

Near-Optimal Signal Detection for Finite-State Markov Signals With Application to Magnetic Resonance Force Microscopy

Michael Ting, *Student Member, IEEE*, Alfred O. Hero, III, *Fellow, IEEE*, Daniel Rugar, *Senior Member, IEEE*, Chun-Yu Yip, *Student Member, IEEE*, and Jeffrey A. Fessler, *Fellow, IEEE*

Abstract—Detection of a finite-state Markov signal in additive white Gaussian noise (AWGN) can be done in an intuitive manner by applying an appropriate filter and using an energy detector. One might not expect this heuristic approach to signal detection to be optimal. However, in this paper, we show that for a certain type of finite-state Markov signal, namely, the discrete-time (DT) random telegraph, this filtered energy detector is approximately optimal under the following conditions of: symmetric transition probabilities, low signal-to-noise ratio (SNR), long observation time, and small probability of transition between two consecutive time instances. When these last three conditions hold, but the transition probabilities are not symmetric, we show that a variant of the filtered energy detector is approximately optimal. It is also shown, under low SNR conditions, that the likelihood ratio test (LRT) for a finite-state DT Markov signal in AWGN reduces to the matched filter statistic with the minimum mean-squared-error (MMSE) predictor signal values used in place of the known signal values. Using this result, we propose a general methodology for obtaining an approximation to the LRT of a finite-state DT Markov signal in AWGN. Specifically, instead of the conditional mean (also MMSE) estimators, affine estimators with lowest mean squared error are used. This work is relevant to magnetic resonance force microscopy, an emerging technology that uses ultrasensitive force sensing to detect magnetic resonance. Sensitivity down to the single spin level was demonstrated in a recent experiment.

Index Terms—Approximation methods, likelihood ratio test, magnetic resonance force microscopy, Markov processes, signal detection.

I. INTRODUCTION

DETECTION of a finite-state discrete-time (DT) Markov signal in additive white Gaussian noise (AWGN) is widespread in many different fields. Detection of a random telegraph signal is used in the study of particle tunnelling [1] and in the study of low-frequency noise characteristics of light-emitting diodes [2]. Markov chains are used in [3] for the purpose of statistical network anomaly detection and in [4] for the purpose

of land mine detection. The focus application of this paper is magnetic resonance force microscopy (MRFM), which is a promising technique for three-dimensional imaging on the nanometer scale. Recent experiments at IBM have shown that MRFM is capable of detecting and localizing individual electron spins associated with subsurface atomic defects in silicon dioxide [5]. This single-spin detection milestone represents a factor of 10^7 improvement over conventional electron spin resonance detection and was achieved using energy detection methods similar to those described in this paper. Other recent MRFM experiments have demonstrated the ability to detect and manipulate naturally occurring statistical fluctuations in small spin ensembles [6]. With further development, single-spin MRFM may eventually lead to atomic-resolution magnetic resonance imaging and find application in quantum computing experiments [7].

The recursive structure of the likelihood ratio test (LRT) for a finite-state DT Markov signal is given in [8]. In this paper, we specialize the noise to AWGN and derive a new interpretation of the optimal LRT for a finite-state DT Markov signal under low signal-to-noise ratio (SNR) conditions. It is shown that, under low SNR, the LRT reduces to the matched filter statistic with the minimum mean-squared-error (MMSE) predictor values used in place of the known signal values. Current single spin experiments operate under conditions of very low SNR; consequently, we are interested in the performance of detectors in the regime of low SNR and long observation time. Our results are applicable to [1]–[4] under conditions of low SNR. When applied to the LRT of the DT random telegraph, the result is an estimator-correlator detector. This estimator-correlator structure appears in the LRT of problems whose probability density functions have an exponential form [9]. In particular, it applies when detecting a Gaussian signal in AWGN. This first result has a continuous-time (CT) analog: in CT, the LRT for detecting a random signal in AWGN has the form of the matched filter statistic with the MMSE predictor used in place of the known signal values. The first difference between CT and DT is that the CT form is exact under all SNR conditions [10]–[12]. There is another difference: in the CT form, the square of the conditional expectation of the random process is used instead of the conditional expectation of the squared value of the random process.

The second result is that, when used to detect the DT random telegraph in AWGN, the filtered energy (FE) detector is approximately optimal under the following four conditions: symmetric transition probabilities, low SNR, long observation time, and

Manuscript received February 7, 2005; revised June 30, 2005. This work was supported in part by the Defense Advanced Research Projects Agency's Mosaic program under ARO Contract DAAD19-02-C-0055. The associate editor coordinating the review of this manuscript and approving it for publication was Dr. Dominic K. C. Ho.

M. Ting, A. O. Hero III, C.-Y. Yip, and J. A. Fessler are with the Department of Electrical Engineering and Computer Science, University of Michigan, Ann Arbor, MI 48109-2108 USA (e-mail: mting@umich.edu; hero@eecs.umich.edu; chunyu@umich.edu; fessler@eecs.umich.edu).

D. Rugar is with the IBM Research Division, Almaden Research Center, San Jose, CA 95120 USA (e-mail: rugar@almaden.ibm.com).

Digital Object Identifier 10.1109/TSP.2006.874413

a small probability of transition between two consecutive instances. The FE detector is no longer approximately optimal when the transition probabilities are asymmetric. We extend the FE detector to a hybrid second-order detector which combines the filtered energy, amplitude, and energy statistics. It is shown that the hybrid detector is approximately optimal for the DT random telegraph model under only the last three conditions. This is an intuitively pleasing result, as the idea of performing detection of a finite-state Markov signal by filtering the noisy observations and applying an energy detector is one that comes naturally.

Third, the first result is used to obtain an approximation to the LRT of a general finite-state DT Markov signal in AWGN. Sub-optimal affine estimators of the random process and the squared value of the random process are used instead of the optimal conditional mean (also MMSE) estimators. When this general methodology is applied to the LRT of the DT random telegraph in AWGN, an approximation is obtained that closely matches the result that was obtained using a straightforward analysis. It has been noted in [13] that linear MMSE estimates of the CT symmetric random telegraph process are as efficient as the non-linear MMSE estimates as the $\text{SNR} \rightarrow 0$. We also present simulations that suggest the optimality of a similar approximation for the DT random walk process. It would be interesting to investigate whether affine MMSE estimators are as efficient as non-linear MMSE estimators for a general finite-state DT Markov signal in AWGN as the $\text{SNR} \rightarrow 0$.

The outline of this paper is as follows. In Section II, we briefly review the basic principles of MRFM. This is followed by a discussion in Section III of two finite-state DT Markov signal models: the random telegraph and random walk models. In Section IV, we describe existing detectors that are commonly used, namely, the amplitude and filtered energy detectors, and compare them to the optimal detectors. We derive a new interpretation of the LRT under low SNR conditions. As well, the FE detector is extended to a hybrid version, and a general methodology to obtain an approximation to the LRT of a finite-state DT Markov process in AWGN is presented. Simulation results are presented in Section V.

II. BASIC PRINCIPLES OF MRFM SPIN DETECTION

MRFM experiments, in general, involve the measurement of magnetic force between a submicrometer-size magnetic tip and spins in a sample. The details of spin manipulation and signal detection depend on the exact MRFM protocol used. One particularly successful protocol is called OSCAR, which stands for oscillating cantilever-driven adiabatic reversal [14], [15]. A variation of this protocol, “interrupted OSCAR” (iOSCAR), was used in recent single spin experiments [5].

A schematic diagram of an OSCAR-type MRFM experiment is shown in Fig. 1. As shown in the figure, a submicrometer ferromagnet is placed on the tip of a cantilever and positioned close to an unpaired electron spin contained within the sample. An applied radio-frequency (RF) field serves to induce magnetic resonance of the spin when the condition $B_0 = \omega_{\text{rf}}/\gamma$ is met. Here, B_0 is the magnitude of the magnetic field from the tip, plus any externally applied static field that may be present. The

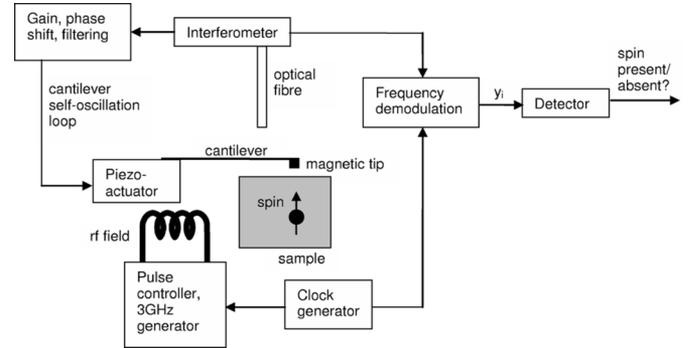


Fig. 1. Schematic of an OSCAR-type MRFM experiment.

constant $\gamma = 5.6\pi \times 10^{10} \text{ T}^{-1}\text{s}^{-1}$ is the gyromagnetic ratio, and ω_{rf} is the (single) frequency of the applied RF field. Because the magnetic field emanating from the tip is highly inhomogeneous, magnetic resonance is confined spatially to a thin bowl-shaped region called the “resonant slice.”

In an OSCAR experiment, a gain-controlled positive feedback loop is used to oscillate the cantilever with a preset amplitude (typically 10–20 nm). The cantilever oscillation frequency is determined by the cantilever itself (specifically, by the fundamental flexural mode eigenfrequency), as well as by tip-sample interactions. As the tip of the cantilever vibrates, the resonant slice passes back and forth through the spin and, as a result, the spin direction is cyclically inverted due to an effect called adiabatic rapid passage [14], [16], [17]. The cyclic inversion is synchronous with the cantilever motion and affects the cantilever dynamics by slightly shifting the cantilever resonant frequency. The frequency shift depends on the angle θ of the spin with respect to a vector called the “effective field in the rotating frame.” See [15] and [16] for further details. The frequency shift can be written as

$$\Delta\omega = \Delta\omega_{\text{max}} \cos\theta \quad (1)$$

where $\Delta\omega_{\text{max}} = 2\omega_0 G\mu/\pi kx_{\text{pk}}$ [15]. Here ω_0 is the unperturbed cantilever frequency, G is the gradient of the magnetic field from the tip as measured at the spin location, μ is the magnetic moment of the spin, k is the cantilever spring constant, and x_{pk} is the peak amplitude of the cantilever vibration. The factor $\cos\theta$ represents the normalized projection of the spin in the direction of the effective field.

