REPORT DOCUMENTATION PAGE			*Form Approved OMB No. 0704-0188*

1. REPORT DATE *(DD-MM-YYYY)*	**2. TYPE REPORT**	**3. DATES COVERED** *(From - To)*
20-08-2007	Journal Article	01 JAN 2007

4. TITLE AND SUBTITLE	**5a. CONTRACT NUMBER**
Importance Sampling for Characterizing STAP Detectors	**5b. GRANT NUMBER**
	5c. PROGRAM ELEMENT NUMBER 61102F
6. AUTHOR(S)	**5d. PROJECT NUMBER** 2311
Rajan Srinivasan, Muralidhar Rangaswamy	**5e. TASK NUMBER** HE
	5f. WORK UNIT NUMBER 02

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)

AFRL/SNHE
80 Scott Drive
Hanscom AFB MA 01731

*Telecommunication Engineering Group,
University of Twente
PO 217, 7500 AE Enschede
The Netherlands

8. PERFORMING ORGANIZATION REPORT

9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES)

Electromagnetics Technology Division SC: 437490
Sensors Directorate
Air Force Research Laboratory
80 Scott Drive
Hanscom AFB MA 01731-2909

10. SPONSOR/MONITOR'S ACRONYM(S)

AFRL-SN-HS

11. SPONSOR/MONITOR'S REPORT NUMBER(S)

AFRL-SN-HS-TP-2007-0007

12. DISTRIBUTION / AVAILABILITY STATEMENT

DISTRIBUTION A. Approved for public release; distribution unlimited.

13. SUPPLEMENTARY NOTES

Supported by European Office of Aerospace Research and Development (EOARD) under contract number FA8655-04-1-3059 for AFRL/SNHE, IN-HOUSE work unit 2311HE02. Published in IEEE Transactions on Aerospace and Electronic Systems, Vol. 43. No. 1 January 2007. Cleared for Public Release by ESC/PA: ESC 06-0672 dtd 7 Jun 06

14. ABSTRACT

This paper describes the development of adaptive importance sampling (IS) techniques for estimating false alarm probabilities of detectors that use space-time adaptive processing (STAP) algorithms. Fast simulation using IS methods has been notably successful in the study of conventional constant false alarm rate (CFAR) radar detectors, and in several other applications. The principal objectives here are to examine the viability of using these methods for STAP detectors, develop them into powerful analysis and design algorithms and, in the long term, use them for synthesizing novel detection structures. The adaptive matched filter (AMF) detector has been analyzed successfully using fast simulation. Of two biasing methods considered, one is implemented and shown to yield good results. The important problem of detector threshold determination is also addressed, with matching outcome. As an illustration of the power of these methods, two variants of the square-law AMF detector that are thought to be robust under heterogeneous clutter conditions have also been successfully investigated. These are the envelope-law and geometric-mean STAP detectors. Their CFAR property is established and performance evaluated. It turns out the variants have detection performances better than those of the AFF detector for training data contaminated by interferers. In summary, the work reported

15. SUBJECT TERMS

AFM, CFAR, Importance Sampling, G-Method

16. SECURITY CLASSIFICATION OF:			**17.LIMITATION OF ABSTRACT**	**18.NUMBER OF PAGES**	**19a. NAME OF RESPONSIBLE PERSON** Muralidhar Rangaswamy
a. REPORT	**b. ABSTRACT**	**c. THIS PAGE**			**19b. TELEPHONE NUMBER** *(include area code)*
Unclassified	Unclassified	Unclassified	UU	15	N/A

Standard Form 298 (Rev. 8-98)
Prescribed by ANSI Std. Z39.18

Importance Sampling for Characterizing STAP Detectors

RAJAN SRINIVASAN
University of Twente
The Netherlands

MURALIDHAR RANGASWAMY
Air Force Research Laboratory Sensors Directorate

This paper describes the development of adaptive importance sampling (IS) techniques for estimating false alarm probabilities of detectors that use space-time adaptive processing (STAP) algorithms. Fast simulation using IS methods has been notably successful in the study of conventional constant false alarm rate (CFAR) radar detectors, and in several other applications. The principal objectives here are to examine the viability of using these methods for STAP detectors, develop them into powerful analysis and design algorithms and, in the long term, use them for synthesizing novel detection structures. The adaptive matched filter (AMF) detector has been analyzed successfully using fast simulation. Of two biasing methods considered, one is implemented and shown to yield good results. The important problem of detector threshold determination is also addressed, with matching outcome. As an illustration of the power of these methods, two variants of the square-law AMF detector that are thought to be robust under heterogeneous clutter conditions have also been successfully investigated. These are the envelope-law and geometric-mean STAP detectors. Their CFAR property is established and performance evaluated. It turns out the variants have detection performances better than those of the AMF detector for training data contaminated by interferers. In summary, the work reported here paves the way for development of advanced estimation techniques that can facilitate design of powerful and robust detection algorithms.

Manuscript received July 25, 2005; revised January 24, 2006; released for publication April 25, 2006.

IEEE Log No. T-AES/43/1/895034.

Refereeing of this contribution was handled by E. S. Chornoboy.

Professor Srinivasan was supported by the European Office of Aerospace Research and Development, London, UK, in collaboration with the Air Force Office of Scientific Research (AFOSR) under Award FA8655-04-1-3025. Dr. Rangaswamy was supported by AFOSR under Project 2304 as well as by in-house research efforts at the Air Force Research Laboratory.

Portions of this paper were presented at the IEEE International Radar Conference, Washington, D.C., May 2005.

Authors' addresses: R. Srinivasan, Telecommunication Engineering Group, University of Twente, PO 217, 7500 AE Enschede, The Netherlands; M. Rangaswamy, Air Force Research Laboratory Sensors Directorate, Hanscom AFB, MA, E-mail: (Muralidhar.Rangaswamy@hanscom.af.mil).

I. INTRODUCTION

Estimation of false alarm probabilities (FAPs) of detection algorithms that employ space-time processing is examined here using forced Monte Carlo (MC) or importance sampling (IS) simulation. Space-time adaptive processing (STAP) algorithms are of much importance for radar detection [1, 2]. They are notoriously intensive from a computational point of view, with the more advanced (and robust) ones being also analytically difficult to quantify. Therefore, it is appropriate to attempt to develop fast simulation methods that could be used in their analysis and design.

In the following we use lessons learned from developing IS techniques for characterizing conventional constant false alarm rate (CFAR) detectors and describe an experiment in applying them to STAP detection. The starting point of this effort is the celebrated adaptive matched filter (AMF) detector derived in [2] and which represents the array version of the workhorse cell-averaging (CA) CFAR detector for conventional radar signal processing algorithms. The FAP performance of the AMF detector is known in integral form under certain conditions and can be numerically computed to any desired accuracy. Thus it forms a suitable basis for validating our simulation experiments. Two specific IS methods (described in the sequel) are presented and the better (and also easier) one is implemented. As a demonstration of the power of IS methods, the envelope-law and geometric-mean detectors are presented and analyzed using fast simulation. These are variants of the AMF detector and their FAPs are not known in analytical form. Their CFAR property is established and FAP behaviour characterized using IS. A brief comparison of detection performances of all 3 detectors is made. In the long term, our aim is to develop and apply these fast simulation techniques to modern STAP detection algorithms (see [3]–[5], and references therein). An important goal in this context is to devise IS biasing methods that result in simulation times that grow slowly with decreasing FAPs. This remains an open problem.

As well known now, IS is the chief simulation methodology for rare-event estimation. It is an enduring method that has had success in several areas of science and engineering, [6]. Briefly, IS works by biasing original probability distributions in ways that accelerate the occurrences of rare events, conducting simulations with these new distributions, and then compensating the obtained results for the changes made. The principal consequence is that unbiased probability estimates with low variances are obtained quickly. The main task in IS therefore is determination of good simulation distributions for an application.

Simulations performed using such distributions can provide enormous speed-ups and, if applied successfully, simulation lengths needed to estimate very low probabilities can become (only) weakly dependent on the actual probabilities. When these probabilities satisfy a large deviations principle, then several asymptotic results are available for devising simulation distributions [6]. For most detection applications however, it appears that adaptive methods which attempt to minimize error variances might be better suited. There have been some recent attempts in the literature, for example [7] and [8], to apply IS for FAP estimation of CFAR detectors with varying degrees of success. Our work uses results developed in [9] and [10], with an adaptive implementation of the so called g-method. Recent theoretical and application work on this method can be found in [11] and [12]. The groundwork for the present results was laid down in [15] which, to our knowledge, is the first article on the topic of using IS for studying STAP detection algorithms.

During the conducting of simulations reported herein, some issues concerning the adaptive IS algorithms used arise, and these are discussed briefly. The positive outcome of the methods used is that excellent match with numerical results is obtained. The succeeding sections provide a brief statement of the detector tests, how IS biasing can be performed to hasten false alarm events, description of the g-method which is a conditional IS technique developed originally for studying sums of random variables [9], how inverse IS can be used to estimate (and choose) detector thresholds, a methodology for using the fast adaptive algorithms, simulation results, and a conclusion.

