_1 : |\langle \Xi \rangle_N - \xi_0| > \tau$, on the basis of observation Y . The hypothesis testing problem is therefore given as:

$$\left\{ \begin{array}{l} \text{Observation : } Y = \Xi + X \in \mathcal{M}(\Omega, \mathbb{R})^{\mathbb{N}} \\ \text{with } \left\{ \begin{array}{l} \Xi = (\Xi_n)_{n \in \mathbb{N}} \in \mathcal{M}(\Omega, \mathbb{R})^{\mathbb{N}}, \\ X_1, X_2, \dots \stackrel{\text{iid}}{\sim} \mathcal{N}(0, 1), \\ \Xi \text{ and } X \text{ are independent.} \end{array} \right. \\ \exists N_0 \in \mathbb{N}, \left\{ \begin{array}{l} \mathcal{H}_0 : \forall N \geq N_0, 0 \leq |\langle \Xi \rangle_N - \xi_0| \leq \tau \text{ (a-s)} \\ \mathcal{H}_1 : \forall N \geq N_0, \tau < |\langle \Xi \rangle_N - \xi_0| \leq \tau_H \text{ (a-s)} \end{array} \right. \end{array} \right. \quad (5)$$

where, $\tau \in [0, \infty)$ is the tolerance and $\tau < \tau_H < \infty$. Note, that the above hypothesis testing model is the same as the BlockRDT model [28] for a fixed sample size, N (see Section III). Here, N_0 and the tolerances τ and τ_H are known *a priori* based on some prior knowledge (or experience) about the signal¹. The algorithms based on formulation (5) have already been used for biomedical signal processing applications, specifically for the detection of Auto-positive end expiratory pressure (Auto-PEEP) [33]. Moreover, for illustration purpose, below we give a simple example where formulation (5) can be easily used.

¹This knowledge can follow from machine learning training procedures or be based on some statistical knowledge of the signal. Discussion of these procedures is beyond the scope of this work.

Example 1 (Bounded regime testing): Given $\xi \in \mathbb{R}$ and $h \in [0, \infty)$, we say Ξ follows the (bounded) regime (ξ, h) and write $\Xi \sim (\xi, h)$ if, for any $N \in \mathbb{N}$, $|\langle \Xi \rangle_N - \xi| \leq h$. A sufficient condition for $\Xi \sim (\xi, h)$ is that $|\Xi_n - \xi| \leq h$ (a-s) for any $n \in \mathbb{N}$. Suppose Ξ satisfies either $\mathcal{H}_0^* : \Xi \sim (\xi_0, h_0)$, with (ξ_0, h_0) known or $\mathcal{H}_1^* : \Xi \sim (\xi_1, h_1)$, with (ξ_1, h_1) possibly unknown. Say that the regimes (ξ_0, h_0) and (ξ_1, h_1) are separate if $|\xi_1 - \xi_0| \geq h_0 + h_1$. When (ξ_0, h_0) and (ξ_1, h_1) are separate, testing \mathcal{H}_0^* against \mathcal{H}_1^* when we observe Y is the particular problem (5) with $h_0 \leq \tau < |\xi_1 - \xi_0| - h_1$, $\tau_H \geq |\xi_1 - \xi_0| + h_1$ and $N_0 = 1$. This example illustrates that only with the knowledge of (ξ_0, h_0) and $|\xi_1 - \xi_0| \geq h_0 + h_1$, the testing of a given bounded regime of Ξ is the hypothesis testing problem (5). Moreover, as shown later for such problems, we can design optimal FSS tests as well as sequential tests guaranteeing PFA and PMD below pre-specified levels and that are, at the same time, faster than the optimal FSS tests. Importantly, these tests guarantee performance irrespective of the underlying signal distributions. ■

Remark 1: Note that the RDT framework of (1) is the same as that given in (5) for $N = 1$. The formulation in (5) generalizes the RDT framework of (1) for testing with multiple samples, i.e., for the FSS test, *BlockRDT*, and sequential hypothesis testing approaches. An alternative testing problem would be to use $|\langle \Xi_n - \xi_0 \rangle_N$ instead of $|\langle \Xi \rangle_N - \xi_0|$ in (5). This would allow a larger class of distortions. However, designing such a test would require stronger assumptions of $|\Xi_n - \xi_0| \leq \tau$ under \mathcal{H}_0 and $|\Xi_n - \xi_0| > \tau$ under \mathcal{H}_1 for all $n \in \mathbb{N}$ in comparison to the condition of (5), where introducing N_0 in (5) gives the designer the flexibility to design the testing problem for models when the condition $|\langle \Xi \rangle_N - \xi_0| \leq \tau$ (resp. $|\langle \Xi \rangle_N - \xi_0| > \tau$) might not hold true for smaller values of $N \in \mathbb{N}$ under \mathcal{H}_0 (resp. \mathcal{H}_1). ■

Following the standard terminology [16] with a slight change of notation, we define a sequential test for the binary hypothesis testing (5) as a pair (T, \mathcal{D}_{N_0}) , where T is the stopping time and \mathcal{D}_{N_0} is a decision rule taking values in $\{0, 1, \infty\}$ such that, for each $1 \leq N \leq T$:

$$\mathcal{D}_{N_0}(N) = \begin{cases} 0 & \mathcal{H}_0 \text{ is accepted} \\ 1 & \mathcal{H}_1 \text{ is accepted} \\ \infty & \text{repeat the test with } N + 1 \text{ samples.} \end{cases} \quad (6)$$

Further, T is defined as

$$T = \inf\{N \in \mathbb{N} : N \leq N_0 + W^* - 1, \mathcal{D}_{N_0}(N) \neq \infty\}. \quad (7)$$

where the condition $N \leq N_0 + W^* - 1$ guarantees that $T \leq N_0 + W^*$ and we refer to W^* as the truncation window. Note that $W^* = \infty$ for non-truncated sequential procedures. It is also worth noticing that FSS tests are particular cases of tests (T, \mathcal{D}_{N_0}) , with stopping time being a deterministic constant $T = N$ and \mathcal{D}_{N_0} valued in $\{0, 1\}$. Given two specified levels α and β in $(0, 1/2)$, we define the class of tests:

$$\mathcal{C}(\alpha, \beta) = \{(T, \mathcal{D}_{N_0}) : \sup_{\Omega_0} \mathbb{P}_{\text{FA}}(\mathcal{D}_{N_0}) \leq \alpha, \sup_{\Omega_1} \mathbb{P}_{\text{MD}}(\mathcal{D}_{N_0}) \leq \beta\} \quad (8)$$

with $\Omega_0 = \{\Xi \in \mathcal{M}(\Omega, \mathbb{R})^{\mathbb{N}} : \forall N \geq N_0, |\langle \Xi \rangle_N - \xi_0| \leq \tau \text{ (a-s)}\}$, $\Omega_1 = \{\Xi \in \mathcal{M}(\Omega, \mathbb{R})^{\mathbb{N}} : \forall N \geq N_0, |\langle \Xi \rangle_N - \xi_0| > \tau \text{ (a-s)}\}$ and where

$$\mathbb{P}_{\text{FA}}(\mathcal{D}_{N_0}) \stackrel{\text{def}}{=} \mathbb{P}[\mathcal{D}_{N_0}(T) = 1], \quad \text{under } \mathcal{H}_0, \quad (9)$$

is the PFA and

$$\mathbb{P}_{\text{MD}}(\mathcal{D}_{N_0}) \stackrel{\text{def}}{=} \mathbb{P}[\mathcal{D}_{N_0}(T) = 0], \quad \text{under } \mathcal{H}_1. \quad (10)$$

is the PMD. Throughout, the levels α and β are chosen in the interval $(0, 1/2)$. The ultimate goal of this work is to design a truncated sequential test belonging to $\mathcal{C}(\alpha, \beta)$. To exhibit elements of $\mathcal{C}(\alpha, \beta)$, we will make use of the following assumption.

Assumption 2.1 (Behavior of $|\langle \Xi \rangle_N - \xi_0|$ under \mathcal{H}_1): There exists $\tau^+ \in (\tau, \infty)$ such that:

$$\text{Under } \mathcal{H}_1 : \forall N \geq N_0, |\langle \Xi \rangle_N - \xi_0| \geq \tau^+ \text{ (a-s)}. \quad \blacksquare$$

Assumption 2.1 states that under \mathcal{H}_1 , the empirical mean of the signal centered around ξ_0 is bounded away from τ . This assumption is similar in nature to that of the indifference zone assumed in [16], [23]. Here the region (τ, τ^+) represents the indifference zone. In Sections III and IV, we first define a FSS test *BlockRDT* and show that with the use of Assumption 2.1 *BlockRDT* can be designed so as to belong to $\mathcal{C}(\alpha, \beta)$. Then, by using *BlockRDT*, we define the truncated sequential test, *T-SeqRDT*, that also belongs to $\mathcal{C}(\alpha, \beta)$.

III. BlockRDT

The *BlockRDT* framework tests the problem defined in (5) for a fixed number of samples $N \geq N_0$. Specifically, suppose that we have only N samples from our observation Y so that $Y = \Xi + X \in \mathcal{M}(\Omega, \mathbb{R})^{[1, N]}$ in (5). To solve this testing problem, the authors in [28], [34] consider all the FSS tests $\mathcal{D}_{N_0}(N) = \mathcal{T}(Y)$, where \mathcal{T} is any (measurable) map $\mathcal{T} : \mathbb{R}^N \rightarrow \{0, 1\}$. All such maps \mathcal{T} are hereafter called N -dimensional tests. In the *BlockRDT* framework [28], [34], we define the size of a given N -dimensional test \mathcal{T} as:

$$\alpha_{\mathcal{T}} = \sup_{\Xi \in \mathcal{M}(\Omega, \mathbb{R})^{[1, N]} : \mathbb{P}[|\langle \Xi \rangle_N - \xi_0| \leq \tau] \neq 0} \mathbb{P}[\mathcal{T}(Y) = 1 \mid |\langle \Xi \rangle_N - \xi_0| \leq \tau]$$

and \mathcal{T} is said to have level $\gamma \in (0, 1)$ if $\alpha_{\mathcal{T}} \leq \gamma$. No Uniformly Most Powerful (UMP) test with level γ exists for *BlockRDT*. By UMP test with level γ , we mean an N -dimensional test \mathcal{T}^* such that $\alpha_{\mathcal{T}^*} \leq \gamma$ and $\mathbb{P}[\mathcal{T}^*(Y) = 1 \mid |\langle \Xi \rangle_N - \xi_0| > \tau] \geq \mathbb{P}[\mathcal{T}(Y) = 1 \mid |\langle \Xi \rangle_N - \xi_0| > \tau]$ for any N -dimensional test \mathcal{T} and any $\Xi \in \mathcal{M}(\Omega, \mathbb{R})^{[1, N]}$. We thus define the subclass of *BlockRDT*-coherent tests [34], among which a “best” test exists. We say that an N -dimensional test \mathcal{T} is *BlockRDT*-coherent if:

[Invariance in mean] Given $y, y' \in \mathbb{R}^N$, if $\langle y \rangle_N = \langle y' \rangle_N$, then $\mathcal{T}(y) = \mathcal{T}(y')$.

[Constant conditional power] For all $\Xi \in \mathcal{M}(\Omega, \mathbb{R})^{[1, N]}$ independent of X , there exists a Borel set \mathcal{B} such that $|\langle \Xi \rangle_N - \xi_0| \in \mathcal{B}$ (a-s) and, for any $\rho \in \mathcal{B} \cap (0, \infty)$, $\mathbb{P}[\mathcal{T}(Y) = 1 \mid |\langle \Xi \rangle_N - \xi_0| = \rho]$ is independent of the distribution of $|\langle \Xi \rangle_N - \xi_0|$.

The rationale behind **[Invariance in mean]** is straightforward and implies that two different observation processes with

the same empirical mean must yield the same decision for \mathcal{T} . **[Constant conditional power]** means that \mathcal{T} should not yield different results for different distributions of $|\langle \Xi \rangle_N - \xi_0|$, conditioned on $|\langle \Xi \rangle_N - \xi_0| = \rho$.

Let the class of all *BlockRDT*-coherent tests with level γ be denoted by \mathcal{K}_γ . This class can be partially pre-ordered as follows: given $\mathcal{T}, \mathcal{T}' \in \mathcal{K}_\gamma$, write that $\mathcal{T} \preceq \mathcal{T}'$ if, for any $\Xi \in \mathcal{M}(\Omega, \mathbb{R})^{[1, N]}$, (i) \mathcal{T} and \mathcal{T}' satisfy **[Constant conditional power]** on the same domain \mathcal{B} and (ii) For all $\rho \in \mathcal{B} \cap (\tau, \infty)$,

$$\mathbb{P}[\mathcal{T}(Y) = 1 | |\langle \Xi \rangle_N - \xi_0| = \rho] \leq \mathbb{P}[\mathcal{T}'(Y) = 1 | |\langle \Xi \rangle_N - \xi_0| = \rho].$$

According to [28], [34], the N -dimensional test defined for every $x \in \mathbb{R}^N$ by:

$$\mathcal{T}_{N, \gamma}^*(x) = \begin{cases} 0 & \text{if } |\langle x \rangle_N - \xi_0| \leq \lambda_\gamma(\tau\sqrt{N})/\sqrt{N} \\ 1 & \text{otherwise.} \end{cases} \quad (11)$$

where $\lambda_\gamma(\tau\sqrt{N})/\sqrt{N}$ is defined using (4), is maximal in \mathcal{K}_γ : for any $\mathcal{T} \in \mathcal{K}_\gamma$, $\mathcal{T} \preceq \mathcal{T}_{N, \gamma}^*$. Let the PFA and PMD of $\mathcal{T}_{N, \gamma}^*$ for *BlockRDT* be denoted by $\mathbb{P}_{\text{FA}}^{\text{B-RDT}}(N, \gamma)$ and $\mathbb{P}_{\text{MD}}^{\text{B-RDT}}(N, \gamma)$, respectively. We have the following proposition [28], [29].

Proposition 3.1: For any $\gamma \in (0, 1)$ and $\tau \geq 0$, we have:

$$\begin{cases} Q_{\frac{1}{2}}\left(0, \lambda_\gamma(\tau\sqrt{N})\right) & \leq \mathbb{P}_{\text{FA}}^{\text{B-RDT}}(N, \gamma) \leq \gamma \\ 1 - Q_{\frac{1}{2}}\left(\tau_H\sqrt{N}, \lambda_\gamma(\tau\sqrt{N})\right) & \leq \mathbb{P}_{\text{MD}}^{\text{B-RDT}}(N, \gamma) \leq 1 - \gamma \end{cases}$$

According to the above proposition, although being optimal for *BlockRDT*, $\mathcal{T}_{N, \gamma}^*$ controls $\mathbb{P}_{\text{FA}}^{\text{B-RDT}}(N, \gamma)$ efficiently but has no control over $\mathbb{P}_{\text{MD}}^{\text{B-RDT}}(N, \gamma)$.

This implies that, without further assumption and for any $\gamma \in (0, 1)$, *BlockRDT* cannot belong to the class $\mathcal{C}(\alpha, \beta)$ (with $\mathbb{P}_{\text{FA}}(\mathcal{D}_{N_0})$ and $\mathbb{P}_{\text{MD}}(\mathcal{D}_{N_0})$ replaced by $\mathbb{P}_{\text{FA}}^{\text{B-RDT}}(N, \gamma)$ and $\mathbb{P}_{\text{MD}}^{\text{B-RDT}}(N, \gamma)$ in (8), respectively) when $\alpha, \beta \in (0, 1/2)$. However, with Assumption 2.1 the next result implies that we can control $\mathbb{P}_{\text{FA}}^{\text{B-RDT}}(N, \gamma)$ such that *BlockRDT* is in $\mathcal{C}(\alpha, \beta)$.

