Sampling Schemes for Sequential Detection With Dependent Observations

Ruixin Niu, *Member, IEEE*, and Pramod K. Varshney, *Fellow, IEEE*

***Abstract*—Several sampling schemes and their corresponding sequential detection procedures in autoregressive noise are presented in this paper. Two of them use uniform sampling procedures with high and low sampling rates, respectively. The other two employ groups of samples, which are separated by long intergroup delays such that the intergroup correlations are negligible. One of the group-sampling schemes also employs optimal signaling waveforms to further improve its energy-efficiency. In all the schemes, data sampling and transformation are designed in such a way that Wald's sequential probability ratio test (SPRT) can still be implemented. The performances of different schemes, in terms of average termination time (ATT), are derived analytically. When all the schemes employ the same sampling interval and under a constant signal amplitude constraint, their performances are compared through analytical and numerical methods. In addition, under a constant power constraint, their ATTs and energy-efficiency are compared. It is theoretically proved that the scheme using groups of samples with the optimal signaling waveform is the most energy-efficient.**

***Index Terms*—Autoregressive noise, colored noise, sampling, sequential detection.**

I. Introduction

SINGLE sample detection may work well when the signal-to-noise ratio (SNR) is relatively high. However, in many practical cases, such as the detection of an aircraft at a very long range based on radar signal returns, the SNR is so low that a decision can not be made with high reliability based on a single sample. In such cases, detection is typically performed based on multiple observation samples to improve the detection performance. A usual multiple-sample procedure is fixed-sample-size (FSS) detection. In a FSS detection system, a predefined number of samples are collected and the detection is based on these samples. The sample size is determined by the performance level to be achieved. Another powerful procedure for the multiple-sample detection problem is Wald's sequential probability ratio test (SPRT) [1]. Contrary to FSS procedures, its sample size is a variable. It is well known that to get a required detection performance, SPRT on the average needs much fewer samples than FSS procedures. As a result, the sequential detection procedure has drawn continued interest for decades.

Manuscript received May 16, 2009; accepted October 08, 2009. First published November 20, 2009; current version published February 10, 2010. The associate editor coordinating the review of this manuscript and approving it for publication was Dr. Marcelo G. S. Bruno. This work was presented in part at ICASSP'04, Montreal, Canada, May 2004.

The authors are with the Department of Electrical Engineering and Computer Science, Syracuse University, Syracuse, NY 13244 USA (e-mail: rniu@ecs.syr.edu; varshney@ecs.syr.edu).

Digital Object Identifier 10.1109/TSP.2009.2037058

In most sequential detection literature, authors have made the assumption of independent and identically distributed (i.i.d.) noise statistics. In practice, this assumption does not always hold. For example, for weak signal detection in underwater acoustics, due to high-rate sampling and temporally and spatially dependent noise and interference, the correlation between adjacent data samples can not be ignored [2]. Hence, the sequential detection problem in colored noise is a very important topic to be studied. Its solution can find many applications, such as target detection in radar/sonar systems and signal detection in communication systems. In [3], the operating characteristic (OC) and the average sample number (ASN) were derived for a SPRT working with dependent observations, which were assumed to form a finite Markov chain. There already exist several publications on nonparametric SPRTs with dependent data. In [4], authors have proposed nonparametric SPRT methods for Markov-dependent input data. A generalized sequential sign detector was studied in [5].

The optimal parametric SPRT with dependent data remains a challenging problem. In [6] and [7], authors have shown that the optimum sequential detector for dependent samples is in the form of a generalized sequential probability ratio test (GSPRT), in which the thresholds are time-varying as opposed to the constant thresholds employed in the SPRT for the i.i.d. noise case. However, the problem of determination of the time-varying thresholds is still not solved. In [8], the SPRT with constant thresholds and the sequential linear detector (SLD) are compared in an autoregressive noise in terms of the ASN. An interesting result is that when the correlation coefficient is positive, the SLD has better performance than the SPRT. This indicates that the SPRT with constant thresholds is not optimal in colored noise. Recently in [9], authors show that under certain regularity conditions, in a multichannel system, a simple generalized SPRT with constant thresholds, which is different from the GSPRT with time-varying thresholds in [6] and [7], is asymptotically optimal for general non-i.i.d. data, when the probabilities of false alarms and missed detections are low.

Our main contribution in this paper is to tackle the challenging problem of sequential detection with correlated data from a system design point of view. To solve the problem of or even take advantage of the correlation between samples, we propose several schemes with different sampling patterns and signal waveforms. To the best of our knowledge, little previous work on sampling design for sequential detection in colored noise has been documented in the literature. Two of the proposed sampling schemes use uniform sampling procedures with high and low sampling rates, respectively. The other two are

based on the group sampling idea, where groups of consecutive samples are transformed to single variables. One of the group sampling schemes adopts the optimal signal waveform to improve its efficiency. In all the sampling schemes proposed in this paper, the signals are sampled and processed in such a way that the resulting samples can be used by the standard framework of a SPRT. We will analyze and compare the performances of the different sampling schemes in terms of the average termination times (ATTs), under the constraints of a constant amplitude and a constant power, respectively. In the process, many theoretical results will be developed. We will show later that the scheme using group sampling with an optimal signal is much more efficient than other schemes, especially when the sampling intervals are small (or the correlation between adjacent samples is high).

The group sampling idea is similar to the one that appeared in [10] and [11], where group sampling has been employed to perform nonparametric FSS detection with dependent data. Other related work includes [12] and [13]. In [12], a memoryless grouped-data sequential (MLGDS) procedure has been proposed. In MLGDS, at each stage, a group of samples is taken to perform a two-threshold test. If either one of the thresholds is crossed, a decision is made in favor of the corresponding hypothesis; otherwise, the group of samples is discarded and the test continues to the next stage. In [13], a multisensor decentralized sequential test utilizing the MLGDS has been proposed. Clearly, by discarding samples from previous stages, the MLGDS entails performance loss. Two of the sequential test procedures proposed in this paper are also based on group sampling, but they are different from the MLGDS, since no samples are discarded in these procedures. As shown later in the paper, the proposed sequential detection approaches based on group sampling require much less ASNs and incur much less energy consumption to achieve the same prespecified detection performance than that based on uniform sampling. The use of the optimal signaling waveform, which maximizes the SNR at the matched filter output under a fixed energy constraint, in conjunction with group sampling can further improve energy-efficiency. Another advantage of group sampling schemes lies in the fact that they could be very flexible and achieve a desired tradeoff between ATT and energy efficiency. The proposed sequential detection approaches are very general and could find applications in communications, biological signal processing and radar/sonar systems.

Note that sampling designs for FSS detection problems have been investigated in [14]–[16], where nonuniform sampling designs or even random sampling schemes [14] have been proposed to improve detection performance. For simplicity, we focus only on the uniform sampling schemes where sample points are either equally spaced over the entire observation interval (in the first two sampling schemes) or equally spaced within sample groups (in the last two sampling schemes).

In Section II, some background about the SPRT including notations and the system model are introduced. In Section III, four sampling schemes along with their associated sequential tests are proposed, and their corresponding average termination times (ATTs) are derived. Under the condition of identical signal amplitude and identical sampling interval, the ATTs for the four procedures are compared in Section IV. Under the constraint of constant signal power, an energy efficiency comparison is made in Section V. Finally, the work is summarized in Section VI.

II. SPRT and System Model

A. Background on Wald's Sequential Probability Ratio Test

In this paper we will investigate different sampling schemes, which are designed by taking into account the correlations between the samples, for sequential detection. Before we describe different sampling schemes, we briefly introduce the standard SPRT here.

Consider the problem of testing a simple hypothesis H_0 versus a simple alternative H_1. The successive observations are denoted by $y_1, y_2, \ldots, y_n, (n \geq 1)$ and they are assumed to be statistically independent and identically distributed (i.i.d.) for the given hypothesis. Wald's SPRT for testing H_1 against H_0 can be described as follows: at each stage n $(n \geq 1)$ of the test, the sum $\sum_{i=1}^{n} \lambda_i$ is computed, and the following test is performed

$$\sum_{i=1}^{n} \lambda_i \begin{cases} \geq a & \text{stop and decide } \mathrm{H}_1 \\ \leq b & \text{stop and decide } \mathrm{H}_0 \\ \text{otherwise} & \text{continue} \end{cases} \tag{1}$$

where

$$\lambda_i \triangleq \ln \frac{f(y_i|\mathrm{H}_1, \theta_1)}{f(y_i|\mathrm{H}_0, \theta_0)} \tag{2}$$

is the log-likelihood ratio of the ith sample y_i, and θ_0 and θ_1 denote the parameters that characterize the likelihood functions under H_0 and H_1, respectively. The choice of the test thresholds a and b depends on the desired values of the error probabilities α and β, where α is the probability of deciding H_1 when H_0 is true, and β is the probability of deciding H_0 when H_1 is the true hypothesis. It has been shown [1] that the thresholds can be approximated as

$$\begin{aligned} a &\approx \ln \frac{1-\beta}{\alpha} \\ b &\approx \ln \frac{\beta}{1-\alpha}. \end{aligned} \tag{3}$$

The performance of an SPRT is characterized in terms of the OC and ASN functions [1]. Ignoring excesses over the test thresholds, the ASN required for decision making is

$$\mathrm{ASN} \approx \frac{L(\theta)b + [1 - L(\theta)]a}{E_\theta[\lambda]} \tag{4}$$

where $L(\theta)$ is the OC function, defined as the probability that the sequential test will terminate with the acceptance of H_0 when θ is the true value of the parameter [1]. Clearly by definition, $L(\theta_0) = 1 - \alpha$ and $L(\theta_1) = \beta$. $E_\theta(\lambda)$ denotes the expected value of the log-likelihood ratio λ when θ is the true value of the parameter. Note that a more accurate estimate for the ASN can be obtained by taking into account the excesses of the likelihood ratio over the thresholds [17]. We will show in Appendix I that (4) is a good approximation when the data sample's SNR is low, a typical scenario for a SPRT to be used.