II. STAP DETECTOR VARIANTS

In a radar system consisting of a linear array of N_s antenna elements a burst of N_t pulses is transmitted, resulting in a return of $N_s N_t = N$ samples in each range gate. The N complex samples are arranged in an $N \times 1$ column vector and there are $L + 1$ such returns. The range gate return to be tested for the presence of target is called the primary data and is denoted by $\mathbf{x}$ while the remaining vectors are known as the training (or secondary) data, assumed to be free of target signal, and denoted by $\mathbf{x}(l)$, $l = 1,\ldots,L$. A target return is modelled as consisting of an $N \times 1$ steering vector $\mathbf{s}$ with an unknown complex amplitude in addition to clutter, interference, and noise. The primary and secondary data vectors are assumed to be jointly independent and zero-mean, spherically symmetric complex Gaussian, sharing the $N \times N$ covariance matrix $\mathbf{R} = E\{\mathbf{X}\mathbf{X}^\dagger\}$, where the superscript $\dagger$ denotes complex conjugate transpose.

A. AMF (Square-Law) STAP Detector

Under these assumptions the AMF detection test, as obtained in [2], is given by

$$\frac{|\mathbf{s}^\dagger \hat{\mathbf{R}}^{-1}\mathbf{x}|^2}{\mathbf{s}^\dagger \hat{\mathbf{R}}^{-1}\mathbf{s}} \underset{H_0}{\overset{H_1}{\gtrless}} \eta \tag{1}$$

where

$$\hat{\mathbf{R}} \equiv \frac{1}{L}\sum_{l=1}^{L} \mathbf{x}(l)\mathbf{x}(l)^\dagger$$

is the estimated covariance matrix of $\mathbf{x}$ based on the secondary data (also referred to as sample matrix), and η is a threshold used to set the FAP at some desired level. This test has the CFAR property. The FAP α of the test is known to be given by

$$\alpha = \frac{L!}{(L-N+1)!(N-2)!}\int_0^1 \frac{x^{L-N+1}(1-x)^{N-2}}{(1+\eta x/L)^{L-N+1}}dx \tag{2}$$

which can be used to numerically determine the threshold setting for a desired FAP. As shown in [2], the test in (1) can be rewritten as

$$|\mathbf{s}^\dagger \hat{\mathbf{R}}^{-1}\mathbf{x}|^2 \underset{H_0}{\overset{H_1}{\gtrless}} \eta \mathbf{s}^\dagger \hat{\mathbf{R}}^{-1}\mathbf{s}$$
$$= \frac{\eta}{L}\sum_{l=1}^{L} |\mathbf{s}^\dagger \hat{\mathbf{R}}^{-1}\mathbf{x}(l)|^2. \tag{3}$$

This is in the form of a vector (or array) version of the usual CA-CFAR test. The left-hand side (LHS) is a square-law detector, being the output of a matched filter (matched to the direction $\mathbf{s}$ in which the array is steered) for incoherent detection using the so-called sample matrix inversion (SMI) beamformer weights $\hat{\mathbf{R}}^{-1}\mathbf{s}$. The right-hand side (RHS) represents a cell-averaging term. Further details on these issues can be found in the references mentioned above.

B. AMF (Envelope-Law) STAP Detector

In contrast to square-law detectors, it is known to radar engineers that envelope detectors possess some robustness properties in terms of detection performance when the training data is contaminated by outliers or inhomogeneities. Accordingly, the envelope-law STAP version of the AMF (abbreviated hereafter as E-AMF) detector is proposed here and its detection properties evaluated. It takes the form

$$|\mathbf{s}^\dagger \hat{\mathbf{R}}^{-1}\mathbf{x}| \underset{H_0}{\overset{H_1}{\gtrless}} \frac{\eta_E}{L}\sum_{l=1}^{L} |\mathbf{s}^\dagger \hat{\mathbf{R}}^{-1}\mathbf{x}(l)| \tag{4}$$

where η_E denotes the threshold multiplier. An analytical expression or approximation for the FAP of this detector is not known.

C. Geometric-Mean STAP Detector

Also proposed here is the geometric-mean STAP version of the AMF detector (abbreviated as GM-STAP). Conventional CFAR detectors that calculate the geometric means of the range cells in the CFAR window (usually referred to as log-t detectors) are known to have robustness properties. The STAP variant[1] takes the form

$$|\mathbf{s}^\dagger \hat{\mathbf{R}}^{-1}\mathbf{x}| \underset{H_0}{\overset{H_1}{\gtrless}} \eta_G \left(\prod_{l=1}^{L} |\mathbf{s}^\dagger \hat{\mathbf{R}}^{-1}\mathbf{x}(l)|\right)^{1/L} \tag{5}$$

where η_G denotes the threshold multiplier. Note that the GM-based square- and envelope-law versions are identical (except for a trivial squaring of threshold multiplier). For the sake of completeness, the value of the multiplier as the number of training vectors $L \to \infty$ is calculated. In such a case, $\hat{\mathbf{R}}^{-1} \xrightarrow{P} \mathbf{R}^{-1}$, $\mathbf{s}^\dagger \hat{\mathbf{R}}^{-1}\mathbf{x} \xrightarrow{D} \mathbf{s}^\dagger \mathbf{R}^{-1}\mathbf{x}$, and $\mathbf{s}^\dagger \hat{\mathbf{R}}^{-1}\mathbf{x}(l) \xrightarrow{D} \mathbf{s}^\dagger \mathbf{R}^{-1}\mathbf{x}(l)$ in the absence of target. As $\mathbf{s}^\dagger \mathbf{R}^{-1}\mathbf{x}$ and $\mathbf{s}^\dagger \mathbf{R}^{-1}\mathbf{x}(l)$ are independent and identically distributed (IID) and distributed as $\mathcal{CN}_1(0, \mathbf{s}^\dagger \mathbf{R}^{-1}\mathbf{s})$, it follows that the squared envelopes $|\mathbf{s}^\dagger \mathbf{R}^{-1}\mathbf{x}|^2$ and $|\mathbf{s}^\dagger \mathbf{R}^{-1}\mathbf{x}(l)|^2$ are exponentially distributed, each with mean $\mathbf{s}^\dagger \mathbf{R}^{-1}\mathbf{s}$. Using convergence arguments based on continuity [14], it is straightforward to show that the asymptotic FAP of the GM detector is $\exp(-\eta_G^2 \exp(-\gamma))$ where γ is the Euler constant. Hence the (asymptotic) threshold η_G required to provide a FAP of 10^{-6} is 4.9605. A similar calculation can be easily made for the E-AMF detector.

Both the E-AMF and GM-STAP detectors have the CFAR property under the assumption of homogeneous Gaussian characterization as stated in the beginning of this section; their FAPs are independent of the true target-free covariance $\mathbf{R}$. This is shown in Appendix A.

III. FAP ESTIMATION USING IS

This section describes procedures to quickly estimate the FAPs and threshold multipliers of STAP detectors using IS. The standard (square-law) AMF detector is used for exposition. The detector variants are handled using parallel arguments.

The actual form of biasing considered here is simple scaling. While other techniques (superior in terms of simulation gains) can be devised, scaling is easy to implement and yields conservative but reasonable results.

[1]This test was actually suggested in [16] but its performance was not evaluated.

A. Square-Law AMF Detection

Two strategies to estimate FAPs using IS are two-dimensional (2-D) biasing and the conditional g-method procedure, described in this section. The 2-D biasing scheme parallels the approach used in previous works on conventional CFAR algorithms, as cited in the introduction. It can be useful in situations wherein the more powerful g-method might be difficult to apply.

1) *2-D Biasing*: To estimate FAP using IS, we make the following observations. Suppose each complex sample of a secondary vector is scaled by a real (and positive) number $\theta^{1/2}$. This has the effect of scaling the covariance matrix estimate $\hat{\mathbf{R}}$ by θ. Therefore, as far as the covariance estimate is concerned, both sides of the test in (3) remain unaffected by the scaling. However, each secondary vector being scaled by $\theta^{1/2}$ results in an additional scaling of the RHS by θ. Hence choosing θ less than unity will have the effect of compressing the density function of the random threshold of the test. Further, a scaling of each complex component of the primary vector by a real positive $a^{1/2}$ will achieve a scaling of the LHS of the test by a. Thus, choosing a larger and θ smaller than unity will achieve an increase in the frequency of occurrence of false alarm events during simulation. The IS optimization problem will be a two-parameter one.