There are several impediments to single spin detection. First, because the force from a single spin is so tiny (a few attonewtons), the maximum cantilever frequency shift is only about one part per million for typical experimental parameters. This small frequency shift must be detected in the presence of the cantilever phase (or frequency) noise that originates from cantilever thermal vibrations and sample-induced force fluctuations. The resulting low SNR necessitates long integration times for signal detection. Second, the detection is complicated by environmental disturbance to the spin (i.e., relaxation effects) that can randomly flip the spin orientation and reverse the signal polarity during the signal integration time. A low operating temperature, on the order of 1 K, can help reduce both the cantilever thermal excitations and the random spin

flip rate. Nevertheless, the signal processing methodology must take these effects into account.

We consider two models for the spin behavior in the presence of environmental disturbances. In a quantum mechanical measurement model, the spin is always found to be either aligned or anti-aligned with the effective field, so that $\cos \theta = \pm 1$. Thus the frequency shift signal has only two levels: $\Delta\omega = \pm\Delta\omega_{\max}$, and the time sequence of the frequency shift is a random telegraph signal with a transition rate that depends on the spin relaxation rate. In a classical measurement model (which we consider for the sake of completeness), $\cos \theta$ can take arbitrary values between $+1$ and -1 . As a result of environmental disturbance, $\Delta\omega$ will be uniformly distributed between $-\Delta\omega_{\max}$ and $+\Delta\omega_{\max}$. For this case, a bounded random walk model is appropriate. Recent results strongly favor the random telegraph model [18]. Other publications of interest include [19] and [20].

III. MRFM SIGNAL MODELS

A. Model 1: Discrete-Time Random Telegraph Model

In the quantum measurement model, the frequency shift is characterized by random transitions between two discrete levels. The transition times are taken to be Poisson distributed [18]. Denote the DT random telegraph signal by ζ_i , where $t_i = iT_s$ are the sampling times, and T_s is the sampling time interval. In this paper, a Markovian process with a finite number of states will have a state space denoted by $\Psi = \{\psi_1, \dots, \psi_d\}$, where d is the number of states. Let the state space of the DT random telegraph be Ψ_{rt} ; it has $d = 2$ states and we shall take $\psi_1 = -A$, $\psi_2 = A$, where A is the amplitude of the random telegraph (A corresponds to $\Delta\omega_{\max}$ for the case of an MRFM signal). As an initial condition, ζ_0 is equally likely to be either $\pm A$. Then, a probability transition matrix \mathbf{P}_{rt} can be associated with ζ_i such that the (j, k) th value of \mathbf{P}_{rt} equals $P(\zeta_i = \psi_k | \zeta_{i-1} = \psi_j)$ for $1 \leq j, k \leq 2$, and $i \geq 1$. Assume that \mathbf{P}_{rt} has the form

$$\mathbf{P}_{\text{rt}} = \begin{pmatrix} q & 1-q \\ 1-p & p \end{pmatrix} \quad (2)$$

where $0 < p, q < 1$. If $p = q$, we say that the transition probabilities are *symmetric*, whereas if $p \neq q$, we shall say that they are *asymmetric*. Define the signal vector $\underline{\zeta} = [\zeta_0, \dots, \zeta_{N-1}]^T$, the noise vector $\underline{w} = [w_0, \dots, w_{N-1}]^T$, and the observation vector $\underline{y} = [y_0, \dots, y_{N-1}]^T$, where the superscript $(\cdot)^T$ denotes the transpose operator, and N is the number of observations. The w_i s are modeled as independent identically distributed Gaussian random variables (r.v.) with zero mean and variance σ^2 . The detection problem is then to decide between

$$\begin{aligned} H_0(\text{spin absent}) : & \quad \underline{y} = \underline{w} \\ H_1(\text{spin present}) : & \quad \underline{y} = \underline{\zeta} + \underline{w}. \end{aligned} \quad (3)$$

Let f_i be the density of \underline{y} induced under hypothesis H_i for $i = 0, 1$. Similarly, let $E_i[\cdot]$ and $\text{var}_i(\cdot)$ denote the expectation and variance respectively under hypothesis H_i for $i = 0, 1$.

In this paper, we shall define the SNR of a finite-state DT Markov process ζ_i in AWGN as

$$\text{SNR} \triangleq N \left\{ -\frac{1}{2} \log \frac{2\sqrt{\text{var}_0(y_i)\text{var}_1(y_i)}}{\text{var}_0(y_i) + \text{var}_1(y_i)} + \frac{(E_1[y_i] - E_0[y_i])^2}{4(\text{var}_0(y_i) + \text{var}_1(y_i))} \right\}. \quad (4)$$

Note that the above definition is a function of i if the random process ζ_i is not wide-sense stationary. In that case, we shall take the SNR to be the steady-state value obtained by letting $i \rightarrow \infty$. Then, using (4), the SNR of the DT random telegraph process in AWGN is

$$\text{SNR} = N \left\{ -\frac{1}{2} \log \frac{2\sqrt{1 + \rho_{\text{rt}}^2}}{2 + \rho_{\text{rt}}^2} + \frac{(AC_m/\sigma)^2}{4(2 + \rho_{\text{rt}}^2)} \right\} \quad (5)$$

where $r \triangleq p + q - 1$, $C_m \triangleq (p - q)/(1 - r)$, and $\rho_{\text{rt}}^2 \triangleq (A/\sigma)^2(1 - C_m^2)$. Note that $p, q \in (0, 1)$ implies that $|r| < 1$. The parameter C_m indicates the mismatch between the transition probabilities p and q . The symmetric case when $p = q$ results in $C_m = 0$. For the DT random telegraph, $E[\zeta_i] = AC_m(1 - r^i)$. So $C_m = 0$ implies that the DT random telegraph is zero mean. The parameter ρ_{rt}^2 is the ratio of the steady-state variance of the random telegraph process ζ_i to the noise variance.

Definition (4) is motivated by [21, (3)] with $\alpha = 1/2$, which is related to the error exponent of the optimum detector in a binary hypothesis test. See [21] and [22] for more details. The SNR in decibels is defined in the usual way as $\text{SNR}_{\text{dB}} \triangleq 10 \log_{10} \text{SNR}$.

Reference [5] uses another definition of SNR. Let $\bar{\zeta}_i \triangleq \zeta_i - E[\zeta_i]$, $i \geq 0$, be the mean-corrected version of ζ_i . The -3 dB bandwidth of the random process $\bar{\zeta}_i$ is

$$W = \arccos \left(\frac{4r - 1 - r^2}{2r} \right). \quad (6)$$

In the symmetric case when $p = q$ and $p \approx 1$, (6) is approximately $2\pi(1 - p)$. The -3 dB bandwidth is then proportional to the mean number of transitions per second, which is $(1 - p)/T_s$. The definition of SNR used in [5] is

$$\text{SNR}' \triangleq \frac{\text{Power of random process } \bar{\zeta}_i \text{ in } [-W, W]}{\text{Power of noise in } [-W, W]}. \quad (7)$$

Under definition (7), the SNR in the single electron spin experiment was reported to be -6.7 dB [5]. In this paper, we shall use the SNR definition of (5) for the DT random telegraph. Under symmetric transition probabilities and $|A/\sigma| \ll 1$, the two SNRs are related as follows: $\text{SNR} \propto N(\text{SNR}')^2$. The condition of low SNR for the DT random telegraph will be taken to mean $|A/\sigma| \ll 1$.

Examples of noiseless and noisy random telegraph signals are given in Fig. 2.

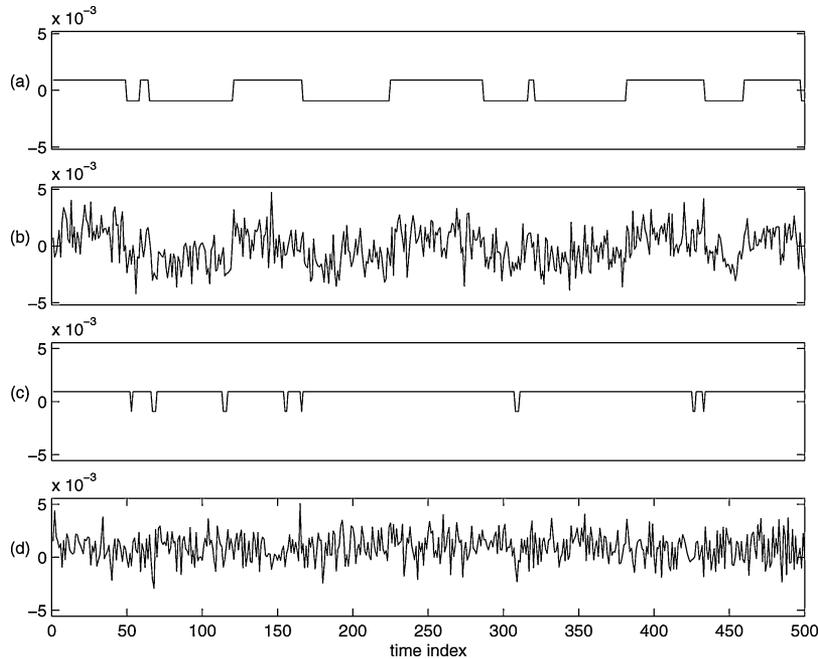


Fig. 2. (a) Noiseless random telegraph signal with symmetric transition probabilities $p = q = 0.98$. (b) Noisy version of (a) at $\text{SNR} = 7.09$ dB. (c) Noiseless random telegraph signal with asymmetric transition probabilities $p = 0.98$, $q = 0.6$. (d) Noisy version of (c) at $\text{SNR} = 13.9$ dB.

B. Model 2: Discrete-Time Random Walk Model

In the classical spin detection model, the frequency shift signal is well approximated by a one dimensional random walk confined to the interval $I = [-A, A]$, where $A = \Delta\omega_{\max}$ for the case of a MRFM signal. We discretize I into $(2M+1)$ states using a step size of s , where $M \in \mathbb{N}$ and $s > 0$, and define ζ_i to be the random walk restricted to the discretized I ; we shall refer to this model as the DT random walk model. The state space Ψ_{rw} of the DT random walk will then have $d = 2M + 1$ states, where $\psi_j = (j - M - 1)s$ for $j = 1, \dots, (2M + 1)$. Associate with ζ_i the probability transition matrix \mathbf{P}_{rw} , so that, as before, the (j, k) th element of \mathbf{P}_{rw} is $P(\zeta_i = \psi_k | \zeta_{i-1} = \psi_j)$ for $1 \leq j, k \leq (2M + 1)$, and $i \geq 1$. \mathbf{P}_{rw} is defined such that, at each time step, ζ_i changes by either $\pm s$. This implies that \mathbf{P}_{rw} is a tridiagonal matrix. We assume reflecting boundary conditions, and ζ_0 is equally likely to be either $\pm s$. The initial condition on ζ_0 was arbitrarily chosen. The regime of interest that we will focus on is large N , and so the effect of the initial condition will not be significant.