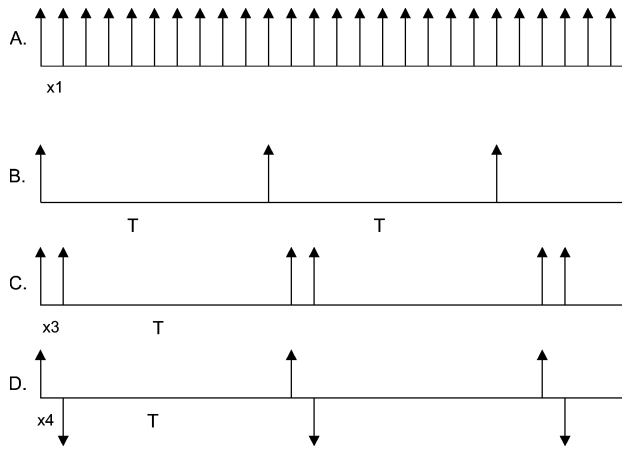


Fig. 1. Four sequential detection sampling schemes in colored noise.

In this paper, we assume that the SNR is very low and (4) is accurate for different sampling schemes.

B. System Model

In this paper, we consider a binary hypothesis testing problem

$$\begin{aligned} H_1 &: y(t) = s(t) + \omega(t) \\ H_0 &: y(t) = \omega(t) \end{aligned} \tag{5}$$

where $y(t)$ is the received noisy observation, $s(t)$ is the transmitted signal, and the noise $\omega(t)$ is a Gaussian random process with zero mean and covariance function $R_\omega(t,\tau)$. We further assume that the Gaussian random process is wide-sense stationary and its autocorrelation function is

$$R_\omega(\tau) = \sigma^2 e^{-|\tau|/\mu} \tag{6}$$

where μ indicates the rate at which the correlation decays and is related to the bandwidth of the noise process $\omega(t)$.

To convert the continuous-time detection problem into a discrete-time one, the signal $y(t)$ is sampled according to a certain sampling procedure. For example, one can uniformly sample the observation $y(t)$, as illustrated in Fig. 1(A). After sampling, the discrete-time hypothesis testing problem can be cast as

$$\begin{aligned} H_1 &: \mathbf{y} = \boldsymbol{s} + \boldsymbol{\omega} \\ H_0 &: \mathbf{y} = \boldsymbol{\omega} \end{aligned} \tag{7}$$

where $\mathbf{y} = [y(x_1)\cdots y(nx_1)]'$, $\mathbf{s} = [s(x_1)\cdots s(nx_1)]'$, $\boldsymbol{\omega} = [\omega(x_1)\cdots\omega(nx_1)]'$, x_1 is the sampling interval, and n is the number of samples in the observation interval. In this paper, sampling schemes and signal design will be discussed and their effects on the performance of the SPRT will be investigated.

From [18], we know that because $\omega(t)$ is a wide-sense stationary Gaussian process, $\boldsymbol{\omega}$ is also a wide-sense stationary Gaussian sequence with autocorrelation

$$R(m) = R_\omega(mx) = \sigma^2 \rho^{|m|} \tag{8}$$

where $\rho \triangleq e^{-x_1/\mu}$ is the correlation coefficient between adjacent samples. Without loss of generality, we assume $\mu = 1$ for notational simplicity. With this assumption, the time interval x_1 could be deemed as being normalized with respect to μ.

If $y(t)$ is sampled at a high rate, or x_1 is small, the correlation between samples can not be ignored. The traditional SPRT, which requires i.i.d. data samples, can not be used directly here. In this case, the optimum detector is in the form of a generalized sequential probability ratio test (GSPRT) [6], [7]

$$\Lambda_n(y_1,\ldots,y_n) \begin{cases} \geq A_n & \text{stop and decide } \mathrm{H}_1 \\ \leq B_n & \text{stop and decide } \mathrm{H}_0 \\ \text{otherwise} & \text{continue} \end{cases} \tag{9}$$

where

$$\Lambda_n(y_1,\ldots,y_n) \triangleq \frac{f_\omega(y_1-\theta, y_2-\theta, \ldots, y_n-\theta)}{f_\omega(y_1, y_2, \ldots, y_n)} \tag{10}$$

is the likelihood ratio, and A_n and B_n are the thresholds that are functions of n. However, the determination of A_n and B_n is still an open problem. In addition, with each new data sample, the GSPRT needs to recalculate the likelihood ratio involving all the previous samples as defined in (9) and (10) contrary to the simple addition operation in a SPRT as illustrated in (1). Hence, signal designs and sampling schemes that generate near-independent data samples are preferred so that the elegant framework of the SPRT can still be implemented.

In this paper, we propose four different signal design and sampling schemes, which are illustrated in Fig. 1. In Scheme A, the signal is transmitted and sampled at a relatively high rate, and data samples are processed by a whitening filter whose output data samples are independent. In Scheme B, the signal is transmitted and sampled at such a low rate that the correlation between adjacent samples is negligible. In both Schemes C and D, adjacent signal sample groups are separated by a large inter-group delay (T), so that the inter-group correlation can be ignored. Therefore, in these schemes the super samples, generated by combining samples within a group, are deemed as independent. Note that Scheme D adopts optimal signal waveform to improve its energy-efficiency, as opposed to the signal with constant amplitude used in other schemes. In summary, all the four sampling schemes generate independent or near-independent samples or super samples, which enable a straightforward application of the SPRT for a sequential hypothesis test.

As shown in Fig. 1, x_3 and x_4 are the intervals between adjacent samples within a group for Schemes C and D, respectively. In this paper, we assume that Schemes A, C, and D use the same sampling interval, namely $x_1 = x_3 = x_4 = x$, in order to obtain theoretical results. However, numerical results for unequal x_1, x_3, and x_4 are obtained and are available in Figs. 5 and 8. Further, to facilitate a comparison between them under a constant power constraint, these schemes may take different signal amplitudes, which are denoted by $\theta_1,\ldots,\theta_4$ for Schemes A through D, respectively.

III. Sampling and Sequential Detection Schemes

In this section, four sampling schemes and their corresponding sequential detection algorithms are introduced.

A. Uniform Sampling at High-Rate

The most common sampling procedure is to uniformly sample the received observation $y(t)$. This sampling scheme (Scheme A) is shown in Fig. 1(A). After sampling, the detection problem has been formulated in Section II-B and the noise

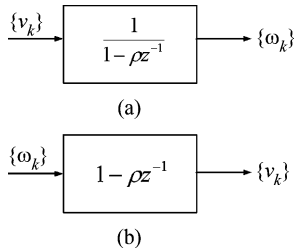


Fig. 2. The autoregressive noise model and the whitening filter for Scheme A.

sequence $\boldsymbol{\omega} = [\omega(x)\ \omega(2x) \cdots \omega(nx)]'$ with an autocorrelation function defined in (8) can be modeled as an autoregressive sequence

$$\omega_i = \nu_i + \rho\omega_{i-1}, \quad i = 1, 2, \ldots \tag{11}$$

where the sequence $\{\nu_i\}$ is i.i.d. and

$$\nu_i \sim \mathcal{N}(0, (1-\rho^2)\sigma^2). \tag{12}$$

In addition, ν_i is uncorrelated with $\{\omega_j\}$ for $j \leq i-1$ and $\omega_0 \sim \mathcal{N}(0,\ \sigma^2)$. Note that due to the assumptions made about the autocorrelation function by (6) and (8), we have $0 < \rho < 1$.

To employ the framework of the SPRT, we use a whitening filter to make the noise sequence $\{\omega_i\}$ i.i.d. The generation of the autoregressive noise and the whitening filter are illustrated in Fig. 2. This is nothing but a first-order finite impulse response (FIR) filtering operation

$$\nu_i = \omega_i - \rho\omega_{i-1}. \tag{13}$$

We define the sample sequence after filtering as

$$z_i \triangleq y_i - \rho y_{i-1}. \tag{14}$$

Therefore, the new sequential detection problem becomes

$$\begin{aligned} H_1 &: z_i = (1-\rho)\theta_1 + \nu_i \\ H_0 &: z_i = \nu_i \end{aligned} \tag{15}$$

where θ_1 denotes the constant signal amplitude for Scheme A. Note that we have assumed the knowledge of the noise autocorrelation function, which is provided in (6). In practice, the autocorrelation function and hence ρ is unknown and needs to be estimated via training data. Study of the impact of the estimation accuracy of ρ on the sequential detection performance is beyond the scope of this paper and could be carried out in future work.