Denoting by $\mathcal{A}$ the false alarm event in (3), the unbiased IS estimator in an IID simulation can be expressed as

$$\hat{\alpha} = \frac{1}{K}\sum_{i=1}^{K}[1(\mathcal{A})W(\mathbf{x},\mathbf{x}_L;a,\theta)]^{(i)}, \qquad \mathbf{x},\mathbf{x}_L \sim f_\star \tag{6}$$

where K is the length (or number of trials) of the IS simulation, $1(\cdot)$ denotes the indicator, $\mathbf{X}_L \equiv (\mathbf{x}(1),\ldots,\mathbf{x}(L))^T$, and the notation $\sim f_\star$ means that all random variables are drawn from biased distributions. The weighting function W is described below. In setting up the joint densities of $\mathbf{X}$ and $\mathbf{X}_L$, we use the fact that the FAP of the AMF has the CFAR property and is independent of the true covariance matrix $\mathbf{R}$. This is so under the assumption of Gaussian distributions for the data. In such a case, the simulation of the AMF test can be carried out for data possessing a diagonal covariance matrix $\mathbf{I}$, denoting the $N \times N$ identity matrix. Therefore, primary and secondary data can be generated as complex vectors with independent components. The unbiased joint densities are

$$f(\mathbf{x}) = \frac{e^{-\mathbf{x}^\dagger\mathbf{x}}}{\pi^N} \qquad \text{and} \qquad f(\mathbf{x}_L) = \frac{e^{-\sum_1^L \mathbf{x}(l)^\dagger\mathbf{x}(l)}}{\pi^{LN}}$$

so that

$$f(\mathbf{x},\mathbf{x}_L) = \frac{e^{-\mathbf{x}^\dagger\mathbf{x}-\sum_1^L \mathbf{x}(l)^\dagger\mathbf{x}(l)}}{\pi^{(L+1)N}}.$$

With scaling performed as described above, the biased joint density takes the form

$$f_\star(\mathbf{x},\mathbf{x}_L) = \frac{e^{-(1/a)\mathbf{x}^\dagger\mathbf{x}-(1/\theta)\sum_1^L \mathbf{x}(l)^\dagger\mathbf{x}(l)}}{\pi^{(L+1)N}a^N\theta^{LN}}$$

resulting in the weighting function

$$\begin{aligned} W(\mathbf{x},\mathbf{x}_L;a,\theta) &\equiv \frac{f(\mathbf{x},\mathbf{x}_L)}{f_\star(\mathbf{x},\mathbf{x}_L)} \\ &= Ca^N\theta^{LN}e^{A/a}e^{B/\theta} \end{aligned} \tag{7}$$

where

$$A \equiv \mathbf{x}^\dagger\mathbf{x}, \qquad B \equiv \sum_1^L \mathbf{x}(l)^\dagger\mathbf{x}(l) \qquad \text{and} \qquad C \equiv e^{-(A+B)}. \tag{8}$$

The variance of the IS estimator $\hat{\alpha}$ can be expressed as

$$\operatorname{var}\hat{\alpha} = \frac{1}{K}[I(\nu) - \alpha^2] \tag{9}$$

with ν denoting the vector biasing parameter $(a,\theta)^T \in [1,\infty)\times(0,1]$. The quantity I is given by

$$\begin{aligned} I(\nu) &= E_\star\{1(\mathcal{A})W^2(\mathbf{x},\mathbf{x}_L;\nu)\} \\ &= E\{1(\mathcal{A})W(\mathbf{x},\mathbf{x}_L;\nu)\} \end{aligned} \tag{10}$$

where the expectation $E_\star$ proceeds over biased distributions. Minimization of $\operatorname{var}\hat{\alpha}$ with respect to the biasing parameters is equivalent to minimization of I and is described in Appendix B.

2) *The g-Method Estimator*: This method exploits knowledge of underlying (input) distributions more effectively, yielding a more powerful estimator. Additional advantages are that only a scalar parameter optimization problem needs to be tackled and the inverse IS problem (for threshold optimization or selection) can be easily solved. The FAP can be written as

$$\begin{aligned} \alpha &= P(\mathcal{A}) \\ &= E\{P(|\mathbf{s}^\dagger\hat{\mathbf{R}}^{-1}\mathbf{x}|^2 > \eta\mathbf{s}^\dagger\hat{\mathbf{R}}^{-1}\mathbf{s} \mid \mathbf{X}_L,H_0)\} \\ &\triangleq E\{g(\mathbf{X}_L)\} \end{aligned} \tag{11}$$

where $g(\mathbf{X}_L)$ denotes the conditional probability in the second step. The conditioning implies that the covariance matrix estimate $\hat{\mathbf{R}}$ is given. We proceed to estimate α using the form in the third step above.

With the conditioning in mind it is easy to show, assuming that $\mathbf{x}$ is rotationally invariant and Gaussian, that the random variable $\mathbf{s}^\dagger\hat{\mathbf{R}}^{-1}\mathbf{x} \triangleq \mathbf{w}^\dagger\mathbf{x}$ is distributed as $\mathcal{CN}_1(0,\mathbf{w}^\dagger\mathbf{R}\mathbf{w})$ with independent real and imaginary components, and the weight vector $\mathbf{w} = \hat{\mathbf{R}}^{-1}\mathbf{s}$. Hence the random variable $Y \triangleq |\mathbf{s}^\dagger\hat{\mathbf{R}}^{-1}\mathbf{x}|^2$ is exponential and has the density function

$$f(y \mid \mathbf{X}_L,H_0) = \frac{e^{-y/\mathbf{w}^\dagger\mathbf{R}\mathbf{w}}}{\mathbf{w}^\dagger\mathbf{R}\mathbf{w}}, \qquad y \geq 0.$$

Therefore

$$\begin{aligned} g(\mathbf{X}_L) &= P(Y \geq \eta\mathbf{s}^\dagger\hat{\mathbf{R}}^{-1}\mathbf{s} \mid \mathbf{X}_L,H_0) \\ &= e^{-\eta D} \end{aligned}$$

where

$$D \equiv \frac{\mathbf{s}^\dagger\hat{\mathbf{R}}^{-1}\mathbf{s}}{\mathbf{w}^\dagger\mathbf{R}\mathbf{w}}.$$

Note that if $\hat{\mathbf{R}} = \mathbf{R}$, then $g(\mathbf{X}_L) = e^{-\eta}$ and this is the FAP of the AMF when the covariance matrix is known. As discussed before, we run simulations with homogeneous data possessing an identity covariance matrix, that is, with $\mathbf{R} = \mathbf{I}$. The g-method IS estimator then takes the form

$$\begin{aligned} \hat{\alpha}_g &= \frac{1}{K}\sum_{i=1}^K [g(\mathbf{X}_L)W(\mathbf{x}_L;\theta)]^{(i)} \\ &= \frac{1}{K}\sum_{i=1}^K [e^{-\eta D}W(\mathbf{x}_L;\theta)]^{(i)}, \qquad \sim f_\star \end{aligned} \tag{12}$$

with D given by

$$\begin{aligned} D &= \frac{\mathbf{s}^\dagger\hat{\mathbf{R}}^{-1}\mathbf{s}}{\|\mathbf{w}\|^2} \\ &= \frac{\mathbf{s}^\dagger\hat{\mathbf{R}}^{-1}\mathbf{s}}{\mathbf{s}^\dagger(\hat{\mathbf{R}}^{-1})^2\mathbf{s}}. \end{aligned} \tag{13}$$

Now, choosing the (single) biasing parameter $\theta < 1$ produces a decrease in D, thereby causing a higher frequency of occurrence of the false alarm event or, more appropriately in this case, a larger value of the g-function. Note that use of the g-method obviates the need to bias primary data vectors. Determination of a good value of θ proceeds as described in Appendix B. The weighting function is simply

$$W(\mathbf{x}_L;\theta) = \theta^{LN}e^{-(1-1/\theta)B} \tag{14}$$

which can be deduced from (7) by setting $a = 1$. The variance of this estimator is given by

$$\operatorname{var}\hat{\alpha}_g = \frac{1}{K}[I_g(\theta) - \alpha^2] \tag{15}$$

where

$$\begin{aligned} I_g(\theta) &\equiv E_\star\{g^2(\mathbf{X}_L)W^2(\mathbf{X}_L;\theta)\} \\ &= E\{g^2(\mathbf{X}_L)W(\mathbf{X}_L;\theta)\}. \end{aligned} \tag{16}$$

The minimization of (an estimate of) I_g with respect to the biasing parameter θ is carried out using the recursion

$$\theta_{m+1} = \theta_m - \delta_\theta\frac{\hat{I}_g'(\theta_m)}{\hat{I}_g''(\theta_m)}, \qquad m = 1,2,\ldots \tag{17}$$

which is just a one-dimensional version of (35) in Appendix B. The reader is again referred to this appendix for definitions of quantities in the summands

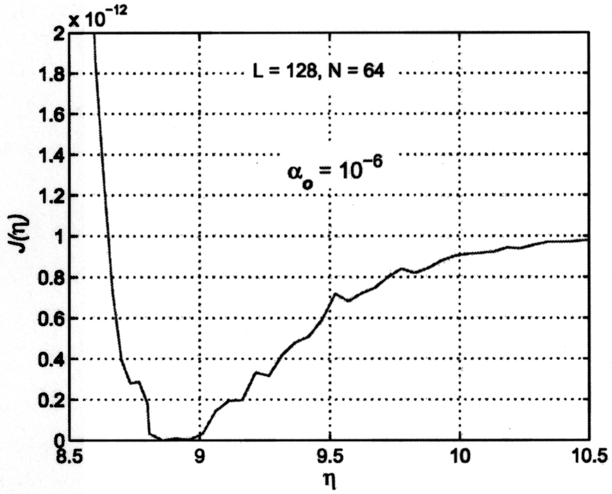

Fig. 1. Example objective function (for E-AMF detector).

of the following estimators for I_g and its derivatives (with respect to θ) that are used in (17). These are respectively given by

$$\hat{I}_g(\theta) = \frac{1}{K}\sum_{i=1}^{K}[g^2(\mathbf{X}_L)W^2(\mathbf{x}_L;\theta)]^{(i)}$$

$$\hat{I}'_g(\theta) = \frac{1}{K}\sum_{i=1}^{K}[g^2(\mathbf{X}_L)W(\mathbf{x}_L;\theta)W_\theta(\mathbf{x}_L;\theta)]^{(i)} \qquad (18)$$

$$\hat{I}''_g(\theta) = \frac{1}{K}\sum_{i=1}^{K}[g^2(\mathbf{X}_L)W(\mathbf{x}_L;\theta)W_{\theta\theta}(\mathbf{x}_L;\theta)]^{(i)}$$

where $\mathbf{X}_L \sim f_*$ in all 3 estimators above.