Proposition 3.2: For any $\gamma \in (0, 1)$, $\mathbb{P}_{\text{FA}}^{\text{B-RDT}}(N, \gamma)$ and $\mathbb{P}_{\text{MD}}^{\text{B-RDT}}(N, \gamma)$ are bounded under Assumption 2.1 as:

$$\begin{cases} \mathbb{P}_{\text{FA}}^{\text{B-RDT}}(N, \gamma) \leq \gamma, \\ \mathbb{P}_{\text{MD}}^{\text{B-RDT}}(N, \gamma) \leq 1 - Q_{\frac{1}{2}}\left(\tau^+\sqrt{N}, \lambda_\gamma(\tau\sqrt{N})\right) \end{cases}$$

and the upper bound on $\mathbb{P}_{\text{MD}}^{\text{B-RDT}}(N, \gamma)$ decreases to 0 with N . **PROOF:** The bound follows from Lemma A.1 and Assumption 2.1. The upper-bound on $\mathbb{P}_{\text{MD}}^{\text{B-RDT}}(N, \gamma)$ decreases with N as a consequence of Lemma B.3. ■

Proposition 3.2 implies that for $\gamma = \alpha$ and a sufficiently large N such that the bound on $\mathbb{P}_{\text{MD}}^{\text{B-RDT}}(N, \gamma)$ is below β , *BlockRDT* is in $\mathcal{C}(\alpha, \beta)$. Since this N might be very large in practice, we introduce a novel truncated sequential algorithm, *T-SeqRDT*, to control the number of samples.

IV. ALGORITHM: T-SeqRDT

In this section, we propose *T-SeqRDT*. In *T-SeqRDT*, if no decision has been reached until a specified time, the decision will be forced using *BlockRDT* [28], since Proposition 3.2 guarantees that we can attain arbitrarily small PMD for a

bounded PFA. The decision variable $\mathcal{D}_{N_0}(N)$ for *T-SeqRDT* is defined as:

$$\begin{cases} \mathcal{D}_{N_0}(1) = \mathcal{D}_{N_0}(2) = \dots = \mathcal{D}_{N_0}(N_0 - 1) = \infty, \\ \text{for } N_0 \leq N < N_0 + W^*, \\ \mathcal{D}_{N_0}(N) = \begin{cases} 0 & \text{if } |\langle Y \rangle_N - \xi_0| \leq \lambda_L(N) \\ 1 & \text{if } |\langle Y \rangle_N - \xi_0| > \lambda_H(N) \\ \infty & \text{if } \lambda_L(N) < |\langle Y \rangle_N - \xi_0| \leq \lambda_H(N) \end{cases} \\ \text{for } N = N_0 + W^*, \\ \mathcal{D}_{N_0}(N) = \begin{cases} 0 & \text{if } |\langle Y \rangle_N - \xi_0| \leq \lambda_{\text{B-RDT}}(N) \\ 1 & \text{if } |\langle Y \rangle_N - \xi_0| > \lambda_{\text{B-RDT}}(N) \end{cases} \end{cases} \quad (12)$$

with decisions taken according to (6) and the stopping time T defined in (7). At time instant $N = N_0 + W^*$, with $W^* \in \mathbb{N}$, the decision is made using *BlockRDT*, if a decision has not been made until then. Recall that W^* is defined as the truncation window. The three thresholds $\lambda_L(N)$, $\lambda_H(N)$ and $\lambda_{\text{B-RDT}}(N)$ must be designed jointly so as to guarantee that *T-SeqRDT* is in $\mathcal{C}(\alpha, \beta)$. In any case, $\lambda_H(N)$ and $\lambda_L(N)$ must be such that $\lambda_L(N) < \lambda_H(N)$. Moreover, we want a decision faster compared to *BlockRDT*, the optimal FSS counterpart of *T-SeqRDT*. The thresholds are chosen with respect to these constraints.

Another sequential algorithm, *SeqRDT*, was proposed in [29], [30] for solving (5). *SeqRDT* was shown to belong to class $\mathcal{C}(\alpha, \beta)$ with the help of a buffer along with the upper and lower thresholds respectively defined as:

$$\lambda_\alpha(\tau\sqrt{N})/\sqrt{N} \text{ and } \lambda_{1-\beta}(\tau\sqrt{N})/\sqrt{N}, \quad (13)$$

to control both PFA and PMD. Along with these thresholds, the buffer size was designed using τ^- and τ^+ along with τ and τ_H defined in (5), where the meaning of τ^- is recalled in Remark 2 below.

Remark 2: *SeqRDT* [29], [30] imposed stricter conditions on the signal compared to *T-SeqRDT*. Beyond Assumption 2.1, it was assumed in the *SeqRDT* framework that under \mathcal{H}_0 , for all $N \geq N_0$, $|\langle \Xi \rangle_N - \xi_0| \leq \tau^-$ (a-s) with $\tau^- \in [0, \tau]$. Therefore, *SeqRDT* requires more parameters than *T-SeqRDT*. In addition, performance bounds were guaranteed by *SeqRDT* via the use of a buffer. The buffer size was selected using τ^- and τ^+ along with τ and τ_H defined in (5). In contrast, *T-SeqRDT* does not need to know τ^- or even τ_H . It requires the knowledge of τ and τ^+ only to guarantee performance. ■

A. Thresholds

SeqRDT [29], [30] is designed to belong to class $\mathcal{C}(\alpha, \beta)$ via the thresholds (13) and thanks to the use of a buffer. *T-SeqRDT* eliminates the need for this buffer, while being in $\mathcal{C}(\alpha, \beta)$. In view of the similarity between the *T-SeqRDT* statistic in (12) to that of *BlockRDT* in (11), we define the thresholds similar in structure to those of *BlockRDT*. The thresholds $\lambda_H(N)$, $\lambda_L(N)$ and $\lambda_{\text{B-RDT}}(N)$ are designed as:

$$\begin{aligned} \lambda_H(N) &= \lambda_H(N, w_H) = \lambda_{\alpha/w_H}(\tau\sqrt{N})/\sqrt{N} \\ \lambda_L(N) &= \lambda_L(N, w_L) = \lambda_{1-\beta/w_L}(\tau\sqrt{N})/\sqrt{N} \\ \lambda_{\text{B-RDT}}(N) &= \lambda_{\text{B-RDT}}(N, w_{BH}) = \lambda_{\alpha/w_{BH}}(\tau\sqrt{N})/\sqrt{N}, \end{aligned} \quad (14)$$

where w_H, w_L and w_{BH} give the algorithm designer control over these thresholds and are equal to or greater than 1. This constraint is necessary to ensure that T-SeqRDT is a valid sequential test, by guaranteeing that $\lambda_L(N, w_L) < \lambda_H(N, w_H)$, as shown in Proposition 4.1 below. In addition, the parameters w_H, w_L and w_{BH} must appropriately be chosen so as to guarantee that T-SeqRDT belongs to $\mathcal{C}(\alpha, \beta)$. To this end, we study the properties of the thresholds (14) and establish that they satisfy suitable properties for T-SeqRDT.

Proposition 4.1: $w_L \geq 1$ and $1 \leq w_{BH} \leq w_H$ given:

- (i) $\lambda_L(N, w_L) < \lambda_{B-RDT}(N, w_{BH}) \leq \lambda_H(N, w_H)$, $\forall N \in \mathbb{N}$.
- (ii) the thresholds $\lambda_H(N, w_H)$ and $\lambda_{B-RDT}(N, w_{BH})$ are decreasing in $N \in \mathbb{N}$ and lower bounded by τ .
- (iii) for N large enough, the threshold $\lambda_L(N, w_L)$ is increasing in N and upper bounded by τ .
- (iv) all the thresholds approach τ as N increases:

$$\lim_{N \rightarrow \infty} \lambda_H(N, w_H) = \lim_{N \rightarrow \infty} \lambda_{B-RDT}(N, w_{BH}) = \lim_{N \rightarrow \infty} \lambda_L(N, w_L) = \tau.$$

PROOF: Since $\alpha, \beta \in (0, 1/2)$, $0 < \frac{\alpha}{w_H} \leq \frac{\alpha}{w_{BH}} < 1/2 < 1 - \frac{\beta}{w_L}$. Thus the proof of (i) follows from Lemma B.4. Statements (ii) and (iii) follow from Lemmas B.5 and B.6, respectively. The proof of (iv) derives from Lemma B.1. ■

As discussed earlier, Proposition 4.1 ensures $\lambda_L(N, w_L) < \lambda_H(N, w_H)$, which is made possible by the assumption that w_H, w_L and $w_{BH} \geq 1$. Moreover, all the thresholds tend to τ as N increases, which intuitively implies that the chance of making a decision should be higher for larger N .

Now, we analyze the behavior of the thresholds with respect to parameters w_H, w_L and w_{BH} , when N is fixed.

Proposition 4.2:

- (i) $\lambda_H(N, w_H)$ increases when w_H increases,
- (ii) $\lambda_L(N, w_L)$ decreases when w_L increases,
- (iii) $\lambda_{B-RDT}(N, w_{BH})$ increases when w_{BH} increases,
- (iv) $\lim_{w_H \rightarrow \infty} \lambda_H(N, w_H) = \infty$ and $\lim_{w_L \rightarrow \infty} \lambda_L(N, w_L) = 0$.

PROOF: The proof of (i), (ii) and (iii) follows from Lemma B.4. Statement (iv) follows from (4) and the fact that the Marcum function (2) is a complementary cdf. ■

According to Proposition 4.2, $\lambda_H(N, w_H)$ and $\lambda_L(N, w_L)$ grow further away as w_H and w_L increase. Therefore, thresholds designed with higher values of w_H and w_L should provide better $\mathbb{P}_{FA}(\mathcal{D}_{N_0})$ and $\mathbb{P}_{MD}(\mathcal{D}_{N_0})$ performance compared to thresholds tuned with lower w_H and w_L values, but at the expense of higher stopping times. For SeqRDT [29], [30], the error probabilities were controlled via the buffer and no control over the stopping time was provided. For T-SeqRDT, the control over the error probabilities is achieved by choosing the parameters w_H, w_L and w_{BH} so as to move the thresholds away from or closer to each other. This gives the designer control over the average stopping time as well. This will be discussed in more detail later.

B. Truncation window

The goal of T-SeqRDT is to make a decision faster compared to its FSS counterpart, BlockRDT, while providing sufficient performance guarantees. Thus, it makes sense to base the choice of the truncation window W^* on BlockRDT as follows. For the threshold $\lambda_{B-RDT}(N, w_{BH})$ given in (14), Proposition

3.2 induces that $\mathbb{P}_{FA}^{B-RDT}(N, \alpha/w_{BH})$ is always upper bounded by α/w_{BH} and hence by α as $w_{BH} \geq 1$. Moreover, the upper bound on $\mathbb{P}_{MD}^{B-RDT}(N, \alpha/w_{BH})$ is a decreasing function of N . We thus propose to choose $W^* = W^*(w_{BH}, w_{BL})$ as:

$$W^* = W^*(w_{BH}, w_{BL}) = \min \left\{ W \in \mathbb{N} : 1 - Q_{\frac{1}{2}} \left(\tau^+ \sqrt{N_0 + W}, \lambda_{\frac{\alpha}{w_{BH}}}(\tau \sqrt{N_0 + W}) \right) \leq \frac{\beta}{w_{BL}} \right\} \quad (15)$$

with $w_{BL} \geq 1$.

Remark 3: Note from Proposition 3.2 that BlockRDT with $\lambda_{B-RDT}(N, w_{BH})$ given in (14) for number of samples $N = N_0 + W^*$ is in $\mathcal{C}(\alpha/w_{BH}, \beta/w_{BL})$.

The parameters w_H and w_L defined earlier are used to control the upper and the lower thresholds, respectively, via (14). On the other hand, the parameters w_{BH} and w_{BL} control the truncation window, $W^*(w_{BH}, w_{BL})$, defined in (15) and the assumption $w_{BL} \geq 1$ is required to make sure that the PMD of T-SeqRDT stays below β (see Theorem 5.2 below). All the thresholds along with the truncation window, which are thus controlled by w_H, w_L, w_{BH} and w_{BL} , govern the performance of T-SeqRDT. Therefore, we next analyse the behavior of $W^*(w_{BH}, w_{BL})$ with w_{BH} and w_{BL} so that w_{BL}, w_H, w_L and w_{BH} can be fixed to guarantee that T-SeqRDT belongs to $\mathcal{C}(\alpha, \beta)$.

Proposition 4.3: We have

- (i) For fixed w_{BL} , $W^*(\bullet, w_{BL})$ does not decrease;
 - (ii) For fixed w_{BH} , $W^*(w_{BH}, \bullet)$ does not decrease;
- PROOF: For any $w_{BH} \geq 1$ and any $W \in \mathbb{N}$, set:

$$\text{UB}_{MD}^{B-RDT}(w_{BH}, W) = 1 - Q_{\frac{1}{2}} \left(\tau^+ \sqrt{N_0 + W}, \lambda_{\frac{\alpha}{w_{BH}}}(\tau \sqrt{N_0 + W}) \right)$$

For any $w_{BL} \geq 1$,

$$W^*(w_{BH}, w_{BL}) = \min \mathcal{A}(w_{BH}, w_{BL}) \quad (16)$$

with:

$$\mathcal{A}(w_{BH}, w_{BL}) = \left\{ W \in \mathbb{N} : \text{UB}_{MD}^{B-RDT}(w_{BH}, W) \leq \frac{\beta}{w_{BL}} \right\} \quad (17)$$

Proof of (i): Consider $w_{BH} \leq w'_{BH}$. According to Lemmas 1.1 and B.4, we have:

$$\text{UB}_{MD}^{B-RDT}(w_{BH}, W) \leq \text{UB}_{MD}^{B-RDT}(w'_{BH}, W) \quad (18)$$

Therefore, from (16) and (17), we have $\mathcal{A}(w'_{BH}, w_{BL}) \subseteq \mathcal{A}(w_{BH}, w_{BL})$ and thus $W^*(w'_{BH}, w_{BL}) \geq W^*(w_{BH}, w_{BL})$.

Proof of (ii): Fix w_{BH} . If $w_{BL} \leq w'_{BL}$, then $\frac{\beta}{w'_{BL}} \leq \frac{\beta}{w_{BL}}$. This implies that $\mathcal{A}(w_{BH}, w'_{BL}) \subseteq \mathcal{A}(w_{BH}, w_{BL})$.

Hence the result. ■

Proposition 4.3 tells us that the smaller the required PFA and PMD for truncation by BlockRDT, the larger the truncation window for T-SeqRDT, which is natural. This will lead to the trade-off pinpointed in the next section between this truncation window and the error probabilities of T-SeqRDT. In addition, the choice of the truncation window using BlockRDT will allow for easier comparison between T-SeqRDT and BlockRDT.

Remark 4: Assumption 2.1 is instrumental in choosing an appropriate truncation window W^* for T-SeqRDT (see Proposition 3.2 and (15)). But, if W^* is known *a priori*, i.e., it is available via some preliminary training procedure or prior

experience, Assumption 2.1 is not needed, while the algorithm will still achieve the same performance. ■

Our next goal is to choose the appropriate thresholds (14) and window size (15), such that T-SeqRDT is in $\mathcal{C}(\alpha, \beta)$. We proceed by noticing that (14) and Proposition 4.3 show that this question is equivalent to choosing appropriate values of w_H , w_L , w_{BH} and w_{BL} .

V. ANALYSIS

We calculate bounds on the PFA and PMD of T-SeqRDT. These bounds are used to derive values for w_H , w_L , w_{BH} and w_{BL} that guarantee the required performance. Then, we study the average stopping time. Finally, we discuss the relationship between the error probabilities and the average stopping time.

A. False alarm and missed detection probabilities

Since closed form expressions for $\mathbb{P}_{\text{FA}}(\mathcal{D}_{N_0})$ and $\mathbb{P}_{\text{MD}}(\mathcal{D}_{N_0})$ cannot be derived, we instead calculate upper and lower bounds on these error probabilities, for the thresholds (14). These bounds provide useful insights into the behavior of T-SeqRDT. We begin with lower bounds.

