Adaptive Subspace Detectors

Shawn Kraut, *Member, IEEE*, Louis L. Scharf, *Fellow, IEEE*, and L. Todd McWhorter

***Abstract*—In this paper, we use the theory of generalized likelihood ratio tests (GLRTs) to adapt the *matched subspace detectors* (MSDs) of [1] and [2] to unknown noise covariance matrices. In so doing, we produce *adaptive* MSDs that may be applied to signal detection for radar, sonar, and data communication. We call the resulting detectors *adaptive subspace detectors* (ASDs). These include Kelly's GLRT and the adaptive cosine estimator (ACE) of [6] and [19] for scenarios in which the scaling of the test data may deviate from that of the training data. We then present a unified analysis of the statistical behavior of the entire class of ASDs, obtaining statistically identical decompositions in which each ASD is simply decomposed into the nonadaptive matched filter, the nonadaptive cosine or t-statistic, and three other statistically independent random variables that account for the performance-degrading effects of limited training data.**

***Index Terms*—Adaptive signal detection, adaptive subspace detector, data communication, matched subspace detector, radar detection, sonar detection.**

I. Introduction

OUR AIM in this paper is to adapt the four matched subspace detectors (MSDs) of [1] and [2] to unknown noise covariance in order to produce *adaptive subspace detectors* (ASDs) that may be applied to signal detection for radar, sonar, and data communication. Whenever we speak of an MSD problem, we assume that the noise covariance matrix $\mathbf{R}$ is known. When we speak of an ASD problem, we assume the covariance matrix is unknown and is estimated from training data.

There are four (nonadaptive) matched subspace detectors that form the basis for the adaptive subspace detectors of interest to us here. They arise from two types of generalizations of the matched filter detector. First, the inner product of the matched filter may be generalized to a projection of the measurement onto a higher dimensional signal subspace, thus producing a *subspace* detector [2]. Second, the detector may be normalized by an estimate of the noise power to make it have a constant false alarm rate (CFAR) with respect to the noise power. The four detectors are thus

1) the *coherent MSD* (i.e., matched filter), which is a normally distributed statistic that detects coherent signals by resolving the inner product of the measurement and signal;
2) the *MSD*, which is a χ^2 statistic that detects subspace signals (including noncoherent signals) by computing the energy of the measurement in the signal subspace;
3) the *coherent CFAR MSD*, which is a t or "cosine" statistic that detects coherent signals in noise of unknown variance by resolving the cosine of the angle the measurement makes with the signal;
4) the *CFAR MSD*, which is an F or "cosine-squared" statistic that detects subspace signals (including noncoherent signals) in noise of unknown variance by measuring the fraction of energy the measurement has in the signal subspace.

Each of the resulting four detectors is a GLRT for a concrete problem, and each is UMP-invariant, uniformly most powerful over the entire class of detectors invariant to an appropriate transformation group. This was one of the main points of [2], namely, that the GLRTs have the same invariances as the UMP-invariant tests of [1], and therefore, they inherit the optimality properties of the UMP-invariant tests for an interesting class of multivariate Gaussian detection problems.

All of these detectors are compelling. They have clearly stated optimalities and invariances, and they have evocative geometrical interpretations. The MSDs use extra knowledge of the noise variance for some performance gain against the CFAR MSDs, which do not assume this knowledge (the gain is slight unless the SNR exceeds the measurement dimension). On the other hand, the CFAR MSDs (or "cosine" statistics) compensate for this lack of knowledge by providing an extra invariance to data scaling, a property that the MSDs do not have. A consequence of this invariance is that the CFAR MSDs are CFAR over the whole class of elliptically contoured distributions (a result that is obvious for the special case of compound-Gaussian noise, multivariate Gaussian with random amplitude scaling; see Section II). The scale invariance sacrifices some high-SNR performance gain in return for robustness to tenuous and changeable prior information about channel noise variances, filter gains, and noise statistics.

The MSDs and CFAR MSDs all assume prior knowledge of noise covariance matrices. However, this information is often not known, meaning that, in practice, it must be estimated and then used correctly in an adaptive detector. In this paper, we address this problem by adapting the MSDs and CFAR MSDs to unknown noise covariance in order to derive ASDs and CFAR ASDs. To adapt the MSDs, we appeal to the fundamental results of Kelly [5] on GLRTs, and to adapt the CFAR MSDs, we use the results of [6].

Manuscript received September 1, 1998; revised August 22, 2000. This work was supported by the Office of Naval Research under Contracts N00014-89-J-1070 and N00014-00-1-0033, and by the National Science Foundation under Contracts MIP-9529050 and ECS 9979400. The associate editor coordinating the review of this paper and approving it for publication was Prof. Victor A. N. Barroso.

S. Kraut is with the Department of Electrical and Computer Engineering, Duke University, Durham, NC 27706 USA.

L. L. Scharf was with the University of Colorado, Boulder, CO 80309 USA. He is now with the Department of Electrical and Computer Engineering, Colorado State University, Fort Collins, CO 80523 USA (e-mail: scharf@engr.colostate.edu).

L. T. McWhorter is with Mission Research Corp., Fort Collins, CO 80522-0466 USA.

Publisher Item Identifier S 1053-587X(01)00066-6.

In order to clarify the adaptive problems we are studying and to establish the meaning of the abbreviations we use, we offer the following taxonomy of adaptive problems and detectors.

1) The *coherent ASD* is the adaptive GLRT generalization of the coherent MSD, wherein the training data for estimating the unknown covariance matrix is scaled the same as the test data.
2) The *ASD* is the adaptive GLRT generalization of the MSD, and it equals the Kelly GLRT [5], wherein the training and test data are scaled the same.
3) The *coherent CFAR ASD* is the adaptive GLRT generalization of the coherent CFAR MSD, and it is used for detecting a coherent signal when the training data is not constrained to be scaled the same as the test data.
4) The *CFAR ASD* is the adaptive GLRT generalization of the CFAR MSD, wherein the training and test data are not constrained to be uniformly scaled, and it equals the adaptive cosine/coherence estimator (ACE) of [7], [8], and [29].

In the nomenclature of the nonadaptive detection literature, "CFAR" is with respect to noise level or variance σ^2 in the test data. In the adaptive detection literature (see, e.g., [9] and [10]), "CFAR" is with respect to the noise covariance $\mathbf{R}$, assumed to be uniform over test and training data. Retaining the assumption that the noise is uniform in covariance structure, we say a detector is "CFAR" when it is insensitive to variation in the overall scale. In other words, we allow the noise level to vary between training and test data with covariance $\mathbf{R}$ and $\sigma^2\mathbf{R}$, respectively. We mean "CFAR" with respect to both the shared noise covariance structure $\mathbf{R}$ and independent scaling σ^2 of the noise in the test data. This generalizes the meaning of "CFAR" in both the nonadaptive and adaptive detection literature, where "CFAR" is respect to a presumed noise level or shared covariance between test and training data, respectively.

The CFAR ASDs have the remarkable property that they are just the "sample versions" of their corresponding CFAR MSD's, with a known covariance $\mathbf{R}$ replaced by the sample covariance $\mathbf{S}$. The coherent ASD and the ASD are *not* sample versions of their corresponding MSDs. However, these two ASDs may be approximated as in [9] and [10] to obtain *adaptive matched filters* (AMFs). The *coherent AMF* is the sample version of the coherent MSD and the *AMF* is a sample version of the MSD. The AMFs are *not* GLRTs, but they may be used when the training and test data are scaled uniformly.

In the following treatment of matched and adaptive subspace detectors, we compare and contrast invariances and performances. Under ideal conditions, the coherent ASD and ASD typically outperform the coherent CFAR ASD and CFAR ASD, although the performance gain is small for low SNR and small sample support. However, the CFAR ASDs enjoy an extra invariance with respect to data scaling that makes them robust against system gains and deviations from the standard Gaussian model. Consequently, we expect the CFAR ASDs to find application in radar, sonar, data communication, time series analysis, and array processing, where this extra invariance to scaling will be desirable for operations in channels and systems with variable gains or non-Gaussian statistics. This point will be developed more fully in the sections to follow.

We conclude the paper with identical statistical decompositions for the ASDs, AMFs, and CFAR ASDs. These decompositions allow us to gain insight into the structure of adaptive detection statistics by decomposing them in terms of their nonadaptive counterparts and corruptive noise terms attributable to finite training-data support. They also allow us to quantitatively characterize the random effects of adaptation, simplify Monte-Carlo simulations, more simply derive exact distributions, and compute receiver operating characteristic (ROC) curves.

II. Problem of Detecting A Subspace Signal

The problems we study are these. The parameter $\underline{\theta}$ locates a signal $\underline{s} = \mathbf{\Psi}\underline{\theta} \in \mathbb{C}^N$ in the signal subspace $\langle\mathbf{\Psi}\rangle$ of dimension p. That is, $\mathbf{\Psi} \in \mathbb{C}^{N\times p}$, which is the linear space of $(N \times p)$ complex matrices. This signal is scaled by μ, and the channel adds scaled noise $\sigma\underline{w}$ to the signal to produce the measurement $\underline{y} = \mu\underline{s} + \sigma\underline{w}$, which is distributed as $CN_N[\mu\mathbf{\Psi}\underline{\theta}, \sigma^2\mathbf{R}]$, where CN_N denotes the complex-normal density of the N-dimensional complex measurement $\underline{y}$. The problem is to test hypotheses about the parameter μ, indicating presence or absence of the signal, under various assumptions about the parameters $\mathbf{\Psi}$, $\mathbf{R}$, σ^2, and $\underline{\theta}$.

Throughout our developments, we define the whitened signal mode matrix $\mathbf{\Phi}$ and the whitened measurement $\underline{z}$ as follows:

$$\mathbf{\Phi} = \mathbf{R}^{-(1/2)}\mathbf{\Psi}, \quad \text{and} \quad \underline{z} = \mathbf{R}^{-(1/2)}\underline{y}. \tag{1}$$

Then, when the detectors are written in terms of whitened variables, these relations may be used to express the detectors in the original coordinates.

Whenever we refer to the "*rank-1 case*," we are discussing the situation where the dimension of the signal subspace $\langle\mathbf{\Psi}\rangle$ is 1. We denote this one-dimensional (1-D) complex subspace by $\langle\underline{\psi}\rangle$, where $\underline{\psi} \in \mathbb{C}^N$. In this case, the parameter $\underline{\theta}$ is the complex phase term $\theta = e^{j\alpha}$, and the signal $\underline{s}$ is the phased vector $\underline{s} = \underline{\psi}e^{j\alpha}$. The whitened signal vector $\underline{\phi}$ is then $\underline{\phi} = \mathbf{R}^{-(1/2)}\underline{\psi}$.