With (12) and $\{\nu_i\}$ being an i.i.d. Gaussian sequence, it is easy to show that the log-likelihood ratio for each sample is

$$\begin{aligned} \lambda_i &= \ln\left[\frac{f_\nu(z_i - (1-\rho)\theta_1)}{f_\nu(z_i)}\right] \\ &= \frac{\theta_1\left[2z_i - (1-\rho)\theta_1\right]}{2(1+\rho)\sigma^2}. \end{aligned} \tag{16}$$

Therefore

$$E[\lambda_i|H_1] = -E[\lambda_i|H_0] = \frac{(1-\rho)\theta_1^2}{2(1+\rho)\sigma^2}. \tag{17}$$

According to Wald's first equation [1], namely (4), the ASNs for this problem are

$$\begin{aligned} E[M_1|H_1] &= \frac{\beta b + (1-\beta)a}{E[\lambda_i|H_1]} \\ E[M_1|H_0] &= \frac{(1-\alpha)b + \alpha a}{E[\lambda_i|H_0]} = -\frac{(1-\alpha)b + \alpha a}{E[\lambda_i|H_1]} \end{aligned} \tag{18}$$

where M_1 denotes the number of samples required by the SPRT to terminate in Scheme A. It is clear that the ASN under either hypothesis is inversely proportional to $E[\lambda_i|H_1]$, up to a certain constant that is determined solely by the type I and type II error probabilities. Therefore, no matter what type I and II error probabilities are, the results and conclusions derived in the paper remain the same. For simplicity and without loss of generality, we assume that the type I and type II error probabilities are the same, namely $\alpha = \beta$. Thus, according to (3), $b = -a$, and the ASNs are the same under both hypotheses H_1 and H_0

$$\begin{aligned} E[M_1|H_1] = E[M_1|H_0] &= \frac{\beta b + (1-\beta)a}{E[\lambda_i|H_1]} \\ &= \frac{2(1-2\beta)a\sigma^2(1+\rho)}{\theta_1^2(1-\rho)}. \end{aligned} \tag{19}$$

With (19), a phenomenon reported in [8], where the SPRT has a better performance for a negative ρ than with a positive ρ, can be explained analytically. As demonstrated in (19), the ASN required by the SPRT is a monotonically increasing function of ρ when $-1 < \rho < 1$. Thus, a positive ρ leads to a larger ASN than a negative ρ. Here in this paper, due to the autocorrelation functions that we have adopted in (6) and (8), we focus only on cases with positive correlation between samples.

To facilitate the comparison between different detection schemes, we have constructed a new metric, the average termination time (ATT), which is the average time needed before either threshold is crossed in a SPRT test. We have defined this new metric due to its flexibility. If time delay is a crucial issue for the system, the ATT is a very natural and fair metric, especially for a system with constraints on sampling rate and signal amplitude. In this case, the ATT measures on the average how fast a decision can be reached by the sequential detector, for a system operating with the highest sampling rate and the largest signal amplitude (equivalently the largest sample energy). On the other hand, if energy-efficiency is a major concern, the ATT is also an appropriate metric when there is a constraint on the average signal power, since in this case the ATT is proportional to the total energy required for the test to stop. Further in Section V-C, the ASNs are derived and compared for different schemes with constraints on sampling rate and signal amplitude, and their relationships with ATTs are investigated.

The ATT for scheme A is

$$E[t_1|H_1] = E[t_1|H_0] = E[M_1|H_1]x = \frac{2(1-2\beta)a\sigma^2}{\theta_1^2} f(x) \tag{20}$$

where

$$f(x) \triangleq \frac{1+e^{-x}}{1-e^{-x}}x = x\coth\left(\frac{x}{2}\right) \tag{21}$$

$\coth(\cdot)$ is the hyperbolic cotangent function, and the identity of $E[M_1|H_1] = E[M_1|H_0]$ in (19) has been used. One property of $f(x)$ is that it is monotonically increasing, which can be proved by checking the first and second derivatives of $f(x)$. As a result, $E[t_1]$ is a monotonically increasing function of x, which means that the higher the sampling rate, on the average, the faster the sequential detection will terminate.

B. Uniform Sampling at Low-Rate

As discussed in [8] and in Section III-A, we know that there is a performance degradation for the SPRT when the noise correlation between adjacent samples is positive. This implies that the positive correlation between samples should be avoided to improve efficiency. This is the motivation behind Scheme B, as shown in Fig. 1(B). In this scheme, the sampling rate is low enough such that the correlation coefficient between adjacent samples is negligible, meaning that,

$$e^{-T} \leq \epsilon \tag{22}$$

where T is the sampling interval and ϵ is a very small constant. Equivalently, we can write

$$T \geq -\ln\epsilon. \tag{23}$$

Data samples collected at this sampling rate can be taken as independent and a standard SPRT can be applied for the hypothesis testing

$$\begin{aligned} H_1 &: y_i = \theta_2 + \omega_i \\ H_0 &: y_i = \omega_i \end{aligned} \tag{24}$$

where θ_2 is the signal amplitude and $\{\omega_i\}$ is assumed to be an i.i.d. sequence with zero mean and variance σ^2. The log-likelihood ratio for each sample is

$$\lambda_i = \ln\left[\frac{f_\omega(y_i-\theta_2)}{f_\omega(y_i)}\right] = \frac{\theta_2(2y_i-\theta_2)}{2\sigma^2}. \tag{25}$$

Therefore

$$E[\lambda_i|H_1] = -E[\lambda_i|H_0] = \frac{\theta_2^2}{2\sigma^2}. \tag{26}$$

The ASNs for this case are

$$E[M_2|H_1] = E[M_2|H_0] = \frac{\beta b + (1-\beta)a}{E[\lambda_i|H_1]} = \frac{2(1-2\beta)a\sigma^2}{\theta_2^2} \tag{27}$$

where M_2 denotes the number of samples required by the SPRT to terminate in Scheme B. The ATT is, therefore

$$E[t_2|H_1] = E[t_2|H_0] = E[M_2|H_1]T = \frac{2(1-2\beta)a\sigma^2 T}{\theta_2^2}. \tag{28}$$

Comparing (20) and (28), it is clear that if $\theta_1 = \theta_2$, as $x \to T$, $E[t_1] \to E[t_2]$, since $e^{-T} \approx 0$. This means that when x is large or equivalently ρ is sufficiently small, the performance of Scheme A converges to that of Scheme B.

C. Group Sampling With Constant Amplitude

The scheme (Scheme C) is illustrated in Fig. 1(C). Within each individual group, uniformly spaced samples are collected. The delay (T) between adjacent groups, identical to that defined in Section III-B, is large enough so that the intergroup correlation is negligible.

The samples within each group are combined to form a super sample and an SPRT is performed for these super samples. Note that in Scheme C, the SPRT incorporates one super sample at each time step, meaning that only an integer number of groups are taken for the test. We denote the amplitude of each sample as θ_3, the sampling interval for samples belonging to the same group as $x_3 = x$, and the number of samples of each group as N. Within the ith group, the hypothesis testing problem is

$$\begin{aligned} H_1 &: \mathbf{y}_i = \boldsymbol{\theta} + \boldsymbol{\omega}_i \\ H_0 &: \mathbf{y}_i = \boldsymbol{\omega}_i \end{aligned} \tag{29}$$

where $\mathbf{y}_i = [y_{i1}\cdots y_{iN}]'$, $\boldsymbol{\theta} = \theta_3[1\cdots 1]'$, and $\boldsymbol{\omega}_i = [\omega_{i1}\cdots\omega_{iN}]'$, which follows a Gaussian distribution:

$$\boldsymbol{\omega}_i \sim \mathcal{N}(\mathbf{0}, \boldsymbol{\Sigma}) \tag{30}$$

where $\mathbf{0} = [0\ 0\ \cdots\ 0]'$

$$\boldsymbol{\Sigma} = \sigma^2 \begin{bmatrix} 1 & \rho & \rho^2 & \cdots & \rho^{N-1} \\ \rho & 1 & \rho & \cdots & \rho^{N-2} \\ \rho^2 & \rho & 1 & \cdots & \rho^{N-3} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \rho^{N-1} & \rho^{N-2} & \rho^{N-3} & \cdots & 1 \end{bmatrix} \tag{31}$$

and $\rho = e^{-x}$.

The log-likelihood ratio of each super sample is

$$\lambda_i = \ln\left\{\frac{\exp\left[-\frac{1}{2}(\mathbf{y}_i-\boldsymbol{\theta})'\boldsymbol{\Sigma}^{-1}(\mathbf{y}_i-\boldsymbol{\theta})\right]}{\exp\left[-\frac{1}{2}\mathbf{y}_i'\boldsymbol{\Sigma}^{-1}\mathbf{y}_i\right]}\right\} = \boldsymbol{\theta}'\boldsymbol{\Sigma}^{-1}\left(\boldsymbol{y}_i - \frac{\boldsymbol{\theta}}{2}\right). \tag{32}$$

Therefore

$$E[\lambda_i|H_1] = -E[\lambda_i|H_0] = \frac{1}{2}\boldsymbol{\theta}'\boldsymbol{\Sigma}^{-1}\boldsymbol{\theta}. \tag{33}$$

The average group number (AGN) and ATT for Scheme C are provided in the following theorem.