Theorem 5.1 (Lower-bounds on $\mathbb{P}_{\text{FA}}(\mathcal{D}_{N_0})$ & $\mathbb{P}_{\text{MD}}(\mathcal{D}_{N_0})$):

$$\begin{cases} \mathbb{P}_{\text{FA}}(\mathcal{D}_{N_0}) \geq Q_{\frac{1}{2}}(0, \lambda_{\alpha/w_H}(\tau\sqrt{N_0})), \\ \mathbb{P}_{\text{MD}}(\mathcal{D}_{N_0}) \geq 1 - Q_{\frac{1}{2}}(\tau_H\sqrt{N_0}, \lambda_{1-\beta/w_L}(\tau\sqrt{N_0})). \end{cases}$$

PROOF: Since $[\mathcal{D}_{N_0}(T) = 1] \supseteq [\mathcal{D}_{N_0}(N_0) = 1]$, (9) implies that, under \mathcal{H}_0 :

$$\begin{aligned} \mathbb{P}_{\text{FA}}(\mathcal{D}_{N_0}) &\geq \mathbb{P}[\mathcal{D}_{N_0}(N_0) = 1] \\ &\stackrel{(a)}{=} \mathbb{E} \left[Q_{\frac{1}{2}} \left(\sqrt{N_0} |\langle \Xi \rangle_{N_0} - \xi_0|, \sqrt{N_0} \lambda_H(N_0, w_H) \right) \right] \\ &\stackrel{(b)}{\geq} Q_{\frac{1}{2}} \left(0, \lambda_{\alpha/w_H}(\tau\sqrt{N_0}) \right) \end{aligned}$$

where (a) follows from Lemma A.1, (b) from (14), Lemma 1.1 and the fact that under \mathcal{H}_0 , $0 \leq |\langle \Xi \rangle_N - \xi_0| \leq \tau$ (a.s). Similarly, consider the event $[\mathcal{D}_{N_0}(T) = 0]$ and follow the same procedure as above to get the lower bound for $\mathbb{P}_{\text{MD}}(\mathcal{D}_{N_0})$. ■

Although the lower bounds play no role in designing the thresholds, note that they decrease with N_0 and approach 0 as $N_0 \rightarrow \infty$, which follows from Lemma B.3 and B.2.

Theorem 5.2 (Upper-bounds on $\mathbb{P}_{\text{FA}}(\mathcal{D}_{N_0})$ & $\mathbb{P}_{\text{MD}}(\mathcal{D}_{N_0})$):

$$\begin{cases} \mathbb{P}_{\text{FA}}(\mathcal{D}_{N_0}) \leq \text{UB}_{\text{FA}} \leq \left(\frac{W^*}{w_H} + \frac{1}{w_{BH}} \right) \alpha, \\ \mathbb{P}_{\text{MD}}(\mathcal{D}_{N_0}) \leq \text{UB}_{\text{MD}} \leq \left(\frac{W^*}{w_L} + \frac{1}{w_{BL}} \right) \beta, \end{cases}$$

where UB_{FA} and UB_{MD} are given in (19) and (20), respectively, and $W^* = W^*(w_{BH}, w_{BL})$.

PROOF: We have

$$\begin{aligned} [\mathcal{D}_{N_0}(T) = 1] &= [\mathcal{D}_{N_0}(N_0) = 1] \bigcup_{N=N_0+1}^{N_0+W^*} \left([\mathcal{D}_{N_0}(N) = 1] \right. \\ &\quad \left. \cap [\mathcal{D}_{N_0}(K) = \infty, \forall K \text{ s.t. } N_0 \leq K \leq N-1] \right). \end{aligned}$$

Since these events are disjoint, we have

$$\begin{aligned} \mathbb{P}[\mathcal{D}_{N_0}(T) = 1] &= \mathbb{P}[\mathcal{D}_{N_0}(N_0) = 1] + \sum_{N=N_0+1}^{N_0+W^*} \mathbb{P} \left([\mathcal{D}_{N_0}(N) = 1] \right. \\ &\quad \left. \cap [\mathcal{D}_{N_0}(K) = \infty, \forall K \text{ s.t. } N_0 \leq K \leq N-1] \right) \\ &\stackrel{(a)}{\leq} \mathbb{P}[\mathcal{D}_{N_0}(N_0) = 1] + \sum_{N=N_0+1}^{N_0+W^*} \mathbb{P}[\mathcal{D}_{N_0}(N) = 1] \\ &\quad \bigwedge \left(\bigwedge_{K=N_0}^{N-1} \mathbb{P}[\mathcal{D}_{N_0}(K) = \infty] \right), \end{aligned} \quad (21)$$

where (a) follows from the Frechet inequality. We bound each individual probability on the right hand side (rhs) of (21) under \mathcal{H}_0 . First, for all $N_0 \leq N \leq N_0 + W^* - 1$, we have:

$$\begin{aligned} \mathbb{P}[\mathcal{D}_{N_0}(N) = 1] &\stackrel{(a)}{=} \mathbb{E} \left[Q_{\frac{1}{2}} \left(\sqrt{N} |\langle \Xi \rangle_N - \xi_0|, \sqrt{N} \lambda_H(N, w_H) \right) \right] \\ &\stackrel{(b)}{\leq} Q_{\frac{1}{2}} \left(\tau\sqrt{N}, \lambda_{\alpha/w_H}(\tau\sqrt{N}) \right) \stackrel{(c)}{=} \alpha/w_H, \end{aligned} \quad (22)$$

where (a) follows from Lemma A.1; (b) results from (14), the fact that under \mathcal{H}_0 : $0 \leq |\langle \Xi \rangle_N - \xi_0| \leq \tau$ and Lemma 1.1; (c) comes from (4).

Second, for $N = N_0 + W^*$, we have under \mathcal{H}_0 :

$$\mathbb{P}[\mathcal{D}_{N_0}(N) = 1] = \mathbb{P}_{\text{FA}}^{\text{B-RDT}} \stackrel{(a)}{\leq} \alpha/w_{BH}, \quad (23)$$

where (a) follows from Proposition 3.2 and (14).

Now, for all $N_0 \leq K \leq N_0 + W^* - 1$, we have:

$$\begin{aligned} \mathbb{P}[\mathcal{D}_{N_0}(K) = \infty] &= \mathbb{P}[| \langle Y \rangle_K - \xi_0 | \geq \lambda_L(K, w_L)] - \mathbb{P}[| \langle Y \rangle_K - \xi_0 | > \lambda_H(K, w_H)] \\ &\stackrel{(a)}{=} \mathbb{E} \left[Q_{\frac{1}{2}} \left(\sqrt{K} |\langle \Xi \rangle_K - \xi_0|, \sqrt{K} \lambda_L(K, w_L) \right) \right] \\ &\quad - \mathbb{E} \left[Q_{\frac{1}{2}} \left(\sqrt{K} |\langle \Xi \rangle_K - \xi_0|, \sqrt{K} \lambda_H(K, w_H) \right) \right] \\ &\stackrel{(b)}{\leq} Q_{\frac{1}{2}} \left(\tau\sqrt{K}, \lambda_{1-\beta/w_L}(\tau\sqrt{K}) \right) - Q_{\frac{1}{2}} \left(0, \lambda_{\alpha/w_H}(\tau\sqrt{K}) \right) \\ &\stackrel{(c)}{=} 1 - \beta/w_L - Q_{\frac{1}{2}} \left(0, \lambda_{\alpha/w_H}(\tau\sqrt{K}) \right), \end{aligned} \quad (24)$$

where: (a) follows from Lemma A.1, (b) from the monotonicity of the Marcum function, (14) and the fact that under \mathcal{H}_0 : $0 \leq |\langle \Xi \rangle_N - \xi_0| \leq \tau$, and (c) from (4). The upper bounds on $\mathbb{P}_{\text{FA}}(\mathcal{D}_{N_0})$ follow by substituting (22), (23) and (24) into (21) and using that $a_1 \wedge a_2 \leq a_1$. The upper bounds for $\mathbb{P}_{\text{MD}}(\mathcal{D}_{N_0})$ result from a similar procedure and the definition of W^* via (15). ■

This theorem justifies the definition of the thresholds in (14). It is clear that $\mathbb{P}_{\text{FA}}(\mathcal{D}_{N_0})$ and $\mathbb{P}_{\text{MD}}(\mathcal{D}_{N_0})$ of T-SeqRDT can be controlled such that T-SeqRDT is in $\mathcal{C}(\alpha, \beta)$ by choosing appropriate parameters w_H , w_L , w_{BH} and w_{BL} , which are independent of the signal model. Moreover, to do so, all these parameters have to be greater than or equal to one. Hereafter, we work with the looser upper bounds stated in Theorem 5.2. They are simpler to analyze as they depend on fewer parameters than UB_{FA} and UB_{MD} and give useful insights into the behavior of T-SeqRDT.

We use threshold $\lambda_{\text{B-RDT}}(N_0 + W^*, w_{BH})$ with $W^* = W^*(w_{BH}, w_{BL})$ to stop T-SeqRDT if a decision has not been taken until $N_0 + W^*$. As pinpointed in Remark 3, the PFA

$$\begin{aligned}
\text{UB}_{\text{FA}} &= \frac{\alpha}{w_H} + \left[\sum_{N=N_0+1}^{N_0+W^*-1} \frac{\alpha}{w_H} \wedge \left(\bigwedge_{K=N_0}^{N-1} \left(\left(1 - \frac{\beta}{w_L} \right) - Q_{\frac{1}{2}} \left(0, \lambda_{\frac{\alpha}{w_H}}(\tau\sqrt{K}) \right) \right) \right) \right] + \frac{\alpha}{w_{BH}} \wedge \left(\bigwedge_{K=N_0}^{N_0+W^*-1} \left(\left(1 - \frac{\beta}{w_L} \right) - Q_{\frac{1}{2}} \left(0, \lambda_{\frac{\alpha}{w_H}}(\tau\sqrt{K}) \right) \right) \right), \\
\text{UB}_{\text{MD}} &= \frac{\beta}{w_L} + \left[\sum_{N=N_0+1}^{N_0+W^*-1} \frac{\beta}{w_L} \wedge \left(\bigwedge_{K=N_0}^{N-1} \left(Q_{\frac{1}{2}} \left(\tau_H \sqrt{K}, \lambda_{1-\frac{\beta}{w_L}}(\tau\sqrt{K}) \right) - \frac{\alpha}{w_H} \right) \right) \right] + \frac{\beta}{w_{BL}} \wedge \left(\bigwedge_{K=N_0}^{N_0+W^*-1} \left(Q_{\frac{1}{2}} \left(\tau_H \sqrt{K}, \lambda_{1-\frac{\beta}{w_L}}(\tau\sqrt{K}) \right) - \frac{\alpha}{w_H} \right) \right).
\end{aligned} \tag{19}$$

$$\tag{20}$$

(resp. PMD) of the corresponding *BlockRDT* is upper-bounded by α/w_{BH} (resp. β/w_{BL}). Therefore, from Theorem 5.2, we see that *T-SeqRDT* may lose some detection performance compared to *BlockRDT*. However, it follows from this same theorem and Subsection IV-B that the upper-bounds on the false alarm and missed detection probabilities are of the same order for *T-SeqRDT* and *BlockRDT*. For example, if $w_{BH} = w_{BL} = 1$ and $w_H = w_L = W^*$, *T-SeqRDT* is in $\mathcal{C}(2\alpha, 2\beta)$ whereas *BlockRDT* is in $\mathcal{C}(\alpha, \beta)$. We can thus increase w_H , w_L , w_{BH} and w_{BL} such that *T-SeqRDT* is in $\mathcal{C}(\alpha, \beta)$. Though this comes at the cost of increasing the average stopping-time, this is the same behavior as observed for SPRT and discussed in the Introduction [15]. We show in the next section that this average stopping time remains always less than $N_0 + W^*$.

B. Stopping time

Similarly to $\mathbb{P}_{\text{FA}}(\mathcal{D}_{N_0})$ and $\mathbb{P}_{\text{MD}}(\mathcal{D}_{N_0})$, a closed form for the average stopping time of *T-SeqRDT* is not derivable. We, however, get an insight into the stopping behavior of *T-SeqRDT* by bounding its average stopping time.

Theorem 5.3 (Bounds on the average stopping time): With $W^* = W^*(w_{BH}, w_{BL})$:

(i) We have

$$\begin{cases} \text{Under } \mathcal{H}_0 : \mathbb{E}[T] \leq \text{UB}_{T_{\mathcal{H}_0}} \\ \text{Under } \mathcal{H}_1 : \mathbb{E}[T] \leq \text{UB}_{T_{\mathcal{H}_1}}, \end{cases}$$

where:

$$\begin{aligned}
\text{UB}_{T_{\mathcal{H}_0}} &= N_0 + W^* - \beta W^* / w_L \\
&\quad - \sum_{N=N_0}^{N_0+W^*-1} Q_{\frac{1}{2}} \left(0, \lambda_{\alpha/w_H}(\tau\sqrt{N}) \right), \\
\text{UB}_{T_{\mathcal{H}_1}} &= N_0 + W^* - \alpha W^* / w_H \\
&\quad - \sum_{N=N_0}^{N_0+W^*-1} \left[1 - Q_{\frac{1}{2}} \left(\tau_H \sqrt{N}, \lambda_{1-\beta/w_L}(\tau\sqrt{N}) \right) \right].
\end{aligned}$$

(ii) $\mathbb{E}[T] < N_0 + W^*$.

PROOF:

Proof of statement (i): Since the random variable T is discrete and valued in $\{N_0, N_0 + 1, \dots, N_0 + W^*\}$ and

$$\mathbb{E}[T] = \sum_{N=0}^{\infty} \mathbb{P}[T > N] = N_0 + \sum_{N=N_0}^{N_0+W^*-1} \mathbb{P}[T > N].$$

By definition of T (7), $[T > N] \subset [\mathcal{D}_{N_0}(N) = \infty]$ for any $N \in \{N_0, N_0 + 1, \dots, N_0 + W^*\}$. Hence, the following inequality:

$$\mathbb{E}[T] \leq N_0 + \sum_{N=N_0}^{N_0+W^*-1} \mathbb{P}[\mathcal{D}_{N_0}(N) = \infty]. \tag{25}$$

According to Lemma A.1, we can write:

$$\begin{aligned}
\mathbb{P}[\mathcal{D}_{N_0}(N) = \infty] &= \mathbb{E} \left[Q_{\frac{1}{2}} \left(\sqrt{N} |\langle \Xi \rangle_N - \xi_0|, \sqrt{N} \lambda_L(N, w_L) \right) \right] \\
&\quad - \mathbb{E} \left[Q_{\frac{1}{2}} \left(\sqrt{N} |\langle \Xi \rangle_N - \xi_0|, \sqrt{N} \lambda_H(N, w_H) \right) \right].
\end{aligned} \tag{26}$$

Under \mathcal{H}_0 , $0 \leq |\langle \Xi \rangle_N - \xi_0| \leq \tau$ (a-s) for all $N \geq N_0$ and thus:

$$\begin{aligned}
\mathbb{P}[\mathcal{D}_{N_0}(N) = \infty] &\stackrel{(a)}{\leq} Q_{\frac{1}{2}} \left(\sqrt{N} \tau, \lambda_{1-\beta/w_L}(\tau\sqrt{N}) \right) \\
&\quad - Q_{\frac{1}{2}} \left(0, \lambda_{\alpha/w_H}(\tau\sqrt{N}) \right) \\
&\stackrel{(b)}{=} 1 - \beta/w_L - Q_{\frac{1}{2}} \left(0, \lambda_{\alpha/w_H}(\tau\sqrt{N}) \right),
\end{aligned}$$

where (a) results from the monotonicity of $Q_{\frac{1}{2}}$ and (b) from (4). The bound on $\mathbb{E}[T]$ under \mathcal{H}_0 follows by substituting the inequality above into (25). Following a similar procedure to bound (26) under \mathcal{H}_1 will yield the bound under \mathcal{H}_1 .