The matched subspace detection problem is as follows.

Measure $\underline{y} \sim CN_N[\mu\mathbf{\Psi}\underline{\theta}, \sigma^2\mathbf{R}]$, with the signal modes $\mathbf{\Psi}$, and the noise covariance $\mathbf{R}$ known. Depending on the problem specification, the noise variance σ^2 and the location parameter $\underline{\theta}$ may be known or unknown. Test the hypothesis H_0: $\mu = 0$ (noise only) versus the alternative H_1: $\mu > 0$ (signal plus noise).

Throughout this paper, we use the notation $\sim$ to mean "is distributed as."

A. MSD for Coherent Detection

A coherent signal can be written as $\underline{s} = \mathbf{\Psi}\underline{\theta} = \underline{\psi}e^{j\alpha}$. In the matched subspace problem for *coherent* detection, the noise variance σ^2 and the location parameter $\underline{\theta}$ are both known. When $\underline{\theta}$ is completely known, then both $\underline{\psi}$ and $e^{j\alpha}$ are known, which is the coherent rank-1 case. A slight modification of standard results (see, for example, [1] and [2]) produces the MSD for coherent detection:

$$\max[0, \operatorname{Re}\{n\}] \gtrless \eta \tag{2}$$

where n is the whitened matched-filter statistic

$$n = \frac{e^{-j\alpha}\underline{\phi}^\dagger \underline{z}}{\sqrt{\underline{\phi}^\dagger \underline{\phi}}\sigma} \sim CN_1\left[\frac{\mu}{\sigma}\sqrt{\underline{\phi}^\dagger \underline{\phi}}, 1\right]. \tag{3}$$

In the original coordinates, this has the form

$$n = \frac{e^{-j\alpha}\underline{\psi}^\dagger \mathbf{R}^{-1}\underline{y}}{\sqrt{\underline{\psi}^\dagger \mathbf{R}^{-1}\underline{\psi}}\sigma} \sim CN_1\left[\frac{\mu}{\sigma}\sqrt{\underline{\psi}^\dagger \mathbf{R}^{-1}\underline{\psi}}, 1\right]. \tag{4}$$

Thus, the whitened matched filter statistic is complex normal, with noncentrality parameter

$$(\mu/\sigma)\sqrt{\underline{\phi}^\dagger \underline{\phi}} = (\mu/\sigma)\sqrt{\underline{\psi}^\dagger \mathbf{R}^{-1}\underline{\psi}}$$

which is the output "voltage" SNR (which equals signal-to-noise ratio when squared). This means the detector statistic n is an unbiased estimator of the output voltage SNR, where the estimator may be thought of as

$$\mathrm{Re}\{n\} = \frac{\hat{\mu}}{\sigma}\sqrt{\underline{\phi}^\dagger \underline{\phi}}; \qquad \hat{\mu} = \max\left[0, \mathrm{Re}\left\{\frac{e^{-j\alpha}\underline{\phi}^\dagger \underline{z}}{\underline{\phi}^\dagger \underline{\phi}}\right\}\right] \tag{5}$$

where $\hat{\mu}$ is the unbiased, maximum-likelihood estimator of the signal level μ under H_1. We call this an SNR representation for n. This detector resolves the whitened measurement $\underline{z}$ in the whitened signal subspace $\underline{\phi}$. The statistic is invariant to translation of $\underline{z}$ in the perpendicular subspace $\langle\underline{\phi}\rangle^\perp$.

B. MSD for Subspace Detection

In the matched subspace detection problem for noncoherent detection, the noise variance σ^2 is known, but the location parameter $\underline{\theta}$ is unknown. The MSD for noncoherent detection is [1], [2]

$$\chi^2 \gtrless \eta \tag{6}$$

where χ^2 is the matched subspace detector statistic

$$\chi^2 = \frac{\underline{z}^\dagger \mathbf{P}_{\mathbf{\Phi}}\underline{z}}{\sigma^2} \sim \chi^2_p\left[\frac{\mu^2}{\sigma^2}\,\underline{\phi}^\dagger \underline{\phi}\right] \tag{7}$$

and $\mathbf{P}_{\mathbf{\Phi}}$ is the orthogonal projection onto the subspace $\langle\mathbf{\Phi}\rangle$: $\mathbf{P}_{\mathbf{\Phi}} = \mathbf{\Phi}(\mathbf{\Phi}^\dagger\mathbf{\Phi})^{-1}\mathbf{\Phi}^\dagger$. In this formula, $\chi^2_p[\cdot]$ denotes the density for a noncentral complex chi-squared (or gamma) random variable. By this, we mean a chi-squared random variable, scaled by $1/2$, with p complex degrees of freedom (or $2p$ real degrees of freedom), and with a noncentrality parameter $(\mu^2/\sigma^2)\underline{\phi}^\dagger\underline{\phi}$. Thus, by measuring the energy $\underline{z}^\dagger\mathbf{P}_{\mathbf{\Phi}}\underline{z}$ in the whitened signal subspace $\langle\mathbf{\Phi}\rangle$, the detector χ^2 is estimating the output SNR $(\mu^2/\sigma^2)\underline{\phi}^\dagger\underline{\phi} = (\mu^2/\sigma^2)\underline{\psi}^\dagger\mathbf{R}^{-1}\underline{\psi}$. It does this by estimating $\mu\mathbf{\Phi}\underline{\theta}$ as $\mathbf{\Phi}\widehat{\mu\underline{\theta}} = \mathbf{\Phi}\mathbf{\Phi}^{\#}\underline{z} = \mathbf{P}_{\mathbf{\Phi}}\underline{z}$, where $\mathbf{\Phi}^{\#} = (\mathbf{\Phi}^\dagger\mathbf{\Phi})^{-1}\mathbf{\Phi}^\dagger$ is the pseudo-inverse of $\mathbf{\Phi}$. This gives the unbiased, maximum-likelihood estimate of $\mu\mathbf{\Phi}\underline{\theta}$, which is then squared to estimate the output SNR. Alternatively, we can write the noncoherent MSD statistic in its SNR representation as

$$\chi^2 = \frac{\|\widehat{\mu\underline{\phi}}\|^2}{\sigma^2}; \quad \widehat{\mu\underline{\phi}} = e^{-j\alpha}\mathbf{\Phi}(\mathbf{\Phi}^\dagger\mathbf{\Phi})^{-1}\mathbf{\Phi}^\dagger\underline{z}. \tag{8}$$

This detector resolves the energy of the measurement $\underline{z}$ that lies in the whitened signal subspace $\langle\mathbf{\Phi}\rangle$. This energy is invariant to rotation of the measurement in the signal subspace $\langle\mathbf{\Phi}\rangle$ and to translation in the perpendicular subspace $\langle\mathbf{\Phi}\rangle^\perp$.

C. CFAR MSD for Coherent Detection

In the CFAR matched subspace detection problem for *coherent* detection, the location parameter $\underline{\theta}$ is known, but the noise variance σ^2 is unknown. The CFAR MSD for coherent detection is

$$\max[0, \mathrm{Re}\{\cos\}] \gtrless \eta \tag{9}$$

where cos is the "cosine" form of the coherent CFAR MSD [1], [2]:

$$\cos = \frac{e^{-j\alpha}\underline{\phi}^\dagger \underline{z}}{\sqrt{\underline{\phi}^\dagger \underline{\phi}}\sqrt{\underline{z}^\dagger \underline{z}}} = \frac{e^{-j\alpha}\underline{\psi}^\dagger \mathbf{R}^{-1}\underline{y}}{\sqrt{\underline{\psi}^\dagger \mathbf{R}^{-1}\underline{\psi}}\sqrt{\underline{y}^\dagger \mathbf{R}^{-1}\underline{y}}}. \tag{10}$$

The square of the cos statistic has a noncentral beta distribution, which is beta under H_0. The cos statistic is a monotone function of the CFAR MSD in its t form [1], [2]. That is

$$\cos = \frac{t}{\sqrt{|t|^2 + 1}} \tag{11}$$

where t is a complex t-distributed statistic

$$\sqrt{N-1}\,t = \frac{e^{-j\alpha}\underline{\phi}^\dagger \underline{z}}{\sqrt{\underline{\phi}^\dagger \underline{\phi}}\sqrt{\underline{z}^\dagger \mathbf{P}^\perp_{\underline{\phi}}\underline{z}/(N-1)}} \sim t_{N-1}\left[\frac{\mu}{\sigma}\sqrt{\underline{\phi}^\dagger \underline{\phi}}\right]$$
$$\mathbf{P}^\perp_{\underline{\phi}} = \mathbf{I} - \underline{\phi}(\underline{\phi}^\dagger\underline{\phi})^{-1}\underline{\phi}^\dagger. \tag{12}$$

In this formula, t_{N-1} denotes a t-distribution with $N-1$ complex degrees of freedom and with a noncentrality parameter $(\mu/\sigma)\sqrt{\underline{\phi}^\dagger\underline{\phi}}$. This so-called t-form of the CFAR MSD for coherent detection uses the estimate $\hat{\sigma} = \sqrt{\underline{z}^\dagger\mathbf{P}^\perp_{\underline{\phi}}\underline{z}/(N-1)}$ in place of the unknown scaling σ. The t-form of this statistic estimates the output voltage SNR $(\mu/\sigma)\sqrt{\underline{\phi}^\dagger\underline{\phi}} = (\mu/\sigma)\sqrt{\underline{\psi}^\dagger\mathbf{R}^{-1}\underline{\psi}}$, by estimating both μ and σ:

$$\mathrm{Re}\{t\} = \frac{1}{\sqrt{N-1}}\,\frac{\hat{\mu}}{\hat{\sigma}}\sqrt{\underline{\phi}^\dagger \underline{\phi}}$$
$$\hat{\mu} = \max\left[0, \mathrm{Re}\left\{\frac{e^{-j\alpha}\underline{\phi}^\dagger \underline{z}}{\underline{\phi}^\dagger \underline{\phi}}\right\}\right]$$
$$\hat{\sigma} = \frac{1}{\sqrt{N-1}}\sqrt{\underline{z}^\dagger \mathbf{P}^\perp_{\underline{\phi}}\underline{z}} \tag{13}$$

where $\hat{\sigma}$ is an unbiased estimate of the noise level σ. The cosine form measures the cosine of the angle that the test vector $\underline{z}$ makes with the signal vector $\langle\underline{\phi}\rangle$ (see [1, p. 141]. This angle is invariant to rotation in the subspace $\langle\underline{\phi}\rangle^\perp$ and to scaling of the test vector.