Theorem 1: For sampling scheme C with a group size N and interval x between adjacent samples within a group, the AGN is

$$E[M_3|H_1] = E[M_3|H_0] = \frac{2(1-2\beta)a\sigma^2(1+\rho)}{\theta_3^2[N-(N-2)\rho]} \quad (34)$$

where M_3 denotes the number of groups required by the SPRT to terminate in Scheme C, and the ATT is

$$E[t_3|H_1] = E[t_3|H_0] = \frac{2(1-2\beta)a\sigma^2}{\theta_3^2} h(x;N) \quad (35)$$

where

$$h(x;N) \triangleq \frac{1+e^{-x}}{N-(N-2)e^{-x}}[(N-1)x+T]. \quad (36)$$

Proof: According to (33), $\mathbf{\Sigma}^{-1}$ needs to be determined in order to derive the AGN and ATT for Scheme C. An analytical solution for the inverse of a matrix in the form of $\mathbf{\Sigma}$ can be obtained [19]

$$\mathbf{\Sigma}^{-1} = \frac{1}{\sigma^2(1-\rho^2)}\mathbf{F} \quad (37)$$

where

$$\mathbf{F} \triangleq \begin{bmatrix} 1 & -\rho & \cdots & 0 & 0 \\ -\rho & 1+\rho^2 & \ddots & 0 & 0 \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ 0 & 0 & \ddots & 1+\rho^2 & -\rho \\ 0 & 0 & \cdots & -\rho & 1 \end{bmatrix}. \quad (38)$$

Substituting (37) into (33), it follows that

$$\begin{aligned} E[\lambda_i|H_1] &= \frac{\theta_3^2}{2\sigma^2(1-\rho^2)} \sum_{i=1}^{N}\sum_{j=1}^{N} \mathbf{F}(i,j) \\ &= \frac{\theta_3^2}{2\sigma^2(1+\rho)}[N-(N-2)\rho]. \end{aligned} \quad (39)$$

The AGNs for this scheme are, therefore

$$\begin{aligned} E[M_3|H_1] = E[M_3|H_0] &= \frac{\beta b + (1-\beta)a}{E[\lambda_i|H_1]} \\ &= \frac{2(1-2\beta)a\sigma^2(1+\rho)}{\theta_3^2[N-(N-2)\rho]}. \end{aligned} \quad (40)$$

Note that within a sample group, it takes the time $(N-1)x$ to collect N samples and the intergroup delay is T. Also, only an integer number of groups are taken for the test. As a result, the ATT is

$$\begin{aligned} E[t_3|H_1] = E[t_3|H_0] &= E[M_3|H_1][(N-1)x+T] \\ &= \frac{2(1-2\beta)a\sigma^2(1+\rho)}{\theta_3^2[N-(N-2)\rho]}[(N-1)x+T]. \end{aligned} \quad (41)$$

□

TABLE I
SOLUTION OF $\eta = f^{-1}(T)$ FOR VARIOUS ϵ AND T

ϵ	10^{-1}	10^{-2}	10^{-3}	10^{-4}	10^{-5}
T	2.30	4.61	6.91	9.21	11.51
η	1.37	4.50	6.89	9.21	11.51

Note that as x increases, $(1+e^{-x})/[N-(N-2)e^{-x}]$ decreases, but $(N-1)x+T$ increases. Hence, it is an optimization problem to find the best sampling interval x for Scheme C. Actually, under certain sufficient conditions, $h(x;N)$ is convex, and its minimum can be found by solving the equation

$$h'(x;N) = 0.$$

This result has been summarized in the following theorem.

Theorem 2: Provided $2 \le N < \infty$ and $T > 4(N-1)/[e^{-2}(N-2)+N]$, as a function of x, $h(x;N)$ is convex for $x \in (0,+\infty)$. Its minimum lies in $(0,+\infty)$, and can be obtained by solving the following equation:

$$(N-2)e^{-2x} + 2(N-1)xe^{-x} + 2(T-1)e^{-x} - N = 0. \quad (42)$$

Proof: See Appendix II.

Note that the sufficient condition, namely $T > 4(N-1)/[e^{-2}(N-2)+N]$, is easy to satisfy. For example, $T = 4/(1+e^{-2})$ corresponds to a correlation coefficient of 0.0295. For a much smaller ϵ, a much larger T is needed in the sampling schemes.

Also, if the sampling interval x falls in a certain range (actually in most of the interval of $(0,\ T)$ for a large T), $h(x;N)$ is a monotonically decreasing function of N, as provided by the following.

Proposition 1: For $0 < x < \eta$, $h(x;N)$ is monotonically decreasing function of N, where $\eta = f^{-1}(T)$.

The proof of Proposition 1 is straightforward, if we check the first derivative of $h(x;N)$ with respect to N and use the property that $f(x)$ is a monotonically increasing function of x.

Note that for a very small ϵ, or equivalently a very large T, $\eta \approx T$. This is due to the fact that

$$f(T) = \frac{1+e^{-T}}{1-e^{-T}}T \approx T.$$

In Table I, η for different values of T is listed. It is clear that as T increases, $\eta = f^{-1}(T)$ quickly converges to T.

In the following numerical examples, we set $\epsilon = 10^{-3}$, and $T = 6.91$. In Fig. 3, $h(x;N)$ as a function of x is plotted. We can see for all the four cases, namely $N = 2,\ 3,\ 5$ and 16, there exists a minimum for $h(x)$, as stipulated by Theorem 2. It is clear that by choosing an optimal sampling interval, the ATT required to achieve a certain detection performance is significantly reduced. We also observe that for $x < T$, as N increases, $h(x;N)$ decreases, as indicated by Proposition 1. This means that a larger group size N leads to a faster termination of

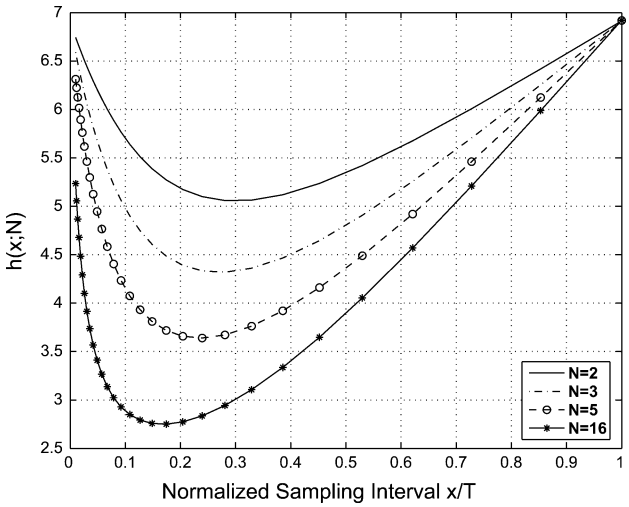


Fig. 3. $h(x;N)$ as a function of x.

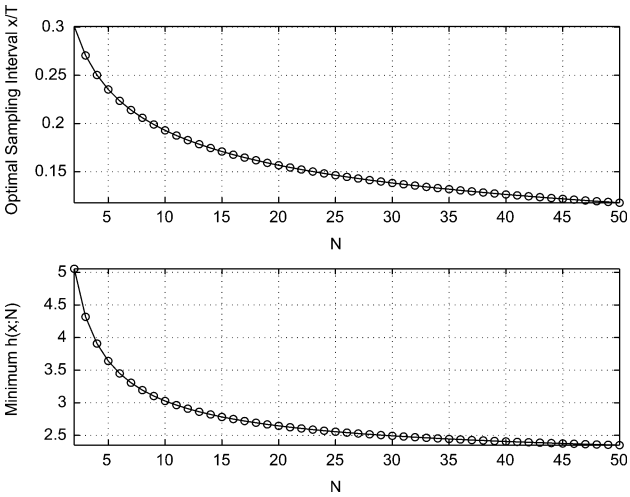


Fig. 4. Top: Optimal sampling interval x. Bottom: Minimum $h(x;N)$ for different N.

the sequential test. Note that this conclusion is based on the assumption that the SNR of a super sample is very low so that the AGN formula of (34) is accurate, as discussed in Appendix I. As shown by (87) in Appendix I, when N increases, the variance and the SNR of a super sample increase, (34) may not be an accurate estimate of the AGN any more, and Proposition 1 should be applied with caution. The optimal sampling interval x as a function of N is shown in the top of Fig. 4. As we can see, for a large group size N, a smaller x will result in faster termination. In the bottom of Fig. 4, the minimum $h(x;N)$, obtained using the optimal sampling interval x, is shown as a function of N. It can be seen that as group size N increases, the ATT decreases as expected.

As will be shown later, Scheme C's performance lies somewhere between Schemes A and B, either in terms of ATT or in terms of energy efficiency. The advantage of this scheme is that it is very flexible. By adjusting the number of samples (N) within a group, a desired tradeoff between the energy efficiency and the ATT can be achieved.