The detection problem is now to test (3) when ζ is modeled by a random walk. Note that the DT random walk model can almost be regarded as a multistate generalization of the DT random telegraph model. There are, however, important differences. The DT random walk process cannot remain in the same state for two consecutive time instances. In contrast, it is possible for the DT random telegraph process to do so. Additionally, the DT random walk process has reflecting boundary conditions. The DT random telegraph process does not have this. In the limit as $s \rightarrow 0$, $M \rightarrow \infty$, the random walk converges to Brownian motion over the interval I [23].

Analogous to the DT random telegraph, the condition of low SNR for the DT random walk will be taken to mean $|\psi_i/\sigma| \ll 1$ for $i = 1, \dots, d$. An example of a noiseless and noisy random

walk signal is given in Fig. 3, where, at each state, a change of $\pm s$ is equally likely.

IV. DETECTION STRATEGIES

The detectors considered here can be placed into three categories: versions of existing detectors that are currently in use for MRFM, LRTs, and approximations to the LRT. The LRT is a most powerful test that satisfies the Neyman–Pearson criterion: it maximizes the probability of detection (P_D) subject to a constraint on the probability of false alarm (P_F) [24], which is set by the user. Consequently, it can be used as a benchmark with which to compare the other detectors. When the initial state value, the random transition times, and all subsequent state values are known, the optimal LRT is the matched filter, called the omniscient matched filter (MF) in this paper. Although unimplementable, the MF detector provides an absolute upper bound when comparing the various detectors' receiver operating characteristic (ROC) curves.

A. Amplitude, Energy, Filtered Energy Detectors

The DT amplitude detector is

$$\left| \frac{1}{N} \sum_{i=0}^{N-1} y_i \right| \begin{matrix} \geq H_1 \\ \leq H_0 \end{matrix} \quad (8)$$

where η is set to satisfy the constraint on P_F . This is the optimal test under the assumption that y_i is the sum of an unknown constant and AWGN. This assumption would be true if there were no random spin flips. However, as the number of random transitions in y_i increases, the performance of the amplitude detector degrades. An intuitive explanation can be obtained by considering the detection of the DT random telegraph process and omitting the absolute value bars in the amplitude detector. Under

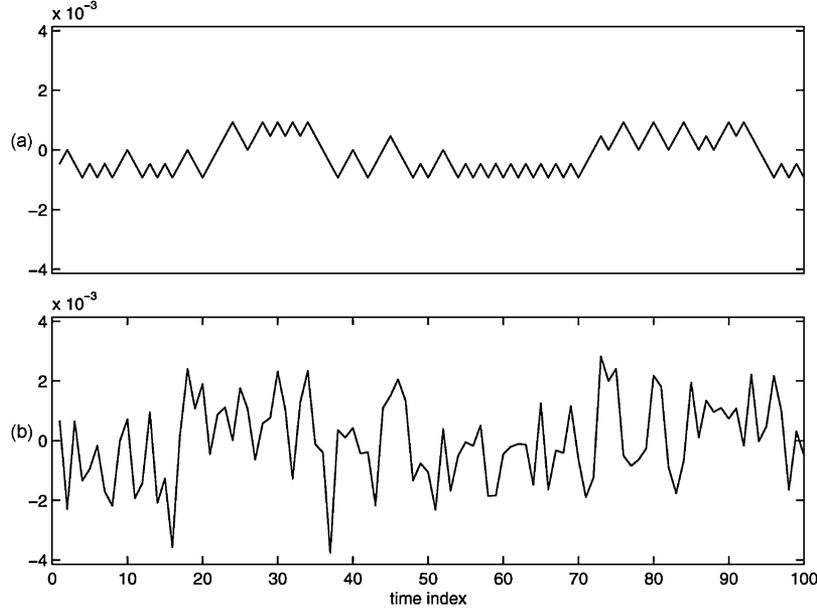


Fig. 3. (a) Noiseless random walk signal with 5 levels. (b) Noisy version of (a) at SNR = -7.33 dB.

H_1 , $\sum_i y_i = \sum_i \zeta_i + \sum_i w_i$. If $\zeta_i = A$ for $i = 0, \dots, N-1$, $E_1[\sum_i y_i] = NA$. In contrast, $E_0[\sum_i y_i] = 0$. However, if ζ_i is equally likely to be $\pm A$, then $E_1[\sum_i y_i] = 0$, which is the same value as $E_0[\sum_i y_i]$. In this case, it would be harder to distinguish between the two hypotheses, as the test statistic has the same expected value under both hypotheses.

An alternative test statistic is the DT signal energy, i.e., the sum of the squares of the y_i instead of the magnitude of the sum in (8). As the signal and noise are assumed to be independent, under hypothesis H_1 , one would expect \underline{y} to have a higher energy on average than under hypothesis H_0 . This can be reliably detected under a sufficiently high SNR. A natural improvement to the energy detector is to reject out-of-band noise by pre-filtering \underline{y} over the signal passband. If the signal ζ_i is baseband, a low-pass filter (LPF) is appropriate. In particular, one might use a simple first-order single-pole filter given by

$$H_{\text{LP}}(z) = \frac{1 - \alpha_{\text{LP}}}{2} \frac{1 + z^{-1}}{1 - \alpha_{\text{LP}} z^{-1}} \quad (9)$$

where we require $|\alpha_{\text{LP}}| < 1$ for stability [25]. The time constant α_{LP} dictates the bandwidth of the LPF. If ω_c is the desired -3 dB bandwidth of the filter, one should set

$$\alpha_{\text{LP}} = \frac{1 - \sin \omega_c}{\cos \omega_c}. \quad (10)$$

The -3 dB bandwidth used depends on the bandwidth of the random process ζ_i . For example, the -3 dB bandwidth of the DT random telegraph is given by (6). Suppose we have symmetric transition probabilities. The CT random telegraph model is typically characterized by the rate parameter λ , which corresponds to the mean number of transitions per second. One can equate the mean number of transitions per second in both DT and CT models to obtain

$$\lambda = \frac{1-p}{T_s}. \quad (11)$$

Since $p > 0$, we require $T_s < \lambda^{-1}$ in order to use (11). In practice, p (or equivalently λ) is only approximately known to the experimenter. As a result, a bank of LPFs with different α_{LP} s is used to perform detection [5].

Let “ $*$ ” be the convolution operator, so that $s = y * h$ is defined to be $s_i \triangleq \sum_k y_k h_{i-k}$. The energy and filtered energy detector can be expressed as

$$\sum_{i=0}^{N-1} (y * h)_i^2 \underset{H_0}{\overset{H_1}{\geq}} \eta \quad (12)$$

where $(y * h)_i$ is taken to be the i th value of $y * h$. For the energy detector, h is taken to be the unit impulse function $\delta[i]$, while for the filtered energy detector $h = h_{\text{LP}}$, the impulse response of $H_{\text{LP}}(z)$ in (9).

Note that the computational complexity for the amplitude, filtered energy, and energy detectors is $\mathcal{O}(N)$.

B. Recursive Equations for the Optimal LRT Detector of a General Finite-State DT Markov Signal in AWGN

In this section, we shall consider the detection of a general finite-state DT Markov process in AWGN and derive the LRT. The formulas that provide an initial starting point are given in [8]. We shall use the notation in [8]: while it is slightly different, the differences are superficial.

The hypothesis test that we consider is (3), where the DT Markov process ζ_i has d possible states. Let $\mathbf{P}^{(i)}$, $i \geq 1$ be the probability transition matrix associated with ζ_i at the i th time step, so that $\mathbf{P}_{jk}^{(i)} = P(X_i = \psi_k | X_{i-1} = \psi_j)$. The noise is denoted by w_i and is independent Gaussian r.v. with mean zero and variance $R_i \in \mathbb{R}$, $i = 0, \dots, N-1$.

Let $y^i \triangleq [y_0, \dots, y_i]^T$ for $i \geq 0$. We shall define $\underline{r}_i \in \mathbb{R}^d$, $i \geq 0$ and $\underline{q}_i \in \mathbb{R}^d$, $i \geq 1$ as

$$\begin{aligned} \underline{r}_i &\triangleq [P(\zeta_i = \psi_1 | y^i), \dots, P(\zeta_i = \psi_d | y^i)]^T \\ \underline{q}_i &\triangleq [P(\zeta_i = \psi_1 | y^{i-1}), \dots, P(\zeta_i = \psi_d | y^{i-1})]^T. \end{aligned} \quad (13)$$

Define $\mathbf{\Omega}^{(i)} \in \mathbb{R}^{d \times d}$, $i \geq 1$ as

$$\mathbf{\Omega}_{jk}^{(i)} \triangleq \begin{cases} \frac{f_1(y_i|\zeta_i=\psi_k, y^{i-1})}{f_1(y_i|y^{i-1})}, & j = k \\ 0, & \text{otherwise} \end{cases} \quad (14)$$

Proposition 1: $\underline{q}_i^T = \underline{r}_{i-1}^T \mathbf{P}^{(i)}$, $i \geq 1$: Examine the j th element of \underline{q}_i . By the Markov assumption

$$\begin{aligned} P(\zeta_i = \psi_j | y^{i-1}) &= \sum_{n=1}^d P(\zeta_i = \psi_j | \zeta_{i-1} = \psi_n, y^{i-1}) \\ &\quad \cdot P(\zeta_{i-1} = \psi_n | y^{i-1}) \\ &= \underline{r}_{i-1}^T \begin{pmatrix} P(\zeta_i = \psi_j | \zeta_{i-1} = \psi_1) \\ \vdots \\ P(\zeta_i = \psi_j | \zeta_{i-1} = \psi_d) \end{pmatrix} \\ &= \underline{r}_{i-1}^T (\text{jth column of } \mathbf{P}^{(i)}). \quad \blacksquare \end{aligned}$$

Proposition 2: $\underline{r}_i^T = \underline{q}_i^T \mathbf{\Omega}^{(i)}$, $i \geq 0$: Using [8, (39)]

$$P(\zeta_i = \psi_j | y^i) = \frac{f_1(y_i | \zeta_i = \psi_j, y^{i-1}) P(\zeta_i = \psi_j | y^{i-1})}{f_1(y_i | y^{i-1})}. \quad \blacksquare$$

Proposition 2 is derived in [8]. From Propositions 1 and 2

$$\underline{q}_i^T = \underline{q}_{i-1}^T \mathbf{\Omega}^{(i-1)} \mathbf{P}^{(i)}. \quad (15)$$

This result is incorrectly given in [8, (38)].