Proof of (ii): The result follows from the bound $\mathbb{P}[\mathcal{D}_{N_0}(N) = \infty] < 1$ for all $N \in \{N_0, \dots, N_0 + W - 1\}$. ■

Theorem 5.3 states that the average stopping time of *T-SeqRDT* is strictly less than the *BlockRDT* block size $N_0 + W^*$. Therefore, on the one hand, Theorem 5.2 suggests that *T-SeqRDT* will lose detection performance compared to *BlockRDT*; but on the other hand, Theorem 5.2 shows that *T-SeqRDT* is faster on average than *BlockRDT*. Moreover, the bounds derived in the two theorems depend on the choice of parameters w_H , w_L , w_{BH} and w_{BL} . As stated earlier, these parameters are used to select the three thresholds and the truncation window required for *T-SeqRDT*. Next, we study the behavior of the error probabilities and the stopping time with these parameters.

C. Trade-off: Error probabilities vs Stopping time

In this subsection, we study how increasing/decreasing $\mathbb{P}_{\text{FA}}(\mathcal{D}_{N_0})$ and $\mathbb{P}_{\text{MD}}(\mathcal{D}_{N_0})$ affect the average stopping time of *T-SeqRDT*. Since $\mathbb{P}_{\text{FA}}(\mathcal{D}_{N_0})$, $\mathbb{P}_{\text{MD}}(\mathcal{D}_{N_0})$ and $\mathbb{E}[T]$ are not available in a closed form, we hereafter study the behavior of the upper bounds for $\mathbb{P}_{\text{FA}}(\mathcal{D}_{N_0})$ and $\mathbb{P}_{\text{MD}}(\mathcal{D}_{N_0})$ with respect to the upper bounds given for $\mathbb{E}[T]$.

Proposition 5.4 (Behavior with w_H and w_L): Given w_{BL} and w_{BH} , we have:

(i) As w_H and w_L tend to ∞ , *T-SeqRDT* approaches *BlockRDT* in the sense that $\lim_{w_L, w_H \rightarrow \infty} \mathbb{E}[T] = N_0 + W^*$;

(ii) As w_H and w_L increase, the upper bounds on $\mathbb{P}_{\text{FA}}(\mathcal{D}_{N_0})$ and $\mathbb{P}_{\text{MD}}(\mathcal{D}_{N_0})$ decrease while the upper bounds on $\mathbb{E}[T]$ increase under each hypothesis.

PROOF:

Proof of (i): Using the definition of the expectation, we have

$$\begin{aligned}
\mathbb{E}[T] &= \sum_{N=N_0}^{N_0+W^*} N \mathbb{P}[T=N] \geq (N_0+W^*) \mathbb{P}[T=N_0+W^*] \\
&= (N_0+W^*) \mathbb{P} \left[\bigcap_{N=N_0}^{N_0+W^*-1} [\mathcal{D}_{N_0}(N) = \infty] \right] \\
&\stackrel{(b)}{\geq} (N_0+W^*) \left(1 - \sum_{N=N_0}^{N_0+W^*-1} \mathbb{P}[\mathcal{D}_{N_0}(N) \neq \infty] \right),
\end{aligned}$$

where (b) follows from the Boole inequality. Moreover, it follows from Proposition 4.2 (iv) that

$$\begin{aligned}
\mathbb{P}[\mathcal{D}_{N_0}(N) = \infty] &= \mathbb{P}[|\langle Y \rangle_N - \xi_0| \leq \lambda_H(N, w_H)] \\
&\quad - \mathbb{P}[|\langle Y \rangle_N - \xi_0| \leq \lambda_L(N, w_L)],
\end{aligned}$$

tends to 1 when both w_H and w_L grow to ∞ . Therefore, the result follows.

Proof of (ii): From Theorem 5.2, the looser upper bound on $\mathbb{P}_{\text{FA}}(\mathcal{D}_{N_0})$ is an inverse function of w_H , whereas the looser upper bound on $\mathbb{P}_{\text{MD}}(\mathcal{D}_{N_0})$ is an inverse function of w_L . Hence, the first part of the statement.

Now, let us look at the upper bounds on the average stopping time from Proposition 5.3. Let us first look at the upper bound under \mathcal{H}_0 . The first term $N_0 + W^*$ is independent of w_L , the second term $\beta W^*/w_L$ decreases when w_L increases and the third term $\sum_{N=N_0}^{N_0+W^*-1} Q_{\frac{1}{2}} \left(0, \lambda_{\alpha/w_H}(\tau\sqrt{N}) \right)$ decreases with increasing w_H , as a consequence of Lemmas 1.1 and B.4. This implies that the upper bound $\mathbb{E}[T]$ under \mathcal{H}_0 will increase with increasing w_H and w_L . Similar reasoning follows for the upper bound under \mathcal{H}_1 . ■

Proposition 5.4 plays an important role in helping us design T-SeqRDT. Proposition 5.4 (i) states that, as w_L and w_H increase, the stopping time of T-SeqRDT approaches the number of samples required by BlockRDT to belong to $\mathcal{C}(\alpha/w_{BH}, \beta/w_{BL})$ (see (15), Remark 3 and discussion afterwards). Moreover, from Theorem 5.2 notice that for all $\epsilon > 0$, there exist w_L and w_H greater than or equal to one such that T-SeqRDT belongs to the class $\mathcal{C}(\alpha/w_{BH} + \epsilon, \beta/w_{BL} + \epsilon)$. This can be achieved by increasing w_H and w_L , which is equivalent to moving the thresholds $\lambda_H(N, w_H)$ and $\lambda_L(N, w_L)$ away from each other (see Proposition 4.2), hence, increasing the average stopping time of T-SeqRDT (Proposition 5.4(ii)). This implies that we can choose larger parameter values w_H and w_L , which moves the thresholds $\lambda_H(N, w_H)$ and $\lambda_L(N, w_L)$ away from each other in order to reduce $\mathbb{P}_{\text{FA}}(\mathcal{D}_{N_0})$ and $\mathbb{P}_{\text{MD}}(\mathcal{D}_{N_0})$. At the same time, this choice of parameters (or thresholds) will increase $\mathbb{E}[T]$ of T-SeqRDT. Next, we analyze the behavior of $\mathbb{E}[T]$ with increasing w_{BH} and w_{BL} .

Proposition 5.5 (Behavior with w_{BH} and w_{BL}): For fixed w_L and w_H , as w_{BH} and w_{BL} increase, the upper bounds on $\mathbb{E}[T]$ increase under each hypothesis.

PROOF: According to Proposition 5.3 (i), we have:

$$\mathbb{E}[T] \leq N_0 + \sum_{N=N_0}^{N_0+W^*-1} \left[1 - \left(\beta/w_L + Q_{\frac{1}{2}} \left(0, \lambda_{\alpha/w_H}(\tau\sqrt{N}) \right) \right) \right]. \quad (27)$$

We have $\beta/w_L < 1/2$ since $\beta < 1/2$ and $w_L \geq 1$. Similarly, since $\alpha < 1/2$ and $w_H \geq 1$, Lemma 1.1, [27, Lemma 2(ii)] and (4) imply that $Q_{\frac{1}{2}} \left(0, \lambda_{\alpha/w_H}(\tau\sqrt{N}) \right) \leq \alpha/w_H < 1/2$. Therefore, the second term on the rhs of (27) is a sum of W^* positive terms. From Proposition 4.3, we know that W^* increases with increasing w_{BH} and w_{BL} . Hence the result under \mathcal{H}_0 . The proof under \mathcal{H}_1 follows similarly. ■

From Proposition 4.3, we know that increasing w_{BH} and w_{BL} will also increase the window size $W^*(w_{BH}, w_{BL})$. Proposition 5.5 above suggests that choosing a larger w_{BH} and w_{BL} , and hence a larger $W^*(w_{BH}, w_{BL})$, while keeping the parameters w_L and w_H fixed, will increase the upper bounds on the average stopping time. But, from Theorem 5.2, we see that, to guarantee that T-SeqRDT belongs to $\mathcal{C}(\alpha, \beta)$, w_L and w_H cannot stay fixed and must satisfy $w_L \geq \frac{w_{BL}W^*(w_{BH}, w_{BL})}{w_{BL}-1}$ and $w_H \geq \frac{w_{BH}W^*(w_{BH}, w_{BL})}{w_{BH}-1}$. Clearly, varying w_H and w_L along with w_{BH} and w_{BL} will also have an impact on the average stopping time. It is not easy to characterize the average stopping time behavior of T-SeqRDT when w_{BH} and w_{BL} increase while maintaining PFA and PMD below levels α and β , respectively. However, Theorem 5.2 suggests that when w_{BH} and w_{BL} are chosen such that $w_{BH} > 1$ and $w_{BL} > 1$, any α and β can be achieved with:

$$w_L = \frac{w_{BL}W^*(w_{BH}, w_{BL})}{w_{BL}-1}, w_H = \frac{w_{BH}W^*(w_{BH}, w_{BL})}{w_{BH}-1}. \quad (28)$$

Now the question that arises is: how should we choose w_H , w_L , w_{BH} and w_{BL} such that $\mathbb{E}[T]$ is minimized and at the same time T-SeqRDT is in $\mathcal{C}(\alpha, \beta)$? The next subsection addresses this question.

D. Tuning T-SeqRDT

We need to choose appropriate thresholds (14) and the window $W^*(w_{BH}, w_{BL})$ (15) such that T-SeqRDT belongs to $\mathcal{C}(\alpha, \beta)$, and at the same time minimizes the average stopping time. The parameters w_H , w_L and w_{BH} fully determine the thresholds (14), whereas w_{BH} and w_{BL} are required to design $W^*(w_{BH}, w_{BL})$. The choice of the appropriate thresholds and window thus boils down to selecting suitable values of parameters w_H , w_L , w_{BH} and w_{BL} . Using (28), we propose to choose the parameters such that the maximum of the two upper bounds on the stopping time derived in Theorem 5.3 is minimized, i.e.,

$$\begin{aligned}
(w_{BH}^*, w_{BL}^*, w_H^*, w_L^*) &= \arg \min_{w_{BH}, w_{BL}, w_H, w_L} \max(\text{UB}_{T_{\mathcal{H}_0}}, \text{UB}_{T_{\mathcal{H}_1}}) \\
\text{s.t. (28), } w_{BH} &> 1, w_{BL} > 1, w_H \geq 1, w_L \geq 1. \quad (29)
\end{aligned}$$

If $w_{BH} = w_{BL}$, which implies that $w_L = w_H$, (29) becomes:

$$\begin{aligned}
(w_{BH}^*, w_H^*) &= \arg \min_{w_{BH}, w_H} \max(\text{UB}_{T_{\mathcal{H}_0}}, \text{UB}_{T_{\mathcal{H}_1}}) \\
\text{s.t. (28), } w_{BH} &> 1, w_H \geq 1. \quad (30)
\end{aligned}$$

The above problem can be further simplified to one-dimensional search via Proposition 5.4(ii), which tells us that for fixed w_{BH} (and w_{BL}), smaller w_H (and w_L) implies smaller bounds on $\mathbb{E}[T]$. Therefore, we can choose w_{BH} (and

w_{BL}), hence $W^*(w_{BH}, w_{BL})$ so as to minimize w_H (and w_L) given by (28) as:

$$w_{BH}^* = \arg \min_{w_{BH}} \frac{w_{BH} W^*(w_{BH}, w_{BL})}{w_{BH} - 1} \text{ s.t. } w_{BH} > 1. \quad (31)$$

Thereby, the upper bounds derived in Theorem 5.2 are maintained equal to α and β and we expect to minimize the stopping time of T-SeqRDT. In the next section, we experimentally show the effect of w_{BH} and w_{BL} on w_L , w_H and $\mathbb{E}[T]$, and pinpoint that the parameters can be chosen in a wide range without significantly impacting $\mathbb{E}[T]$. Since we have a method to choose appropriate values for w_H , w_L , w_{BH} and w_{BL} from which derives $W^* = W^*(w_{BH}, w_{BL})$, we can calculate the thresholds according to (14) and, then, perform T-SeqRDT. Algorithm 1 lists the steps of T-SeqRDT. Next, we perform some simulations to get insights into the behavior of the algorithm.

Algorithm 1: T-SeqRDT

Initialize Given N_0 , τ , τ^+ , α and β .

- 1) Choose w_{BL} and w_{BH} , thereby $W^* = W^*(w_{BH}, w_{BL})$, $w_L = w_{BL} W^* / (w_{BL} - 1)$ and $w_H = w_{BH} W^* / (w_{BH} - 1)$ using (29), (30) or (31).
- 2) Compute $\lambda_H(N, w_H)$, $\lambda_L(N, w_L)$ and $\lambda_{B-RDT}(N, w_{BH})$ from (14)

While $\lambda_L(N, w_L) < |\langle Y \rangle_N - \xi_0| \leq \lambda_H(N, w_H)$ & $N_0 \leq N < N_0 + W^*$
 $N = N + 1$

End

If $|\langle Y \rangle_N - \xi_0| \leq \lambda_L(N, w_L)$ & $N < N_0 + W^*$

Accept \mathcal{H}_0

else if $|\langle Y \rangle_N - \xi_0| > \lambda_H(N, w_H)$ & $N < N_0 + W^*$

Reject \mathcal{H}_0

else if $|\langle Y \rangle_N - \xi_0| \leq \lambda_{B-RDT}(N, w_{BH})$ & $N = N_0 + W^*$

Accept \mathcal{H}_0

else if $|\langle Y \rangle_N - \xi_0| > \lambda_{B-RDT}(N, w_{BH})$ & $N = N_0 + W^*$

Reject \mathcal{H}_0

End If

VI. EXPERIMENTAL RESULTS AND DISCUSSION

In this section, we perform some simulations to highlight the advantages of T-SeqRDT compared to BlockRDT and SPRT as proposed in [28] and [11], [12], respectively. Moreover, we compare T-SeqRDT to a composite hypothesis test, WSPRT defined in [12] and SeqRDT to highlight the differences between the three. We first present the detection problem considered in these experiments. Then we carry out the comparison of the algorithms.

A. Detection with signal distortions

We address the problem of testing the mean of a signal. Let us first consider the case when $Y_n = \Xi_n + X_n$, for $n \in \mathbb{N}$ with $\Xi_n = \xi_0$ under \mathcal{H}_0 and $\Xi_n \neq \xi_0$ under \mathcal{H}_1 . Here, ξ_0 is a deterministic constant and the noise is Gaussian, i.e., $X_n \sim \mathcal{N}(0, 1)$ for all $n \in \mathbb{N}$. This model can be formulated in the framework defined in (5) with $\tau = 0$ and $N_0 = 1$. This is the classical Gaussian mean shift detection problem.

However, in many practical systems, there might unfortunately be a mismatch between the model and the actual signal.

In many practical applications, the underlying signal Ξ_n will not be a constant ξ_0 under \mathcal{H}_0 , but a perturbed version of this value. These unavoidable perturbations are difficult to model in a parametric setup. Therefore, likelihood ratio based tests fail to guarantee reliable performance [5], [16], [27]. However, the BlockRDT [28] and the SeqRDT setup [29], [30] are not limited by these drawbacks. Therefore, instead of dealing with a perfect model as described above, we consider the case when $\Xi_n = \xi_i + \Delta_n$ under \mathcal{H}_i for $i \in \{0, 1\}$ and all $n \in \mathbb{N}$. Here, the Δ_n s model possible perturbations with unknown distribution. We thus want to experimentally assess different algorithms for testing $\Xi = (\Xi_n)_{n \in \mathbb{N}}$ when we observe $Y = (Y_n)_{n \in \mathbb{N}}$ (5). We focus on algorithms in class $\mathcal{C}(\alpha, \beta)$. If the distributions in play are perfectly known, SPRT is optimal in the sense that it makes a faster decision on average, compared to all other algorithms in class $\mathcal{C}(\alpha, \beta)$. Otherwise, if the distributions are not completely known and only partial knowledge of the distortions is assumed, the above hypothesis testing problem can easily be formulated in the framework of (5). Then the problem can efficiently be solved by BlockRDT, SeqRDT or T-SeqRDT. In this respect, we hereafter benchmark T-SeqRDT against WSPRT, SPRT, BlockRDT and SeqRDT under experimental settings described below.