3) *Threshold Determination*: The converse problem, namely that of finding by fast simulation the value of detector threshold η satisfying a prescribed FAP, is an important task that can be readily carried out using the estimator of (12). The genesis of the method lies in the so-called inverse IS problem first formulated and solved in [9]. The solution is to minimize a "squared performance error" stochastic objective function

$$J(\eta) = [\hat{\alpha}_g(\eta) - \alpha_o]^2$$

where α_o is a desired FAP. A typical example is shown in Fig. 1 for an E-AMF detector. All detection algorithms that involve a threshold crossing will possess objective functions that have the general behaviour shown for $\alpha_o < 1$, assuming of course that the FAP estimate is (in a stochastic sense) monotone in the threshold η. Minimization of J with respect to η is carried out by the algorithm

$$\eta_{m+1} = \eta_m + \delta_\eta \frac{\alpha_o - \hat{\alpha}_g(\eta_m)}{\hat{\alpha}'_g(\eta_m)}, \qquad m = 1,2,\ldots \quad (19)$$

where δ_η is a step-size parameter and the derivative estimator in the denominator is given by

$$\hat{\alpha}'_g(\eta_m) = -\frac{1}{K}\sum_{i=1}^{K}[De^{-\eta D}W(\mathbf{x}_L;\theta)]^{(i)}, \qquad \sim f_* \quad (20)$$

with the prime indicating derivative with respect to η. Note that this derivative estimator actually estimates (negative of) the probability density function of the AMF statistic on the LHS of (1) under the hypothesis that target is absent. The threshold-finding algorithm of (19) is a key component of the fast simulation methodology.

4) *Simulation Gain*: A useful measure of the effectiveness of any IS procedure is the simulation gain Γ. It is the ratio of simulation lengths required by conventional MC and IS estimators to achieve the same error variance. The variance of an MC estimator ($W = 1$) is given by $(\alpha - \alpha^2)/K_{\mathrm{MC}}$ for a simulation length K_{MC}. Hence, using (9), the IS gain for the estimator of (6) is

$$\Gamma = \frac{K_{\mathrm{MC}}}{K} = \frac{\alpha - \alpha^2}{I(\nu) - \alpha^2} \qquad (21)$$

with $I(\nu)$ defined in (10). Similarly, using (15), the g-method estimator of (12) has gain Γ_g over MC estimation given by

$$\Gamma_g = \frac{\alpha - \alpha^2}{I_g(\theta) - \alpha^2} \qquad (22)$$

where $I_g(\theta)$ is given by (16). Note that setting $\theta = 1$ in the above provides gain of the g-method estimator without IS. The estimator always has a smaller variance and consequently, $\Gamma_g > \Gamma$. Comparing (21) and (22), it is sufficient to show that $I_g < I$ to prove this. This is done in Appendix C.

As seen in the sequel, estimation of gain during simulation plays an important role in mechanizing an IS methodology that allows rare events to be studied quickly. In particular, Γ_g is estimated as

$$\hat{\Gamma}_g = \frac{\hat{\alpha}_g - \hat{\alpha}_g^2}{\hat{I}_g(\theta) - \hat{\alpha}_g^2} \qquad (23)$$

wherein the required estimates are given in (12) and (18). The denominator of the above equation is related to an estimate of the IS variance, denoted as $\widehat{\mathrm{var}}\,\hat{\alpha}_g$. Substituting (12) and (18) into (23) and performing a little algebra yields

$$\widehat{\mathrm{var}}\,\hat{\alpha}_g \equiv \frac{\hat{I}_g(\theta) - \hat{\alpha}_g^2}{K} = \frac{1}{2K^3}\sum_{i=1}^{K}\sum_{\substack{j=1\\ i\neq j}}^{K}[y_i - y_j]^2 > 0 \quad \text{(with probability 1)} \qquad (24)$$

where $y_i \equiv [g(\mathbf{X}_L)W(\mathbf{x}_L;\theta)]^{(i)}$. The variance estimate is asymptotically unbiased and it is easy to show, using (15) and (24), that $\widehat{\mathrm{var}}\,\hat{\alpha}_g \xrightarrow{P} \mathrm{var}\,\hat{\alpha}_g \to 0$ as $K \to \infty$, for any fixed θ.

If the IS estimator of (6) is used instead of (12), then $y_i \equiv [1(\mathcal{A})W(\mathbf{x},\mathbf{x}_L;a,\theta)]^{(i)}$ and the inequality in (24) will not be strict, leading to a non-zero probability of instability in gain estimation. One way of trying to prevent this from happening is to overbias the simulation, but this can result in underestimation of the FAP. Such instabilities do not occur with the g-method estimator and this is one of its implementation advantages.

B. E-AMF and GM-STAP Detection

False alarm probability IS estimators for these two detector variants are set up in completely parallel fashion, starting with the expression in (12). The only difference is in the definition of the quantity D (given in (13) for the AMF detector), while the threshold multipliers get squared. This follows by observing from (3), (4), and (5) that all 3 detectors studied here differ, primarily, on the RHSs of their respective tests. Consequently, denoting by D_E and D_G, respectively, the D values for the two variants, this leads to

$$D_E \equiv \frac{\left((1/L)\sum_{l=1}^{L} |\mathbf{s}^\dagger \hat{\mathbf{R}}^{-1}\mathbf{x}(l)|\right)^2}{\mathbf{s}^\dagger(\hat{\mathbf{R}}^{-1})^2\mathbf{s}}$$

and

$$D_G \equiv \frac{\left(\prod_{l=1}^{L} |\mathbf{s}^\dagger \hat{\mathbf{R}}^{-1}\mathbf{x}(l)|^2\right)^{1/L}}{\mathbf{s}^\dagger(\hat{\mathbf{R}}^{-1})^2\mathbf{s}}.$$

The corresponding FAP estimators are obtained by using the above definitions in (12) with $g(\mathbf{X}_L)$ replaced by $\exp(-\eta_E^2 D_E)$ and $\exp(-\eta_G^2 D_G)$, respectively. The threshold-finding algorithm in (19) is altered accordingly.

IV. MECHANIZING THE IS ALGORITHM

A. Methodology

The various issues of the preceding sections are summarized in an IS simulation methodology that outlines the principal steps required for implementation of the adaptive algorithms. It is a cautious methodology which has been used in previous works on applications of IS.

Of interest is the extreme but realistic (and often encountered) situation where we have no knowledge whatsoever of the FAP α of a detection algorithm for a given value of threshold η. In such a case, referring to the objective function of Fig. 1, the basic idea of fast adaptive simulation is to travel down the curve from its left side. An initial (α_o,η_o) pair is needed. It is easily obtained using conventional MC simulation for a high value of α_o, say 0.1 (and a correspondingly low value of η_o), using $K = 100/\alpha_o = 1000$ trials according to the well-known MC thumb rule. This can be accomplished quickly with a few experiments. The (α_o,η_o) pair (whose accuracy need not be very high) provides the starting values for the IS procedure. It begins by forming the estimate $\hat{I}_g(\theta)$ at the threshold η_o (using 1000 trials) and locating its minimum (possibly graphically) as a function of the biasing parameter θ. It is advisable to search for the optimum bias, denoted as θ_o, starting from $\theta = 1$ to avoid locating an overbiasing minimizer. This leads, via (23), to an estimate $\hat{\Gamma}$ of the maximum IS gain available (with a possible correction for α) and results in the quadruplet $(\alpha_o,\eta_o,\theta_o,\hat{\Gamma}(\theta_o))$. At this stage the IS adaptations are implemented. The threshold-finding algorithm of (19) is implemented with initial value η_o for a new prespecified α_o slightly smaller than the initial α_o and using a conservative number of $K = 100/(\alpha_o\hat{\Gamma}(\theta_o))$ for the IS trials. It is conservative because we know that simulation gain increases with decreasing rare-event probability and hence this number is guaranteed to provide better than the rule-of-thumb accuracy at the new α_o. Note that most of the IS gain can be leveraged if the FAP decrements are kept small. The biasing algorithm of (17) is implemented simultaneously using θ_o as initial value. At the end of the adaptations the resulting gain is estimated for updating K and a new quadruplet is obtained. This is continued until we have a complete characterization of false-alarm behaviour of the detector down to the desired α_o. The procedure is summarized below.

1) *Implementation*: Define the set $\{\alpha_o^{(p)}\}_1^P$ such that $0.1 = \alpha_o^{(1)} > \alpha_o^{(2)} > \cdots > \alpha_o^{(P)} = 10^{-7}$, for example.