Define $\varphi(x; \mu, \sigma^2) \triangleq \exp[-(x - \mu)^2 / 2\sigma^2] / \sqrt{2\pi\sigma}$, i.e., a Gaussian r.v. with mean μ and variance σ^2 . Let π_k , $k = 1, \dots, d$, denote the initial probability density of ζ_0 and $\underline{\pi} \triangleq [\pi_1, \dots, \pi_d]^T$. Define

$$\underline{n}_i = [\varphi(y_i; \psi_1, R_i), \dots, \varphi(y_i; \psi_d, R_i)]^T. \quad (16)$$

For a vector $\underline{x} \in \mathbb{R}^d$, $\text{diag}(\underline{x})$ is defined to be the d -by- d matrix with \underline{x} along its main diagonal. Since $\mathbf{\Omega}^{(i)} = \text{diag}(\underline{n}_i) / (\underline{q}_i^T \underline{n}_i)$, (15) can be written as

$$\underline{q}_i^T = \underline{q}_{i-1}^T \frac{\text{diag}(\underline{n}_{i-1})}{\underline{q}_{i-1}^T \underline{n}_{i-1}} \mathbf{P}^{(i)}. \quad (17)$$

If we define $\underline{q}_0 \triangleq \underline{\pi}$, the log LRT for (3) is given by

$$\log \Lambda(\underline{y}) = \sum_{i=0}^{N-1} \log \frac{\underline{q}_i^T \underline{n}_i}{\varphi(y_i; 0, R_i)}. \quad (18)$$

Equations (17) and (18) are a recursive way to compute the LRT for a general finite-state DT Markov signal in AWGN. See [22] for more details; note that this result does not appear in [8]. We see that the running time of (17) and (18) is of order $\mathcal{O}(Nd^2)$ for general matrices $\mathbf{P}^{(i)}$. If each $\mathbf{P}^{(i)}$ were tridiagonal, for example, the running time would be $\mathcal{O}(Nd)$.

The detection test would then be

$$\Lambda(\underline{y}) \underset{H_0}{\overset{H_1}{\geq}} \eta. \quad (19)$$

One can also take the log of both sides of (19). As the log function is strictly monotone increasing, the threshold would be altered but the performance of the detector would be unaffected. The threshold η can be determined via simulation, if all of the parameters of the signal and noise models were known. The ROC curve could be generated, and the value of η that corresponded to the desired P_F could be obtained. Alternatively, the ROC curve could be generated via experimentation and the threshold η selected in a similar fashion.

Define the transition likelihood ratio $l(y_i | y^{i-1})$, $i \geq 1$, as $l(y_i | y^{i-1}) \triangleq f_1(y_i | y^{i-1}) / f_0(y_i | y^{i-1})$. Note that

$$\log \Lambda(\underline{y}) = \sum_{i=1}^{N-1} \log l(y_i | y^{i-1}) + \log \left(\frac{f_1(y_0)}{f_0(y_0)} \right). \quad (20)$$

In [9], $\log l(y_i | y^{i-1})$, $i \geq 1$ is given when $P(y_i | y^{i-1}, \zeta^i)$ belongs to a class of exponential functions of the form $K(\zeta^i) \exp[\zeta_i g(y^i) + B(y^i)]$ for functions $K(\cdot), g(\cdot), B(\cdot)$ which give rise to a valid density. The expression for $\log l(y_i | y^{i-1})$ involves the conditional mean estimate

$$\hat{\zeta}_i(y^i) = \int_{\Psi} \zeta_i dP(\zeta_i | y^i) \quad (21)$$

and the function

$$V(y^i) = \int \hat{\zeta}_i(\xi^i) d\xi_i \Big|_{\xi^i = y^i}. \quad (22)$$

For the hypothesis testing problem considered here

$$\hat{\zeta}_i(y^i) = \underline{r}_i^T \mathbf{D}_{\zeta} \underline{1} \quad (23)$$

where $\mathbf{D}_{\zeta} \triangleq \text{diag}(\psi_1, \dots, \psi_d)$. It is possible to apply Propositions 1 and 2 to obtain a recursive equation for \underline{r}_i and solve for a closed-form expression. It would be difficult, however, to evaluate (22).

C. Approximation of the LRT Under Low SNR

Let us consider the log LRT under low SNR. Each transition likelihood ratio can be simplified as follows: for $i \geq 1$

$$\begin{aligned} l(y_i | y^{i-1}) &= \sum_{n=1}^d P(\zeta_i = \psi_n | y^{i-1}) \cdot \exp \left[\frac{2y_i \psi_n - \psi_n^2}{2R_i} \right] \\ &\approx \sum_{n=1}^d P(\zeta_i = \psi_n | y^{i-1}) \left(1 + \frac{1}{R_i} y_i \psi_n - \frac{1}{2R_i} \psi_n^2 \right) \\ &= 1 + \frac{1}{R_i} y_i E_1[\zeta_i | y^{i-1}] - \frac{1}{2R_i} E_1[\zeta_i^2 | y^{i-1}] \quad (24) \end{aligned}$$

where the approximation $e^\delta \approx 1 + \delta$ for small δ was used. Next, we shall use the approximation $\log(1 + \delta) \approx \delta$ for small δ . This is justified if the SNR is low so that $|\psi_n / \sqrt{R_i}| \ll 1$ for all $n = 1, \dots, d$. So

$$\log l(y_i | y^{i-1}) \approx \frac{1}{R_i} y_i E_1[\zeta_i | y^{i-1}] - \frac{1}{2R_i} E_1[\zeta_i^2 | y^{i-1}]. \quad (25)$$

The same approximation can be applied to $\log[f_1(y_0)/f_0(y_0)]$, so that

$$\log\left(\frac{f_1(y_0)}{f_0(y_0)}\right) \approx \frac{1}{R_0}y_0E[\zeta_0] - \frac{1}{2R_0}E[\zeta_0^2]. \quad (26)$$

Define the conditional mean (also MMSE) predictor of ζ_i under H_1 as follows: $\hat{\zeta}_i = E_1[\zeta_i|y^{i-1}]$ for $i \geq 1$ and $\hat{\zeta}_0 = E[\zeta_0]$. Use a similar notation for ζ_i^2 , so that $\hat{\zeta}_i^2 = E_1[\zeta_i^2|y^{i-1}]$ for $i \geq 1$ and $\hat{\zeta}_0^2 = E[\zeta_0^2]$. Using (20), (25), and (26), the log LRT can be approximately written under low SNR as

$$\log \Lambda(\underline{y}) \approx \sum_{i=0}^{N-1} \frac{1}{R_i} y_i \hat{\zeta}_i - \frac{1}{2} \sum_{i=0}^{N-1} \frac{1}{R_i} \hat{\zeta}_i^2 \quad (27)$$

The right-hand side (RHS) of (27) is similar to the matched filter statistic, but with the MMSE predictors of ζ_i and ζ_i^2 used instead of the known values. Note that the conditional mean estimate of ζ_i used in (27) is different from that defined in (21). Schwartz's version in (21) includes the observation y_i , whereas ours does not. This makes our estimate of ζ_i a one-step predictor.

In [8], it is shown that in detecting a finite-state DT Markov signal, the LRT is in general not expressible as the known form LRT with an estimator of ζ_i used, i.e., the RHS of (27) but with $(E_1[\zeta_i|y^{i-1}])^2$ used instead of $E_1[\zeta_i^2|y^{i-1}]$ for $i \geq 1$.

D. Comparison to the CT Analog

Consider the CT analog of the hypothesis test problem (3) when the noise variances R_i are all equal. One has to decide between the following two hypotheses:

$$\begin{aligned} H_0: & \dot{y}(t) = \dot{w}(t), \quad t \in [0, T] \\ H_1: & \dot{y}(t) = \zeta(t) + \dot{w}(t), \quad t \in [0, T]. \end{aligned} \quad (28)$$

Define $I \triangleq [0, T]$. Here, $\zeta(t)$ is a random process (not necessarily finite state) such that $\int_I E[\zeta^2(t)]dt < \infty$ and $\dot{w}(t)$ is AWGN with

$$E[\dot{w}(t)] = 0, \quad E[\dot{w}(t)\dot{w}(s)] = \sigma^2\delta(t-s). \quad (29)$$

The LRT is given by [10]–[12]

$$\exp\left(\frac{1}{\sigma^2} \int_I \hat{\zeta}_1(t)y(t)dt - \frac{1}{2\sigma^2} \int_I \hat{\zeta}_1^2(t)dt\right) \quad (30)$$

where $\hat{\zeta}_1(t) = E_1[\zeta(t)|\{y(\xi), \xi < t\}]$ is the conditional mean estimate under hypothesis H_1 given the previous observations, and the first integral in (30) is an Itô stochastic integral.

There are three noteworthy differences between (27) and (30). First, (27) approximately holds only under low SNR, whereas (30) is exact under all SNR conditions. Secondly, $\zeta(t)$ is not constrained to be a Markov process, whereas its DT counterpart ζ_i is Markovian. Lastly, in the second term of (27), the expected value of ζ_i^2 conditioned on the past observations is used. On the other hand, the square of the expected value of $\zeta(t)$ conditioned on the past observations is used in the CT version. In general, for $i \geq 1$, $E_1[\zeta_i^2|y^{i-1}] \neq (E_1[\zeta_i|y^{i-1}])^2$. Indeed, for an r.v. X

$$E[X^2] = (E[X])^2 \quad \text{iff} \quad \text{var}(X) = 0 \quad (31)$$

By the Chebyshev inequality, for $\delta > 0$, $P[|X - E[X]| \geq \delta] \leq \text{var}(X)/\delta^2 = 0$. So (31) holds iff X is some value $c \in \mathbb{R}$ with probability (w.p.) 1. As a result $E_1[\zeta_i^2|y^{i-1}] = (E_1[\zeta_i|y^{i-1}])^2$ iff ζ_i is a function of y^{i-1} w.p. 1.

E. Application to the Detection of the DT Random Telegraph Process

Under the regime of low SNR, long observation time ($N \gg 1$), and $r = p + q - 1 \approx 1$ (the probability of transition between consecutive samples is small), the second-order expansion of $\log \Lambda_{\text{rt}}(\underline{y})$ is approximately equal to the hybrid detector with the test statistic

$$\sum_i (y * h_{\text{LP}})_i^2 + K_a \sum_i y_i + K_e \sum_i y_i^2 \quad (32)$$

where the constants $K_a = K_a(p, q, A, \sigma)$ and $K_e = K_e(p, q)$ are given in (59) of the Appendix and $\alpha_{\text{LP}} = p + q - 1$. Therefore, in the aforementioned regime, one expects the hybrid detector to have performance similar to the optimal LRT. When $p = q$, the second-order expansion of the LRT is approximately equal to the filtered energy detector, which is given by (12) with $h = h_{\text{LP}}$ and $\alpha_{\text{LP}} = 2p - 1$. See the Appendix for more details.