B. Experimental setup

We first list the parameters required to design each algorithm. BlockRDT only requires τ , but guarantees \mathbb{P}_{FA}^{B-RDT} only, with no control over \mathbb{P}_{MD}^{B-RDT} . With additional knowledge of τ^+ , BlockRDT can control both \mathbb{P}_{FA}^{B-RDT} and \mathbb{P}_{MD}^{B-RDT} as illustrated in Proposition 3.2. Likewise, T-SeqRDT also requires τ and τ^+ , whereas SeqRDT requires τ^- , τ , τ^+ and τ_H . On the other hand, SPRT requires complete knowledge of the signal distributions under each hypothesis. Similarly, WSPRT also requires complete knowledge of the signal distributions at least up to an unknown (possibly vector) parameter. Note that BlockRDT is a FSS algorithm whereas the rest of the algorithms are sequential and belong to class $\mathcal{C}(\alpha, \beta)$. For the experimental setup, let us assume τ^- to be some positive real value. We consider ξ_1 and ξ_0 such that $|\xi_1 - \xi_0| \geq 4\tau^-$. We set $\tau^+ = |\xi_1 - \xi_0| - \tau^-$ and $\tau_H \in [|\xi_1 - \xi_0| + \tau^-, \infty)$. Suppose that the empirical mean of the distortion $\Delta = (\Delta_n)_{n \in \mathbb{N}}$ exhibits the following bounded behavior: there exists some $N_0 \in \mathbb{N}$ such that, for all $N \geq N_0$, $0 \leq |\langle \Delta \rangle_N| \leq \tau^-$ and $\tau^+ \leq |\langle \Delta \rangle_N + \xi_1 - \xi_0| \leq \tau_H$. The first inequality captures the signal behavior under \mathcal{H}_0 , whereas the second inequality captures the signal behavior under \mathcal{H}_1 . The problem of testing the mean of Ξ can be rewritten as:

$$\begin{cases} \text{under } \mathcal{H}_0 : \forall N \geq N_0, 0 \leq |\langle \Xi \rangle_N - \xi_0| \leq \tau^- < \tau \quad (\text{a-s}), \\ \text{under } \mathcal{H}_1 : \forall N \geq N_0, \tau < \tau^+ \leq |\langle \Xi \rangle_N - \xi_0| \leq \tau_H \quad (\text{a-s}). \end{cases} \quad (32)$$

We can choose $\tau \in (\tau^-, \tau^+)$. For simulation purposes, we set $\tau = 2\tau^-$. Note that (32) is a special case of the testing problem (5) and can thus be tested using the BlockRDT, SeqRDT and T-SeqRDT frameworks. None of these algorithms need the complete knowledge of the distortion (or signal) distributions under either hypothesis, unlike SPRT and WSPRT, which require the complete knowledge of these distributions under both

hypotheses. We consider three different types of distortions, two when (32) is only required to be satisfied with high probability and the third when it is satisfied with probability 1 (in (a-s) sense).

Case 1: Gaussian distortion: We assume $\Delta_n \stackrel{\text{iid}}{\sim} \mathcal{N}(0, \sigma^2)$ for $n \in \mathbb{N}$. For simulation purposes, we choose $\tau^- = \sigma/4$. For this distortion type, the inequalities in 32 will only be satisfied with high-probability. Below, we list the probabilities corresponding to the Gaussian distortion. We have $\mathbb{P}[|\langle \Delta \rangle_N| \leq \tau] \geq 0.9545$, $\mathbb{P}[|\langle \Delta \rangle_N + \xi_1 - \xi_0| > \tau] \geq 0.9772$, $\mathbb{P}[|\langle \Delta \rangle_N| \leq \tau^-] \geq 0.6827$ and $\mathbb{P}[|\langle \Delta \rangle_N + \xi_1 - \xi_0| > \tau^+] \geq 0.8413$ for all $N \geq N_0$ with $N_0 = 16$ and $|\xi_1 - \xi_0| \geq 2\tau$. Note that these probabilities increase with N .

Case 2: Heavy-Tailed distortion: We model Δ_n as an $\bar{\alpha}$ -stable random variable denoted as $\Delta_n \stackrel{\text{iid}}{\sim} \mathcal{S}(\bar{\alpha}, \bar{\beta}, \bar{\gamma}, \bar{\delta})$ for $n \in \mathbb{N}$ [35]. The parameters $\bar{\alpha} \in (0, 2]$, $\bar{\beta} \in [-1, 1]$, $\bar{\gamma} > 0$ and $\bar{\delta} \in (-\infty, \infty)$ are the tail-index, location, dispersion and skewness parameters, respectively. In general, an $\bar{\alpha}$ -stable distribution does not admit a closed-form probability density function, except in a few special cases like the Cauchy ($\bar{\alpha} = 1, \bar{\beta} = 0$) and Gaussian ($\bar{\alpha} = 2$) distributions. Moreover, for the Cauchy distribution and for $\bar{\alpha} \in (0, 1]$, none of the moments of the $\bar{\alpha}$ -stable distribution exist. For $\bar{\alpha} \in (1, 2)$ the distribution is sometimes referred to as the Pareto-Lévy distribution and for this class of distributions, all higher moments beyond the mean do not exist. For simulation purposes, we consider the following two types of heavy-tailed distortions:

Case 2(i) [Pareto-Lévy distortion ($\bar{\alpha} \in (1, 2)$):] We assume the distortion to be Pareto-Lévy distributed with $\Delta_n \stackrel{\text{iid}}{\sim} \mathcal{S}(1.5, 0, \tau^-, 0)$ for $n \in \mathbb{N}$. We thus have: $\mathbb{P}[|\langle \Delta \rangle_N| \leq \tau] \geq 0.9885$, $\mathbb{P}[|\langle \Delta \rangle_N + \xi_1 - \xi_0| > \tau] \geq 0.9953$, $\mathbb{P}[|\langle \Delta \rangle_N| \leq \tau^-] \geq 0.9646$, $\mathbb{P}[|\langle \Delta \rangle_N + \xi_1 - \xi_0| > \tau^+] \geq 0.9832$ for all $N \geq N_0$ with $N_0 = 30$ and $|\xi_1 - \xi_0| \geq 2\tau$. Again, note that these probabilities increase with N .

Case 2(ii) [Cauchy distortion ($\bar{\alpha} = 1$):] Note that, unlike in the cases involving Gaussian and Pareto-Lévy distortions, the empirical mean of i.i.d Cauchy distributed random variables is again Cauchy distributed [35] and none of the moments exist for the empirical mean as well. Therefore, the empirical mean of a Cauchy distorted signal does not converge in the neighborhood of ξ_0 and ξ_1 under \mathcal{H}_0 and \mathcal{H}_1 , respectively, in contrast to the Gaussian and Pareto-Lévy distortions. Below, we show that, although the Cauchy distortion does not exhibit the desired convergence properties, the proposed algorithms guarantee performance if (32) holds with sufficiently high probabilities. To experimentally show this, we assume the distortion to be Cauchy distributed as $\Delta_n \stackrel{\text{iid}}{\sim} \mathcal{S}(1, 0, \tau^-/10, 0)$ for $n \in \mathbb{N}$ with the associated probabilities given as: $\mathbb{P}[|\langle \Delta \rangle_N| \leq \tau] = 0.9682$ and $\mathbb{P}[|\langle \Delta \rangle_N| \leq \tau^-] = 0.9365$ for all $N \in \mathbb{N}$. Also, $\mathbb{P}[|\langle \Delta \rangle_N + \xi_1 - \xi_0| > \tau] \geq 0.9894$ and $\mathbb{P}[|\langle \Delta \rangle_N + \xi_1 - \xi_0| > \tau^+] \geq 0.9728$ for all $N \in \mathbb{N}$ and $|\xi_1 - \xi_0| \geq 2\tau$. Note that, unlike the Cases 1 and 2(i) these probabilities do not increase with N as the distribution of the empirical mean of a Cauchy distribution remains the same. As a consequence, the probabilities stay the same for all $N \in \mathbb{N}$.

$\alpha = \beta = 0.01$					
SNR = $ \xi_1 - \xi_0 $		0.4	0.5	0.6	0.8
BlockRDT	$N_{\text{B-RDT}}$	2165	542	241	87
SeqRDT	$\mathbb{E}[T_{\text{SeqRDT}}]$	171.34	145.83	141.12	140.26
T-SeqRDT	$\mathbb{E}[T]$	567.73	349.42	192.82	73.92

$\alpha = \beta = 0.001$					
BlockRDT	$N_{\text{B-RDT}}$	3820	955	425	153
SeqRDT	$\mathbb{E}[T_{\text{SeqRDT}}]$	252.55	198.51	185.17	181.80
T-SeqRDT	$\mathbb{E}[T]$	720.08	481.75	298.65	114.88

TABLE I: T-SeqRDT vs SeqRDT and BlockRDT for Gaussian distortion.

$\alpha = \beta = 0.01$					
SNR = $ \xi_1 - \xi_0 $		0.4	0.5	0.6	0.8
T-SeqRDT	$\mathbb{E}[T]$	567.73	349.42	192.82	73.92
	$\mathbb{P}_{\text{FA}}(\mathcal{D}_{N_0})$	0.0004	0.0001	0.0001	0.0002
	$\mathbb{P}_{\text{MD}}(\mathcal{D}_{N_0})$	0.0001	0.0003	0.0002	0.0003
SeqRDT	$\mathbb{E}[T_{\text{SeqRDT}}]$	171.34	145.83	141.12	140.26
	$\mathbb{P}_{\text{FA}}(\mathcal{D}_M)$	0.00015	0.0002	0.0002	0.00021
	$\mathbb{P}_{\text{MD}}(\mathcal{D}_M)$	0.00044	0.00013	4×10^{-5}	$< 10^{-5}$
SPRT	$\mathbb{E}[T_{\text{SPRT}}]$	58.98	38.40	27.18	15.84
	$\mathbb{P}_{\text{FA}}^{\text{SPRT}}$	0.0150	0.0146	0.0131	0.0114
	$\mathbb{P}_{\text{MD}}^{\text{SPRT}}$	0.0149	0.0143	0.0142	0.0124
WSPRT	$\mathbb{E}[T_{\text{WSPRT}}]$	209.77	198.55	191.85	184.07
	$\mathbb{P}_{\text{FA}}^{\text{WSPRT}}$	0.0192	0.0185	0.0187	0.0182
	$\mathbb{P}_{\text{MD}}^{\text{WSPRT}}$	$< 10^{-5}$	$< 10^{-5}$	$< 10^{-5}$	$< 10^{-5}$
GSPRT	$\mathbb{E}[T_{\text{GSPRT}}]$	41.47	26.01	18.07	10.68
	$\mathbb{P}_{\text{FA}}(\mathcal{D}_M)$	0.1307	0.1339	0.1377	0.1292
	$\mathbb{P}_{\text{MD}}(\mathcal{D}_M)$	0.1430	0.1481	0.1433	0.1325

$\alpha = \beta = 0.001$					
SNR = $ \xi_1 - \xi_0 $		0.4	0.5	0.6	0.8
T-SeqRDT	$\mathbb{E}[T]$	720.08	481.75	298.65	114.88
	$\mathbb{P}_{\text{FA}}(\mathcal{D}_{N_0})$	3×10^{-5}	2×10^{-5}	5×10^{-5}	1×10^{-5}
	$\mathbb{P}_{\text{MD}}(\mathcal{D}_{N_0})$	1×10^{-5}	3×10^{-5}	2×10^{-5}	$< 10^{-5}$
SeqRDT	$\mathbb{E}[T_{\text{SeqRDT}}]$	252.55	198.51	185.17	181.80
	$\mathbb{P}_{\text{FA}}(\mathcal{D}_M)$	3×10^{-5}	1×10^{-5}	4×10^{-5}	2×10^{-5}
	$\mathbb{P}_{\text{MD}}(\mathcal{D}_M)$	0.00011	3×10^{-5}	1×10^{-5}	$< 10^{-5}$
SPRT	$\mathbb{E}[T_{\text{SPRT}}]$	89.24	57.82	40.56	23.32
	$\mathbb{P}_{\text{FA}}^{\text{SPRT}}$	0.0022	0.0019	0.0018	0.0016
	$\mathbb{P}_{\text{MD}}^{\text{SPRT}}$	0.0023	0.0020	0.0019	0.0018
WSPRT	$\mathbb{E}[T_{\text{WSPRT}}]$	304.37	288.09	278.55	267.46
	$\mathbb{P}_{\text{FA}}^{\text{WSPRT}}$	0.0033	0.0033	0.0030	0.0036
	$\mathbb{P}_{\text{MD}}^{\text{WSPRT}}$	$< 10^{-5}$	$< 10^{-5}$	$< 10^{-5}$	$< 10^{-5}$
GSPRT	$\mathbb{E}[T_{\text{GSPRT}}]$	68.75	42.39	28.76	16.05
	$\mathbb{P}_{\text{FA}}^{\text{GSPRT}}$	0.0937	0.0970	0.0981	0.0994
	$\mathbb{P}_{\text{MD}}^{\text{GSPRT}}$	0.1021	0.1059	0.1047	0.1003

TABLE II: T-SeqRDT vs SeqRDT, SPRT and WSPRT. Here, $\text{PFA} < 10^{-5}$ and $\text{PMD} < 10^{-5}$ indicate that probabilities of errors are at most of the order of 10^{-5} .

To ensure that (32) is satisfied with high probability, we need the dispersion parameter, $\bar{\gamma}$, to be small enough. Later in the section we show how the above probabilities, PFA and PMD vary with $\bar{\gamma}$ for T-SeqRDT.

Case 3 [Deterministic unknown distortion]: The distortion is assumed to be unknown deterministic with $|\Delta_n| \leq \tau^-$ for all $n \in \mathbb{N}$. For simulation purposes, we choose $\Delta_n = \tau^-$. With this choice, the inequalities in (32) are satisfied with probability 1. However, not all types of distortions satisfy (32) with probability 1 as shown in Cases 1 and 2. Next, we discuss different algorithms.

C. Algorithms

Here, we discuss the algorithms we use to solve the above mean testing problem. Certainly, using (32), we can cast the problem in the *BlockRDT*, *SeqRDT* and *T-SeqRDT* frameworks. Next, we discuss likelihood ratio based parametric and semiparametric approaches for comparison purposes.

Sequential probability ratio test (SPRT): For SPRT, we assume that the probability density function, f_i , of the observations is known under \mathcal{H}_i for $i = 0, 1$. For $\alpha, \beta \in (0, 1/2)$, and with initialization $\Lambda_N = 1$, SPRT with stopping time and decision pair $(T_{\text{SPRT}}, \mathcal{D})$ is defined as:

$$T_{\text{SPRT}} = \inf\{N \geq 0 : \Lambda_N \notin (\lambda_L^{\text{SPRT}}, \lambda_H^{\text{SPRT}})\}$$

$$\mathcal{D}(N) = \begin{cases} 1 & \text{if } \Lambda_N \geq \lambda_H^{\text{SPRT}} \\ 0 & \text{if } \Lambda_N \leq \lambda_L^{\text{SPRT}} \\ \infty & \text{if } \lambda_L^{\text{SPRT}} < \Lambda_N < \lambda_H^{\text{SPRT}} \end{cases}$$

where $\Lambda_N = \sum_{n=1}^N \frac{f_1(Y_i)}{f_0(Y_i)}$ is the likelihood ratio based on the observations, $\lambda_L^{\text{SPRT}} = \frac{\beta}{1-\alpha}$ and $\lambda_H^{\text{SPRT}} = \frac{1-\beta}{\alpha}$ are the lower and upper thresholds, respectively. We denote the stopping time, PFA and PMD of SPRT as T_{SPRT} , $\mathbb{P}_{\text{FA}}^{\text{SPRT}}$ and $\mathbb{P}_{\text{MD}}^{\text{SPRT}}$, respectively. For the model described above, SPRT for detecting the mean with unknown distortions we have $\Lambda_N = \exp\left(N \frac{\xi_0^2 - \xi_1^2}{2} + (\xi_1 - \xi_0) \sum_{n=1}^N Y_n\right)$.