D. Group Sampling With Optimal Signal Waveform

Scheme D is still based on the group sampling idea as in Scheme C. The only difference is that the constant-amplitude signal within each group is replaced with the optimal signal that is designed to take the best advantage of the correlation between samples. It is well known that the eigenvalues of a positive definite covariance matrix are all real and positive. According to [20, Ch. 4.4.1], for a fixed energy constraint, the optimal signal that maximizes the SNR at the matched filter output is the eigenvector of the noise covariance $\mathbf{\Sigma}$ that corresponds to the minimum eigenvalue. For example, if group size $N = 2$, we have

$$\mathbf{\Sigma} = \sigma^2 \begin{bmatrix} 1 & \rho \\ \rho & 1 \end{bmatrix} \tag{43}$$

where $\rho = e^{-x}$ is the correlation coefficient. The eigenvalues are $\sigma^2(1+\rho)$ and $\sigma^2(1-\rho)$. The corresponding eigenvectors are $[1\ \ 1]'$ and $[1\ \ -1]'$, respectively. Therefore, the optimal signal is in the form of $\theta[1\ \ -1]'$, as shown in Fig. 1(D) for the case of $N = 2$.

Let us assume that each group has N samples. The corresponding covariance matrix $\mathbf{\Sigma}$ has been provided in (31). We denote the minimum eigenvalue and corresponding eigenvector of this covariance matrix as $\lambda_{\min}$ and $\mathbf{v}_{\min}$, respectively. We further assume that $\mathbf{v}_{\min}$ has an energy of $N\theta_4^2$, meaning that

$$\mathbf{v}_{\min}{}'\mathbf{v}_{\min} = N\theta_4^2. \tag{44}$$

Different from that described in (29), the hypothesis testing problem with the optimal signal is now

$$\begin{aligned} H_1 &: \mathbf{y}_i = \mathbf{v}_{\min} + \boldsymbol{\omega}_i \\ H_0 &: \mathbf{y}_i = \boldsymbol{\omega}_i. \end{aligned} \tag{45}$$

Similar to the derivation of (32), we have the log-likelihood ratio of each super sample given as

$$\lambda_i = \mathbf{v}'_{\min}\mathbf{\Sigma}^{-1}\left(\boldsymbol{y}_i - \frac{1}{2}\mathbf{v}_{\min}\right). \tag{46}$$

Hence

$$E[\lambda_i|H_1] = -E[\lambda_i|H_0] = \frac{1}{2}\mathbf{v}'_{\min}\mathbf{\Sigma}^{-1}\mathbf{v}_{\min}. \tag{47}$$

By the definitions of eigenvalue and eigenvector, it can be readily shown that

$$\mathbf{\Sigma}^{-1}\mathbf{v}_{\min} = \frac{\mathbf{v}_{\min}}{\lambda_{\min}}. \tag{48}$$

Substituting (48) and (44) into (47), we have

$$E[\lambda_i|H_1] = -E[\lambda_i|H_0] = \frac{N\theta_4^2}{2\lambda_{\min}}. \tag{49}$$

Therefore, the AGNs for Scheme D are

$$\begin{aligned} E[M_4|H_1] = E[M_4|H_0] &= \frac{\beta b + (1-\beta)a}{E[\lambda_i|H_1]} \\ &= \frac{2(1-2\beta)a\lambda_{\min}}{N\theta_4^2} \end{aligned} \tag{50}$$

where M_4 denotes the number of groups required by the SPRT to terminate in Scheme D, and the ATT is

$$E[t_4|H_1] = E[t_4|H_0] = E[M_4|H_1][(N-1)x+T] = \frac{2(1-2\beta)a\lambda_{\min}}{N\theta_4^2}[(N-1)x+T]. \quad (51)$$

IV. Performance Comparison in Terms of ATT

In this section, the performances of different schemes are compared in terms of their ATTs. We assume that for Schemes A, C, and D, $x_1 = x_3 = x_4 = x$, and that Schemes B, C, and D use the same T. In addition, the amplitudes of the signal are all the same for different schemes, meaning that $\theta_1 = \theta_2 = \theta_3 = \theta_4 = \theta$. ATT for different schemes is evaluated next.

A. Analytical Comparison

For Scheme A, we rewrite (20) as

$$E[t_1] = \kappa f(x) \quad (52)$$

where

$$\kappa \triangleq \frac{2(1-2\beta)a\sigma^2}{\theta^2}. \quad (53)$$

Similarly, for Scheme B, (28) is rewritten

$$E[t_2] = \kappa T. \quad (54)$$

For Scheme C, from (35), we have

$$E[t_3] = \kappa h(x;N). \quad (55)$$

Let us now compare the ATT of Scheme C with those of Schemes A and B. It turns out that the ATT of Scheme C lies between those of Schemes A and B, as stated here.

Corollary 1: Provided $2 \le N < \infty$ and $0 < x < \eta$, where η has been defined in Proposition 1, $f(x) < h(x;N) < T$, implying that the ATT of Scheme C is less than that of Scheme B, but more than that of Scheme A.

Proof: This corollary can be deduced from Proposition 1. Scheme A can be deemed as a special case of Scheme C, as when $N = \infty$ we have

$$h(x;N=\infty) = \frac{1+e^{-x}}{1-e^{-x}}x = f(x). \quad (56)$$

A direct application of Proposition 1 results in

$$f(x) < h(x;N) \quad (57)$$

as long as $x < \eta$ and $N < \infty$.

Now let us prove $h(x;N) < T$. Scheme B is actually another special case of Scheme C, since

$$h(x;N=1) = T.$$

Based on Proposition 1, we have

$$h(x;N) < h(x;1) = T \quad (58)$$

for $N \ge 2$ and $x < \eta$. □

From Corollary 1, we gain very interesting insights that Schemes A and B are nothing but two special cases of Scheme C, when $N \to \infty$ and $N = 1$, respectively. This enables Scheme C to be very flexible. We will show later that Scheme A is not very energy-efficient. Therefore, by adjusting N, Scheme C can achieve a suitable tradeoff between the ATT and energy efficiency.

Now let us consider Scheme D. We know that when $N = 2$, the optimal signal is in the form of $\theta[1 \ -1]'$ and the signal has a constant amplitude. For cases where $N > 2$, the optimal signal for Scheme D does not have a constant amplitude and we will investigate them in Section V. Using the fact that $N\theta_4^2 = 2\theta^2$, $\lambda_{\min} = \sigma^2(1-\rho)$ and (51), we have

$$E[t_4] = \frac{\kappa(1-e^{-x})(x+T)}{2}. \quad (59)$$

Obviously, this is a monotonically increasing function of x. This means a smaller x leads to a smaller ATT of the sequential test.

Another observation is that when $x = T$, all the schemes become Schemes B, and their ATTs approach that of Scheme B, κT. This can be proved mathematically. Plugging $x = T$ and $e^{-T} \approx 0$ into (52), (55), and (59), respectively, we can show that $E[t_1]$, $E[t_3]$ and $E[t_4]$ are approximately κT.

B. Numerical Results

Note that the ATTs of four schemes that we have derived have a common factor, $\kappa = 2(1-2\beta)a\sigma^2/\theta^2$, which is inversely proportional to the SNR (θ^2/σ^2), meaning that the higher the SNR, on the average the faster the schemes will terminate their corresponding sequential tests. We take $\epsilon = 10^{-3}$ and hence $T = 6.91$. The ATT normalized with respect to κT for different schemes are plotted in Fig. 5. It is clear that the curve for Scheme C lies between those of Schemes A and B, as predicted by Corollary 1. Also, when $N = 2$, Scheme D always has less ATT than Scheme C. This is because

$$E[t_3|N=2] = \frac{\kappa}{2}(1+e^{-x})(x+T)$$

which is always greater than

$$E[t_4|N=2] = \frac{\kappa}{2}(1-e^{-x})(x+T).$$

From this figure, it is clear that when the sampling interval is small, Scheme D $(N = 2)$ has the best performance. This is because the high correlation between samples benefits Scheme D and compromises Scheme A to some extent. When sampling interval is medium or large, Scheme A results in the smallest ATT. However, because all the schemes have identical signal amplitude, Scheme A requires the highest power, with its uniform sampling procedure. We will discuss the energy efficiency of the schemes in the next section. When the sampling interval tends to T, basically the performances of all the schemes tend to that of Scheme B. This is because the correlation between samples is so weak that all the schemes become a standard SPRT.