The previous result did not make use of the approximation under low SNR obtained in Section IV-C. Equation (27), however, provides a general approach to deriving an approximation to the LRT for a finite-state DT Markov signal in AWGN. We shall illustrate by deriving an approximation to the LRT for the DT random telegraph using this general approach.

Let us specialize the result of Section IV-C to the DT random telegraph processes, so that $R_i = \sigma^2$, $i = 0, \dots, N - 1$ and $\mathbf{P}^{(i)} = \mathbf{P}_{\text{rt}}$, $i = 1, \dots, N - 1$. Since $E[\zeta_0^2] = A^2$ and $E_1[\zeta_i^2|y^{i-1}] = A^2$ for $i \geq 1$, the second term of (27) will be a constant and can be omitted. Then

$$\log \Lambda_{\text{rt}}(\underline{y}) \approx \frac{1}{\sigma^2} \sum_{i=0}^{N-1} y_i \hat{\zeta}_i \quad (33)$$

under low SNR conditions. We see that the log LRT for the random telegraph is an estimator-correlator detector. The estimator-correlator structure is known to be optimal for detecting Gaussian signals in AWGN [9].

Now, the conditional mean predictor $\hat{\zeta}_i = E_1[\zeta_i|y^{i-1}]$, $i \geq 1$ is a function of y_0, \dots, y_{i-1} . Its exact form will be dictated by the conditional probability mass function $P(\zeta_i|y^{i-1})$ and will be nonlinear in general. Suppose that we would like to find the predictor of ζ_i with smallest MSE that lies in the linear span of $\{1, y_0, \dots, y_{i-1}\}$. Let $\hat{\zeta}_i^A = c + \sum_{n=0}^{i-1} \gamma_n y_n$ be this predictor. It is also known as the best affine estimator of ζ_i given y_0, \dots, y_{i-1} . One can apply the projection theorem [26] to obtain the well-known result that

$$\hat{\zeta}_i^A = E[\zeta_i] + \text{cov}_1(y^{i-1}, \zeta_i) \mathbf{\Gamma}_{i-1}^{-1} (y^{i-1} - E_1[y^{i-1}]) \quad (34)$$

for $i \geq 1$, where $\text{cov}_1(a, b) = E_1[ab^T] - E_1[a]E_1[b]^T$ is the covariance of the random vectors a, b under hypothesis H_1 and $\mathbf{\Gamma}_{i-1} \triangleq \text{cov}_1(y^{i-1}, y^{i-1})$ is the covariance matrix of y^{i-1} under hypothesis H_1 . The MSE achieved with $\hat{\zeta}_i^A$ cannot be smaller than the MSE achieved with $\hat{\zeta}_i = E_1[\zeta_i|y^{i-1}]$, as the

conditional mean achieves the smallest MSE out of all possible estimators.

For $0 \leq n < i$, $\text{cov}_1(y_n, \zeta_i) = \text{cov}(\zeta_n, \zeta_i)$ since w_n is independent of ζ_i . As well, $E_1[y^{i-1}] = E[\zeta^{i-1} + w^{i-1}] = E[\zeta^{i-1}]$ and $\mathbf{\Gamma}_{i-1} = \text{cov}_1(y^{i-1}, y^{i-1}) = \sigma^2 \mathbf{I} + \text{cov}(\zeta^{i-1}, \zeta^{i-1})$. Substituting these into (34) results in

$$\hat{\zeta}_i^A = E[\zeta_i] + \text{cov}(\zeta^{i-1}, \zeta_i) [\sigma^2 \mathbf{I} + \text{cov}(\zeta^{i-1}, \zeta^{i-1})]^{-1} \cdot (y^{i-1} - E[\zeta^{i-1}]), \quad i \geq 1. \quad (35)$$

It remains to compute the various quantities in (35). Recalling that $r = p + q - 1$ and $C_m = (p - q)/(1 - r)$

$$\begin{aligned} E[\zeta_i] &= AC_m(1 - r^i), \quad i \geq 0 & (36) \\ \text{cov}(\zeta_i, \zeta_j) &= 4A^2 \frac{(1-p)(1-q)}{(1-r)^2} r^{|i-j|} \\ &\quad + A^2 C_m^2 \left(2r^{\max(i,j)} - r^{i+j} \right), \\ &\quad i, j \geq 0. & (37) \end{aligned}$$

Using (36) and (37) in (35) and plugging the resulting expression of $\hat{\zeta}_i^A$ into (33), one obtains

$$\log \Lambda_{\text{rt}}(\underline{y}) \approx \frac{1}{\sigma^2} \sum_{i=0}^{N-1} y_i \hat{\zeta}_i^A \quad (38)$$

and we shall take $\hat{\zeta}_0^A = E[\zeta_0] = 0$.

Unfortunately, (38) is not in a form that can be compared to the approximation derived previously without the benefit of (37). In order to do this, we have to explicitly evaluate (35). As the inversion of a matrix without any special structure is required, this will be difficult. Make the approximation that

$$\text{cov}(\zeta_i, \zeta_j) \approx 4A^2 \frac{(1-p)(1-q)}{(1-r)^2} r^{|i-j|}. \quad (39)$$

This will be a good approximation when $\max(i, j)$ is large. Moreover, since we are assuming a low SNR condition, $|A/\sigma| \ll 1$, so

$$\sigma^2 \mathbf{I} + \text{cov}(\zeta^{i-1}, \zeta^{i-1}) \approx \sigma^2 \mathbf{I}. \quad (40)$$

With (39) and (40)

$$\begin{aligned} \hat{\zeta}_i^A &\approx AC_m \left[1 - r^i - C_f \left(\frac{r(1 - r^{i+1})}{1 - r} - ir^i \right) \right] \\ &\quad + C_f \sum_{n=0}^{i-1} r^{i-n} y_n, \quad i \geq 1 \end{aligned} \quad (41)$$

where $C_f \triangleq 4(1-p)(1-q)A^2/[\sigma^2(1-r)^2]$.

As $i \rightarrow \infty$, $ir^i \rightarrow 0$ as $|r| < 1$, and $(1 - r^{i+1})/(1 - r) \rightarrow 1/(1 - r)$. Apply these to (41); then, using the subsequent result in (38)

$$\log \Lambda_{\text{rt}}(\underline{y}) \approx L_1 + L_2 \quad (42)$$

$$\text{where } L_1 = \frac{A}{\sigma^2} C_m \sum_{i=1}^{N-1} \left(1 - r^i - C_f \frac{r}{1 - r} \right) y_i \quad (43)$$

$$L_2 = \frac{1}{\sigma^2} C_f \sum_{i=1}^{N-1} \sum_{n=0}^{i-1} r^{i-n} y_i y_n. \quad (44)$$

The expressions (42)–(44) are similar to those in (49)–(51).

F. General Methodology for Obtaining an Approximation to the LRT of a Finite State DT Markov Signal in AWGN Under Low SNR

In the previous section, the affine predictor of ζ_i , $i \geq 1$, with the smallest MSE was used in place of the conditional mean $E_1[\zeta_i | y^{i-1}]$, $i \geq 1$, in order to obtain an approximation to the LRT of the DT random telegraph. This was obtained by finding the estimator of ζ_i with smallest MSE that lay in the linear span of $\{1, y_0, \dots, y_{i-1}\}$. The best affine predictor of ζ_i is necessarily suboptimal, as the conditional mean achieves the lowest MSE out of all estimators.

For the general case when the second term of (27) is present, one could also find a suboptimal predictor of ζ_i^2 , $i \geq 1$, in terms of the previous observations y_0, \dots, y_{i-1} . Following the idea with regard to $\hat{\zeta}_i$, a suboptimal predictor can be obtained by finding the estimator of ζ_i^2 with smallest MSE that lies in the linear span of $\{1, y_0^2, \dots, y_{i-1}^2\}$.

The suboptimal predictors of ζ_i and ζ_i^2 can then be used in (27).

V. SIMULATION RESULTS

The objective in this section is to compare the detection methods discussed in the previous section. The class of LRT detectors is optimal for their respective signal models and provides a good comparison benchmark. Comparison of the various detectors is done using 1) ROC curves, each of which is a plot of P_D versus P_F , and 2) power curves, each of which is a plot of P_D versus SNR at a fixed P_F . Recall that P_D is the probability of detection and P_F is the probability of false alarm. To generate each ROC curve, 20 simulations were generated. Then, the average and the error bars of one standard deviation were plotted. In a similar fashion, the data plotted in each power curve are the average over the 20 simulations at each SNR value, along with error bars of one standard deviation.

Some of the parameters used in the simulation of the DT random telegraph and random walk models are as follows: $k = 10^{-3} \text{ Nm}^{-1}$, $\omega_0 = 2\pi \times 10^4 \text{ rad s}^{-1}$, $G = 2 \times 10^6 \text{ T m}^{-1}$. The sampling period was $T_s = 1 \text{ ms}$, and signal durations of $T = 60 \text{ s}$ and $T = 150 \text{ s}$ were used. The amplitude of the RF field was $B_1 = 0.2 \text{ mT}$. The performance of the detectors varies as a function of T ; in general, a larger T results in better performance. Values of T used in iOSCAR MRFM experiments are on the order of tens of hours [5]. Nevertheless, the comparative results obtained from using the two values of T above are representative of larger values. Indeed, our approximations to the optimal detectors improve with increased T .