Composite hypothesis test, GSPRT: A simple GSPRT can be designed for the case of Gaussian distortions when the means under \mathcal{H}_0 and \mathcal{H}_1 are known but the variances are unknown. Specifically, the algorithm is aware that the distortion is zero mean Gaussian distributed, but is unaware of its variance [5]. The generalized log likelihood ratio for such a test is given as: $\log \hat{\Lambda}_N = \frac{\xi_1 - \xi_0}{s_n^2} \sum_{n=1}^N (Y_n - \frac{1}{2}(\xi_0 + \xi_1))$ with $s_n^2 = \frac{1}{n-1} \sum_{i=1}^n (Y_i - \langle Y \rangle_i)$, for $N \geq 2$. GSPRT uses the same thresholds as SPRT [5]. We denote the stopping time, PFA and PMD of GSPRT as $\mathbb{E}[T_{\text{GSPRT}}]$, $\mathbb{P}_{\text{FA}}^{\text{GSPRT}}$ and $\mathbb{P}_{\text{MD}}^{\text{GSPRT}}$, respectively.

Composite hypothesis test, WSPRT: WSPRT considers the problem of testing $\mathcal{H}_0 : |\xi - \xi_0| < \tau$ vs $\mathcal{H}_1 : |\xi - \xi_0| > \tau$ [12, Chapter 4]. In *T-SeqRDT*, the signal Ξ is assumed to be a corrupted version of ξ and the distribution of Ξ is assumed to be unknown. In contrast, for WSPRT, ξ is deterministic and the test can only handle the case when the observations, Y_i s, are Gaussian distributed [12], [16]. We denote PFA, PMD and stopping time of WSPRT as $\mathbb{P}_{\text{FA}}^{\text{WSPRT}}$, $\mathbb{P}_{\text{MD}}^{\text{WSPRT}}$ and T_{WSPRT} , respectively. WSPRT uses the same thresholds as SPRT. However, the likelihood ratio for WSPRT is given as: $\hat{\Lambda}_N = \prod_{n=1}^N \frac{e^{-(Y_n - \xi_0 + \tau)^2/2} + e^{-(Y_n - \xi_0 - \tau)^2/2}}{2e^{-(Y_n - \xi_0)^2/2}}$. The distribution under \mathcal{H}_1 is thus replaced by a weighted average of two distributions.

D. Comparison: *T-SeqRDT*, *SeqRDT*, *BlockRDT*, *SPRT*, *WSPRT* and *GSPRT*

We define $|\xi_1 - \xi_0|$ as the SNR and for simulation purposes, we assume $\tau^- = 0.1$. For *T-SeqRDT*, the thresholds (4) and the truncation window $W^*(w_{BH}, w_{BL})$ (15) are selected via parameters w_H , w_L , w_{BH} and w_{BL} , using Algorithm 1. For the simulations, we assume $w_{BL} = w_{BH}$, which implies

$\alpha = \beta = 0.01$					
SNR = $ \xi_1 - \xi_0 $		0.4	0.5	0.6	0.8
<i>T-SeqRDT</i>	$\mathbb{E}[T]$	576.14	350.71	192.296	75.17
	$\mathbb{P}_{\text{FA}}(\mathcal{D}_{N_0})$	0.0097	0.0095	0.0079	0.0075
	$\mathbb{P}_{\text{MD}}(\mathcal{D}_{N_0})$	0.0003	0.0002	0.0004	0.0013
<i>SeqRDT</i>	$\mathbb{E}[T_{\text{SeqRDT}}]$	171.71	146.85	142.97	141.81
	$\mathbb{P}_{\text{FA}}(\mathcal{D}_M)$	0.0088	0.0085	0.0086	0.0089
	$\mathbb{P}_{\text{MD}}(\mathcal{D}_M)$	0.0044	0.0013	0.0004	0.0006
SPRT	$\mathbb{E}[T_{\text{SPRT}}]$	58.56	37.83	27.10	15.82
	$\mathbb{P}_{\text{FA}}^{\text{SPRT}}$	0.0211	0.0204	0.0167	0.0132
	$\mathbb{P}_{\text{MD}}^{\text{SPRT}}$	0.0200	0.0199	0.0165	0.0148
WSPRT	$\mathbb{E}[T_{\text{WSPRT}}]$	207.63	195.07	189.20	182.07
	$\mathbb{P}_{\text{FA}}^{\text{WSPRT}}$	0.0454	0.0495	0.0517	0.0505
	$\mathbb{P}_{\text{MD}}^{\text{WSPRT}}$	0.0001	$< 10^{-4}$	$< 10^{-4}$	$< 10^{-4}$

TABLE III: *T-SeqRDT* vs *SeqRDT*, SPRT and WSPRT for heavy-tailed Pareto-Lévy Distortion.

$\alpha = \beta = 0.05$					
SNR = $ \xi_1 - \xi_0 $		0.4	0.5	0.6	0.8
<i>T-SeqRDT</i>	$\mathbb{E}[T]$	453.05	229.33	117.30	44.59
	$\mathbb{P}_{\text{FA}}(\mathcal{D}_{N_0})$	0.0471	0.0424	0.0357	0.0271
	$\mathbb{P}_{\text{MD}}(\mathcal{D}_{N_0})$	0.0047	0.0040	0.0052	0.0069
<i>SeqRDT</i>	$\mathbb{E}[T_{\text{SeqRDT}}]$	124.73	115.10	113.75	111.23
	$\mathbb{P}_{\text{FA}}(\mathcal{D}_M)$	0.0303	0.0299	0.0283	0.0286
	$\mathbb{P}_{\text{MD}}(\mathcal{D}_M)$	0.0087	0.0037	0.0025	0.0013
SPRT	$\mathbb{E}[T_{\text{SPRT}}]$	35.51	23.50	17.03	10.07
	$\mathbb{P}_{\text{FA}}^{\text{SPRT}}$	0.0570	0.0491	0.0481	0.0408
	$\mathbb{P}_{\text{MD}}^{\text{SPRT}}$	0.0612	0.0531	0.0510	0.0419
WSPRT	$\mathbb{E}[T_{\text{WSPRT}}]$	132.75	124.89	119.00	114.63
	$\mathbb{P}_{\text{FA}}^{\text{WSPRT}}$	0.1198	0.1162	0.1210	0.1115
	$\mathbb{P}_{\text{MD}}^{\text{WSPRT}}$	0.0006	0.0002	$< 10^{-4}$	$< 10^{-4}$

TABLE IV: *T-SeqRDT* vs *SeqRDT*, SPRT and WSPRT for heavy-tailed Cauchy distortion.

$\alpha = \beta = 0.01$					
SNR = $ \xi_1 - \xi_0 $		0.4	0.5	0.6	0.8
<i>T-SeqRDT</i>	$\mathbb{E}[T]$	387.86	177.24	106.44	48.42
	$\mathbb{P}_{\text{FA}}(\mathcal{D}_{N_0})$	0.0059	0.0062	0.0063	0.0063
	$\mathbb{P}_{\text{MD}}(\mathcal{D}_{N_0})$	$< 10^{-4}$	$< 10^{-4}$	$< 10^{-4}$	$< 10^{-4}$
<i>SeqRDT</i>	$\mathbb{E}[T_{\text{SeqRDT}}]$	220.10	210.61	206.25	202.16
	$\mathbb{P}_{\text{FA}}(\mathcal{D}_M)$	0.0050	0.0065	0.0061	0.0058
	$\mathbb{P}_{\text{MD}}(\mathcal{D}_M)$	9×10^{-4}	$< 10^{-4}$	$< 10^{-4}$	$< 10^{-4}$
SPRT	$\mathbb{E}[T_{\text{SPRT}}]$	70.77	43.40	29.75	16.73
	$\mathbb{P}_{\text{FA}}^{\text{SPRT}}$	0.0802	0.0537	0.0333	0.0222
	$\mathbb{P}_{\text{MD}}^{\text{SPRT}}$	0.0006	0.0012	0.0012	0.0014
WSPRT	$\mathbb{E}[T_{\text{WSPRT}}]$	317.05	310.01	305.50	303.89
	$\mathbb{P}_{\text{FA}}^{\text{WSPRT}}$	0.4566	0.4439	0.4538	0.4569
	$\mathbb{P}_{\text{MD}}^{\text{WSPRT}}$	$< 10^{-4}$	$< 10^{-4}$	$< 10^{-4}$	$< 10^{-4}$
$\alpha = \beta = 0.05$					
SNR = $ \xi_1 - \xi_0 $		0.4	0.5	0.6	0.8
<i>T-SeqRDT</i>	$\mathbb{E}[T]$	232.18	109.74	63.85	29.34
	$\mathbb{P}_{\text{FA}}(\mathcal{D}_{N_0})$	0.0244	0.0259	0.0232	0.0275
	$\mathbb{P}_{\text{MD}}(\mathcal{D}_{N_0})$	0.0002	0.0004	0.0005	0.0010
<i>SeqRDT</i>	$\mathbb{E}[T_{\text{SeqRDT}}]$	120.23	105.07	101.30	98.78
	$\mathbb{P}_{\text{FA}}(\mathcal{D}_M)$	0.0441	0.0442	0.0440	0.0466
	$\mathbb{P}_{\text{MD}}(\mathcal{D}_M)$	0.0052	6×10^{-4}	2×10^{-4}	$< 10^{-4}$
SPRT	$\mathbb{E}[T_{\text{SPRT}}]$	39.56	25.38	17.87	10.50
	$\mathbb{P}_{\text{FA}}^{\text{SPRT}}$	0.1653	0.1265	0.1029	0.0694
	$\mathbb{P}_{\text{MD}}^{\text{SPRT}}$	0.0070	0.0100	0.0098	0.0113
WSPRT	$\mathbb{E}[T_{\text{WSPRT}}]$	160.72	156.14	152.06	149.89
	$\mathbb{P}_{\text{FA}}^{\text{WSPRT}}$	0.4299	0.4272	0.4253	0.4202
	$\mathbb{P}_{\text{MD}}^{\text{WSPRT}}$	$< 10^{-4}$	$< 10^{-4}$	$< 10^{-4}$	$< 10^{-4}$

TABLE V: *T-SeqRDT* vs *SeqRDT*, SPRT and WSPRT for deterministic distortion.

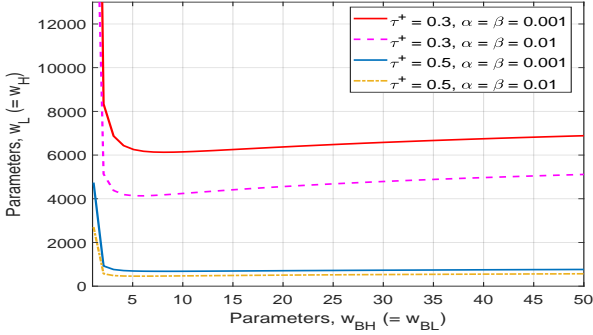


Fig. 1: $w_H = w_L$ vs $w_{BH} = w_{BL}$ such that UB_{FA} and UB_{MD} in Theorem 5.2 stay equal to α and β , respectively.

that $w_H = w_L$ and make use of the simpler of the three methods proposed in (31) to choose the parameters. In Figures 1 and 2, we plot w_H against w_{BH} and $\mathbb{E}[T]$, respectively, for different α , β and τ^+ . We notice that w_H and w_L capture the behavior of $\mathbb{E}[T]$ as was suggested by Proposition 5.4(ii). Moreover, we see that we can choose w_{BH} and w_{BL} (hence w_H and w_L) over a wide range. For simulation purposes, we select $w_{BH} = w_{BL} = 2$, which is then used to choose an appropriate W^* via (15). We then set $w_H = w_L = 2W^*$. Albeit not unique, these choices guarantee that T-SeqRDT is in $\mathcal{C}(\alpha, \beta)$ by Theorem 5.2. In contrast to T-SeqRDT and BlockRDT, SeqRDT requires additional knowledge of τ^- and τ_H to design the test: τ along with levels α and β is used to design the thresholds; τ^+ , τ^- and τ_H are used to choose an appropriate buffer size, M . The thresholds, $\lambda_\alpha(\tau\sqrt{N})/\sqrt{N}$ and $\lambda_{1-\beta}(\tau\sqrt{N})/\sqrt{N}$ are defined in (13). We denote by $\mathbb{P}_{\text{FA}}(\mathcal{D}_M)$ and $\mathbb{P}_{\text{MD}}(\mathcal{D}_M)$ the PFA and PMD, respectively, and by T_{SeqRDT} the stopping time of SeqRDT.

Case 1: For Gaussian distortion, we compare in Table I the average stopping times of T-SeqRDT and SeqRDT to the block-size of BlockRDT, for different SNR values, $\alpha = \beta = 0.01$ and $\alpha = \beta = 0.001$. For SeqRDT, the buffer size $M = 90$ is selected. From Table I, SeqRDT is the fastest on average, especially at low SNR values, but needs the most amount of information (all of τ^- , τ , τ^+ and τ_H to design M) about the signal. BlockRDT is the slowest and requires the same information (τ and τ^+ only) as T-SeqRDT. However, T-SeqRDT is considerably faster on average. Moreover, at moderate to high SNRs, T-SeqRDT is the fastest among the three algorithms and considerably outperforms SeqRDT as well. It must be noted that the stopping time of SeqRDT is limited by the need and the choice of the buffer size, which makes SeqRDT relatively slower compared to T-SeqRDT, especially at higher SNRs. In Table II, we compare the average stopping times, PFAs and PMDs of T-SeqRDT, SeqRDT, SPRT, WSPRT and GSPRT. From Table II, we notice that, because of the distortion, WSPRT and SPRT do not belong to $\mathcal{C}(\alpha, \beta)$ as $\mathbb{P}_{\text{FA}}^{\text{WSPRT}}$, $\mathbb{P}_{\text{MD}}^{\text{WSPRT}}$ and $\mathbb{P}_{\text{FA}}^{\text{GSPRT}}$ are above the pre-specified levels α and β . Moreover, GSPRT, even with prior knowledge of the distortion, does not belong to $\mathcal{C}(\alpha, \beta)$, as the $\mathbb{P}_{\text{FA}}^{\text{GSPRT}}$ and $\mathbb{P}_{\text{MD}}^{\text{GSPRT}}$ are orders of magnitude higher compared to α and β , respectively. In contrast, both SeqRDT and T-SeqRDT, with only limited knowledge about the signal under

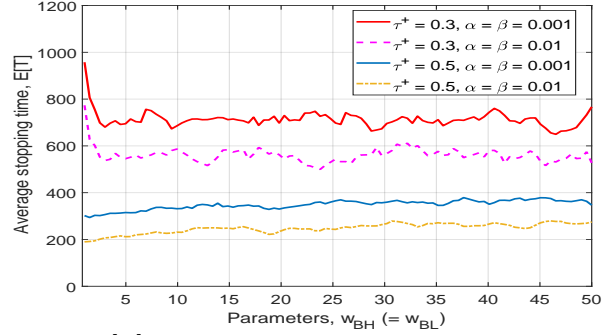


Fig. 2: $\mathbb{E}[T]$ vs $w_{BH} = w_{BL}$ such that UB_{FA} and UB_{MD} in Theorem 5.2 stay equal to α and β , respectively.

each hypothesis, are in $\mathcal{C}(\alpha, \beta)$. Importantly, by design, T-SeqRDT eliminates the need for buffer M , whereas SeqRDT does need such a buffer to guarantee the pre-specified levels α and β . Moreover, it seems that the bounds on PFA and PMD are loose for Gaussian distortion. Therefore, we next consider different types of distortions to see if the bounds are tight for some other scenarios.