Preprocessing:

Step 1 $p = 1$. Use 1000 conventional MC trials to obtain $(\alpha_o^{(1)},\eta_o^{(1)})$ pair.

Step 2 Find $\theta_o^{(p)} = \arg\min_\theta \hat{I}_g(\theta)$ and calculate $\hat{\Gamma}^{(p)}(\theta_o^{(p)})$ using (23).

Adaptive IS:

Step 3 $p = p + 1$.

Step 4 Let $m = 1$. Set $\theta_m^{(p)} = \theta_o^{(p-1)}$, $\eta_m^{(p)} = \eta_o^{(p-1)}$ and $K = 100/(\alpha_o^{(p)}\hat{\Gamma}^{(p-1)}(\theta_o^{(p-1)}))$.

Step 5 Generate K secondary vectors and compute B and D using (8) and (13).

Step 6 Implement (12), (18) using biased secondary vectors with parameter $\theta_m^{(p)}$.

Step 7 Compute $\theta_{m+1}^{(p)}$ and $\eta_{m+1}^{(p)}$ using (17) and (19); if both algorithms have converged, set $\theta_o^{(p)} = \theta_{m+1}^{(p)}$, $\eta_o^{(p)} = \eta_{m+1}^{(p)}$, $\hat{\Gamma}^{(p)}(\theta_o^{(p)}) = \hat{\Gamma}^{(p)}(\theta_{m+1}^{(p)})$ using (23), and go to step 3 or stop if $p = P$; otherwise go to step 5 with $m = m + 1$.

At the end of simulations we have the P quadruplets $\{\alpha_o^{(p)},\eta_o^{(p)},\theta_o^{(p)},\hat{\Gamma}^{(p)}(\theta_o^{(p)})\}_1^P$.

B. Complexity Reduction for IS Adaptations

The methodology is simple enough but certain observations can be made which lead to an artifice that further reduces computational effort to a considerable extent. In rare-event simulation of highly reliable systems that involve complex signal processing operations, two factors contribute to simulation time. The first is the rare event itself that is under study; this is handled by suitable IS biasing techniques. The second factor is the computational intensity of the signal processing of the system. In STAP detectors the chief processing burden is from inversion of large matrices. Several millions of MC trials with as many matrix inversions are needed to estimate low FAPs. Using an IS scheme the number of trials can be reduced appreciably, for a single rare-event probability estimation. However, there remains the important matter of executing the IS adaptations of (17) and (19), using this small number of IS trials for each iteration. These can be computationally burdensome if the recursions are to be carried out for several α_o values. This is where the artifice comes in. It turns out, in contrast with conventional MC simulation, that adaptations can be performed by reusing the unbiased random quantities generated for any one set of trials. Consequently it is not necessary to run truly randomized adaptive IS algorithms. The result will be a large savings in computational effort.

This idea is illustrated by rewriting the FAP estimator of (12) in the form

$$\hat{\alpha}_g = \frac{1}{K}\sum_{i=1}^{K}[e^{-\eta\theta D_{\mathrm{U}}}\theta^{LN}e^{-(\theta-1)B_{\mathrm{U}}}]^{(i)}, \qquad \sim f \quad (25)$$

obtained by substituting (14) into (12), replacing D and B, respectively, by their unbiased versions denoted as D_{U} and B_{U}, and observing that whereas the estimator in (12) uses samples drawn from biased distributions f_*, the mathematically equivalent estimator of (25) operates with only unbiased quantities, obtained via f. Biasing is handled by the (explicit) presence of θ. The number of trials remains unchanged, at K. The estimators required in (17) and (19) can be rewritten in a similar manner. It is a straightforward exercise and is omitted. Assuming then that the value of K is large enough for estimation of the target α_o (and this is guaranteed by the methodology described above), it follows that the set of K instances of D_{U} and B_{U} employed in (25) can be repeatedly used in the adaptations. With the complexity of frequent regeneration thus removed, these adaptations (such as minimization of $\hat{I}_g(\theta)$ in (18)) can be extremely fast, often requiring only a few iterations. It is assumed that a sufficiently large pool of unbiased variables is pregenerated to accommodate K trials.

There are two related issues that are briefly discussed here in qualitative terms. In IS simulations carried out here (and elsewhere), it is tacitly assumed that the number of trials K is also large enough to accurately estimate $I_g(\theta)$. This quantity is admittedly not a rare-event probability and its estimator $\hat{I}_g(\theta)$, given in (18), is clearly unbiased. Nevertheless, its accuracy should be studied in terms of its error variance, irrespective of whether we employ the above described reuse strategy or not. In this regard it must be pointed out that the biasing densities used for IS simulations are being chosen to minimize $I_g(\theta)$ (or its estimate $\hat{I}_g(\theta)$), and not to estimate it accurately. Indeed, attempting to estimate $I_g(\theta)$ accurately would lead us to consider further biasing of the original biasing distribution. As probably known to IS researchers, the more powerful a biasing distribution is (in the sense of being "close" to the unrealizable optimal biasing density that estimates the rare-event probability exactly), the more "constant" will be the IID random variables y_i defined after (24). Interestingly this implies that even for small K, the estimated variance $\widehat{\mathrm{var}}\,\hat{\alpha}_g$ can become small for such biasing distributions. The latter observations accentuate the need to search for good simulation distributions. Unfortunately, it is beyond the remit of this paper to investigate the variance and rate of convergence of $\hat{I}_g(\theta)$ and the dependencies of these on biasing distributions and number of trials K. Some remarks on this issue can be found in [16]. We have confirmed the legitimacy of the reuse strategy in several cases in two ways. Using an overkill, that is, employing a large number of IS trials for varying θ, it has been ascertained that the correct value of the optimum bias does not differ noticeably from the θ_o determined as above. Secondly, truly randomized optimization algorithms have been used, with the same conclusion.

The second issue concerns the ease with which the form of the estimator in (25) is obtained. The reason lies in the biasing technique used. Biasing in IS can be classed, for the purpose of this discussion, into two types. One kind derives biased quantities by direct parameterized transformations of unbiased variables, such as the scaling employed here. The second is just the complementary class, where such transformations do not exist. Examples of the latter can include new distributions for conducting simulations and some instances of distributions based on large deviations theory arguments. Within the first class, only the subset of transformations that are many-to-one will allow placing the IS estimator in a form such as that of (25).

V. SIMULATION RESULTS

Implementation results of the fast simulation procedures using the g-method are described here.

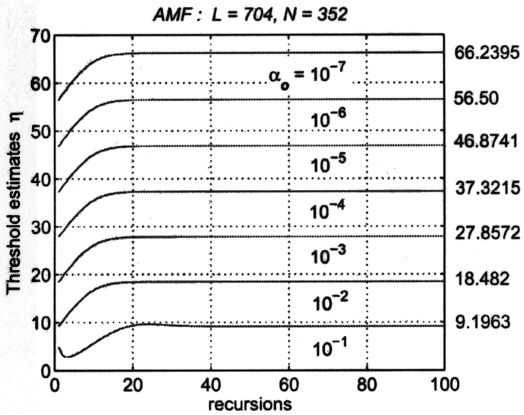

Fig. 2. Threshold optimization for AMF detector ($L = 704, N = 352$) using inverse IS.

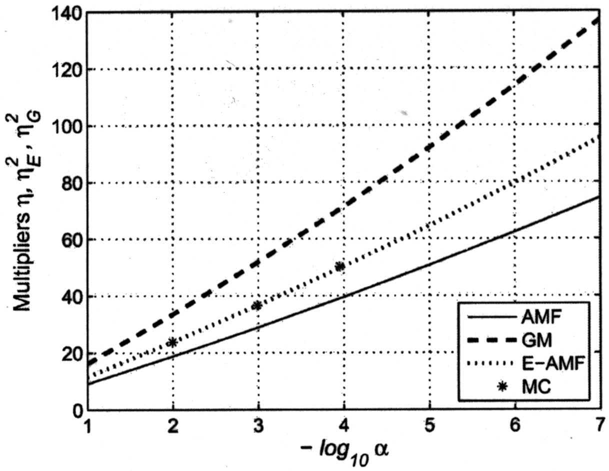

Fig. 3. Multipliers for $L = 128$, and $N = 64$. Starred dots indicate validations for E-AMF detector using conventional MC simulations.

A typical example of evolution of the threshold-finding algorithm is shown in Fig. 2 for the square-law AMF detector with $L = 704$ training vectors and space-time product $N = 352$. Clearly, about 20 recursions of (19) appear sufficient for convergence. Note also from this figure that the "converged" value of the threshold for an α_o acts as the initial value for the next lower α_o, as described in Section IVA. Although not shown here, similar estimation results have been obtained for the E-AMF and GM-STAP detectors. The results for the AMF detector almost "coincide" with numerical computations of (2). As analytical expressions or approximations for the FAPs are not available, such a comparison is not possible for the variants. Threshold multiplier values for the 3 detectors are summarized in Fig. 3 for $L = 128$ and $N = 64$. For the E-AMF and GM-STAP detectors the interpolations

$$\eta_E = 0.0095x^3 - 0.1713x^2 + 1.8869x + 1.6873$$

$$\eta_G = 0.0111x^3 - 0.1985x^2 + 2.2351x + 1.9899$$

for $L = 128$ and $N = 64$, and

$$\eta_E = 0.01x^3 - 0.1865x^2 + 1.8822x + 1.7255$$

$$\eta_G = 0.0118x^3 - 0.2198x^2 + 2.2292x + 2.0408$$

for $L = 704$ and $N = 352$, obtained from inverse IS results where $x = -\log_{10}\alpha$, can be used for estimating multipliers for FAPs in the range 10^{-1} to 10^{-7}.