V. Energy Efficiency Comparison

We have compared the ATTs for different schemes under the equality constraint on the signal amplitude. It is of much impor-

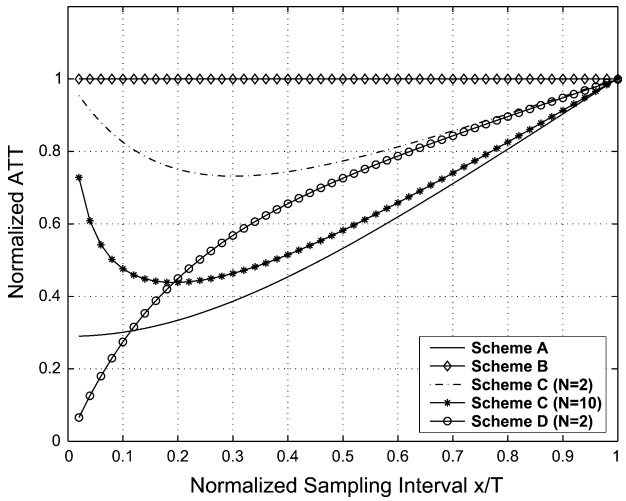


Fig. 5. Average termination times for different schemes.

tance to compare their ATTs under a power constraint as well, since in many practical applications, such as target detection at a long range with a radar, the SPRT works with a very low SNR and the power is a major issue. In this section, we assume a constant average system power P for all the sampling schemes. The energy efficiency is measured in terms of the ATT again, since the scheme that has a smaller ATT will also consume less energy under the constant average power constraint. Also, let us assume that $x_1 = x_3 = x_4 = x$. We analyze and compare their ATTs both theoretically and numerically. We will also show that the ATTs for different schemes under a constant average power constraint are proportional to the corresponding ASNs under a constant signal amplitude constraint, up to a common factor. Hence, all the comparison conclusions made for the energy-efficiency are also valid for the ASNs under a constant signal amplitude constraint.

A. Schemes A, B and C

For Scheme A, the average power is

$$P = \frac{\theta_1^2}{x}. \tag{60}$$

Substituting the above into (20), it follows that

$$E[t_1] = \gamma \frac{1+e^{-x}}{1-e^{-x}} \tag{61}$$

where

$$\gamma \triangleq \frac{2(1-2\beta)a\sigma^2}{P}. \tag{62}$$

For Scheme B, the average power is

$$P = \frac{\theta_2^2}{T}. \tag{63}$$

According to (28), it follows that the ATT is

$$E[t_2] = \gamma. \tag{64}$$

For Scheme C, the average power is

$$P = \frac{N\theta_3^2}{(N-1)x+T}. \tag{65}$$

Plugging the above into (35), we have

$$E[t_3] = \gamma \frac{1+e^{-x}}{1-(1-\frac{2}{N})e^{-x}}. \tag{66}$$

Obviously, both $E[t_1]$ and $E[t_3]$ are monotonically decreasing functions of x, meaning that the longer the sampling interval x is, the more efficient the two schemes are. When $x \longrightarrow \infty$ (or approximately $x \longrightarrow T$), both $E[t_1]$ and $E[t_3]$ tend to $E[t_2] = \gamma$. This implies that both Scheme A and Scheme C are always less efficient than Scheme B. It is also very easy to show that

$$\frac{1+e^{-x}}{1-e^{-x}} > \frac{1+e^{-x}}{1-(1-\frac{2}{N})e^{-x}} \tag{67}$$

for $N \geq 2$. Therefore, if they have the same sampling interval x, Scheme C is always more efficient than Scheme A. Also, $E[t_3]$ in (66) is a monotonically increasing function of N and as $N \longrightarrow \infty$

$$\frac{1+e^{-x}}{1-(1-\frac{2}{N})e^{-x}} \rightarrow \frac{1+e^{-x}}{1-e^{-x}}. \tag{68}$$

Therefore, as the group size increases, the energy efficiency decreases. When N is very large, the performance of Scheme C will converge to that of Scheme A.

In summary, we have

$$\frac{1+e^{-x}}{1-e^{-x}} > \frac{1+e^{-x}}{1-(1-\frac{2}{N})e^{-x}} > 1 \tag{69}$$

meaning that Scheme C's ATT again lies between those of Schemes A and B. This is not surprising since Schemes A and B are two extreme cases of Scheme C for $N \longrightarrow \infty$ and $N = 1$, respectively.

B. Scheme D

For Scheme D, the average power is

$$P = \frac{N\theta_4^2}{(N-1)x+T}. \tag{70}$$

Using the above equation and (51), we have

$$E[t_4] = \gamma \frac{\lambda_{\min}}{\sigma^2}. \tag{71}$$

We have shown that Scheme D is more efficient than Scheme B, which is summarized in the following theorem.

Theorem 3: The smallest eigenvalue of the noise covariance matrix $\mathbf{\Sigma}$ defined in (31) is always less than or equal to σ^2, or $\lambda_{\min}/\sigma^2 \leq 1$.

Proof: In [19], three approximations of the eigenvalues of the matrix $\mathbf{F}$ defined in (38), as well as their error bounds are provided. We use one of the approximations to prove Theorem 3. As shown in [19], for $\rho > 0$, the largest eigenvalue of $\mathbf{F}$ can be estimated by

$$\hat{\xi}_{\max} = (1+\rho)^2 - \frac{2}{N}\rho(1+\rho) \tag{72}$$

with an error bound given by

$$|\xi_{\max} - \hat{\xi}_{\max}| \leq \left(\frac{2}{N}\right)^{1/2} \rho(1+\rho) \tag{73}$$

where $\xi_{\max}$ is the true value of the largest eigenvalue of $\mathbf{F}$. Since $\mathbf{\Sigma} = \sigma^2(1-\rho^2)\mathbf{F}^{-1}$, its smallest eigenvalue is given by $\lambda_{\min} = \sigma^2(1-\rho^2)/\xi_{\max}$. To complete the proof, we need to show that

$$\xi_{\max} \geq 1 - \rho^2. \tag{74}$$

Substituting (72) into (73), and after simplification, we get

$$\xi_{\max} \geq (1+\rho)\left[1 + \left(1 - \frac{2}{N} - \sqrt{\frac{2}{N}}\right)\rho\right]. \tag{75}$$

Since $N \geq 2$, we have

$$\frac{2}{N} + \sqrt{\frac{2}{N}} \leq 2. \tag{76}$$

Plugging (76) back to (75), the inequality (74) can be readily established. □

With Theorem 3, we finally have

$$\frac{1+e^{-x}}{1-e^{-x}} > \frac{1+e^{-x}}{1-(1-\frac{2}{N})e^{-x}} > 1 \geq \frac{\lambda_{\min}}{\sigma^2}. \tag{77}$$

Namely, sampling Scheme D is the most energy-efficient among all the schemes.

The monotonicity of $\lambda_{\min}(x, N)$ is very difficult to establish since no closed-form solution exists. We instead study its estimated value using the approximation provided in [19], and the results are provided in Proposition 2.

Proposition 2: $\hat{\lambda}_{\min}(x, N)$, the estimated value of the minimum eigenvalue of $\mathbf{\Sigma}$, is a monotonically increasing function of x, and a monotonically decreasing function of N, where

$$\hat{\lambda}_{\min}(x; N) = \frac{\sigma^2(1-\rho^2)}{\hat{\xi}_{\max}(x; N)} = \frac{\sigma^2(1-e^{-x})}{1+(1-\frac{2}{N})e^{-x}}. \tag{78}$$

Proposition 2 can be shown based on (78), and we skip the proof for brevity.

This proposition implies that the longer the sampling interval x is, the less efficient Scheme D is. As $x \longrightarrow \infty$, $\hat{\lambda}_{\min}(x, N)/\sigma^2 \longrightarrow 1$, indicating that Scheme D converges to Scheme B for a very large x.

We use numerical methods to find the minimum eigenvalue of the noise covariance matrix $\mathbf{\Sigma}$, which is then compared with the estimated value obtained through the approach discussed in Proposition 2. For this purpose, we take $\epsilon = 10^{-3}$. The normalized minimum eigenvalues $\lambda_{\min}/\sigma^2$ as functions of sampling interval x and group size N are shown in Figs. 6 and 7, respectively. As we can see, $\lambda_{\min}$ is a monotonically increasing function of x, meaning that the smaller the sampling interval, the better the performance is. This is not surprising, since smaller sampling interval gives rise to higher correlation between samples, which helps Scheme D perform a better noise cancellation. Note that $\lambda_{\min}$ is a monotonically decreasing function of N, meaning that the larger the group size, the better the performance. However, when group size is greater than 10, the improvement becomes much less significant.

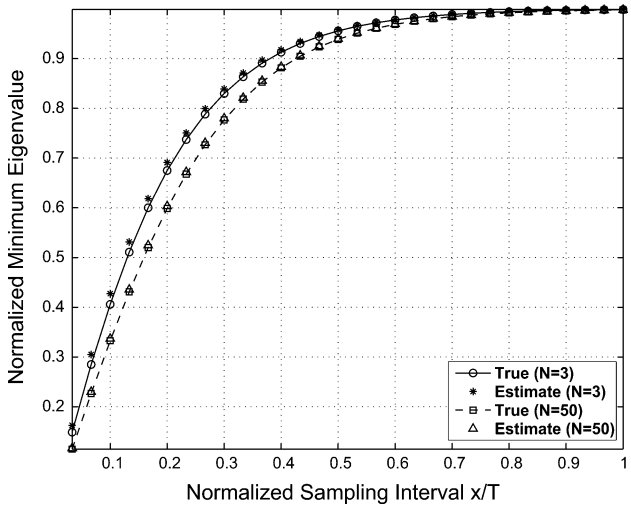


Fig. 6. Minimum eigenvalues for Scheme D.