Case 2: For heavy-tailed distortions, we again compare T-SeqRDT, SeqRDT, SPRT and WSPRT. For Cases 2(i) and 2(ii), we simulate PFA, PMD and average stopping times for $\alpha = \beta = 0.01$ and $\alpha = \beta = 0.05$ to obtain Tables III and IV, respectively. The average stopping time of T-SeqRDT and SeqRDT stay similar to those obtained in the Gaussian distortion case. However, the bounds on PFA and PMD are tight, as a consequence of the heavy-tailed distribution of the distortion. Moreover, similar to Case 1, SPRT and WSPRT do not belong to $\mathcal{C}(\alpha, \beta)$ for both Cases 2(i) and 2(ii).

Case 3: Finally, we consider the unknown deterministic distortion case. In this case, (32) is satisfied with probability 1, unlike in Cases 1 and 2. We choose $\Delta_n = \tau^-$, and simulate PFA, PMD and average stopping times for $\alpha = \beta = 0.05$ and $\alpha = \beta = 0.01$. From Table V, T-SeqRDT and SeqRDT belong to $\mathcal{C}(\alpha, \beta)$, whereas SPRT and WSPRT fail to. Also, note that, similar to Case 2, the bounds on PFA are tight.

A note of caution: The above simulation results show that T-SeqRDT is robust to mismatches and can guarantee performance even in the cases when (32) is not always satisfied with probability 1, as shown in Cases 1 and 2 above. Now, the question that arises is: “How high do these probabilities need to be so that T-SeqRDT belongs to $\mathcal{C}(\alpha, \beta)$?” The simulation results of Case 1 suggest that Gaussian distortions allow for large mismatches, i.e., T-SeqRDT works even when the probabilities are not very high. On the other hand, Case 2, involving heavy-tailed distortions, requires these probabilities to be high, i.e., a relatively smaller mismatch. In the following, for two different cases, we show how high these probabilities need to be for T-SeqRDT to belong to $\mathcal{C}(\alpha, \beta)$.

Case A: We mentioned earlier in Case 2(ii) that, with Cauchy distortions, we needed the dispersion parameter $\bar{\gamma}$ to be small. In Fig. 3, we show how PFA and PMD of T-SeqRDT and the probabilities associated in (32), when they are not satisfied in (a-s) sense, vary with increasing $\bar{\gamma}$, for $\alpha = \beta = 0.05$ and SNR = 0.8. Notice that there exists a

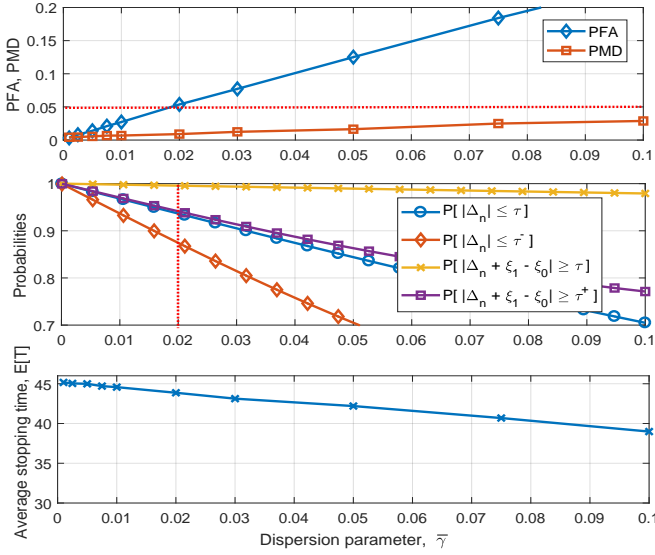


Fig. 3: $\mathbb{P}_{FA}(\mathcal{D}_{N_0})$, $\mathbb{P}_{MD}(\mathcal{D}_{N_0})$, Probabilities in (32) and $\mathbb{E}[T]$ against $\bar{\gamma}$ for T-SeqRDT.

threshold $\bar{\gamma} = 0.02$ above which T-SeqRDT does not belong to $\mathcal{C}(\alpha, \beta)$ and that we need probabilities in (32) as high as 95% for T-SeqRDT to guarantee the required performance.

Case B: We perform further simulations for a simple model of impulsive distortion. We assume that Δ_n is Bernoulli distributed as

$$\Delta_n \sim \begin{cases} 10\tau^- & \text{with probability } p \\ 0 & \text{with probability } 1 - p \end{cases}$$

This implies that $\mathbb{P}[|\langle \Delta \rangle_N| \leq \tau] \geq 1 - p$ and $\mathbb{P}[|\langle \Delta_n - \xi_0 + \xi_1 \rangle_N| > \tau] = 1$ for all $N \geq N_0$ with $N_0 = 1$. In Fig. 4, we show PFA and PMD of T-SeqRDT for $\alpha = \beta = 0.05$ and SNR = 0.8. As expected, as the mismatch grows, PFA grows and crosses the level α if p increases beyond 20%. The probabilities in the above cases depend on a multitude of parameters like, SNR, tolerances, levels α and β and, most importantly, on the underlying signal distribution, as shown in all the above cases. The above discussion shows the flexibility as well as the robustness of T-SeqRDT.

VII. CONCLUSION AND PERSPECTIVES

In this work, we proposed an alternative approach to hypothesis testing. The proposed formulation can be useful when the signal distributions in play are unknown as it is completely data dependent rather than model dependent. We introduced a new non-parametric algorithm, T-SeqRDT, for sequential hypothesis testing. The work builds on earlier work where the authors proposed SeqRDT [29], [30]. SeqRDT, although marginally faster compared to T-SeqRDT, required more assumptions on the signal along with a buffer to ensure the desired performance. In contrast, T-SeqRDT guarantees the required performance and eliminates the need for a buffer by designing thresholds and truncation window appropriately. We studied the properties of these thresholds along with the trade-off between the error probabilities and the stopping time. Finally, simulations showed that T-SeqRDT, even with little

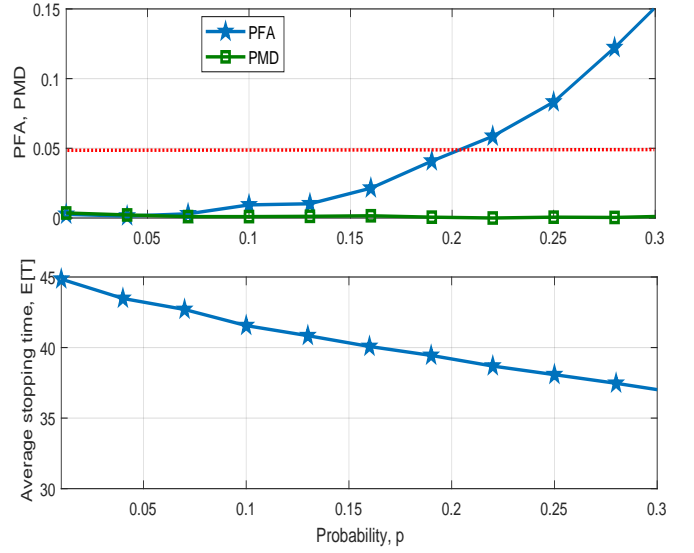


Fig. 4: $\mathbb{P}_{FA}(\mathcal{D}_{N_0})$, $\mathbb{P}_{MD}(\mathcal{D}_{N_0})$ and $\mathbb{E}[T]$ against probability of impulse, p for T-SeqRDT.

knowledge of the signal, is able to provide sufficient performance guarantees while making a decision faster on average compared to BlockRDT. Furthermore, with the knowledge of a few parameters only, instead of the complete distributions, the proposed framework is capable of carrying out hypothesis testing. Another critical feature of the proposed approach is that it gives the algorithm designer freedom to choose these parameters, thus making it possible to test signals with arbitrarily low SNRs. Also, the algorithm is robust to mismatches in signal distributions as it does not rely on the underlying signal distributions. In the future, the optimality properties of the proposed tests remain to be studied. Also, extension of T-SeqRDT to multi-dimensional signals may be addressed. Moreover, the simulations indicate that the bounds derived on PFA and PMD might be loose in some cases, which highlights the need for novel assumptions and analysis in the future. Generalization of the model to distributed systems might be another future endeavor. In conclusion, we believe that the proposed hypothesis testing approach provides an appealing alternative to likelihood ratio based sequential frameworks.

ACKNOWLEDGMENT

We gratefully acknowledge the insightful comments of the reviewers and the Associate Editor whose suggestions were extremely valuable to improve the clarity of the paper.

APPENDIX A

Lemma A.1: For any $N \in \mathbb{N}$ and any $\eta \geq 0$, we have:

$$\mathbb{P}[|\langle \Xi \rangle_N + \langle X \rangle_N - \xi_0| > \eta] = \mathbb{E} \left[Q_{\frac{1}{2}} \left(\sqrt{N} |\langle \Xi \rangle_N - \xi_0|, \eta \sqrt{N} \right) \right]$$

PROOF: By property of a conditional and taking the independence of $\langle \Xi \rangle_N$ and $\langle X \rangle_N$ into account, we have:

$$\begin{aligned} \mathbb{P}[|\langle \Xi \rangle_N + \langle X \rangle_N - \xi_0| \leq \eta] \\ = \int_0^\infty \mathbb{P}[|\rho + \langle X \rangle_N| \leq \eta] \mathbb{P}[|\langle \Xi \rangle_N - \xi_0|^{-1}(\mathrm{d}\rho)] \end{aligned}$$

It follows from $X \sim \mathcal{N}(0, 1)$ that, for all $\rho \in [0, \infty)$:

$$\mathbb{P}[|\rho + \langle X \rangle_N| \leq \eta] = \Phi(\sqrt{N}(\eta - \rho)) - \Phi(-\sqrt{N}(\eta + \rho))$$

The foregoing and (3) imply the result through the equality:

$$\mathbb{P}[|\langle \Xi \rangle_N + \langle X \rangle_N - \xi_0| \leq \eta] = 1 - \mathbb{E}\left[Q_{\frac{1}{2}}\left(\sqrt{N}|\langle \Xi \rangle_N - \xi_0|, \eta\sqrt{N}\right)\right].$$

APPENDIX B

Lemma B.1: For any $\gamma \in (0, 1)$:

- (i) $\lim_{\rho \rightarrow \infty} (\lambda_\gamma(\rho) - \rho) = \Phi^{-1}(1 - \gamma)$;
- (ii) $\lim_{\rho \rightarrow \infty} \lambda_\gamma(\rho)/\rho = 1$;

PROOF: We prove (i) only since it straightforwardly implies (ii). Pose $g_\gamma(\rho) = \lambda_\gamma(\rho) - \rho$ and $\theta = \Phi^{-1}(1 - \gamma)$. Since $\Phi(x) + \Phi(-x) = 1$, (3) and the definition of $\lambda_\gamma(\tau)$ induce that:

$$\Phi(g_\gamma(\rho)) + \Phi(g_\gamma(\rho) + 2\rho) = 1 + \Phi(\theta). \quad (33)$$

To prove that $g_\gamma(\rho)$ tends to θ when $\rho \rightarrow \infty$, we proceed by contradiction. If $g_\gamma(\rho)$ does not tend to θ when $\rho \rightarrow \infty$, there exists some positive real number ε such that, for all $n \in \mathbb{N}$, there exists some real number $\rho_n > n$ such that either $g_\gamma(\rho_n) > \theta + \varepsilon$ or $g_\gamma(\rho_n) < \theta - \varepsilon$. Basically, $\lim_{n \rightarrow \infty} \rho_n = \infty$. Consider any $\eta \in (0, \Phi(\theta) - \Phi(\theta - \varepsilon))$. Since $\lim_{n \rightarrow \infty} \Phi(2\rho_n + \theta + \varepsilon) = 1$, there exists $N_0 \in \mathbb{N}$ such that, for all $n \geq N_0$:

$$\Phi(2\rho_n + \theta + \varepsilon) > 1 - \eta. \quad (34)$$

Similarly, since $\lim_{n \rightarrow \infty} \Phi(2\rho_n + \theta - \varepsilon) = 1$, there exists $N_1 \in \mathbb{N}$ such that, for all $n \geq N_1$:

$$\Phi(2\rho_n + \theta - \varepsilon) < 1 + \eta. \quad (35)$$

Let n be any integer above $\max(N_0, N_1)$. If $g_\gamma(\rho_n) < \theta - \varepsilon$, we then have $\Phi(g_\gamma(\rho_n)) < \Phi(\theta - \varepsilon)$ and $\Phi(2\rho_n + g_\gamma(\rho_n)) < \Phi(2\rho_n + \theta - \varepsilon)$. Eqs. (33) and (35) then imply that:

$$1 + \Phi(\theta) < \Phi(\theta - \varepsilon) + \Phi(2\rho_n + \theta - \varepsilon) < \Phi(\theta - \varepsilon) + 1 + \eta,$$

which is impossible because of our choice for η . Therefore, we cannot have $g_\gamma(\rho_n) < \theta - \varepsilon$. We cannot have $g_\gamma(\rho_n) > \theta + \varepsilon$ either because, via (33) and (34), this inequality implies:

$$1 + \Phi(\theta) > \Phi(\theta + \varepsilon) + \Phi(2\rho_n + \theta + \varepsilon) > \Phi(\theta + \varepsilon) + 1 - \eta, \quad (36)$$

which is contradictory to our choice for η . ■

Lemma B.2 (Behavior of $Q_{\frac{1}{2}}$ in vanishing noise): Consider $\tau \in [0, \infty)$ and $\rho \in (0, \infty)$ such that $\rho \neq \tau$.

$$\forall \gamma \in (0, 1), \lim_{\sigma \rightarrow 0} Q_{\frac{1}{2}}(\rho/\sigma, \lambda_\gamma(\tau/\sigma)) = \mathbb{1}_{(\tau, \infty)}(\rho).$$

PROOF: Let $(\sigma_n)_{n \in \mathbb{N}}$ be a sequence of positive real values such that $\lim_{n \rightarrow \infty} \sigma_n = 0$ and set $\rho_n = \tau/\sigma_n$ for each $n \in \mathbb{N}$. According to (2), $Q_{\frac{1}{2}}((\rho/\tau)\rho_n, \lambda_\gamma(\rho_n)) = \mathbb{P}[|(\rho/\tau) + X/\rho_n| > \lambda_\gamma(\rho_n)/\rho_n]$ for any $X \sim \mathcal{N}(0, 1)$. It follows from Lemma B.1 (ii) that $|(\rho/\tau) + (X/\rho_n)| - \lambda_\gamma(\rho_n)/\rho_n = (\rho/\tau) - 1$ (a.s). Therefore, the cdf of $|(\rho/\tau) + (X/\rho_n)| - \lambda_\gamma(\rho_n)/\rho_n$ converges weakly to $\mathbb{1}_{[(\rho/\tau)-1, \infty)}$. Since $\rho \neq \tau$, this weak convergence implies that $\lim_{n \rightarrow \infty} \mathbb{P}[|(\rho/\tau) + X/\rho_n| > \lambda_\gamma(\rho_n)/\rho_n] = \mathbb{1}_{(\tau, \infty)}(\rho)$. Thence the result since $(\sigma_n)_{n \in \mathbb{N}}$ is arbitrary. ■

Lemma B.3 (Non-Asymptotic behavior of $Q_{\frac{1}{2}}$): Consider $\tau \in [0, \infty)$, $\rho \in (0, \infty)$ and $\gamma \in (0, 1)$, the map:

$$\sigma \in [0, \infty) \mapsto Q_{\frac{1}{2}}(\rho\sigma, \lambda_\gamma(\tau\sigma)) \text{ is } \begin{cases} \text{constant equal to } \gamma \text{ for } \rho = \tau \\ \text{decreasing for } \rho < \tau \\ \text{increasing for } \rho > \tau \end{cases}$$