Performance graphs of the IS algorithms are in Figs. 4 and 5, which show the estimated IS gain Γ obtained over MC simulation and the number of trials required, respectively. The number of IS trials, although appreciably reduced in comparison with MC, increases rapidly with decrease in FAP. From these two figures it can be observed that, for a FAP of 10^{-6}, the square-law AMF with $L = 128$ and $N = 64$ requires 1500 IS trials whereas only 100 trials are needed for $L = 704$ and $N = 352$. This can be attributed to the fact that D and B, the only random

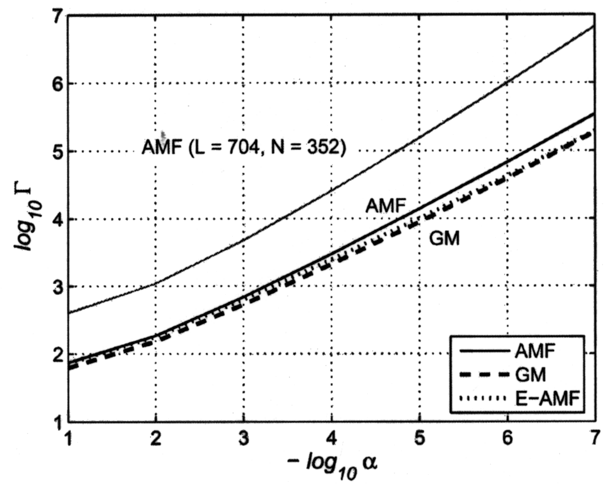

Fig. 4. IS gains. Set of 3 lower graphs correspond to $L = 128$, $N = 64$.

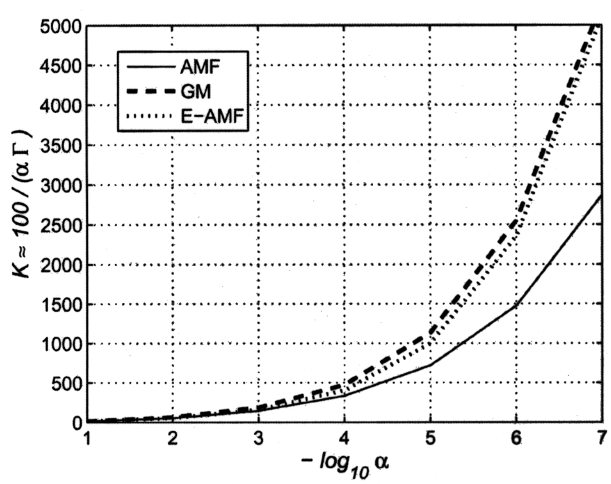

Fig. 5. Simulation lengths for $L = 128$, $N = 64$.

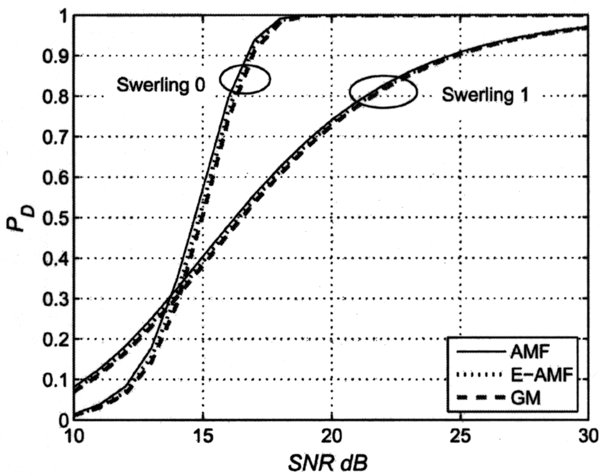

Fig. 6. Detection performances in homogeneous clutter. $L = 128$, $N = 64$, $\alpha_o = 10^{-6}$.

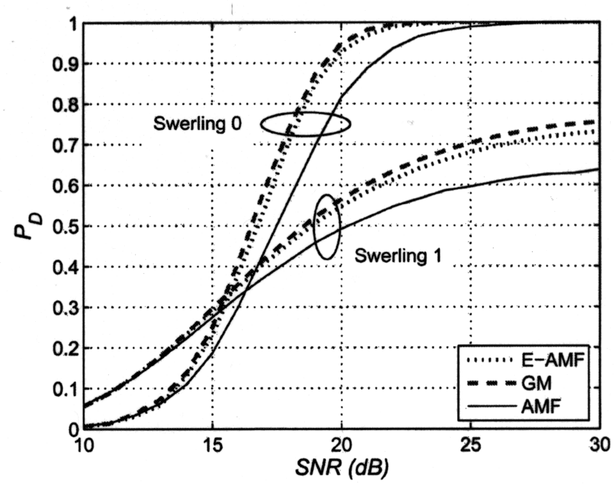

Fig. 7. Detection with 2 interfering targets in training data. $L = 128$, $N = 64$, $\alpha_o = 10^{-6}$.

quantities that appear in (12), have more concentrated density functions in the latter case, thus causing the IS biasing to be more effective during simulation.

Detection probabilities for all three detectors have been estimated and compared at a FAP of 10^{-6}. For homogeneous clutter backgrounds and nonfluctuating (Swerling 0) and fluctuating (Swerling 1) targets, the results are in Fig. 6. The SNR loss of the two variants compared with the AMF detector is very slight, the maximum being 0.3 dB for a Swerling 0 target. Performance degradations for training data contaminated by nonhomogeneities consisting of interfering targets that are assumed to have the same Doppler-angle properties and characteristics as the primary target are shown in Fig. 7 for two Swerling 0 and Swerling 1 targets. Each interferer is assumed to have the same steering vector and power as the primary target. Similar results (not shown here) have been obtained for different numbers of interferers. It is evident that the GM-STAP detector is most robust in the presence of interferers, enjoying (in some cases) several decibels of power advantage over the square-law AMF detector. Consequently, its FAP performance in these situations (though not evaluated here) should also be relatively robust.

A. Comments

Despite the accuracy of our results, biasing by scaling has not done as well for the variants as for the AMF detector. Furthermore, even for the latter, IS performance should be improved by a better choice of biasing scheme in order to achieve the goal of having the required number of trials to be constant, or at least increasing very slowly, with decrease of FAP. This is certainly a matter for further investigation that may pay dividends while examining other detector configurations. Scaling was used in this work in the interest of expediency. Some details regarding the behaviour of the scaling parameter θ are available in [15]. Briefly, it is close to unity and has a small spread over the range of FAPs considered. This is due to the shape of the density functions of B and D.

VI. CONCLUSION

A humble inroad into the use of adaptive IS algorithms to characterize a STAP detector has been made. The AMF detector was used for validation and results have been good. The chief reasons for this are that we were able to invoke the g-method and inverse IS, use a biasing strategy that could be easily optimized adaptively, and find a way around the difficult task of inverting large matrices several times during simulations. As a small demonstration of the potential of IS, two AMF detector variants that are known to be relatively obdurate to mathematical analysis have been suggested and characterized. Sufficient numerical results have been provided here in order to motivate interested researchers to confirm our findings and possibly carry out their own investigations into the subject. There are some technical issues regarding IS estimators that remain to be settled; however, the development of better biasing methods is an important matter. In any case, our hope is that application of these fast simulation techniques to more advanced STAP configurations will also meet with success. But this remains to be seen as we are certainly not in position to predict what subtleties (and difficulties) these other detection algorithms can throw up. It is clear that IS is still in its infancy, especially insofar as its use for characterizing modern detection algorithms is concerned.