D. CFAR MSD for Subspace Detection

In the CFAR matched subspace detection problem for noncoherent detection, the location parameter $\underline{\theta}$ and the noise variance σ^2 are both unknown. The CFAR MSD is [1], [2]

$$\cos^2 \gtrless \eta \tag{14}$$

where $\cos^2$ is the cosine-squared form of the noncoherent CFAR MSD:

$$\cos^2 = \frac{\underline{z}^\dagger \mathbf{P}_{\mathbf{\Phi}} \underline{z}}{\underline{z}^\dagger \underline{z}}. \tag{15}$$

This statistic has a noncentral beta distribution, which is complex central beta under H_0. It may be written in terms of an F-statistic by constructing the monotone function

$$\cos^2 = \frac{F}{F+1} \tag{16}$$

where

$$\frac{N-p}{p} F = \frac{\underline{z}^\dagger \mathbf{P}_{\mathbf{\Phi}} \underline{z}/p}{\underline{z}^\dagger \mathbf{P}_{\mathbf{\Phi}}^{\perp} \underline{z}/(N-p)} \sim F_{p,\,N-p}\left[\frac{\mu^2}{\sigma^2}\, \underline{\phi}^\dagger \underline{\phi}\right]. \tag{17}$$

In this formula, $F_{p,\,N-p}$ denotes an F-distribution with p and $N-p$ degrees of freedom and noncentrality parameter $(\mu^2/\sigma^2)\underline{\phi}^\dagger \underline{\phi}$. This F-form of the detector statistic uses the estimator $\widehat{\sigma^2} = (1/N-p)\underline{z}^\dagger \mathbf{P}_{\mathbf{\Phi}}^{\perp} \underline{z}$ in place of the unknown σ^2. By measuring the ratio of energies per dimension, in the signal subspace $\langle \mathbf{\Phi} \rangle$ and the orthogonal subspace $\langle \mathbf{\Phi} \rangle^\perp$, the detector estimates the output SNR $(\mu^2/\sigma^2)\underline{\phi}^\dagger \underline{\phi} = (\mu^2/\sigma^2)\underline{\psi}^\dagger \mathbf{R}^{-1} \underline{\psi}$. It does this by estimating $\mu^2 \underline{\phi}^\dagger \underline{\phi}$ as $\underline{z}^\dagger \mathbf{P}_{\mathbf{\Phi}} \underline{z}$, and σ^2 as $(\underline{z}^\dagger \mathbf{P}_{\mathbf{\Phi}}^{\perp} \underline{z}/N-p)$. Alternatively, we can write the noncoherent CFAR MSD statistic in the SNR representation

$$F = \frac{1}{N-p} \frac{\|\widehat{\mu \underline{\phi}}\|^2}{\hat{\sigma}^2}. \tag{18}$$

The $\cos^2$ form estimates the squared direction cosine between the measurement $\underline{z}$ and the signal subspace $\langle \mathbf{\Phi} \rangle$, as illustrated in Fig. 1. This direction cosine is invariant to rotations in $\langle \mathbf{\Phi} \rangle$ and $\langle \mathbf{\Phi} \rangle^\perp$, and to scaling of the measurement, as illustrated.

E. Summary and General Comments

There are four detectors of interest for detecting subspace signals, which establish the basis for the adaptive detectors of the next section. They are summarized in Table I. The "CFAR" detectors were obtained in [1] and [11] as UMP-invariant detectors and are invariant to scaling of the test data. They were rederived in [2] as generalized likelihood ratio tests by incorporating maximum-likelihood estimates for the noise standard deviation, i.e., by estimating the scaling of test data. The noncoherent, rank-1 CFAR MSD was obtained in [12] and [13] as an asymptotic GLRT when the scaling of the test measurement is treated as a random parameter, i.e., under compound-Gaussian noise (multivariate-Gaussian with random amplitude scaling). Compound Gaussian noise is a special case of "elliptically contoured" (EC) random vectors, whose distribution depends on the measurement $\underline{y}$ only through the quadratic form $\underline{y}^\dagger \mathbf{R}^{-1} \underline{y}$. Then, contours of constant probability density for $\underline{y}$ are ellipsoids, and contours of the density for $\underline{z} = \mathbf{R}^{-(1/2)} \underline{y}$ are spheres, meaning that $\underline{z}$ is a spherically invariant random vector.

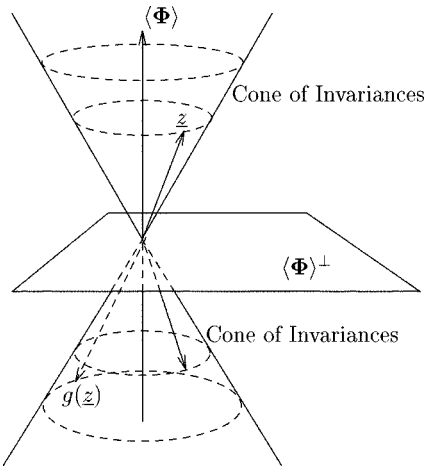

Fig. 1. *Geometry and invariances* of the CFAR matched subspace detector (CFAR MSD) for *noncoherent* detection. The statistic is invariant to transformations $g(\underline{z})$ that include scaling and rotations in the subspaces $\langle \mathbf{\Phi} \rangle$ and $\langle \mathbf{\Phi} \rangle^\perp$.

A result from the study of spherically invariant distributions is that the F (or cosine-squared) and t (or cosine) statistics presented in this section have the *same* distribution for any zero-mean EC distribution on $\underline{y}$ and are therefore CFAR across the class of such distributions (see, for example, [14, pp. 38–39]). This result is a consequence of the scale-invariance property of these statistics; that is, a zero-mean elliptical random vector has the stochastic representation $\underline{y} \stackrel{d}{=} \mathbf{R}^{(1/2)} \underline{u} \sqrt{\rho}$, where $\stackrel{d}{=}$ means "equal in distribution to," $\underline{u}$ is uniformly distributed on the N-dimensional unit spherical shell ($\|\underline{u}\|^2 = 1$), and ρ incorporates the radial dependence that distinguishes different EC distributions (see, for example, [15, p. 55–57]). Any scale-invariant statistic will not depend on ρ and will thus be distribution free within the class of zero-mean EC distributions on $\underline{y}$ (note that this is not true of the n and χ^2 statistics). This point is critical and not widely appreciated in the detection literature; with the CFAR MSDs, a given threshold will give the same the probability of false alarm for *any* multivariate density with ellipsoidal contours parameterized by $\underline{y}^\dagger \mathbf{R}^{-1} \underline{y} = constant$. The density could be Gaussian, compound-Gaussian, uniform on an ellipsoid, multivariate-Cauchy, etc.

Richmond uses this type of scale invariance argument to show that the *adaptive* Kelly GLRT (ASD) and AMF statistics are distribution free under H_0 over a whole class of EC distributions on the test and training data [15, p. 70], in which the concatenated vector consisting of stacked training and test data vectors is EC distributed, meaning that these vectors are uncorrelated but statistically dependent. He uses the fact that the ASD and AMF are invariant to uniform scaling of the test and training data. Richmond's argument can also be applied to the adaptive cosine (ACE) statistic of the next section because it has this invariance, in addition to being more generally invariant to independent scaling of test and training data.

TABLE I
TAXONOMY OF RESULTS FOR MATCHED SUBSPACE DETECTORS

Problem	Invariances with respect to transformations of $\underline{z}$	Detector Statistic $\mathbf{\Phi} = \mathbf{R}^{-\frac{1}{2}}\mathbf{\Psi}$; $e^{j\alpha}\underline{\phi} = \mathbf{\Phi}\underline{\theta}$; $\underline{z} = \mathbf{R}^{-\frac{1}{2}}\underline{y}$	SNR Representations $\hat{\mu} = e^{-j\alpha}(\underline{\phi}^\dagger\underline{\phi})^{-1}\underline{\phi}^\dagger\underline{z}$; $\widehat{\mu\underline{\phi}} = e^{-j\alpha}\mathbf{P}_{\mathbf{\Phi}}\underline{z}$; $\sqrt{N-p}\,\hat{\sigma} = \sqrt{\underline{z}^\dagger\mathbf{P}^\perp_{\mathbf{\Phi}}\underline{z}}$
Coherent MSD known: $\mathbf{\Psi}, \mathbf{R}, \sigma^2, \underline{\theta}$ unknown: μ	Translations in $\langle\underline{\phi}\rangle^\perp$	$n = \dfrac{e^{-j\alpha}\underline{\phi}^\dagger\underline{z}}{\sqrt{\underline{\phi}^\dagger\underline{\phi}}\,\sigma}$ (Equation 3)	$n = \dfrac{\hat{\mu}}{\sigma}\sqrt{\underline{\phi}^\dagger\underline{\phi}}$ (Equation 5)
MSD known: $\mathbf{\Psi}, \mathbf{R}, \sigma^2$ unknown: $\mu, \underline{\theta}$	Rotations in $\langle\mathbf{\Phi}\rangle$ Translations in $\langle\mathbf{\Phi}\rangle^\perp$	$\chi^2 = \dfrac{\underline{z}^\dagger\mathbf{P}_{\mathbf{\Phi}}\underline{z}}{\sigma^2}$ (Equation 7)	$\chi^2 = \dfrac{\|\widehat{\mu\underline{\phi}}\|^2}{\sigma^2}$ (Equation 8)
Coherent CFAR MSD known: $\mathbf{\Psi}, \mathbf{R}, \underline{\theta}$ unknown: μ, σ^2	Scaling Rotations in $\langle\underline{\phi}\rangle$	$cos = \dfrac{e^{-j\alpha}\underline{\phi}^\dagger\underline{z}}{\sqrt{\underline{\phi}^\dagger\underline{\phi}}\sqrt{\underline{z}^\dagger\underline{z}}} = \dfrac{t}{\sqrt{\|t\|^2+1}}$ (Equation 10)	$t = \frac{1}{\sqrt{N-1}}\dfrac{\hat{\mu}}{\hat{\sigma}}\sqrt{\underline{\phi}^\dagger\underline{\phi}}$ (Equation 13)
CFAR MSD known: $\mathbf{\Psi}, \mathbf{R}$ unknown: $\mu, \sigma^2, \underline{\theta}$	Scaling Rotations in $\langle\mathbf{\Phi}\rangle, \langle\mathbf{\Phi}\rangle^\perp$ (Figure 1)	$cos^2 = \dfrac{\underline{z}^\dagger\mathbf{P}_{\mathbf{\Phi}}\underline{z}}{\underline{z}^\dagger\underline{z}} = \dfrac{F}{F+1}$ (Equation 15)	$F = \frac{1}{N-p}\dfrac{\|\widehat{\mu\underline{\phi}}\|^2}{\hat{\sigma}^2}$ (Equation 18)

III. PROBLEM OF ADAPTIVELY DETECTING A SUBSPACE SIGNAL

How does one make the MSDs and CFAR MSDs of the previous section adaptive to unknown covariance structure $\mathbf{R}$ in order to derive ASDs and CFAR ASDs? To make this question meaningful, we must be clear about the assumptions under which the detector is applied. There are many variations on the experiment we describe, and each produces a different detector. We will depart slightly, but significantly, from Kelly's lead [5] and design the experiment as follows.