PROOF: Given ρ and τ , we want to study the behavior of

$$\mathcal{Q}(\sigma) = Q_{\frac{1}{2}}(\rho\sigma, \lambda_\gamma(\tau\sigma)) = 1 - \Phi(r_-(\sigma)) + \Phi(-r_+(\sigma)) \quad (37)$$

with $r_+ = \lambda_\gamma(\tau\sigma) + \rho\sigma$ and $r_- = \lambda_\gamma(\tau\sigma) - \rho\sigma$. For $\rho = \tau$, it follows from (4) that \mathcal{Q} is constant equal to γ . We thus have $1 - \Phi(\lambda_\gamma(\tau\sigma) - \rho\sigma) + \Phi(-\lambda_\gamma(\tau\sigma) - \rho\sigma) = \gamma$. After differentiating the two members of the equality above and after some routine algebra, we obtain:

$$\lambda'_\gamma(\tau\sigma) = (1 - e^{-2\tau\sigma\lambda_\gamma(\tau\sigma)}) / (1 + e^{-2\tau\sigma\lambda_\gamma(\tau\sigma)}) \quad (38)$$

where λ'_γ is the first derivative of λ_γ . We now differentiate \mathcal{Q} defined by (37). Some easy computation yields:

$$\mathcal{Q}'(\sigma) = \frac{1}{\sqrt{2\pi}} \left(e^{-r_-^2(\sigma)/2} - e^{-r_+^2(\sigma)/2} \right) \left(\rho - \tau \lambda'_\gamma(\tau\sigma) \frac{1 + e^{-2\rho\sigma\lambda_\gamma(\tau\sigma)}}{1 - e^{-2\rho\sigma\lambda_\gamma(\tau\sigma)}} \right)$$

By injecting (38) into the equality above, we obtain:

$$\mathcal{Q}'(\sigma) = \frac{\tau}{\sqrt{2\pi}} \left(e^{-r_-^2(\sigma)/2} - e^{-r_+^2(\sigma)/2} \right) \left(\frac{\rho}{\tau} - \frac{\Delta_-(\rho, \tau)}{\Delta_+(\rho, \tau)} \right) \quad (39)$$

with $\Delta_\varepsilon(\rho, \tau) = (1 + \varepsilon e^{-2\tau\sigma\lambda_\gamma(\tau\sigma)}) / (1 + \varepsilon e^{-2\rho\sigma\lambda_\gamma(\tau\sigma)})$ and $\varepsilon \in \{-1, +1\}$. For all $\sigma \geq 0$, the sign of \mathcal{Q}' is therefore that of $(\rho/\tau) - (\Delta_-(\rho, \tau)/\Delta_+(\rho, \tau))$. We verify easily that:

$$\begin{cases} \rho < \tau \Leftrightarrow \Delta_-(\rho, \tau) > 1 \Leftrightarrow \Delta_+(\rho, \tau) < 1 \\ \rho = \tau \Leftrightarrow \Delta_-(\rho, \tau) = \Delta_+(\rho, \tau) = 1 \end{cases}$$

Therefore, if $\rho < \tau$, $\rho/\tau < 1 < \Delta_-(\rho, \tau)/\Delta_+(\rho, \tau)$, which implies that $\mathcal{Q}'(\sigma) \leq 0$ and, thus, that \mathcal{Q} is decreasing. On the other hand, if $\rho > \tau$, we have $\rho/\tau > 1 > \Delta_-(\rho, \tau)/\Delta_+(\rho, \tau)$, so that \mathcal{Q} is increasing in this case. ■

Lemma B.4: Given $\rho \in (0, \infty)$, the map $\gamma \in (0, 1) \mapsto \lambda_\gamma(\rho)$ is decreasing.

PROOF: It is a straightforward consequence of (4) and the decreasing nature of $Q_{\frac{1}{2}}$ with its second argument given in Lemma 1.1. ■

Lemma B.5:

(P1) For any $\tau \in (0, \infty)$ and any $\eta \in (\tau, \infty)$, the map $\sigma \in (0, \infty) \mapsto Q_{\frac{1}{2}}(\tau/\sigma, \eta/\sigma)$ is increasing.

(P2) The map $\rho \in (0, \infty) \mapsto Q_{\frac{1}{2}}(\rho, \rho)$ is decreasing, lower-bounded by 1/2 and $\lim_{\rho \rightarrow \infty} Q_{\frac{1}{2}}(\rho, \rho) = 1/2$

(P3) For any $\gamma \in (0, 1/2)$, the map $\rho \in (0, \infty) \mapsto \lambda_\gamma(\rho)/\rho$ is decreasing, lower bounded by 1.

PROOF: *Proof of statement (P1):* Using (3), define $\mathcal{Q}(\sigma)$ as:

$$\mathcal{Q}(\sigma) = Q_{\frac{1}{2}}(\tau/\sigma, \eta/\sigma) = 1 - \Phi(\eta/\sigma - \tau/\sigma) + \Phi(-\eta/\sigma - \tau/\sigma).$$

We now differentiate \mathcal{Q} and some easy computation yields:

$$\mathcal{Q}'(\sigma) = (1/\sqrt{2\pi}\sigma^2) e^{-\frac{(\eta-\tau)^2}{2\sigma^2}} \left[(\eta - \tau) + (\eta + \tau) e^{-\frac{2\eta\tau}{\sigma^2}} \right].$$

Thence the result, since $\eta \in (\tau, \infty)$ implies that $\mathcal{Q}'(\sigma) > 0$.

Proof of statement (P2): The map $\rho \in (0, \infty) \mapsto Q_{\frac{1}{2}}(\rho, \rho)$ is decreasing as a consequence of (P1). Given $\rho \in (0, \infty)$, $Q_{\frac{1}{2}}(\rho, \rho) = 1/2 + \Phi(-2\rho)$ from (3). Whence the result. *Proof of statement (P3):* Let ρ and ρ' be two positive real numbers such that $0 < \rho < \rho'$. According to (4), we have:

$$Q_{\frac{1}{2}}(\rho, \lambda_\gamma(\rho)) = Q_{\frac{1}{2}}(\rho', \lambda_\gamma(\rho')) = \gamma. \quad (40)$$

Since $\gamma < 1/2$ so that $1/2 < 1 - \gamma$, it follows from (P2) and (40) that: $Q_{\frac{1}{2}}(\rho, \rho) > 1/2 > Q_{\frac{1}{2}}(\rho, \lambda_\gamma(\rho))$. The decreasing behavior of $Q_{\frac{1}{2}}$ with its second argument implies that $\lambda_\gamma(\rho) > \rho$, so that $\lambda_\gamma(\rho)/\rho$ is lower bounded by 1. We then derive from (P1) that $x \in (0, \infty) \mapsto Q_{\frac{1}{2}}(\rho/x, \lambda_\gamma(\rho)/x)$ is an increasing map. Since $\rho/\rho' < 1$, we thus have $Q_{\frac{1}{2}}(\rho, \lambda_\gamma(\rho)) > Q_{\frac{1}{2}}(\rho', \rho' \lambda_\gamma(\rho)/\rho)$. This inequality and (40) induce that $Q_{\frac{1}{2}}(\rho', \lambda_\gamma(\rho')) > Q_{\frac{1}{2}}(\rho', \rho' \lambda_\gamma(\rho)/\rho)$. The decreasing nature of $Q_{\frac{1}{2}}(\rho', \cdot)$ then implies that $\lambda_\gamma(\rho') < \rho' \lambda_\gamma(\rho)/\rho$. Thereby, $\rho \in (0, \infty) \mapsto \lambda_\gamma(\rho)/\rho$ is decreasing in ρ . Since the map $\rho \in (0, \infty) \mapsto \lambda_\gamma(\rho)/\rho$ is decreasing and lower bounded by 1, this map has a limit $\ell \geq 1$ when ρ tends to ∞ . The result then follows as a consequence of Lemma B.1 (ii). ■

Lemma B.6: For $\gamma \in (1/2, 1)$ and ρ large enough, the map $\rho \in (0, \infty) \mapsto \lambda_\gamma(\rho)/\rho$ is increasing, upper bounded by 1.

PROOF: According to statement (i) of Lemma B.1, $\lambda_\gamma(\rho) - \rho = \Phi^{-1}(1 - \gamma) + \kappa(\rho)$ where κ is such that $\lim_{\rho \rightarrow \infty} \kappa(\rho) = 0$. Since $\gamma > 1/2$, $\Phi^{-1}(1 - \gamma) < 0$. Given η such that $0 < \eta < -\Phi^{-1}(1 - \gamma)$, there exists ρ_0 such that, for all $\rho \geq \rho_0$, $\kappa(\rho) \leq \eta$. Therefore, for all $\rho \geq \rho_0$, $\lambda_\gamma(\rho) - \rho \leq \Phi^{-1}(1 - \gamma) + \eta < 0$. We have hence proved that $\lambda_\gamma(\rho) < \rho$ for ρ large enough.

With $h_\gamma(\rho) = \lambda_\gamma(\rho)/\rho$, $\Phi(\rho(h_\gamma(\rho) - 1)) - \Phi(-\rho(h_\gamma(\rho) + 1)) = 1 - \gamma$. By differentiation of this equality with respect to ρ and since h_γ is differentiable via the implicit function theorem, we find that $h'_\gamma(\rho)$ has the same sign as $\Upsilon(\rho) = (1 - h_\gamma(\rho)) \left(e^{2\rho\lambda_\gamma(\rho)} + \frac{\lambda_\gamma(\rho) + \rho}{\lambda_\gamma(\rho) - \rho} \right)$. For ρ large enough, $h_\gamma(\rho) < 1$ by the first part of the proof and Lemma B.1 implies that $\lim_{\rho \rightarrow \infty} \Upsilon(\rho) = \infty$. Therefore, $\Upsilon(\rho) > 0$ for ρ large enough and the proof is complete. ■

REFERENCES

- [1] P. Khanduri, D. Pastor, V. Sharma, and P. K. Varshney, "On random distortion testing based sequential non-parametric hypothesis testing," in *2018 56th Annual Allerton Conference on Communication, Control, and Computing (Allerton)*, Oct 2018.
- [2] P. Bühlmann and S. Geer, "Statistics for big data: A perspective," *Statistics & Probability Letters*, vol. 136, pp. 37 – 41, 2018.
- [3] V. Chandrasekaran and M. I. Jordan, "Computational and statistical tradeoffs via convex relaxation," *Proceedings of the National Academy of Sciences*, vol. 110, no. 13, pp. E1181–E1190, 2013.
- [4] P. Bühlmann and N. Meinshausen, "Magging: Maximin aggregation for inhomogeneous large-scale data," *Proceedings of the IEEE*, vol. 104, no. 1, pp. 126–135, Jan 2016.
- [5] B. K. Ghosh and P. K. Sen, *Handbook of Sequential Analysis*, Statistics: A Series of Textbooks and Monographs. Taylor & Francis, 1991.
- [6] E. L. Lehmann and J. P. Romano, *Testing Statistical Hypotheses*, 3rd edition, Springer, 2005.
- [7] D. J. Wheeler and D. S. Chambers, *Understanding Statistical Process Control*, SPC Press, 2010.
- [8] J. Whitehead, *The Design and Analysis of Sequential Clinical Trials*, Statistics in Practice. Wiley, 1997.
- [9] Thomas S. Ferguson, "Who solved the secretary problem?," *Statist. Sci.*, vol. 4, no. 3, pp. 282–289, 08 1989.
- [10] E. Brodsky and B. S. Darkhovsky, *Nonparametric Methods in Change Point Problems*, Mathematics and Its Applications. Springer Netherlands, 1993.
- [11] A. Wald, "Sequential tests of statistical hypotheses," *The Annals of Mathematical Statistics*, vol. 16, no. 2, pp. 117–186, 06 1945.
- [12] A. Wald, *Sequential Analysis*, John Wiley and Sons, New York, 1948.
- [13] R. Bechhofer, "A note on the limiting relative efficiency of the wald sequential probability ratio test," *Journal of the American Statistical Association*, vol. 55, no. 292, pp. 660–663, 1960.
- [14] S. Tantarana and J. B. Thomas, "Relative efficiency of the sequential probability ratio test in signal detection," *IEEE Transactions on Information Theory*, vol. 24, no. 1, pp. 22–31, January 1978.
- [15] S. Tantarana and J. B. Thomas, "Truncated sequential probability ratio test," *Information Sciences*, vol. 13, no. 3, pp. 283 – 300, 1977.
- [16] A. G. Tartakovsky, M. Basseville, and I. V. Nikiforov, *Sequential Analysis: Hypothesis Testing and Changepoint Detection*, Chapman & Hall/CRC, 2014.
- [17] P. K. Sen, *Sequential Nonparametrics: Invariance Principles and Statistical Inference*, Wiley Series in Probability and Statistics - Applied Probability and Statistics Section. Wiley, 1981.
- [18] P. K. Sen, Conference Board of the Mathematical Sciences, Society for Industrial, and Applied Mathematics, *Theory and Applications of Sequential Nonparametrics*, CBMS-NSF Regional Conference Series in Applied Mathematics. Society for Industrial and Applied Mathematics, 1985.
- [19] E. L. Lehmann and J. P. Romano, *Testing statistical hypotheses*, Springer, 3rd edition, 2005.
- [20] T. L. Lai, "Sequential analysis: Some classical problems and new challenges," *Statistica Sinica*, vol. 11, no. 2, pp. 303–351, 2001.
- [21] X. Li, J. Liu, and Z. Ying, "Generalized sequential probability ratio test for separate families of hypotheses," *Sequential Analysis*, vol. 33, no. 4, pp. 539–563, 2014.
- [22] A. G. Tartakovsky, "Nearly optimal sequential tests of composite hypotheses revisited," *Proceedings of the Steklov Institute of Mathematics*, vol. 287, no. 1, pp. 268–288, Dec 2014.
- [23] T. L. Lai, "Nearly optimal sequential tests of composite hypotheses," *Ann. Statist.*, vol. 16, no. 2, pp. 856–886, 06 1988.
- [24] T. L. Lai and L. Zhang, "A modification of schwarz's sequential likelihood ratio tests in multivariate sequential analysis," *Sequential Analysis*, vol. 13, no. 2, pp. 79–96, 1994.
- [25] T. L. Lai, "Likelihood ratio identities and their applications to sequential analysis," *Sequential Analysis*, vol. 23, no. 4, pp. 467–497, 2004.
- [26] G. Fellouris and A. G. Tartakovsky, "Almost optimal sequential tests of discrete composite hypotheses," *Statistica Sinica*, vol. 23, no. 4, pp. 1717–1741, 2013.
- [27] D. Pastor and Q-T. Nguyen, "Random distortion testing and optimality of thresholding tests," *IEEE Transactions on Signal Processing*, vol. 61, no. 16, pp. 4161 – 4171, Aug. 2013.
- [28] D. Pastor and Q-T. Nguyen, "Robust statistical process control in Block-RDT framework," in *2015 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, April 2015, pp. 3896–3900.
- [29] P. Khanduri, D. Pastor, V. Sharma, and P. K. Varshney, "Sequential random distortion testing of non-stationary processes," *Manuscript submitted for publication*, 2018.
- [30] P. Khanduri, D. Pastor, V. Sharma, and P. K. Varshney, "On sequential random distortion testing of non-stationary processes," in *2018 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, April 2018.
- [31] Y. Sun, A. Baricz, and S. Zhou, "On the monotonicity, log-concavity, and tight bounds of the generalized marcum and nuttall Q-functions," *IEEE Transactions on Information Theory*, vol. 56, no. 3, pp. 1166 – 1186, Mar. 2010.
- [32] D. Pastor, R. Gay, and A. Gronenboom, "A sharp upper bound for the probability of error of likelihood ratio test for detecting signals in white gaussian noise," *IEEE Transactions on Information Theory*, vol. 48, no. 1, pp. 228 – 238, Jan. 2002.
- [33] Q-T. Nguyen, D. Pastor, and E. L'Her, "Automatic detection of autopeep during controlled mechanical ventilation," *BioMedical Engineering OnLine*, vol. 11, no. 1, pp. 32, 2012.
- [34] D. Pastor and F-X. Socheleau, "Random distortion testing with linear measurements," *Signal Processing*, vol. 145, pp. 116 – 126, 2018.
- [35] G. Samorodnitsky and M. S. Taqqu, *Stable Non-Gaussian Random Processes: Stochastic Models with Infinite Variance*, Stochastic Modeling Series. Taylor & Francis, 1994.