A. Discrete-Time Random Telegraph Model

First, consider the DT random telegraph. Fig. 4 depicts the simulated ROC curves at $\text{SNR} = -34.3 \text{ dB}$, $\lambda = 0.5 \text{ s}^{-1}$, and with symmetric transition probabilities ($p = q$). With $T_s = 1 \text{ ms}$, this results in $p = q = 0.9995$. We examine the matched filter, DT random telegraph LRT (RT-LRT), filtered energy, hybrid [given by (32)], amplitude, and unfiltered energy detectors. The RT-LRT, filtered energy, and hybrid detector curves are virtually identical, which is consistent with our analysis. The unfiltered energy and amplitude detectors have performance that

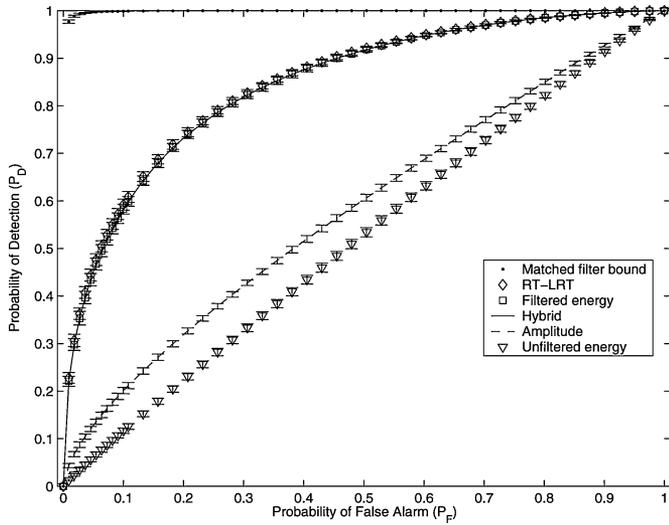


Fig. 4. Simulated ROC curves (P_D versus P_F) for the DT random telegraph model with symmetric transition probabilities at SNR = -34.3 dB, $T = 60$ s, and $\lambda = 0.5$ s $^{-1}$ for the omniscient matched filter, DT random telegraph LRT (RT-LRT), filtered energy, hybrid, amplitude, and unfiltered energy detectors. The RT-LRT is theoretically optimal.

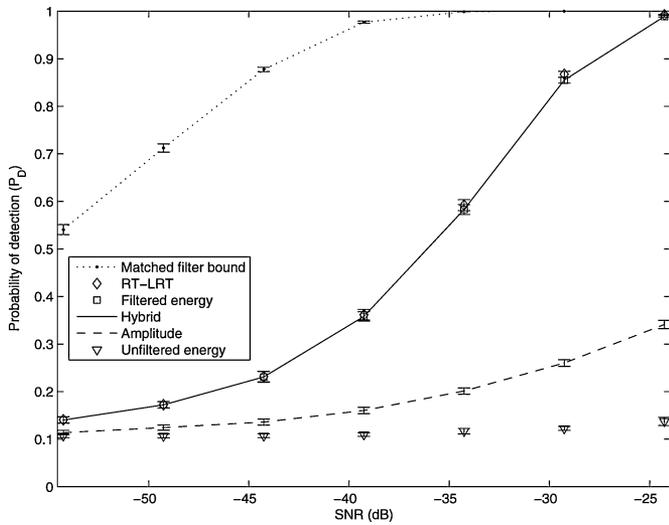


Fig. 5. Simulated power curves (P_D versus SNR) for the DT random telegraph model with P_F fixed at 0.1 and $\lambda = 0.5$ s $^{-1}$, $T = 60$ s. The RT-LRT is theoretically optimal.

is poorer than the RT-LRT, as it should be since the RT-LRT is the optimal detector. The unfiltered energy detector has the worst performance out of the five detector methods considered. Lastly, the omniscient MF detector has the best performance.

A power curve was generated over a range of SNR's under the same conditions as before with a fixed $P_F = 0.1$; it is illustrated in Fig. 5. For spin detection, an acceptable range for P_F is on the order of 0.05 to 0.1. The RT-LRT, filtered energy, and hybrid detector have similar performance from -25 to -55 dB. With this particular value of P_F and λ , the RT-LRT, filtered energy, and hybrid detector perform from 10 to 20 dB worse than the MF detector. Although the amplitude detector has worse performance than the RT-LRT and filtered energy detector, all three have comparable performance at -55 dB.

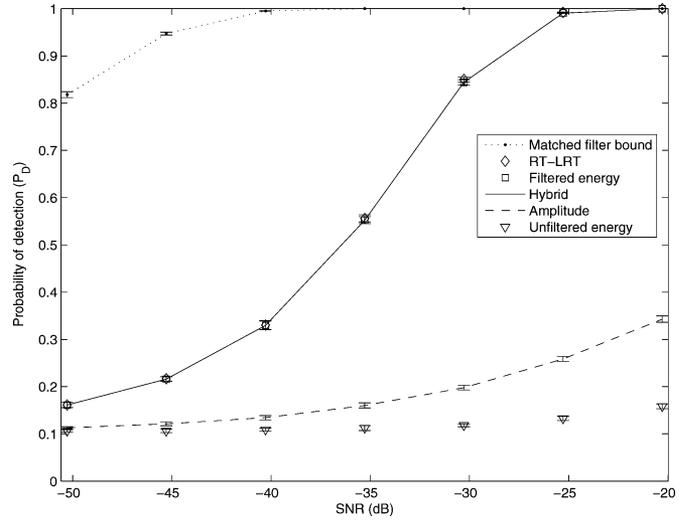


Fig. 6. Simulated power curves for the DT random telegraph model with P_F fixed at 0.1 and $\lambda = 0.5$ s $^{-1}$, $T = 150$ s. The RT-LRT is theoretically optimal.

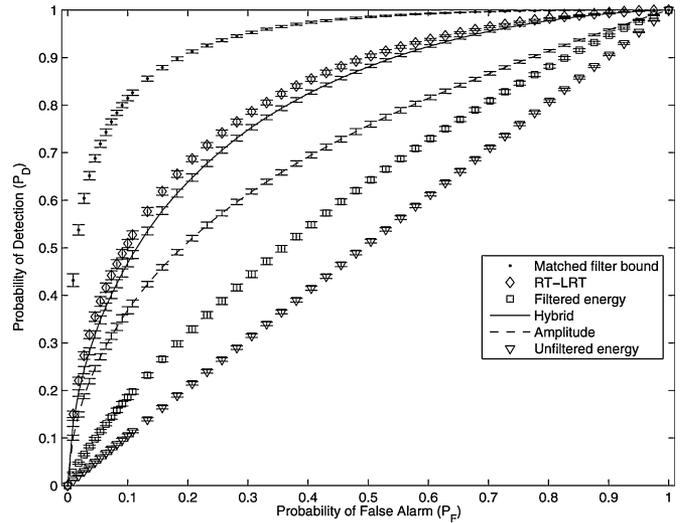


Fig. 7. Simulated ROC curves for the DT random telegraph model with asymmetric transition probabilities ($p = 0.9998$, $q = 0.9992$) at SNR = -6.71 dB, $T = 150$ s. The RT-LRT is theoretically optimal.

Fig. 6 shows the power curves generated using the bigger value of $T = 150$ s. The RT-LRT, filtered energy, and hybrid detectors have the same performance from -20 to -50 dB. Note that the definition of SNR we use scales with $N = T/T_s$. As a result, a larger T increases the SNR. It is intuitively pleasing that the same SNR results in similar P_D values in Figs. 5 and 6.

The ROC and power curve simulations were repeated with different values of λ , and the same relative performance was observed. In the interest of space, however, they will not be shown. Note that performance degrades as λT_s increases. From (11), the probability of transition between consecutive time samples is $1 - p = \lambda T_s$. A higher value of λT_s results in a higher probability of transition, which decreases performance.

In the second set of simulations, we investigate the case in which the transition probabilities are asymmetric, i.e., $p \neq q$. Consider the scenario where $p = 0.9998$, $q = 0.9992$, and

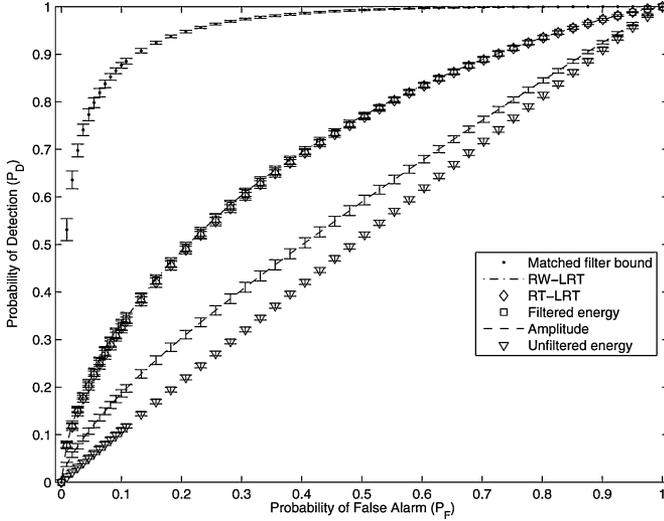


Fig. 9. Simulated ROC curves for the DT symmetric random walk $p_{l,1} = p_{l,2} = p_{u,1} = p_{u,2} = 0.5$ at SNR = -44.0 dB, $T = 60$ s for the matched filter, RW-LRT, RT-LRT, filtered energy, amplitude, and unfiltered energy detector. The RW-LRT is theoretically optimal.

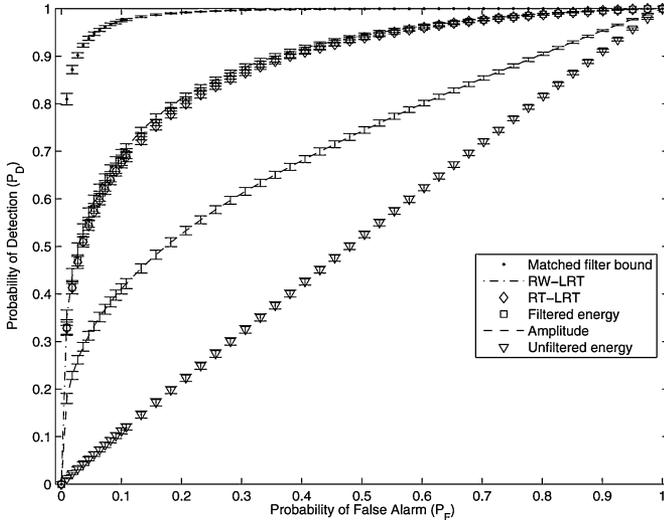


Fig. 10. Simulated ROC curves for the DT symmetric random walk $p_{l,1} = p_{u,2} = 0.52$, $p_{l,2} = p_{u,1} = 0.48$ at SNR = -39.1 dB, $T = 60$ s. The RW-LRT is theoretically optimal.

electron spin. This is strong evidence that the random telegraph signal model accurately describes the cantilever-single spin interaction.