A sequence of K independent and identically distributed *training* vectors $\underline{x}_1, \underline{x}_2, \cdots \underline{x}_K$, each distributed as $CN_N[\underline{0}, \mathbf{R}]$, is measured in the training phase of the experiment. In the detection phase, a statistically independent *test* vector $\underline{y} \sim CN_N[\mu\mathbf{\Psi}\underline{\theta}, \sigma^2\mathbf{R}]$ is measured, and from it, the hypothesis H_0: $\mu = 0$ (noise only) is tested against the alternative H_1: $\mu > 0$ (signal plus noise).

In this experiment, we generalize Kelly's original experiment by allowing the covariance matrix for the test vector, namely $\sigma^2\mathbf{R}$, to differ by a scale constant σ^2 from the covariance matrix for the training data $\mathbf{R}$. This generalization produces new adaptive detectors, with extra invariances that Kelly's detector [5], and its derivative AMF forms [9], [10], do not have.

We organize the training vectors into the data matrix $\mathbf{X} = [\underline{x}_1, \cdots \underline{x}_K]$ and call $(\mathbf{X}, \underline{y})$ the *composite* measurement. The joint density function for the composite measurement, under the alternative H_1, is then

$$f_1(\mathbf{X}, \underline{y}) = f_1(\underline{y})\prod_{i=1}^{K} f(\underline{x}_i)$$
$$= \frac{1}{\pi^N \det(\sigma^2\mathbf{R})}\exp\left\{-\frac{1}{\sigma^2}(\underline{y}-\mu\mathbf{\Psi}\underline{\theta})^\dagger\mathbf{R}^{-1}(\underline{y}-\mu\mathbf{\Psi}\underline{\theta})\right\}$$
$$\cdot\prod_{i=1}^{K}\frac{1}{\pi^N\det(\mathbf{R})}\exp\left\{-\underline{x}_i^\dagger\mathbf{R}^{-1}\underline{x}_i\right\}. \quad (19)$$

Under H_0, the density is $f_0(\mathbf{X}, \underline{y}) = f_1(\mathbf{X}, \underline{y})|_{\mu=0}$. The density f_1 may be rewritten as

$$f_1(\mathbf{X}, \underline{y}) = \left\{\frac{1}{\pi^N\det(\mathbf{R})\sigma^{(2N/K+1)}} \cdot \exp\{-tr(\mathbf{R}^{-1}\mathbf{T}_1)\}\right\}^{K+1} \quad (20)$$

where $\mathbf{T}_1$ is the composite *sample* covariance matrix constructed from both the training and test data:

$$\mathbf{T}_1 = \frac{K}{K+1}\mathbf{S} + \frac{1}{K+1}\frac{1}{\sigma^2}(\underline{y}-\mu\mathbf{\Psi}\underline{\theta})(\underline{y}-\mu\mathbf{\Psi}\underline{\theta})^\dagger \quad (21)$$

$$\mathbf{S} = \frac{1}{K}\sum_{i=1}^{K}\underline{x}_i\underline{x}_i^\dagger. \quad (22)$$

In the section to follow, we extend the GLRT methodologies of [2], [5], and [6] to determine the GLRT tests

$$\hat{\lambda}(\mathbf{X}, \underline{y}) = \frac{\hat{f}_1(\mathbf{X}, \underline{y})}{\hat{f}_0(\mathbf{X}, \underline{y})} \quad (23)$$

where the carets denote the generalized likelihood ratios that result from maximizing the likelihood with respect to parameters which are unknown, such as $\mathbf{R}$, σ^2, μ, and $\underline{\theta}$.

When the noise scaling σ^2 is assumed to be known, this GLRT procedure yields coherent and *multirank* versions of the noncoherent detector of Kelly [5]. These detectors are *not* sample-matrix versions of the coherent MSD and MSD, meaning that a sample covariance does not simply replace a known covariance in the detector statistic. However, when σ^2 is unknown, maximizing the likelihood functions over this additional parameter yields CFAR ASDs that *are* sample-matrix versions of the CFAR MSDs [6]. Thus, we have the remarkable fact that the CFAR ASDs retain the *same form* as the CFAR

MSDs when σ^2 is unknown. We proceed case by case to outline these results.

In our discussion of the GLRT, we will find that the following *approximately* whitened signal modes $\hat{\mathbf{\Phi}}$, whitened signal vector $\underline{\hat{\phi}}$, and measurement $\underline{\hat{z}}$ arise naturally in the theory:

$$\hat{\mathbf{\Phi}} = \mathbf{S}^{-(1/2)}\mathbf{\Psi}, \quad \underline{\hat{\phi}} = \mathbf{S}^{-(1/2)}\underline{\psi}$$

and

$$\underline{\hat{z}} = \mathbf{S}^{-(1/2)}\underline{y}. \tag{24}$$

A. ASD for Coherent Detection

The adaptive subspace detection problem for coherent detection is the MSD problem outlined in Section II-A, with the modification that the noise covariance matrix $\mathbf{R}$ is unknown. When the maximum-likelihood estimates of $\mathbf{R}$ are used (under H_0 and H_1), together with the maximum-likelihood estimate of μ (constrained to be real and positive), the resulting generalized likelihood ratio (GLR) is [6]

$$\hat{\lambda}(\mathbf{X}, \underline{y}) = \left(\frac{1}{1 - [\max[0, \operatorname{Re}\{\hat{n}\}]]^2}\right)^{K+1}. \tag{25}$$

Constructing a monotonic function of this gives the GLRT

$$\max[0, \operatorname{Re}\{\hat{n}\}] \gtrless \eta \tag{26}$$

where $\hat{n}$ is the ASD statistic for coherent detection

$$\hat{n} = \frac{e^{-j\alpha}\underline{\hat{\phi}}^{\dagger}\underline{\hat{z}}}{\sqrt{\underline{\hat{\phi}}^{\dagger}\underline{\hat{\phi}}}\sqrt{K + \underline{\hat{z}}^{\dagger}\underline{\hat{z}}}} = \frac{e^{-j\alpha}\underline{\psi}^{\dagger}\mathbf{S}^{-1}\underline{y}}{\sqrt{\underline{\psi}^{\dagger}\mathbf{S}^{-1}\underline{\psi}}\sqrt{K + \underline{y}^{\dagger}\mathbf{S}^{-1}\underline{y}}} \tag{27}$$

where we have set the known parameter σ equal to unity to be consistent with the adaptive detection literature. This statistic may also be written as a monotone function of $\hat{\kappa}$, which is defined through

$$\hat{n} = \frac{\hat{\kappa}}{\sqrt{|\hat{\kappa}|^2 + 1}}. \tag{28}$$

Then, an alternate representation for the coherent ASD is

$$\operatorname{Re}\{\hat{\kappa}\} = \frac{\hat{\hat{\mu}}}{\sqrt{K + \underline{\hat{z}}^{\dagger}\mathbf{P}^{\perp}_{\underline{\hat{\phi}}}\underline{\hat{z}}}}\sqrt{\underline{\hat{\phi}}^{\dagger}\underline{\hat{\phi}}}$$

$$\hat{\hat{\mu}} = \max\left[0, \operatorname{Re}\left\{\frac{e^{-j\alpha}\underline{\hat{\phi}}^{\dagger}\underline{\hat{z}}}{\sqrt{\underline{\hat{\phi}}^{\dagger}\underline{\hat{\phi}}}}\right\}\right] \tag{29}$$

where we employ the double-hat notation to indicate that $\hat{\hat{\mu}}$ is an *adaptive* estimator of the signal amplitude μ affected by the training data. The quadratic form $\underline{\hat{z}}^{\dagger}\mathbf{P}^{\perp}_{\underline{\hat{\phi}}}\underline{\hat{z}}$ would be proportional to the maximum-likelihood estimator of the relative noise scaling σ^2 were σ^2 not constrained to be unity (we will discuss this further in Section III-C).

The coherent ASD statistic $\hat{n}$ extends Kelly's detector [5] to coherent problems. It is not quite the *sample-matrix* detector that one obtains by just replacing $\mathbf{R}$ with $\mathbf{S}$ in (4). However, it does become this detector if the denominator term $K + \underline{\hat{z}}^{\dagger}\underline{\hat{z}}$ is well approximated by K. Then, it generalizes the "adaptive matched filter" (AMF) of [9], [10] to coherent problems, as follows:

$$\hat{r} = \frac{e^{-j\alpha}\underline{\hat{\phi}}^{\dagger}\underline{\hat{z}}}{\sqrt{K\underline{\hat{\phi}}^{\dagger}\underline{\hat{\phi}}}} = \frac{e^{-j\alpha}\underline{\psi}^{\dagger}\mathbf{S}^{-1}\underline{y}}{\sqrt{K\underline{\psi}^{\dagger}\mathbf{S}^{-1}\underline{\psi}}} = \frac{1}{\sqrt{K}}\hat{\hat{\mu}}\sqrt{\underline{\hat{\phi}}^{\dagger}\underline{\hat{\phi}}}. \tag{30}$$

The detector $\hat{r}$ resolves the projection of the adaptively whitened measurement $\underline{\hat{z}}$ in the adaptively whitened signal subspace $\langle\underline{\hat{\phi}}\rangle$. The interpretation and invariances remain those of Section II-A.