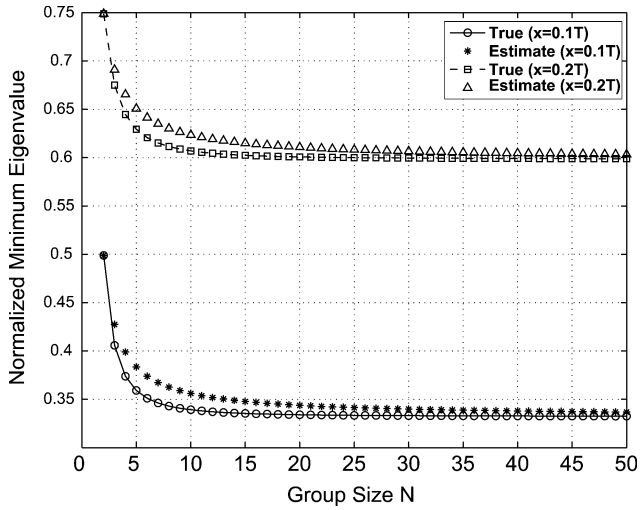


Fig. 7. Minimum eigenvalues for Scheme D.

It is also clear that (78) provides a reasonably good estimate of $\lambda_{\min}$, especially when N is large. The estimation accuracy improves as N increases, a trend that is consistent with the error bound given by (73). The estimation approach proposed in [19] is very useful when N is large, since the numerical calculation of the eigenvalues of $\mathbf{\Sigma}$, a $N \times N$ matrix, will become more complex and costly.

C. ASN Comparison Under a Constant Amplitude Constraint

Here as in Section IV, we assume that the amplitudes of the signal are all the same for different schemes, namely $\theta_1 = \theta_2 = \theta_3 = \theta_4 = \theta$. Let us derive the ASNs for different schemes.

For Scheme A, based on (19), the ASN is

$$\mathrm{ASN}_1 = E[M_1] = \kappa\frac{1+e^{-x}}{1-e^{-x}} \tag{79}$$

where $\kappa = 2(1-2\beta)a\sigma^2/\theta^2$ has been defined in (53). Now the relationship between $E[t_1]$ defined in (61), which is the ATT under the constant average power constraint, and $E[M_1]$ is very clear. $E[t_1]$ is proportional to $E[M_1]$, and vice versa. This is not surprising since both ATT under a constant average power constraint and the ASN under a constant signal amplitude constraint are measures of the total energy required for the SPRT to reach a decision.

Following a similar procedure, it is straightforward to derive that for Scheme B,

$$\mathrm{ASN}_2 = \kappa. \tag{80}$$

Now let us study the ASN for Scheme C. Note that in this case, the ASN is the AGN multiplied by the group size N. Based on (34), we have

$$\mathrm{ASN}_3 = E[M_3] \times N = \kappa \frac{1+e^{-x}}{1-(1-\frac{2}{N})e^{-x}}. \tag{81}$$

Similarly, for Scheme D, we have

$$\mathrm{ASN}_4 = \kappa \frac{\lambda_{\min}}{\sigma^2}. \tag{82}$$

Comparing these equations with (64), (66), and (71), it is clear that they are all proportional to their corresponding ATTs under a constant average power constraint. Therefore, the result derived in the last subsection remains the same, and we rewrite (77)

$$\frac{1+e^{-x}}{1-e^{-x}} > \frac{1+e^{-x}}{1-(1-\frac{2}{N})e^{-x}} > 1 \geq \frac{\lambda_{\min}}{\sigma^2}. \tag{83}$$

This implies that Scheme D requires the least number of samples with the constant amplitude constraint.

D. Numerical Comparison

In the numerical example, we take $\epsilon = 10^{-3}$ and hence $T = 6.91$. The ATTs normalized with respect to γ, under the assumption of identical signal average power for different schemes, are plotted in Fig. 8. One observation is that as group size N increases, Scheme C becomes less efficient (or needs a larger ATT), since $E[t_3]$ is a monotonically increasing function of N, as discussed in Section V-A. From this figure, it is evident that Scheme A has the worst efficiency. On the other hand, Scheme D has the best performance, especially when sampling rate is high and group size is large. For a numerical example, when $x = T/50$, the ATTs required by Schemes A, C (N = 5), C (N = 2), B, D (N = 2), and D (N = 20) are $[14.50,\ 3.92,\ 1.87,\ 1,\ 0.13,\ 0.069]\gamma$, respectively. It is also noteworthy that as $x \longrightarrow T$, all the schemes' performances converge to that of Scheme B again.

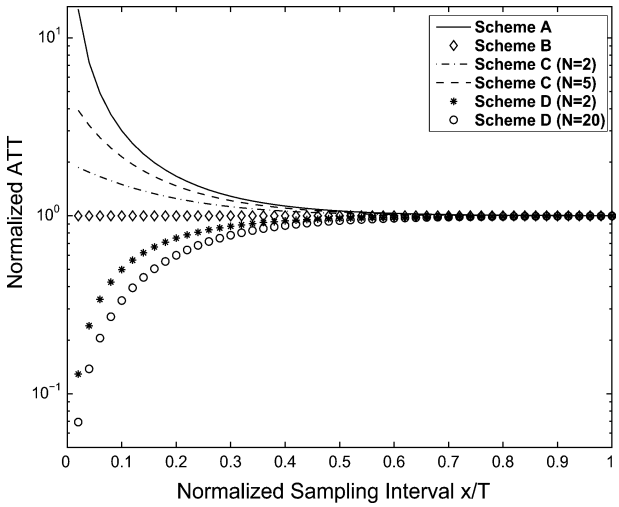


Fig. 8. Average termination times for different schemes.

VI. Conclusion

In this paper, we have proposed four different sampling schemes for sequential detection in autoregressive noise. Scheme A adopts a uniform sampling procedure with a high sampling rate. The correlated data are processed by a whitening filter to generate a white sequence. Scheme B also uses uniform sampling, but it employs a low sampling rate such that the correlation between samples can be ignored. Both Schemes C and D are based on the group sampling idea. The intergroup delay is large enough to avoid intergroup correlation. In Scheme C, within each group, the signal amplitude is constant; whereas, Scheme D uses an optimal signal waveform for each group.

Based on the assumptions of constant signal amplitude and identical sampling rates, the schemes have been compared in terms of ATT. Scheme D has the best performance when sampling rate is very high. Otherwise, Scheme A has the smallest ATT. However, Scheme A requires larger signal power than any other scheme. Therefore, a comparison of energy efficiency is conducted. Under a constant power constraint, Scheme B, C, and D are all more energy-efficient than Scheme A. Taking full advantage of the correlation between data samples, Scheme D has the highest efficiency.

In all the proposed sampling schemes, the samples are equally spaced either over the entire observation interval or within a group. In our future work, this restriction will be relaxed by considering nonuniform sampling within a group, similar to the nonuniform sampling designs for FSS detection that have been investigated in [14]–[16]. In the paper, the problem is formulated as detecting a deterministic signal in the presence of additive Gaussian noise. In practical radar and communication systems, the signal is typically random due to fluctuating targets or multipath channel fading. The current work will be extended to take into consideration of the randomness of the signal.

Appendix I
Improved ASN or AGN Approximation

The ASN provided in (4) is typically underestimated when the excesses of the likelihood ratio over the test thresholds are neglected. By approximating rather than neglecting the excesses, the accuracy of estimating ASN can be improved [17], [21]. In [21, Theorem 2], a more accurate ASN expression

$$h'(x) = -\frac{(N-1)}{\left[N-(N-2)e^{-x}\right]^2} \times \left[(N-2)e^{-2x} + 2(N-1)xe^{-x} + 2(T-1)e^{-x} - N\right]. \tag{89}$$

$$(N-1)(N-2)[(x+2)e^{-x}+x] + T(N-2)e^{-x} + N(T+2-2N) > 0. \tag{94}$$

is derived by applying corrected Brownian motion approximations

$$\mathrm{ASN}|H_j = \frac{P_j(a+\vartheta\sigma_j) + (1-P_j)(b-\vartheta\sigma_j)}{E[\lambda|H_j]} \tag{84}$$

where $P_0 = \alpha, P_1 = 1-\beta, \sigma_j \triangleq \sqrt{\mathrm{Var}(\lambda|H_j)}$, and $\vartheta \approx 0.583$ when λ is a Gaussian random variable [17].

It is easy to show that for Schemes A to D, we have

$$\mathrm{Var}(\lambda_1|H_1) = \mathrm{Var}(\lambda_1|H_0) = \frac{(1-\rho)\theta_1^2}{(1+\rho)\sigma^2} \tag{85}$$

$$\mathrm{Var}(\lambda_2|H_1) = \mathrm{Var}(\lambda_2|H_0) = \frac{\theta_2^2}{\sigma^2} \tag{86}$$

$$\mathrm{Var}(\lambda_3|H_1) = \mathrm{Var}(\lambda_3|H_0) = \frac{\theta_3^2[N-(N-2)\rho]}{\sigma^2(1+\rho)} \tag{87}$$

$$\mathrm{Var}(\lambda_4|H_1) = \mathrm{Var}(\lambda_4|H_0) = \frac{N\theta_4^2}{\lambda_{\min}}. \tag{88}$$

It is clear from these equations that the correction terms $(\vartheta\sigma_j)$ are related to the SNR. When the SNR is very small, which is the typical situation that necessitates a sequential detector, the correction terms have a negligible effect in (84). For example, if we set $\alpha = \beta = 10^{-4}$, and use Scheme C with $N = 2$, $\rho = 0.5$, and 0 dB SNR ($\theta_3^2/\sigma^2 = 1$), we have $a = -b = 9.2$ and $\vartheta\sigma_j = 0.67$.