The results of this paper lend strong theoretical and practical support to the use of the simple filtered energy detector for the current MRFM single spin research community. It has been shown that the existing baseband filtered energy detector that is in current use is approximately optimal in the case of the symmetric DT random telegraph model under the regime of low SNR, long observation time, and p close to one. The last condition can be achieved by sampling at a sufficiently fast rate as compared to the rate of random transitions. This result has been extended to the case of the asymmetric DT random telegraph by using a hybrid filtered energy/amplitude/energy detector. Simulations were presented showing that the near optimality of the

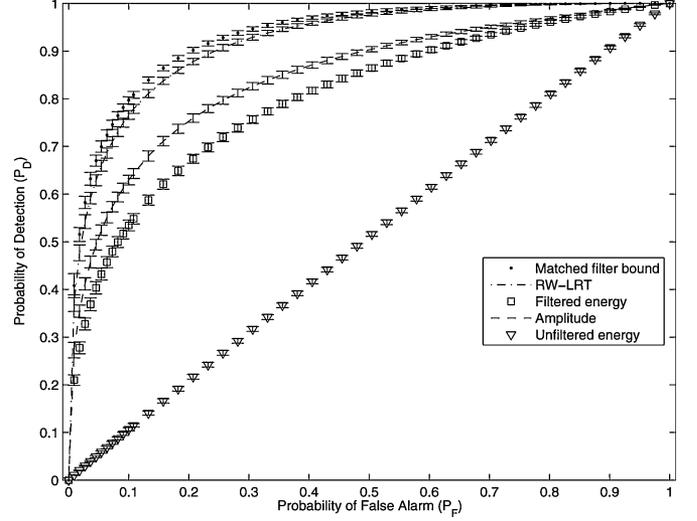


Fig. 11. Simulated ROC curves for the DT asymmetric random walk $p_{l,1} = p_{u,1} = 0.45$, $p_{l,2} = p_{u,2} = 0.55$ at SNR = -2.50 dB, $T = 60$ s. The RW-LRT is theoretically optimal.

baseband filtered energy detector extends to the case of the symmetric DT random walk model. In the case of the asymmetric DT random walk, the filtered energy detector does not perform as well as the optimal LRT. We expect that a hybrid detector along the lines of that formulated for the DT random telegraph will perform close to the optimal for the asymmetric DT random walk. Mathematical analysis of the DT random walk model will be presented in a future paper.

A new interpretation of the LRT for a finite-state DT Markov signal in AWGN under low SNR conditions was presented. Specifically, the LRT is approximately the matched filter statistic with the MMSE predictor signal values used in place of the known signal values.

The previous result can be used to obtain an approximation to the LRT for a general finite-state DT Markov signal in AWGN under low SNR conditions. For ease of computation, we have proposed the use of affine estimators as suboptimal versions of the conditional mean predictors in (27). In particular, the suboptimal estimators of ζ_i and ζ_i^2 for $i \geq 1$ can be computed as the affine estimators with lowest MSE that lie in the linear span of $\{1, y_0, \dots, y_{i-1}\}$ and $\{1, y_0^2, \dots, y_{i-1}^2\}$, respectively. We apply this methodology to compute an approximation to the LRT for the DT random telegraph process in AWGN. The approximation matches the result previously obtained using a straightforward analysis.

APPENDIX

Define $q_i(u)$, $u \in \Psi_{rt}$ to be the element of q_i that corresponds to the state u . From (17) and (18), one can obtain the LRT for the DT random telegraph as

$$\Lambda_{rt}(\underline{y}) = \prod_{i=0}^{N-1} \left[q_i(A) e^{\frac{A}{\sigma^2} y_i} + q_i(-A) e^{-\frac{A}{\sigma^2} y_i} \right]. \quad (46)$$

Let $f(y_0, \dots, y_{N-1})$ denote the log LRT function of the DT random telegraph, i.e., the log of (46). Let $g(y_0, \dots, y_{N-1})$ be

the filtered energy detector function in (12). Let us analyze the two functions f and g under the regime of low SNR ($|A/\sigma| \ll 1$) and long observation time ($N \gg 1$).

We want to obtain the approximate Taylor series expansion of f about $\underline{y} = \underline{0}$ and compare that with g . Define

$$\theta_i \triangleq \frac{q_i(A)e^{\frac{A}{\sigma^2}y_i}}{q_i(A)e^{\frac{A}{\sigma^2}y_i} + q_i(-A)e^{-\frac{A}{\sigma^2}y_i}}$$

for $i \geq 0$. A recursive equation for θ_i can be derived based on (17). Its approximate solution is

$$\begin{aligned} \theta_i &\approx \beta_i + \frac{qA}{\sigma^2} \sum_{j=0}^i \xi_{ij} y_j, \quad i \geq 0 \\ \text{where } \beta_i &= \frac{1-q}{1-r} + \left(\frac{1}{2} - \frac{1-q}{1-r}\right) r^i, \quad i \geq 0 \\ \xi_{ij} &= \frac{2(1-q)r^{i-j} + (2q-r-1)r^i}{1-r}, \\ &\quad 0 \leq j \leq i-1 \\ \xi_{ii} &= \frac{2(1-q)}{1-r} + \frac{r^i(2q-r-1)}{1-r} = 2\beta_i, \\ &\quad i \geq 0 \end{aligned} \quad (47)$$

and $r = p + q - 1$. Define $s_i \triangleq (A/\sigma^2)y_i$. Then

$$f \approx \sum_i \left\{ \left[s_i(2q_i(A) - 1) + \frac{1}{2}s_i^2 \right] - \frac{1}{2} \left[s_i(2q_i(A) - 1) + \frac{1}{2}s_i^2 \right]^2 \right\}. \quad (48)$$

By solving for $q_i(A)$ in terms of θ_i and using (47), one obtains the approximate Taylor series expansion of f as

$$f \approx L_1 + L_{2a} + L_{2b} + \text{h.o.t.} \quad (49)$$

with

$$L_1 = \frac{A}{\sigma^2} C_m \sum_i (1-r^i) y_i \quad (50)$$

$$L_{2a} = 2q \left(\frac{A}{\sigma^2} \right)^2 \sum_i \sum_{j=0}^{i-1} \left[\frac{2(1-q)}{1-r} r^{i-j} - r^i C_m \right] \times y_i y_j \quad (51)$$

$$L_{2b} = \left(\frac{A}{\sigma^2} \right)^2 \sum_i \left\{ 4r \left(\frac{1-q}{1-r} \right)^2 + 2 \frac{(q-r)(1-q)}{(1-r)^2} - C_m(2q + C_m)r^i + \frac{1}{2} C_m^2 r^{2i} \right\} y_i^2. \quad (52)$$

Recall that $C_m = (p-q)/(1-r)$. In (49), ‘‘h.o.t.’’ denotes the higher order terms; specifically, terms of degree three or higher.

Let f_{sym} be the function f under symmetric transition probabilities, i.e., $p = q$. Then

$$f_{\text{sym}} \approx 2p \left(\frac{A}{\sigma^2} \right)^2 \left\{ \sum_{i=1}^{N-1} \sum_{j=0}^{i-1} (2p-1)^{i-j} y_i y_j + \sum_{i=0}^{N-1} \left(1 - \frac{1}{4p} \right) y_i^2 \right\}. \quad (53)$$

For sufficiently large N , it can be shown that

$$g \approx D \left\{ \sum_{i=1}^{N-1} \sum_{j=0}^{i-1} \alpha_{\text{LP}}^{i-j} y_i y_j + \frac{\alpha_{\text{LP}}}{1 + \alpha_{\text{LP}}} \sum_{i=0}^{N-1} y_i^2 \right\} \quad (54)$$

where $D = (1 - \alpha_{\text{LP}}^2)/2\alpha_{\text{LP}}$ is a constant. Note that D plays no role in the performance of the test statistic. Indeed, the detection test

$$\begin{aligned} &H_1 \\ &g \geq \eta \\ &H_0 \end{aligned}$$

has the same performance as

$$\begin{aligned} &H_1 \\ &\gamma g \geq \gamma \eta \\ &H_0 \end{aligned}$$

for a constant $\gamma \in \mathbb{R}_+$ not dependent on the observations.

In order to compare the performance of f_{sym} and g , we consider their normalized versions $(1/2p)(A/\sigma^2)^{-2} f_{\text{sym}}$ and $(1/D)g$. Denote these statistics by \tilde{f}_{sym} and \tilde{g} , respectively. Both \tilde{f}_{sym} and \tilde{g} are a weighted sum of two terms: an energy term of the form $\sum_i y_i^2$ and a second-order term of the form $\sum_{j < i} \gamma^{i-j} y_i y_j$, where $\gamma = 2p - 1$ for \tilde{f}_{sym} and $\gamma = \alpha_{\text{LP}}$ for \tilde{g} . If $\alpha_{\text{LP}} = 2p - 1$, then

$$|\tilde{f}_{\text{sym}} - \tilde{g}| \approx \frac{1}{4p} \sum_i y_i^2. \quad (55)$$

Now, $E_1[\sum_{i=0}^{N-1} y_i^2] - E_0[\sum_{i=0}^{N-1} y_i^2] = A^2 N$. On the other hand, for large N

$$E_1 \left[\sum_{i=1}^{N-1} \sum_{j=0}^{i-1} \alpha_{\text{LP}}^{i-j} y_i y_j \right] - E_0 \left[\sum_{i=1}^{N-1} \sum_{j=0}^{i-1} \alpha_{\text{LP}}^{i-j} y_i y_j \right] \approx GA^2 N \quad (56)$$

where $G = \alpha_{\text{LP}}(2p-1)/[1 - \alpha_{\text{LP}}(2p-1)]$. When $\alpha_{\text{LP}} = 2p - 1$, $G = (2p-1)^2/[1 - (2p-1)^2] = (4(1-p))^{-1} + (4p)^{-1} - 1$. For p close to one, $G \gg (4p)^{-1}$ and $GA^2 N \gg (4p)^{-1} A^2 N$. So to the first moment, the difference of $(1/4p) \sum_i y_i^2$ between \tilde{f}_{sym} and \tilde{g} does not represent a significant difference when $p \approx 1$. Under these conditions, we expect the performance of the filtered energy detector and the DT random telegraph LRT to be similar.