B. ASD for Subspace Detection

The adaptive MSD problem for noncoherent detection is the problem outlined in Section II-B, with the modification that the noise covariance matrix $\mathbf{R}$ is unknown. The maximum-likelihood estimates of $\mathbf{R}$ (under H_0 and H_1), and the product $\mu\underline{\theta}$, yield the GLR [6]

$$\hat{\lambda}(\mathbf{X}, \underline{y}) = \left(\frac{1}{1 - \widehat{\chi^2}}\right)^{K+1}. \tag{31}$$

A monotone function of this gives the GLRT

$$\widehat{\chi^2} \gtrless \eta \tag{32}$$

where $\widehat{\chi^2}$ is the ASD statistic for noncoherent detection

$$\widehat{\chi^2} = \frac{\underline{\hat{z}}^{\dagger}\mathbf{P}_{\hat{\mathbf{\Phi}}}\underline{\hat{z}}}{K + \underline{\hat{z}}^{\dagger}\underline{\hat{z}}} = \frac{\widehat{\kappa^2}}{\widehat{\kappa^2} + 1}. \tag{33}$$

In terms of $\widehat{\kappa^2}$, this statistic has the SNR representation

$$\widehat{\kappa^2} = \frac{\left\|\widehat{\mu\underline{\phi}}\right\|^2}{K + \underline{\hat{z}}^{\dagger}\mathbf{P}^{\perp}_{\hat{\mathbf{\Phi}}}\underline{\hat{z}}}; \quad \widehat{\mu\underline{\phi}} = e^{-j\alpha}\hat{\mathbf{\Phi}}(\hat{\mathbf{\Phi}}^{\dagger}\hat{\mathbf{\Phi}})^{-1}\hat{\mathbf{\Phi}}^{\dagger}\underline{\hat{z}}. \tag{34}$$

This detector generalizes Kelly's test [5] to multidimensional subspaces [16], [17] and when the denominator is well-approximated by K it generalizes the AMF of Robey *et al.* [10] and Chen and Reed [9] to multidimensional subspaces [15], [18]

$$\widehat{r^2} = \frac{1}{K}\underline{\hat{z}}^{\dagger}\mathbf{P}_{\hat{\mathbf{\Phi}}}\underline{\hat{z}} = \frac{1}{K}\left\|\widehat{\mu\underline{\phi}}\right\|^2. \tag{35}$$

The AMF $\widehat{r^2}$ measures the energy of $\underline{\hat{z}}$ in the subspace $\langle\hat{\mathbf{\Phi}}\rangle$. The interpretations and invariances remain those of Section II-B.

C. CFAR ASD for Coherent Detection

The CFAR ASD problem for coherent detection is the problem outlined in Section II-C, with the modification that the noise covariance matrix $\mathbf{R}$ is unknown. The maximum-likelihood estimates of $\mathbf{R}$ and σ (under H_0 and H_1), as well as μ (constrained to be real and positive), yield the GLR [6]

$$\hat{\lambda}(\mathbf{X}, \underline{y}) = \left(\frac{1}{1 - [\max[0, \operatorname{Re}\{\widehat{\cos}\}]]^2}\right)^{N}. \tag{36}$$

A monotone function of this gives the GLRT

$$\max[0, \operatorname{Re}\{\widehat{\cos}\}] \gtrless \eta \tag{37}$$

where $\widehat{\cos}$ is the CFAR ASD statistic for coherent detection:

$$\widehat{\cos} = \frac{e^{-j\alpha}\hat{\underline{\phi}}^{\dagger}\hat{\underline{z}}}{\sqrt{\hat{\underline{\phi}}^{\dagger}\hat{\underline{\phi}}}\sqrt{\hat{\underline{z}}^{\dagger}\hat{\underline{z}}}} = \frac{e^{-j\alpha}\underline{\psi}^{\dagger}\mathbf{S}^{-1}\underline{y}}{\sqrt{\underline{\psi}^{\dagger}\mathbf{S}^{-1}\underline{\psi}}\sqrt{\underline{y}^{\dagger}\mathbf{S}^{-1}\underline{y}}}. \tag{38}$$

This detector generalizes the detector of [12] and [19] to coherent problems. It is just the "sample-matrix" version of the coherent CFAR MSD detector, and its $\hat{t}$ version is the sample version of t

$$\hat{t} = \frac{e^{-j\alpha}\hat{\underline{\phi}}^{\dagger}\hat{\underline{z}}}{\sqrt{\hat{\underline{\phi}}^{\dagger}\hat{\underline{\phi}}}\sqrt{\hat{\underline{z}}^{\dagger}\mathbf{P}^{\perp}_{\hat{\underline{\phi}}}\hat{\underline{z}}}}. \tag{39}$$

This statistic may be written as

$$\hat{t} = \frac{\hat{\hat{\mu}}}{\sqrt{\hat{\underline{z}}^{\dagger}\mathbf{P}^{\perp}_{\hat{\underline{\phi}}}\hat{\underline{z}}}}\sqrt{\hat{\underline{\phi}}^{\dagger}\hat{\underline{\phi}}}; \quad \hat{\hat{\mu}} = \max\left[0, \operatorname{Re}\left\{\frac{e^{-j\alpha}\hat{\underline{\phi}}^{\dagger}\hat{\underline{z}}}{\hat{\underline{\phi}}^{\dagger}\hat{\underline{\phi}}}\right\}\right] \tag{40}$$

where $\hat{\hat{\mu}}$ is an *adaptive* estimator of the signal amplitude μ. The quadratic form $\hat{\hat{\sigma}}^2 = (K-N+1/KN)\hat{\underline{z}}^{\dagger}\mathbf{P}^{\perp}_{\hat{\underline{\phi}}}\hat{\underline{z}}$ is the maximum-likelihood estimate of σ^2 under H_1, which is inserted into the likelihood-ratio to obtain the CFAR ASD, or ACE detector, as a GLRT in [6]. The interpretation and invariances remain those of Section II-C.

D. CFAR ASD for Subspace Detection

The CFAR ASD problem for noncoherent detection is the problem outlined in Section II-D, with the modification that the noise covariance matrix $\mathbf{R}$ is unknown. The maximum-likelihood estimates of $\mathbf{R}$ and σ (under H_0 and H_1), and the product $\mu\underline{\theta}$, yield the GLR [6]

$$\hat{\lambda}(\mathbf{X}, \underline{y}) = \left(\frac{1}{1-\widehat{\cos^2}}\right)^{M+1}. \tag{41}$$

A monotone function of this gives the GLRT

$$\widehat{\cos^2} \gtrless \eta \tag{42}$$

where $\widehat{\cos^2}$ is the CFAR ASD statistic for noncoherent detection

$$\widehat{\cos^2} = \frac{\hat{\underline{z}}^{\dagger}\mathbf{P}_{\hat{\mathbf{\Phi}}}\hat{\underline{z}}}{\hat{\underline{z}}^{\dagger}\hat{\underline{z}}}. \tag{43}$$

This detector generalizes the ACE of [12] and [19] to multidimensional subspaces. It is just the sample version of the CFAR MSD, and its $\hat{F}$ version is just the sample version of F

$$\hat{F} = \frac{\hat{\underline{z}}^{\dagger}\mathbf{P}_{\hat{\mathbf{\Phi}}}\hat{\underline{z}}}{\hat{\underline{z}}^{\dagger}\mathbf{P}^{\perp}_{\hat{\mathbf{\Phi}}}\hat{\underline{z}}}. \tag{44}$$

This statistic may be rewritten in its SNR form as

$$\hat{F} = \frac{K-N+1}{KN}\frac{\left\|\widehat{\hat{\mu}\underline{\phi}}\right\|^2}{\hat{\hat{\sigma}}^2}. \tag{45}$$

The interpretation and invariances remain those of Fig. 1.

E. Summary and Taxonomy

The results for ASDs and CFAR ASDs are summarized in Table II. This time, the noise structure $\mathbf{R}$ is unknown. Each of the ASD and CFAR ASD detectors is a GLRT. The CFAR ASDs are sample versions of their CFAR MSD counterparts and thus enjoy the same invariances. The approximations to the ASDs, which are termed "adaptive matched filters" (AMFs) to be consistent with the terminology of [10], are not GLRT. However, they are sample versions of their MSD counterparts.

The ASD and AMF statistics are invariant to the transformation group $g(\mathbf{X}, \underline{y}) = (c\mathbf{X}, c\underline{y})$, where c is a positive scalar for the coherent detectors, complex for the subspace detectors, meaning the test and training data may be scaled *identically* without changing these statistics. In contrast, the CFAR ASD statistics are invariant to the transformation group $g(\mathbf{X}, \underline{y}) = (c_1\mathbf{X}, c_2\underline{y})$, which means the training data and the test data may be scaled *differently* without affecting them. This is the key point when comparing ASDs, AMFs, and CFAR ASDs.

IV. Stochastic Representations for Multirank Adaptive Subspace Detectors

In this section, we will analyze how the ASDs in Section III are distributed by using statistically identical decompositions [3]. Using this approach, it is possible to represent any one of the adaptive detectors in Table II as a simple function of the same set of five independent random variables. We will carry out our derivation, in detail, for the AMF, when the signal subspace Ψ has dimension p. For this analysis, we make use of insights obtained from a similar, and simpler, analysis of the rank-1 case in [3] and [4]. To compare the robustness of the adaptive statistics, their distribution is analyzed in the general case where the measurement $\underline{y}$ has covariance $\sigma^2\mathbf{R}$, even though σ^2 is assumed to be unity in the standard derivations of the ASD and AMF detection statistics.

The derivation of statistical decompositions for the multirank ASD, AMF, and CFAR ASD can be outlined in six steps. The first four steps are analogous to those presented in [3] and [4] for the rank-1 detectors.

1) Apply the whitening transformation $\mathbf{R}^{-(1/2)}$ to the training and test data to generate the transformed signal modes $\mathbf{\Phi} = \mathbf{R}^{-(1/2)}\mathbf{\Psi}$ and test vector $\underline{z} = \mathbf{R}^{-(1/2)}\underline{y}$.
2) Next, apply a unitary transformation to rotate to a coordinate system in which the first $p+1$ basis vectors are set in the direction of $\mathbf{\Phi}$ and $\mathbf{P}^{\perp}_{\mathbf{\Phi}}\underline{z}$, where p is the number of signal modes or columns of $\mathbf{\Phi}$.
3) Resolve the inverse of the sample covariance matrix $\mathbf{S}$ onto the $(p+1)\times(p+1)$ subspace $\langle\mathbf{\Phi}, \mathbf{P}^{\perp}_{\mathbf{\Phi}}\underline{z}\rangle$.
4) Perform a change of variables on the elements of the resulting $(p+1)\times(p+1)$ covariance matrix so that these variables are now *statistically independent*.

For the 1-D case, where $p=1$, the procedure terminates here. For higher rank ASDs, these steps must be followed by two more steps: 5) rotation and 6) matrix partitioning. Those less interested in the details of the derivation may wish to skip to (62).