Appendix II
Proof of Theorem 2

We first derive the first order derivative of $h(x;N)$ with respect to x as shown in (89) at the top of the page. The second order derivative can been readily derived also

$$h''(x) = \frac{2(N-1)e^{-x}m(x)}{\left[N-(N-2)e^{-x}\right]^3} \tag{90}$$

where

$$m(x) \triangleq (N-1)(N-2)(x+2)e^{-x} + T(N-2)e^{-x} + xN(N-1) + N(T+2-2N). \tag{91}$$

It is obvious that $h''(x) > 0$ if the sum of the last two terms in the right-hand side (RHS) of (91) is positive, or equivalently

$$x > 2 - \frac{T}{N-1}. \tag{92}$$

The above inequality implies that when $T > 2(N-1)$, $h''(x) > 0$ always holds; for the case $T < 2(N-1)$, if $x \geq 2$, $h''(x) > 0$ still holds. Let us consider the case where $x < 2$. It suffices to prove that

$$m(x) > 0. \tag{93}$$

Since $xN(N-1) > x(N-1)(N-2)$, we can prove instead that [see (94), shown at the top of the page]. Define $k(x) \triangleq x + (x+2)e^{-x}$, we have

$$k'(x) = 1 - (1+x)e^{-x} \tag{95}$$

and

$$k''(x) = xe^{-x} > 0 \tag{96}$$

which means that $k'(x) > k'(0) = 0$, and $k(\cdot)$ is a monotonically increasing function. A direct result is

$$k(x) > k(0) = 2. \tag{97}$$

Also note that $e^{-x} > e^{-2}$ for $0 < x < 2$. Plugging (97) and $e^{-x} > e^{-2}$ into (94) and after some manipulations, we have

$$T > \frac{4(N-1)}{e^{-2}(N-2)+N}. \tag{98}$$

Since $N \geq 2$, it is easy to show that

$$\frac{4(N-1)}{e^{-2}(N-2)+N} \leq 2(N-1). \tag{99}$$

Hence $T > 4(N-1)/[e^{-2}(N-2)+N]$ is a weaker condition than $T > 2(N-1)$, and once the former is satisfied, $h(x)$ is convex even for $x \in (0,\ 2)$.

Now let us define $l(N) \triangleq 4(N-1)/[e^{-2}(N-2)+N]$, it follows that

$$\frac{\partial l(N)}{\partial N} = \frac{4(1-e^{-2})}{\left[(1+e^{-2})N - 2e^{-2}\right]^2} > 0. \tag{100}$$

Therefore, $l(2) \leq l(N) < l(\infty)$, or $l(N) \in \left[2, 4/(1+e^{-2})\right)$.

According to (89), and considering $T > l(N) \geq 2$, we have

$$h'(0) = -\frac{(N-1)(T-2)}{2} < 0.$$

Also

$$h'(\infty) = \frac{N-1}{N} > 0.$$

Therefore, the minimum of $h(x)$, or the solution of $h'(x) = 0$, lies in $(0,\ \infty)$.

REFERENCES

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

[2] R. Dwyer and L. Kurz, "Characterizing partition detectors with stationary and quasi-stationary Markov dependent data," *IEEE Trans. Inf. Theory*, vol. 32, pp. 471–482, Jul. 1986.

[3] R. M. Phatarfod, "Sequential analysis of dependent observations. I," *Biometrika*, vol. 52, no. 1/2, pp. 157–165, 1965.

[4] B. Dimitriadis and D. Kazakos, "A nonparametric sequential test for data with Markov dependence," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 19, pp. 338–347, 1983.

[5] R. Chandramouli and N. Ranganathan, "A generalized sequential sign detector for binary hypothesis testing," *IEEE Signal Process. Lett.*, vol. 5, pp. 295–297, 1998.

[6] B. Eisenberg, B. K. Ghosh, and G. Simons, "Properties of generalized sequential probability ratio tests," *The Ann. Statist.*, vol. 4, no. 2, pp. 237–251, Mar. 1976.

[7] J. Cochlar and I. Vrana, "On the optimum sequential test of two hypotheses for statistically dependent observations," *Kybernetika*, vol. 14, no. 1, pp. 57–69, 1978.

[8] S. Tantaratana, "Comparison of the SPRT and the sequential linear detector in autoregressive noise," *IEEE Trans. Inf. Theory*, vol. 31, pp. 693–697, Sep. 1985.

[9] A. G. Tartakovsky, X. R. Li, and G. Yaralov, "Sequential detection of targets in multichannel systems," *IEEE Trans. Inf. Theory*, vol. 49, pp. 425–445, 2003.

[10] M. Woinsky and L. Kurz, "Nonparametric detection using dependent samples (Corresp.)," *IEEE Trans. Inf. Theory*, vol. 16, pp. 355–358, May 1970.

[11] S. A. Kassam and J. B. Thomas, "A class of nonparametric detectors for dependent input data," *IEEE Trans. Inf. Theory*, vol. 21, pp. 431–437, Jul. 1975.

[12] C. C. Lee and J. B. Thomas, "A modified sequential detection procedure," *IEEE Trans. Inf. Theory*, vol. IT-30, pp. 16–23, Jan. 1984.

[13] M. M. Al-Ibrahim and P. K. Varshney, "A simple multi-sensor sequential detection procedure," in *Proc. 27th Conf. Decision Control*, Austin, TX, Dec. 1988.

[14] S. Cambanis and E. Masry, "Sampling designs for the detection of signals in noise," *IEEE Trans. Inf. Theory*, vol. 29, pp. 83–104, Jan. 1983.

[15] R. K. Bahr and J. A. Bucklew, "Optimal sampling schemes for the Gaussian hypothesis testing problem," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. 38, no. 10, pp. 1677–1686, Oct. 1990.

[16] C. T. Yu and P. K. Varshney, "Sampling design for Gaussian detection problems," *IEEE Trans. Signal Process.*, vol. 45, pp. 2328–2337, Sep. 1997.

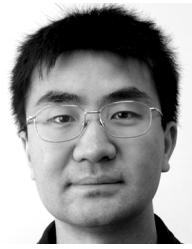
[17] D. Siegmund, *Sequential Analysis: Tests and Confidence Intervals*. New York: Springer-Verlag, 1985.

[18] A. Papoulis, *Probability, Random Variables, and Stochastic Processes*. New York, NY: McGraw-Hill, 1984.

[19] R. J. Stroeker, "Approximations of the eigenvalues of the covariance matrix of a first-order autoregressive process," *J. Econometr.*, vol. 22, no. 3, pp. 269–279, Aug. 1983.

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

[21] Q. Cheng, P. K. Varshney, K. Mehrotra, and C. Mohan, "Bandwidth management in distributed sequential detection," *IEEE Trans. Inf. Theory*, vol. 51, pp. 2954–2961, 2005.



Ruixin Niu (M'04) received the B.S. degree from Xi'an Jiaotong University, Xi'an, China, in 1994, the M.S. degree from the Institute of Electronics, Chinese Academy of Sciences, Beijing, in 1997, and the Ph.D. degree from the University of Connecticut, Storrs, in 2001, all in electrical engineering.

He is currently a Research Assistant Professor with Syracuse University, Syracuse, NY. His research interests are in the areas of statistical signal processing and its applications, including detection, estimation, data fusion, sensor networks, communications, and image processing.

Dr. Niu received the Best Paper award at the Seventh International Conference on Information Fusion in 2004. He is the managing editor of the *Journal of Advances in Information Fusion*. He serves as a Lead Guest Editor for a special issue of EURASIP *Journal on Advances in Signal Processing*.

Pramod K. Varshney (F'97) was born in Allahabad, India, on July 1, 1952. He received the B.S. degree in electrical engineering and computer science (with highest honors) and the M.S. and Ph.D. degrees in electrical engineering from the University of Illinois at Urbana-Champaign in 1972, 1974, and 1976, respectively.

Since 1976, he has been with Syracuse University, Syracuse, NY, where he is currently a Distinguished Professor of electrical engineering and computer science. His current research interests are in distributed sensor networks and data fusion, detection and estimation theory, wireless communications, image processing, radar signal processing, and remote sensing. He has published extensively. He is the author of *Distributed Detection and Data Fusion* (New York: Springer-Verlag, 1997).

Dr. Varshney was a James Scholar, a Bronze Tablet Senior, and a Fellow while with the University of Illinois. He is a member of Tau Beta Pi and is the recipient of the 1981 ASEE Dow Outstanding Young Faculty Award. He was the Guest Editor of the Special Issue on Data Fusion of the PROCEEDINGS OF THE IEEE, January 1997. In 2000, he received the Third Millenium Medal from the IEEE and the Chancellor's Citation for Exceptional Academic Achievement at Syracuse University. He serves as a Distinguished Lecturer for the IEEE Aerospace and Electronic Systems (AES) Society. He is on the editorial boards of the *International Journal of Distributed Sensor Networks* and the IEEE TRANSACTIONS ON SIGNAL PROCESSING. He was the President of the International Society of Information Fusion during 2001.