APPENDIX A. CFAR PROPERTY

An invariance property is established that can be used to show that certain STAP detection algorithms have FAPs that do not depend on the data covariance **R**. The proposition given here follows quite simply from arguments contained in the exposition of the generalized likelihood ratio STAP detector test (Kelly's GLRT) given in [1]. They are outlined here for convenience of the reader.[2] Assume, as before, that the primary and training data vectors have the same covariance. Consider the variables

$$G \equiv \mathbf{s}^\dagger \hat{\mathbf{R}}^{-1}\mathbf{x} \qquad \text{and} \qquad G(l) \equiv \mathbf{s}^\dagger \hat{\mathbf{R}}^{-1}\mathbf{x}(l) \qquad (26)$$

for $l = 1,\ldots,L$. Using the transformations $\mathbf{u} = \mathbf{R}^{-1/2}\mathbf{s}$, $\mathbf{y} = \mathbf{R}^{-1/2}\mathbf{x}$, and $\mathbf{y}(l) = \mathbf{R}^{-1/2}\mathbf{x}(l)$, leads to

$$G = \mathbf{u}^\dagger \tilde{\mathbf{R}}^{-1}\mathbf{y} \qquad \text{and} \qquad G(l) = \mathbf{u}^\dagger \tilde{\mathbf{R}}^{-1}\mathbf{y}(l) \qquad (27)$$

where $\tilde{\mathbf{R}} \equiv \mathbf{R}^{-1/2}\hat{\mathbf{R}}\mathbf{R}^{-1/2}$. The whitened vectors **Y** and **Y**(l) are both distributed $\mathcal{CN}_N(0,\mathbf{I})$. It turns out that $\tilde{\mathbf{R}}$ has the complex Wishart distribution $\mathcal{CW}(L,N;(1/L)\mathbf{I})$, [13]. Further, a unitary transformation **U** can be found which rotates the new signal vector **u** into an elementary vector **e** as $d\mathbf{e} = \mathbf{U}^\dagger\mathbf{u}$, such that $\mathbf{e} = [1,0,\ldots,0]^\dagger$ and where $d^2 = \mathbf{u}^\dagger\mathbf{u} = \mathbf{s}^\dagger\mathbf{R}^{-1}\mathbf{s}$. The first column of **U** is the new signal vector $\mathbf{u}/d$. The remaining columns comprise an orthonormal basis determined, for example, by a Gram-Schmidt procedure. Let $\mathbf{z} = \mathbf{U}^\dagger\mathbf{y}$ and $\mathbf{z}(l) = \mathbf{U}^\dagger\mathbf{y}(l)$. Applying these to (27) yields the variables

$$G = \frac{d}{L}\mathbf{e}^\dagger \mathcal{S}^{-1}\mathbf{z} \qquad \text{and} \qquad G(l) = \frac{d}{L}\mathbf{e}^\dagger \mathcal{S}^{-1}\mathbf{z}(l) \qquad (28)$$

where $\mathcal{S} \equiv L\mathbf{U}^\dagger\tilde{\mathbf{R}}\mathbf{U}$. While **Z** and **Z**$(l)$ are distributed as $\mathcal{CN}_N(0,\mathbf{I})$ and are independent, $\mathcal{S}$ has the distribution $\mathcal{CW}(L,N;\mathbf{I})$. The vectors **z** and **z**(l) are decomposed as $\mathbf{z} = [\mathbf{z}_A \ \mathbf{z}_B]^T$ and $\mathbf{z}(l) = [\mathbf{z}_A(l) \ \mathbf{z}_B(l)]^T$ where the A components are scalar and B components $(N-1)$-vector. Correspondingly, $\mathcal{S}$ is decomposed as

$$\mathcal{S} = \sum_{l=1}^{L}\mathbf{z}(l)\mathbf{z}(l)^\dagger = \begin{bmatrix} \mathcal{S}_{AA} & \mathcal{S}_{AB} \\ \mathcal{S}_{BA} & \mathcal{S}_{BB} \end{bmatrix} \qquad (29)$$

so that $\mathcal{S}_{AB} = \sum_{l=1}^{L}\mathbf{z}_A(l)\mathbf{z}_B(l)^\dagger$, $\mathcal{S}_{BB} = \sum_{l=1}^{L}\mathbf{z}_B(l)\mathbf{z}_B(l)^\dagger$ and so on. Also

$$\mathbf{e}^\dagger\mathcal{S}^{-1} = [\mathcal{P}_{AA} \quad \mathcal{P}_{AB}] \qquad (30)$$

where

$$\begin{aligned} \mathcal{P}_{AA} &\equiv (\mathcal{S}_{AA} - \mathcal{S}_{AB}\mathcal{S}_{BB}^{-1}\mathcal{S}_{BA})^{-1} \\ \mathcal{P}_{AB} &\equiv -\mathcal{P}_{AA}\mathcal{S}_{AB}\mathcal{S}_{BB}^{-1} \end{aligned} \qquad (31)$$

which follow from the Frobenius relations for inversion of block matrices, [1]. Substituting (30) and (31) in (28) gives

$$G = \frac{d}{L}\mathcal{P}_{AA}y \qquad \text{and} \qquad G(l) = \frac{d}{L}\mathcal{P}_{AA}y(l) \qquad (32)$$

where

$$\begin{aligned} y &\equiv \mathbf{z}_A - \sum_{l=1}^{L}\mathbf{z}_A(l)\mathbf{z}_B(l)^\dagger\mathcal{S}_{BB}^{-1}\mathbf{z}_B \\ y(l) &\equiv \mathbf{z}_A(l) - \sum_{i=1}^{L}\mathbf{z}_A(i)\mathbf{z}_B(i)^\dagger\mathcal{S}_{BB}^{-1}\mathbf{z}_B(l). \end{aligned} \qquad (33)$$

Conditioned on the vectors $\mathbf{z}_B$ and $\mathbf{z}_B(l)$, it follows that the random variables Y and $Y(l)$ in (33) are (in the absence of target) zero mean Gaussian. With a little more algebra it can be shown, as in [1], that they are uncorrelated with variances

$$E_B\{|Y|^2\} = 1 + \mathbf{z}_B^\dagger\left(\sum_{l=1}^{L}\mathbf{z}_B(l)\mathbf{z}_B(l)^\dagger\right)^{-1}\mathbf{z}_B$$

and

$$E_B\{|Y(l)|^2\} = 1 - \mathbf{z}_B(l)^\dagger\left(\sum_{i=1}^{L}\mathbf{z}_B(i)\mathbf{z}_B(i)^\dagger\right)^{-1}\mathbf{z}_B(l)$$

for $l = 1,\ldots,L$, with E_B denoting conditional expectation. Further, the conditional covariance of the variables $Y(l)$ is given by

$$E_B\{Y(k)Y(n)^*\} = -\mathbf{z}_B(n)^\dagger\left(\sum_{l=1}^{L}\mathbf{z}_B(l)\mathbf{z}_B(l)^\dagger\right)^{-1}\mathbf{z}_B(k)$$

for $k \neq n$.

Hence the set of conditionally jointly Gaussian zero mean random variables Y and $\{Y(l)\}_1^L$ have individual variances and covariances that are functions of the random vectors $\mathbf{z}_B$ and $\{\mathbf{z}_B(l)\}_1^L$. The latter are all jointly independent, each being distributed as $\mathcal{CN}_{N-1}(0,\mathbf{I})$. The probability of any event defined on the random variables Y and $\{Y(l)\}_1^L$ in (33) can thus be determined by performing an averaging operation over the distributions of $\mathbf{z}_B$ and $\{\mathbf{z}_B(l)\}_1^L$ and this probability will be independent of the data covariance **R**. This statement is also true for the random variables G and $\{G(l)\}_1^L$ in (32) with the caveat that any constant scaling of these variables should leave the event unchanged. The preceding arguments therefore constitute proof of the following.

PROPOSITION *Any STAP detection algorithm that uses only the random variables G and $\{G(l)\}_1^L$ defined in* (26) *for its description such that the algorithm itself is unchanged by arbitrary but equal scaling of all these variables, has a FAP which is independent of the target-free data covariance* **R**.

It follows immediately from this that both the E-AMF and GM-STAP detectors have FAPs that are independent of the covariance matrix **R**.

[2]It would be helpful for the reader to refer to [1] (see also [2]). We have used several results from this now classic paper and have attempted to maintain the same notation.

APPENDIX B. ADAPTIVE ALGORITHMS FOR 2-D BIASING

From the first line of (10), the I-function is estimated as

$$\hat{I}(\boldsymbol{\nu}) = \frac{1}{K}\sum_{i=1}^{K}[1(\mathcal{A})W^2(\mathbf{x},\mathbf{x}_L;\boldsymbol{\nu})]^{(i)}, \qquad \sim f_* \quad (34)$$

and its minimization can be carried out using the 2-D adaptive algorithm

$$\boldsymbol{\nu}_{m+1} = \boldsymbol{\nu}_m - \delta\hat{\mathbf{J}}_m^{-1}\hat{\nabla}I(\boldsymbol{\nu}_m), \qquad m = 1,2,\ldots \quad (35)$$

Here, δ is a step-size parameter used to control convergence, and m is the index of iteration. This is a stochastic Newton recursion. It achieves minimization of $\hat{I}$ by estimating a solution of

$$\hat{\nabla}I(\boldsymbol{\nu}) \equiv (\hat{I}_a \quad \hat{I}_\theta)^T = \mathbf{0}$$

where $I_a \triangleq \partial I(\boldsymbol{\nu})/\partial a$ and $I_\theta \triangleq \partial I(\boldsymbol{\nu})/\partial\theta$. The estimate of the Jacobian $\mathbf{J}$ (which is the Hessian matrix of I) is given by

$$\hat{\mathbf{J}} = \begin{pmatrix} \hat{I}_{aa} & \hat{I}_{a\theta} \\ \hat{I}_{a\theta} & \hat{I}_{\theta\theta} \end{pmatrix}$$

where $I_{xy} \equiv \partial I_x/\partial y$. Starting with the second line of (10), the various I-functions defined above can be obtained by using the notational equations

$$\begin{aligned} I_x &= E\{1(\mathcal{A})\partial W/\partial x\} \\ &= E_*\{1(\mathcal{A})WW_x\} \end{aligned}$$

and similarly

$$\begin{aligned} I_{xx} &= E_*\{1(\mathcal{A})WW_{xx}\} \\ I_{xy} &= E_*\{1(\mathcal{A})WW_{xy}\} \end{aligned}$$

with the corresponding weighting function derivatives easily calculated, using (7), as $W_x = \partial W/\partial x$, $W_{xy} = \partial^2 W/\partial x\partial y$, and so on. These I-functions can be estimated as in (34).