The analysis will be based on the statistical behavior of the *sample* correlation matrix $\mathbf{S} = (1/K)\sum_{i=1}^{K}\underline{x}_i\underline{x}_i^{\dagger}$, where the

TABLE II
TAXONOMY OF RESULTS FOR ADAPTIVE SUBSPACE DETECTORS. NOTE THAT THE LOWER FOUR PANELS HAVE THE SAME FORM AS THOSE OF TABLE I, WITH $\mathbf{S}$ REPLACING $\mathbf{R}$. THE UPPER TWO PANELS ARE GENERALIZATIONS OF THE KELLY DETECTOR [5]. ALL FOUR ASDS ARE GLRTS. THE AMFS ARE NOT

Problem	Invariances with respect to transformations of $\hat{\underline{z}}$	Detector Statistic $\hat{\mathbf{\Phi}} = \mathbf{S}^{-\frac{1}{2}}\mathbf{\Psi}$; $e^{j\alpha}\hat{\underline{\phi}} = \hat{\mathbf{\Phi}}\underline{\theta}$; $\hat{\underline{z}} = \mathbf{S}^{-\frac{1}{2}}\underline{y}$	SNR Representations $\hat{\mu} = e^{-j\alpha}(\hat{\underline{\phi}}^\dagger\hat{\underline{\phi}})^{-1}\hat{\underline{\phi}}^\dagger\hat{\underline{z}}$; $\widehat{\mu\underline{\phi}} = e^{-j\alpha}\mathbf{P}_{\hat{\mathbf{\Phi}}}\hat{\underline{z}}$; $\sqrt{\frac{KN}{K-N+1}}\hat{\sigma} = \sqrt{\hat{\underline{z}}^\dagger\mathbf{P}^\perp_{\hat{\mathbf{\Phi}}}\hat{\underline{z}}}$
Coherent ASD known: $\mathbf{\Psi}, \sigma^2, \underline{\theta}$ unknown: $\mathbf{R}, \mu$	Rotations in $\langle\hat{\underline{\phi}}\rangle^\perp$	$\hat{n} = \frac{e^{-j\alpha}\hat{\underline{\phi}}^\dagger\hat{\underline{z}}}{\sqrt{\hat{\underline{\phi}}^\dagger\hat{\underline{\phi}}}\sqrt{K+\hat{\underline{z}}^\dagger\hat{\underline{z}}}} = \frac{\hat{\kappa}}{\sqrt{|\hat{\kappa}|^2+1}}$ (Equation 27)	$\hat{\kappa} = \frac{\hat{\hat{\mu}}}{\sqrt{K+\hat{\underline{z}}^\dagger\mathbf{P}^\perp_{\hat{\underline{\phi}}}\hat{\underline{z}}}}\sqrt{\hat{\underline{\phi}}^\dagger\hat{\underline{\phi}}}$ (Equation 29)
ASD known: $\mathbf{\Psi}, \sigma^2$ unknown: $\mathbf{R}, \mu, \underline{\theta}$	Rotations in $\langle\hat{\mathbf{\Phi}}\rangle$ Rotations in $\langle\hat{\mathbf{\Phi}}\rangle^\perp$	$\widehat{\chi^2} = \frac{\hat{\underline{z}}^\dagger\mathbf{P}_{\hat{\mathbf{\Phi}}}\hat{\underline{z}}}{K+\hat{\underline{z}}^\dagger\hat{\underline{z}}} = \frac{\widehat{\kappa^2}}{\widehat{\kappa^2}+1}$ (Equation 33)	$\widehat{\kappa^2} = \frac{\|\widehat{\mu\underline{\phi}}\|^2}{K+\hat{\underline{z}}^\dagger\mathbf{P}^\perp_{\hat{\underline{\phi}}}\hat{\underline{z}}}$ (Equation 34)
Coherent AMF known: $\mathbf{\Psi}, \sigma^2, \underline{\theta}$ unknown: $\mathbf{R}, \mu$	Translations in $\langle\hat{\underline{\phi}}\rangle^\perp$	$\hat{r} = \frac{e^{-j\alpha}\hat{\underline{\phi}}^\dagger\hat{\underline{z}}}{\sqrt{K\hat{\underline{\phi}}^\dagger\hat{\underline{\phi}}}}$ (Equation 30)	$\hat{r} = \frac{1}{\sqrt{K}}\hat{\mu}\sqrt{\hat{\underline{\phi}}^\dagger\hat{\underline{\phi}}}$ (Equation 30)
AMF known: $\mathbf{\Psi}, \sigma^2$ unknown: $\mathbf{R}, \mu, \underline{\theta}$	Rotations in $\langle\hat{\mathbf{\Phi}}\rangle$ Translations in $\langle\hat{\mathbf{\Phi}}\rangle^\perp$	$\widehat{r^2} = \frac{1}{K}\hat{\underline{z}}^\dagger\mathbf{P}_{\hat{\mathbf{\Phi}}}\hat{\underline{z}}$ (Equation 35)	$\widehat{r^2} = \frac{1}{K}\|\widehat{\mu\underline{\phi}}\|^2$ (Equation 35)
Coherent CFAR ASD known: $\mathbf{\Psi}, \underline{\theta}$ unknown: $\mathbf{R}, \mu, \sigma^2$	Scaling Rotations in $\langle\hat{\underline{\phi}}\rangle^\perp$	$\widehat{cos} = \frac{e^{-j\alpha}\hat{\underline{\phi}}^\dagger\hat{\underline{z}}}{\sqrt{\hat{\underline{\phi}}^\dagger\hat{\underline{\phi}}}\sqrt{\hat{\underline{z}}^\dagger\hat{\underline{z}}}} = \frac{\hat{t}}{\sqrt{|\hat{t}|^2+1}}$ (Equation 38)	$\hat{t} = \sqrt{\frac{K-N+1}{KN}}\frac{\hat{\mu}}{\hat{\sigma}}\sqrt{\hat{\underline{\phi}}^\dagger\hat{\underline{\phi}}}$ (Equation 40)
CFAR ASD known: $\mathbf{\Psi}$ unknown: $\mathbf{R}, \mu, \sigma^2, \underline{\theta}$	Scaling Rotations in $\langle\hat{\mathbf{\Phi}}\rangle, \langle\hat{\mathbf{\Phi}}\rangle^\perp$ (Figure 1)	$\widehat{cos^2} = \frac{\hat{\underline{z}}^\dagger\mathbf{P}_{\hat{\mathbf{\Phi}}}\hat{\underline{z}}}{\hat{\underline{z}}^\dagger\hat{\underline{z}}} = \frac{\hat{F}}{\hat{F}+1}$ (Equation 43)	$\widehat{F} = \frac{K-N+1}{KN}\frac{\|\widehat{\mu\underline{\phi}}\|^2}{\hat{\sigma}^2}$ (Equation 45)

training data used to build $\mathbf{S}$ is complex Gaussian distributed as $\underline{x}_i \sim CN_N[\underline{0}, \mathbf{R}]$. Then, the scaled covariance estimate $\mathbf{A} = K\mathbf{S}$ has a complex Wishart distribution $\mathbf{A} \sim CW[K, N; \mathbf{R}]$

$$f(\mathbf{A}) = \frac{\det(\mathbf{A})^{K-N}}{J(\mathbf{R})}\, e^{-tr(\mathbf{R}^{-1}\mathbf{A})}$$

where

$$J(\mathbf{R}) = \pi^{(1/2)N(N-1)}\Gamma(K)\Gamma(K-1) \cdots \Gamma(K-N+1)\det(\mathbf{R})^K. \quad (46)$$

This puts the noncoherent AMF statistic into the normalized form

$$\widehat{r^2} = \underline{y}^\dagger\mathbf{A}^{-1}\mathbf{\Psi}(\mathbf{\Psi}^\dagger\mathbf{A}^{-1}\mathbf{\Psi})^{-1}\mathbf{\Psi}^\dagger\mathbf{A}^{-1}\underline{y}. \quad (47)$$

Throughout the derivation, we will follow the lead of Reed *et al.* [20] by making use of the following theorem mentioned in [21]: If $\mathbf{A} \sim CW[K, N; \mathbf{R}]$ and $\mathbf{C} \in \mathbb{C}^{\mathbb{N}\times\mathbb{N}}$ is nonsingular, then $\mathbf{B} = \mathbf{CAC}^\dagger \sim CW[K, N; \mathbf{CRC}^\dagger]$. We will also make use of the following theorem, which describes how to construct statistically independent random matrices from partitioned Wishart matrices.

A. Partitioned Wishart Matrices

Theorem 1: Consider the whitened random vector $\underline{z} \sim CN[\underline{0}, \mathbf{I}_N]$, which we partition as $\underline{z}^\dagger = [\underline{x}^\dagger, \underline{y}^\dagger]$, where $\underline{x}^\dagger \in \mathbb{C}^q$ and $\underline{y}^\dagger \in \mathbb{C}^r$. Now, suppose we construct a data matrix $\mathbf{Z} = [\underline{z}_1, \underline{z}_2 \cdots \underline{z}_K] = \left[\begin{matrix}\mathbf{X}\\ \mathbf{Y}\end{matrix}\right]$ consisting of K realizations of $\underline{z}$. Partition the Wishart matrix $\mathbf{B} = \mathbf{ZZ}^\dagger \sim CW[K, N; \mathbf{I}_N]$ as follows:

$$\mathbf{C} = = \left[\begin{array}{c|c}\mathbf{C}_{xx} & \mathbf{C}_{xy}\\ \hline \mathbf{C}_{yx} & \mathbf{C}_{yy}\end{array}\right]$$

$$f(\mathbf{C}) = \frac{\det(\mathbf{C})^{K-N}}{J(\mathbf{I})}\, e^{-tr(\mathbf{C})}. \quad (48)$$

Then, the following change of variables produces random matrices that are *statistically independent*

$$\mathbf{H}_{xx} = \mathbf{C}_{xx} - \mathbf{C}_{xy}\mathbf{C}_{yy}^{-1}\mathbf{C}_{yx} \sim CW[K-r, q; \mathbf{I}_q]$$
$$\mathbf{H}_{yy} = \mathbf{C}_{yy} \sim CW[K, r, \mathbf{I}_r]$$
$$\mathbf{H}_{xy} = \mathbf{C}_{xy}\mathbf{C}_{yy}^{-(1/2)} \sim CN[\mathbf{0}, \mathbf{I}_r \otimes \mathbf{I}_q]. \quad (49)$$

$\mathbf{H}_{xx}$ and $\mathbf{H}_{yy}$ are Wishart with $(K-r)$ and K degrees of freedom, and $\mathbf{H}_{xy}$ is a matrix with complex normal columns.