It is possible to obtain an approximation to the DT random telegraph LRT that holds when we make no assumption about p being equal to q . When $p \neq q$, we have $C_m \neq 0$, and there are terms of the form $r^i C_m$ and $r^{2i} C_m^2$ in (50)–(52). Since $|r| < 1$, $r^i \rightarrow 0$ in the limit as $i \rightarrow \infty$. So drop these terms to get

$$f \approx C \left\{ \frac{(p-q)\sigma^2}{4q(1-r)A} \sum_i y_i + \sum_i \sum_{j<i} r^{i-j} y_i y_j + \left[\frac{1}{2} + \frac{r(1-q)}{2q(1-r)} \right] \sum_i y_i^2 \right\} \quad (57)$$

where $C = 4q(1-q)(A/\sigma^2)^2/(1-r)$ is a constant. Define $C_a \triangleq (p-q)\sigma^2/[4q(1-r)A]$ and $C_e \triangleq r(1-q)/[2q(1-r)]$. In order to equate the coefficients of the cross-terms $y_i y_j$ between (57) and (54), we require $\alpha_{LP} = r = p + q - 1$. In g , the ratio of the energy terms to the cross-terms is $\alpha_{LP}/(1 + \alpha_{LP})$. For $r = \alpha_{LP} \approx 1 \Rightarrow \alpha_{LP}/(1 + \alpha_{LP}) \approx 1/2$. The idea is to add the energy and amplitude statistics to g so that all three statistics are in the same ratio as in (57). Let g_{hyb} be the “extended” version of g , which we shall call the hybrid filtered energy/amplitude/energy detector

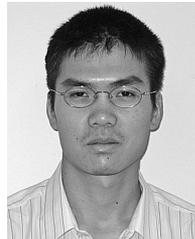
$$g_{\text{hyb}} \triangleq g + \frac{1 - \alpha_{LP}^2}{2\alpha_{LP}} \left[C_a \sum_i y_i + C_e \sum_i y_i^2 \right]. \quad (58)$$

We expect g_{hyb} to perform similar to f under the conditions of large N , low SNR, and $r \approx 1$. The constants in (58) can be further simplified. Let $K_a \triangleq C_a(1 - \alpha_{LP}^2)/2\alpha_{LP}$ and $K_b \triangleq C_e(1 - \alpha_{LP}^2)/2\alpha_{LP}$. As $\alpha_{LP} = p + q - 1$, one obtains after some algebra

$$K_a = \frac{p^2 - q^2}{8q(p+q-1)} \left(\frac{A}{\sigma^2} \right)^{-1} \quad \text{and} \quad K_e = \frac{(p+q)(1-q)}{4q}. \quad (59)$$

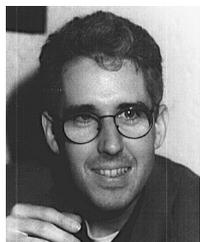
REFERENCES

- [1] K. Chun and N. O. Birge, “Dissipative quantum tunneling of a single defect in a disordered metal,” *Phys. Rev. B*, vol. 54, no. 7, pp. 4629–4637, 1996.
- [2] L. J. Vandamme, A. V. Belyakov, M. Y. Perov, and A. V. Yakimov, “Difference in dependence of $1/f$ and RTS noise on current in quantum dots light emitting diodes,” in *Proc. SPIE Noise Devices Circuits*, vol. 5113, 2003, pp. 368–378.
- [3] S. Jha, K. Tan, and R. A. Maxion, “Markov chains, classifiers, and intrusion detection,” in *Proc. 14th IEEE Comput. Security Foundations Workshop*, 2001, pp. 206–219.
- [4] P. D. Gader, M. Mystkowski, and Y. Zhao, “Landmine detection with ground penetrating radar using hidden Markov models,” *IEEE Trans. Geosci. Remote Sensing*, vol. 39, no. 6, pp. 1231–1244, 2001.
- [5] D. Rugar, R. Budakian, H. J. Mamin, and B. W. Chui, “Single spin detection by magnetic resonance force microscopy,” *Nature*, vol. 430, no. 6997, pp. 329–332, 2004.
- [6] R. Budakian, H. J. Mamin, B. W. Chui, and D. Rugar, “Creating order from random fluctuations in small spin ensembles,” *Science*, vol. 307, no. 5708, pp. 408–411, 2005.
- [7] G. P. Berman, G. D. Doolen, P. C. Hammel, and V. I. Tsifrinovich, “Solid-state nuclear-spin quantum computer based on magnetic resonance force microscopy,” *Phys. Rev. B*, vol. 61, no. 21, pp. 14 694–14 699, 2000.
- [8] L. L. Scharf and L. W. Nolte, “Likelihood ratios for sequential hypothesis testing on Markov sequences,” *IEEE Trans. Inf. Theory*, vol. IT-23, no. 1, pp. 101–109, 1977.
- [9] S. C. Schwartz, “The estimator-correlator for discrete-time problems,” *IEEE Trans. Inf. Theory*, vol. IT-23, no. 1, pp. 93–100, 1977.
- [10] T. Kailath, “A general likelihood-ratio formula for random signals in Gaussian noise,” *IEEE Trans. Inf. Theory*, vol. IT-15, no. 3, pp. 350–361, 1969.
- [11] —, “A further note on a general likelihood formula for random signals in Gaussian noise,” *IEEE Trans. Inf. Theory*, vol. IT-16, no. 4, pp. 393–396, 1970.
- [12] T. Kailath and M. Zakai, “Absolute continuity and Radon-Nikodym derivatives for certain measures relative to Wiener measure,” *Ann. Math. Statist.*, vol. 42, no. 1, pp. 130–140, 1971.
- [13] Y.-C. Yao, “Estimation of noisy telegraph processes: Nonlinear filtering versus nonlinear smoothing,” *IEEE Trans. Inf. Theory*, vol. IT-31, no. 3, pp. 444–446, 1985.
- [14] B. C. Stipe, H. J. Mamin, C. S. Yannoni, T. D. Stowe, T. W. Kenny, and D. Rugar, “Electron spin relaxation near a micron-size ferromagnet,” *Phys. Rev. Lett.*, vol. 87, no. 27, pp. 277 602/1–277 602/4, 2001.
- [15] H. J. Mamin, R. Budakian, B. W. Chui, and D. Rugar, “Detection and manipulation of statistical polarization in small spin ensembles,” *Phys. Rev. Lett.*, vol. 91, no. 20, pp. 207 604/1–207 604/4, 2003.
- [16] C. P. Slichter, *Principles of Magnetic Resonance*, 3rd ed. Berlin, Germany: Springer-Verlag, 1989.
- [17] K. Wago, D. Botkin, C. S. Yannoni, and D. Rugar, “Force-detected electron-spin resonance: adiabatic inversion, nutation, and spin echo,” *Phys. Rev. B*, vol. 57, no. 2, pp. 1108–1114, 1998.
- [18] G. P. Berman, F. Borgonovi, and V. I. Tsifrinovich, “A model for quantum jumps in magnetic resonance force microscopy,” *Phys. Lett. A*, vol. 337, no. 3, pp. 161–165, 2005.
- [19] —, “Wave function collapses in a single spin magnetic resonance force microscopy,” *Phys. Lett. A*, vol. 331, no. 3–4, pp. 187–192, 2004.
- [20] G. P. Berman, F. Borgonovi, V. N. Gorshkov, and V. I. Tsifrinovich, “Modeling and simulations of a single-spin measurement using MRFM,” *IEEE Trans. Nanotechnol.*, vol. 4, no. 1, pp. 14–20, 2005.
- [21] A. O. Hero, B. Ma, O. J. J. Michel, and J. Gorman, “Applications of entropic spanning graphs,” *IEEE Signal Process. Mag.*, vol. 19, no. 5, pp. 85–95, 2002.
- [22] M. Ting, “Signal processing for magnetic resonance force microscopy,” Univ. of Michigan, Ann Arbor, 2005.
- [23] H. Stark and J. W. Woods, *Probability, Random Processes, and Estimation Theory for Engineers*, 2nd ed. Englewood Cliffs, NJ: Prentice-Hall, 1994.
- [24] H. L. Van Trees, *Detection, Estimation, and Modulation Theory*. New York: Wiley, 1968, vol. 1.
- [25] S. K. Mitra, *Digital Signal Processing: A Computer-Based Approach*, 2nd ed. New York: McGraw-Hill, 2001.
- [26] D. G. Luenberger, *Optimization by Vector Space Methods*. New York: Wiley, 1969.



Michael Ting (S’99) received the B.A.Sc. degree (first-class honors) in computer engineering from the University of Waterloo, Waterloo, ON, Canada, in 2001 and the M.S.E. degree in electrical engineering from the University of Michigan, Ann Arbor, in 2003, where he is currently pursuing the Ph.D. degree.

His research interests lie in detection, estimation theory, and inverse problems.



Alfred O. Hero, III (S'79–M'80–SM'96–F'98) received the B.S. degree (*summa cum laude*) from Boston University, Boston, MA, in 1980 and the Ph.D. degree from Princeton University, Princeton, NJ, in 1984, both in electrical engineering.

Since 1984 he has been with the University of Michigan, Ann Arbor, where he is a Professor in the Department of Electrical Engineering and Computer Science and, by courtesy, in the Department of Biomedical Engineering and the Department of Statistics. His recent research interests have been in

areas including inference on sensor networks, bioinformatics, and statistical signal and image processing.

Prof. Hero has received an IEEE Signal Processing Society Meritorious Service Award (1998), an IEEE Signal Processing Society Best Paper Award (1998), and the IEEE Third Millennium Medal (2000). He is currently President-Elect of the IEEE Signal Processing Society (2004–2005) and Associate Editor of the IEEE TRANSACTIONS ON COMPUTATIONAL BIOLOGY AND BIOINFORMATICS (2004–2006).



Daniel Rugar (M'87–SM'05) received the B.A. degree in physics from Pomona College, Claremont, CA, in 1975 and the Ph.D. degree in applied physics from Stanford University, Stanford, CA, in 1982.

He joined the IBM Research Division in 1984 and has worked on ultra-high-density data storage and new approaches to scanning probe microscopy. His present position is Manager of nanoscale studies, leading a group investigating magnetic resonance force microscopy, detection of ultrascale forces, and applications of nanomechanics. He has published

more than 100 papers and has received 19 patents.

Dr. Rugar was the 1999–2000 Distinguished Lecturer for the IEEE Magnetic Society. He has received IBM internal awards for work on magnetic force microscopy, magnetic resonance force microscopy, near-field optical data storage, and ultra-high-density data storage based on the atomic force microscope.



Chun-Yu Yip (S'03) received the B.S. degree in electrical engineering from Cornell University, Ithaca, NY, in 2001. He is currently pursuing the Ph.D. degree in the Department of Electrical Engineering and Computer Science in University of Michigan, Ann Arbor.

His current research interests include methodology of functional magnetic resonance imaging and application of neuroimaging to cognitive psychology.

Mr. Yip is a member of the International Society of Magnetic Resonance in Medicine.



Jeffrey A. Fessler (S'83–M'90–SM'00–F'06) received the Ph.D. degree in electrical engineering from Stanford University, Stanford, CA, in 1990.

He has since been with the University of Michigan, first as a U.S. Department of Energy Alexander Hollaender Postdoctoral Fellow and then as an Assistant Professor in the Division of Nuclear Medicine. Since 1995, he has been with the Electrical Engineering and Computer Science Department, the Biomedical Engineering Department, and the Radiology Department. His research interests are in statistical aspects

of imaging problems. He has supervised doctoral research in PET, SPECT, X-ray CT, MRI, and optical imaging.