APPENDIX C. GAIN OF THE g-METHOD ESTIMATOR

The proof given here is in the context of STAP detectors and is a generalization of the one in [10] for conventional CFAR detection algorithms. A more complete generalization for estimators in an arbitrary rare-event setting is available in [11].

When no IS is used ($W = 1$), then $I_g = E\{g^2(\mathbf{X}_L)\} < E\{g(\mathbf{X}_L)\} = \alpha$. To show that $I_g < I \equiv I(\boldsymbol{\nu})$ (the biasing vector is omitted for convenience) with IS, some additional notation is useful. Let $U \equiv |\mathbf{s}^\dagger\hat{\mathbf{R}}^{-1}\mathbf{X}|^2$ and let V denote the RHS of any of the three STAP detectors, excluding the multiplier. We bear in mind that U is a function of $\mathbf{X}$ and $\mathbf{X}_L$, and V a function of $\mathbf{X}_L$. Then, in the absence of target, $\alpha = P(U \geq \eta V)$ and (10) can be written as

$$I = E\{1(U \geq \eta V)W(\mathbf{x},\mathbf{x}_L)\}$$

with W defined in the first line of (7). For the g-method estimator, we can write

$$\begin{aligned} g(\mathbf{x}_L) &= P(U \geq \eta V \mid \mathbf{X}_L = \mathbf{x}_L) \\ &= E\{1(U \geq \eta V) \mid \mathbf{X}_L = \mathbf{x}_L\} \\ &= \int 1(u \geq \eta v) f(\mathbf{x} \mid \mathbf{x}_L)d\mathbf{x} \\ &= \int 1(u \geq \eta v)W_c(\mathbf{x},\mathbf{x}_L)f_*(\mathbf{x} \mid \mathbf{x}_L)d\mathbf{x} \end{aligned}$$

where $W_c(\mathbf{x},\mathbf{x}_L) \equiv f(\mathbf{x} \mid \mathbf{x}_L)/f_*(\mathbf{x} \mid \mathbf{x}_L)$. Then

$$\begin{aligned} g^2(\mathbf{x}_L) &\leq \int 1(u \geq \eta v)W_c^2(\mathbf{x},\mathbf{x}_L)f_*(\mathbf{x} \mid \mathbf{x}_L)d\mathbf{x} \\ &= \int 1(u \geq \eta v)W_c(\mathbf{x},\mathbf{x}_L)f(\mathbf{x} \mid \mathbf{x}_L)d\mathbf{x} \end{aligned}$$

by Jensen's inequality. Therefore I_g, defined in (16), is given by

$$\begin{aligned} I_g &= \int g^2(\mathbf{x}_L)W(\mathbf{x}_L)f(\mathbf{x}_L)d\mathbf{x}_L \\ &\leq \int\int 1(u \geq \eta v)W_c(\mathbf{x},\mathbf{x}_L)W(\mathbf{x}_L)f(\mathbf{x} \mid \mathbf{x}_L)f(\mathbf{x}_L)d\mathbf{x}d\mathbf{x}_L \\ &= \int\int 1(u \geq \eta v)W(\mathbf{x},\mathbf{x}_L)f(\mathbf{x},\mathbf{x}_L)d\mathbf{x}d\mathbf{x}_L \\ &= I. \end{aligned}$$

ACKNOWLEDGMENT

The authors are grateful to Dr. Michael Milligan of EOARD and Dr. Arje Nachman of AFOSR for their support and encouragement during the course of this work. Thanks are also due to Rusdha Muharar and Laura Anitori for software support and discussions.

REFERENCES

[1] Kelly, E. J.
An adaptive detection algorithm.
IEEE Transactions on Aerospace and Electronic Systems, **AES-22**, 1 (Mar. 1986), pp. 115–127.

[2] Robey, F. C., Fuhrmann, D. R., Kelly, E. J., and Nitzberg, R.
A CFAR adaptive matched filter detector.
IEEE Transactions on Aerospace and Electronic Systems, **28**, 1 (Jan. 1992), 208–216.

[3] Gerlach, K. R.
Outlier resistant adaptive matched filtering.
IEEE Transactions on Aerospace and Electronic Systems, **38**, 3 (July 2002), 885–901.

[4] Michels, J. H., Rangaswamy, M., and Himed, B.
Peformance of parametric and covariance based STAP tests in compound-Gaussian clutter.
Digital Signal Processing, **12**, 2, 3 (Apr.–July 2002), 307–328.

[5] Rangaswamy, M.
Statistical analysis of the non-homogeneity detector for non-Gaussian interference backgrounds.
IEEE Transactions on Signal Processing, **53**, 6 (June 2005), 2101–2111.

[6] Bucklew, J. A.
Introduction to Rare Event Simulation.
New York: Springer, 2004.

[7] Gerlach, K. R.
New results in importance sampling.
IEEE Transactions on Aerospace and Electronic Systems, **35**, 3 (July 1999), 917–925.

[8] Stadelman, D. L., Weiner, D. D., and Keckler, A. D.
Efficient determination of thresholds via importance sampling for Monte Carlo evaluation of radar performance in non-Gaussian clutter.
Proceedings of the IEEE Radar Conference, Long Beach, CA, Apr. 2002, 272–277.

[9] Srinivasan, R.
Some results in importance sampling and an application to detection.
Signal Processing, **65**, 1 (Feb. 1998), 73–88.

[10] Srinivasan, R.
Simulation of CFAR detection algorithms for arbitrary clutter distributions.
IEE Proceedings of Radar, Sonar and Navigation, **147** (Feb. 2000), 31–40.

[11] Bucklew, J. A.
Conditional importance samplng estimators.
IEEE Transactions on Information Theory, **51**, 1 (Jan. 2005).

[12] Weinberg, G. V.
Estimation of false alarm probabilities in cell averaging constant false alarm rate detectors via Monte Carlo methods.
DSTO Systems Science Laboratory, Report DSTO-TR-1624, Australia, Nov. 2004.

[13] Maiwald, D., and Kraus, D.
Calculation of moments of complex Wishart and complex inverse Wishart distributed matrices.
IEE Proceedings of Radar, Sonar and Navigation, **147**, 4 (Aug. 2000).

[14] Billingsley, P.
Convergence of Probability Measures.
New York: Wiley, 1999.

[15] Srinivasan, R., and Rangaswamy, M.
Fast estimation of false alarm probabilities of STAP detectors—the AMF.
Proceedings of the IEEE International Radar Conference, Arlington, VA, May 9–12, 2005.

[16] Srinivasan, R.
Importance Sampling—Applications in Communications and Detection.
Berlin: Springer-Verlag, 2002.

Rajan Srinivasan (SM'94) received his first two degrees from the Indian Institute of Technology, Delhi, India, and the doctorate from Aston University, Birmingham, UK, all in electrical engineering.

He is an educator and researcher, having worked in various academic institutions worldwide including some years spent as a senior research scientist in the Ministry of Defence of the Government of India. His principal interests are in statistical signal processing covering the fields of communications and radar. During his career he received a Young Scientist award from the International Union of Radio Science, in Florence, Italy, for his work on distributed detection which resulted in seminal publications. More recently he wrote "Importance Sampling—Applications in communications and Detection" the first monograph on the subject in the world, published by Springer-Verlag in 2002. A widely published author, Srinivasan is a Fellow of the IEE and currently teaches at the University of Twente in the Netherlands, carrying out research mainly focused on the development of fast simulation techniques for design and analysis of highly reliable processing systems. He was listed in the Millennium 2000 Edition of Marquis Who's Who in the World.

Muralidhar Rangaswamy (S'89—M'93—SM'98—F'06) received the B.E. degree in electronics engineering from Bangalore University, Bangalore, India in 1985 and the M.S. and Ph.D. degrees in electrical engineering from Syracuse University, Syracuse, NY, in 1992.

He is presently employed as a senior electronics engineer at the Sensors Directorate of the Air Force Research Laboratory (AFRL), Hanscom Air Force Base, MA. Prior to this he has held industrial and academic appointments. His research interests include radar signal processing, spectrum estimation, modeling non-Gaussian interference phenomena, and statistical communication theory. He has coauthored more than 70 refereed journal and conference record papers in the areas of his research interests. Additionally, he is a contributor to 3 books and is a coinventor on 2 U.S. patents.

Dr. Rangaswamy is an associate editor for the *IEEE Transactions on Aerospace and Electronic Systems* and is a member of the sensor array and multichannel processing technical committee (SAM-TC) of the IEEE Signal Processing Society. He was a coinstructor with Dr. W. Melvin for two short courses on space-time adaptive processing for the IEEE Boston section (April 2003) and for the IEEE-AESS Atlanta section at the Southeastern Symposium on System Theory (March 2004). He has served on the organizing committee of numerous IEEE AESS and IEEE Signal Processing Society sponsored conferences. He received the 2004 Fred Nathanson memorial radar award from the IEEE AES Society, the 2006 Distinguished Member award from the IEEE Boston Section, and the Charles Ryan basic research award from the Sensors Directorate of AFRL, in addition to 20 AFRL scientific achievement awards.