This can be verified by substituting the change of variables into the density of (48) and computing the Jacobian of the transformation by using the techniques employed in [20], for example. Then, the joint density of $\mathbf{H}_{xx}$, $\mathbf{H}_{yy}$, and $\mathbf{H}_{xy}$ is

$$f(\mathbf{H}_{xx}, \mathbf{H}_{yy}, \mathbf{H}_{xy}) = \frac{1}{J(\mathbf{I})}\det(\mathbf{H}_{xx})^{(K-r)-q}e^{-tr(\mathbf{H}_{xx})} \cdot \det(\mathbf{H}_{yy})^{K-r}e^{-tr(\mathbf{H}_{yy})} \cdot e^{-tr(\mathbf{H}_{xy}\mathbf{H}_{xy}^\dagger)}. \quad (50)$$

The interpretation of this is as follows. The data matrices $\mathbf{X}$ and $\mathbf{Y}$ are uncorrelated (and because they are Gaussian, statistically independent). Further, we can apply the following projection operator to the rows of $\mathbf{X}$, which projects onto the row space of $\mathbf{Y}$: $\mathbf{P}_{\mathbf{Y}^\dagger} = \mathbf{Y}^\dagger(\mathbf{Y}\mathbf{Y}^\dagger)^{-1}\mathbf{Y}$. This allows us to decompose $\mathbf{X}$ as a superposition of two parts ($\mathbf{X} = \hat{\mathbf{X}} + \hat{\mathbf{E}}$), where $\hat{\mathbf{X}}$ is a *sample* version of the Wiener-filter estimator of $\mathbf{X}$, and $\hat{\mathbf{E}}$ is the residual adaptive error

$$\begin{aligned}\hat{\mathbf{X}} &= \mathbf{X}\mathbf{P}_{\mathbf{Y}^\dagger} = \mathbf{X}\mathbf{Y}^\dagger(\mathbf{Y}\mathbf{Y}^\dagger)^{-1}\mathbf{Y} = \hat{\mathbf{R}}_{xy}\hat{\mathbf{R}}_{yy}^{-1}\mathbf{Y} \\ \hat{\mathbf{E}} &= \mathbf{X}\mathbf{P}_{\mathbf{Y}^\dagger}^{\perp} = \mathbf{X} - \hat{\mathbf{X}}\end{aligned} \tag{51}$$

where $\hat{\mathbf{R}}_{xy}$ and $\hat{\mathbf{R}}_{yy}$ are the adaptive estimators of $\mathbf{R}_{xy}$ and $\mathbf{R}_{yy}$, respectively. Then $\mathbf{H}_{xx} = \hat{\mathbf{E}}\hat{\mathbf{E}}^\dagger = K(\hat{\mathbf{R}}_{xx} - \hat{\mathbf{R}}_{xy}\hat{\mathbf{R}}_{yy}^{-1}\hat{\mathbf{R}}_{yx}^\dagger)$ gives the adaptive estimate of the error covariance, and $\mathbf{H}_{yy} = K\hat{\mathbf{R}}_{yy}$ gives the estimate of $\mathbf{R}_{yy}$. The remaining matrix $\mathbf{H}_{xy}$ can be attributed to $\hat{\mathbf{X}}$ by constructing the unitary transformation $\mathbf{U} = [\mathbf{U}_{\mathbf{Y}}, \mathbf{U}_\perp]$, where $\mathbf{U}_{\mathbf{Y}} = \mathbf{Y}^\dagger(\mathbf{Y}\mathbf{Y}^\dagger)^{-(1/2)}$, and $\mathbf{Y}\mathbf{U}_\perp = \mathbf{0}$. Then, $\hat{\mathbf{X}}\mathbf{U}_{\mathbf{Y}} = \mathbf{X}\mathbf{U}_{\mathbf{Y}} = \mathbf{X}\mathbf{Y}^\dagger(\mathbf{Y}\mathbf{Y}^\dagger)^{-(1/2)} = \mathbf{H}_{xy}$. (Theorem 1 is a little more general than the standard theorem on Schur complements of Wishart matrices in its treatment of the matrix $\mathbf{H}_{xy}$. It generalizes to arbitrary $\mathbf{R}$, but the $\mathbf{R} = \mathbf{I}$ case is all that is needed for the results of this paper. Please note that $\underline{x}$ and $\underline{y}$ are used differently in this section than in the rest of this paper.)

B. Step 1: Whitening

We now return to (47) and apply a series of transformations to simplify its form. The whitening transformation $\mathbf{R}^{-(1/2)}$ generates the transformed signal modes $\mathbf{\Phi}$, test vector $\underline{z}$, and whitened sample covariance matrix $\mathbf{B} = \mathbf{R}^{-(1/2)}\mathbf{A}(\mathbf{R}^{-(1/2)})^\dagger \sim CW[K, N; \mathbf{I}]$. The multirank AMF can now be written as

$$\widehat{r^2} = \underline{z}^\dagger\mathbf{B}^{-1}\mathbf{\Phi}(\mathbf{\Phi}^\dagger\mathbf{B}^{-1}\mathbf{\Phi})^{-1}\mathbf{\Phi}^\dagger\mathbf{B}^{-1}\underline{z}. \tag{52}$$

C. Step 2: Rotation Into the Signal and Measurement Subspace

Next, we rotate into a new coordinate system by applying the unitary transformation

$$\mathbf{U} = \left[\mathbf{\Phi}(\mathbf{\Phi}^\dagger\mathbf{\Phi})^{-(1/2)}, \quad \frac{\mathbf{P}_{\mathbf{\Phi}}^{\perp}\underline{z}}{\sqrt{\underline{z}^\dagger\mathbf{P}_{\mathbf{\Phi}}^{\perp}\underline{z}}}, \quad \tilde{\mathbf{U}}\right] \tag{53}$$

where $\tilde{\mathbf{U}}^\dagger\mathbf{\Phi} = \mathbf{0}$, and $\tilde{\mathbf{U}}^\dagger\mathbf{P}_{\mathbf{\Phi}}^{\perp}\underline{z} = \underline{0}$. In the new coordinate system, the signal matrix $\mathbf{\Phi}$ determines the direction of the first p basis vectors: $(\mathbf{U}^\dagger\mathbf{\Phi})^\dagger = \mathbf{\Delta}^\dagger = [(\mathbf{\Phi}^\dagger\mathbf{\Phi})^{(1/2)}, \mathbf{0}^\dagger]$. The $(p+1)$th basis vector is determined by that part of the test vector $\underline{z}$ that does not lie in the signal basis; therefore, the rotated test vector $\underline{\xi} = \mathbf{U}^\dagger\underline{z}$ has only $p+1$ nonzero components: $(\mathbf{U}^\dagger\underline{z})^\dagger = \underline{\xi}^\dagger = \sigma[\underline{n}^\dagger, g, \underline{0}^\dagger]$. These vectors are illustrated in Fig. 2. The vector $\underline{n} = (1/\sigma)(\mathbf{\Phi}^\dagger\mathbf{\Phi})^{-(1/2)}\mathbf{\Phi}^\dagger\underline{z} \sim CN[(\mu/\sigma)(\mathbf{\Phi}^\dagger\mathbf{\Phi})^{(1/2)}\underline{\theta}, \mathbf{I}_p]$ is the same vector that arises in the (nonadaptive) MSD statistic: $\chi^2 = \|\underline{n}\|^2$. The component g contributes to the estimated noise scaling, that is, $\sigma g = \sqrt{\underline{z}^\dagger\mathbf{P}_{\mathbf{\Phi}}^{\perp}\underline{z}}$ is an estimate of $\sqrt{N-p}\,\sigma$. Then, g^2 has a complex chi-squared distribution with $N-p$ degrees of freedom: $g^2 \sim \chi^2_{N-p}[0]$. Note that the ratio $(\|\underline{n}\|^2/g^2) = (\underline{z}^\dagger\mathbf{P}_{\mathbf{\Phi}}\underline{z}/\underline{z}^\dagger\mathbf{P}_{\mathbf{\Phi}}^{\perp}\underline{z})$ is the F-test version of the CFAR MSD in (17):

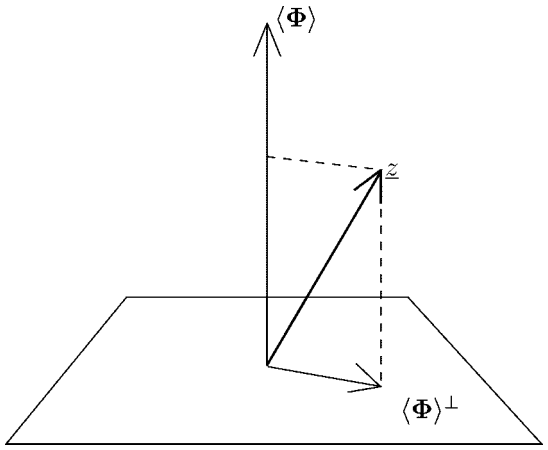

Fig. 2. *The rotation into the signal and measurement* subspace $\mathbf{U}$ sets the first $p+1$ basis vectors in the directions of $\mathbf{\Phi}$ and $\mathbf{P}_{\mathbf{\Phi}}^{\perp}\underline{z}$.

The transformed sample covariance $\mathbf{C} = \mathbf{U}^\dagger\mathbf{B}\mathbf{U}$ is also an estimate of identity and is Wishart distributed as $\mathbf{C} \sim CW[K, N; \mathbf{I}]$. In this new coordinate system, the multirank AMF statistic is

$$\widehat{r^2} = \underline{\xi}^\dagger\mathbf{C}^{-1}\mathbf{\Delta}(\mathbf{\Delta}^\dagger\mathbf{C}^{-1}\mathbf{\Delta})^{-1}\mathbf{\Delta}^\dagger\mathbf{C}^{-1}\underline{\xi}.$$

D. Step 3: Partitioning the Covariance Matrix

Since the signal and test vectors in the new coordinate system have, at most, $p+1$ nonzero components, we are really only concerned with the upper left $(p+1)\times(p+1)$ block of $\mathbf{C}^{-1}$. We partition $\mathbf{C}$ as

$$\mathbf{C} = \left[\begin{array}{c|c} \mathbf{C}_{11} & \mathbf{C}_{21}^\dagger \\ \hline \mathbf{C}_{21} & \mathbf{C}_{22} \end{array}\right] \tag{54}$$

and use Theorem 1 to identify $\mathbf{D} = \mathbf{C}_{11} - \mathbf{C}_{21}^\dagger\mathbf{C}_{22}^{-1}\mathbf{C}_{21}$ as a Wishart distributed $(p+1)\times(p+1)$ block, with reduced degrees of freedom, $\mathbf{D} \sim CW[K - N + p + 1, p+1, \mathbf{I}_{p+1}]$.

The multirank AMF can now be written in terms of the $(p+1)\times(p+1)$ Wishart matrix $\mathbf{D}$

$$\widehat{r^2} = \underline{\xi}_{p+1}^\dagger\mathbf{D}^{-1}\mathbf{\Delta}_{p+1}(\mathbf{\Delta}_{p+1}^\dagger\mathbf{D}^{-1}\mathbf{\Delta}_{p+1})^{-1}\mathbf{\Delta}_{p+1}^\dagger\mathbf{D}^{-1}\underline{\xi}_{p+1} \tag{55}$$

where $\underline{\xi}_{p+1}^\dagger = \sigma[\underline{n}^\dagger \quad g]$ contains the first $p+1$ elements of $\underline{\xi}^\dagger$, and $\mathbf{\Delta}_{p+1}^\dagger = [(\mathbf{\Phi}^\dagger\mathbf{\Phi})^{(1/2)}, \quad \underline{0}_p^\dagger]$.

We now apply Theorem 1 for a second time to identify the upper-left $p\times p$ block of $\mathbf{D}^{-1}$

$$\mathbf{D} = \left[\begin{array}{c|c} \mathbf{D}_{11} & \mathbf{D}_{21}^\dagger \\ \hline \mathbf{D}_{21} & \mathbf{D}_{22} \end{array}\right]$$

$$\mathbf{D}^{-1} = \left[\begin{array}{c|c} \mathbf{E}^{-1} & -\mathbf{E}^{-1}\mathbf{D}_{21}^\dagger\mathbf{D}_{22}^{-1} \\ \hline -\mathbf{D}_{22}^{-1}\mathbf{D}_{21}\mathbf{E}^{-1} & \mathbf{D}_{22}^{-1}\mathbf{D}_{21}\mathbf{E}^{-1}\mathbf{D}_{21}^\dagger\mathbf{D}_{22}^{-1} + \mathbf{D}_{22}^{-1} \end{array}\right]$$

where $\mathbf{E} = \mathbf{D}_{11} - \mathbf{D}_{21}^{\dagger}\mathbf{D}_{22}^{-1}\mathbf{D}_{21} \sim CW[K-N+p, p, \mathbf{I}_p]$. Then, we can write the multirank AMF as the quadratic form

$$\begin{aligned}\widehat{r^2} &= \sigma^2[\underline{n}^{\dagger} \mid g]\begin{bmatrix}\mathbf{E}^{-1} \\ \hline -\mathbf{D}_{22}^{-1}\mathbf{D}_{21}\mathbf{E}^{-1}\end{bmatrix}(\mathbf{\Phi}^{\dagger}\mathbf{\Phi})^{(1/2)} \\ &\quad \cdot \left[(\mathbf{\Phi}^{\dagger}\Phi)^{(1/2)}\mathbf{E}^{-1}(\mathbf{\Phi}^{\dagger}\mathbf{\Phi})^{(1/2)}\right]^{-1} \\ &\quad \cdot (\mathbf{\Phi}^{\dagger}\mathbf{\Phi})^{(1/2)}[\mathbf{E}^{-1} \mid -\mathbf{E}^{-1}\mathbf{D}_{21}^{\dagger}\mathbf{D}_{22}^{-1}]\begin{bmatrix}\underline{n} \\ \hline g\end{bmatrix} \\ &= \sigma^2[\underline{n}^{\dagger} \mid g] \\ &\quad \cdot \left[\begin{array}{c|c}\mathbf{E}^{-1} & -\mathbf{E}^{-1}\mathbf{D}_{21}^{\dagger}\mathbf{D}_{22}^{-1} \\ \hline -\mathbf{D}_{22}^{-1}\mathbf{D}_{21}\mathbf{E}^{-1} & \mathbf{D}_{22}^{-1}\mathbf{D}_{21}\mathbf{E}^{-1}\mathbf{D}_{21}^{\dagger}\mathbf{D}_{22}^{-1}\end{array}\right]\begin{bmatrix}\underline{n} \\ \hline g\end{bmatrix} \\ &= \sigma^2(\underline{n} - \mathbf{D}_{21}^{\dagger}\mathbf{D}_{22}^{-1}g)^{\dagger}\mathbf{E}^{-1}(\underline{n} - \mathbf{D}_{21}^{\dagger}\mathbf{D}_{22}^{-1}g). \end{aligned} \tag{56}$$

E. Step 4: Transformation to Independent Covariance Coefficients

While $\mathbf{D}$ is an estimate of identity, the covariance partitions ($\mathbf{D}_{11}$, $\mathbf{D}_{21}$, $\mathbf{D}_{22}$) are not independent, as can be seen from the Wishart distribution of $\mathbf{D}$. However, application of Theorem 1 generates new variables ($\mathbf{E}$, $\underline{h}_3$, h_2^2) that *are* statistically independent:

$$\begin{aligned}\mathbf{E} &= \mathbf{D}_{11} - \mathbf{D}_{21}^{\dagger}\mathbf{D}_{22}^{-1}\mathbf{D}_{21} \sim CW[K-N+p, p, \mathbf{I}_p] \quad \text{(a matrix)} \\ \underline{h}_3 &= -\frac{\mathbf{D}_{12}}{\sqrt{\mathbf{D}_{22}}} \sim CN[\underline{0}, \mathbf{I}_p] \quad \text{(a vector)} \\ h_2^2 &= \mathbf{D}_{22} \sim \chi^2_{K-N+p+1}[0] \quad \text{(a scalar).}\end{aligned} \tag{57}$$

This gives the following representation for the multirank AMF $\widehat{r^2}$ as a quadratic form in the Wishart matrix $\mathbf{E}$:

$$\widehat{r^2} = \sigma^2\left(\underline{n} + g\frac{\underline{h}_3}{h_2}\right)^{\dagger}\mathbf{E}^{-1}\left(\underline{n} + g\frac{\underline{h}_3}{h_2}\right) \tag{58}$$

where g and $h_2 = \sqrt{h_2{}^2}$ are Rayleigh-distributed random variables (square-root of a chi-squared).

F. Step 5: Another Rotation

For 1-D subspaces with $p = 1$, the stochastic representation of (58) is completely simplified since $\mathbf{E}$, $\underline{h}_3$, and $\underline{n}$ will be scalars, which we will later denote as $h_1{}^2$, h_3, and n. However, for multirank subspaces, rotation and matrix partitioning operations need to be applied once again in order to reduce $\mathbf{E}$ to a scalar. Letting $\underline{\lambda} = \underline{n} + g(\underline{h}_3/h_2)$, we apply the unitary transformation

$$\mathbf{V} = \left[\underline{\lambda}/\sqrt{\underline{\lambda}^{\dagger}\underline{\lambda}}, \quad \tilde{\mathbf{V}}\right] \tag{59}$$

where $\tilde{\mathbf{V}}^{\dagger}\underline{\lambda} = \underline{0}$. The transformation $\mathbf{V}$ rotates to a coordinate system in which the first basis vector is set in the direction of $\underline{\lambda}$, that is, $\mathbf{V}^{\dagger}\underline{\lambda} = \sqrt{\underline{\lambda}^{\dagger}\underline{\lambda}}\underline{\delta}$, where $\underline{\delta}^{\dagger} = [1 \quad 0 \quad \cdots \quad 0]$. This operation transforms the multirank AMF to

$$\widehat{r^2} = \sigma^2\underline{\lambda}^{\dagger}\underline{\lambda}\underline{\delta}^{\dagger}\mathbf{G}^{-1}\underline{\delta} \tag{60}$$

where $\mathbf{G}$ is the transformed sample covariance $\mathbf{G} = \mathbf{V}^{\dagger}\mathbf{E}\mathbf{V} \sim CW[K-N+p, p, \mathbf{I}_p]$.

G. Step 6: Another Matrix Partitioning

We can again use Theorem 1 to identify the upper-left element of $\mathbf{G}^{-1}$

$$\mathbf{G}^{-1} = \left[\begin{array}{c|c}h_1{}^{-2} & * \\ \hline * & *\end{array}\right] \tag{61}$$

where the scalar $h_1{}^2 = \mathbf{G}_{11} - \mathbf{G}_{21}^{\dagger}\mathbf{G}_{22}^{-1}\mathbf{G}_{21} \sim \chi^2_{K-N+1,1}[0]$. This gives the final decomposition for the multirank AMF statistic $\widehat{r^2}$ compared with the multirank MSD:

$$\widehat{r^2} = \frac{\sigma^2}{h_1{}^2}\left\|\underline{n} + g\frac{\underline{h}_3}{h_2}\right\|^2 \quad \text{vs.} \quad \chi^2 = \|\underline{n}\|^2. \tag{62}$$

(Recall from Section IV-C that the nonadaptive MSD is $\chi^2 = \|\underline{n}\|^2$.) In this formula, the normal random vectors $\underline{n}$ and $\underline{h}_3$, and the Rayleigh random variables g, h_1, and h_2, are all statistically independent.

The decompositions for the remaining multirank adaptive detectors, namely, the ASD $\widehat{\chi^2}$ and the CFAR ASD $\widehat{\cos^2}$, can be found by following the same sequence of six steps. The results are summarized in Table III.

H. Observations

Table III is organized so that decompositions for the MSD and CFAR MSD statistics are recorded in the left-hand column,

TABLE III
STATISTICAL DECOMPOSITIONS FOR THE MULTIRANK ASD, AMF, AND CFAR ASD STATISTICS

MSD	ASD
$\chi^2 = \|\underline{n}\|^2$	$\widehat{\chi^2} = \dfrac{\widehat{\kappa^2}}{\widehat{\kappa^2}+1}, \quad \widehat{\kappa^2} = \dfrac{\left\|\underline{n} + \frac{g}{h_2}\underline{h}_3\right\|^2}{h_1{}^2\left(\frac{1}{\sigma^2} + \frac{g^2}{h_2{}^2}\right)}$
	AMF
	$\widehat{r^2} = \dfrac{\sigma^2}{h_1{}^2}\left\|\underline{n} + \dfrac{g}{h_2}\underline{h}_3\right\|^2$
CFAR MSD	CFAR ASD
$\beta = \dfrac{F}{F+1}, \quad F = \left\|\dfrac{\underline{n}}{g}\right\|^2$	$\hat{\beta} = \dfrac{\hat{F}}{\hat{F}+1}, \quad \hat{F} = \dfrac{1}{h_1{}^2}\left\|\underline{n}\dfrac{h_2}{g} + \underline{h}_3\right\|^2$
$\underline{n}$, g, h_1, h_2, $\underline{h}_3$ all $\perp\!\!\!\perp$ (independent)	
$\underline{n} \sim CN[\frac{\mu}{\sigma}(\mathbf{\Phi}^{\dagger}\mathbf{\Phi})^{\frac{1}{2}}\underline{\theta}, \mathbf{I}_p], \quad g^2 \sim \chi^2_{N-p}[0]$	
$h_1{}^2 \sim \chi^2_{K-N+1}[0], \quad h_2{}^2 \sim \chi^2_{K-N+p+1}[0], \quad \underline{h}_3 \sim CN[\underline{0}, \mathbf{I}_p]$	

TABLE IV
STATISTICAL DECOMPOSITIONS FOR THE COHERENT ASD, AMF, AND CFAR ASD STATISTICS

Coh. MSD n	Coh. ASD $\hat{n}=\dfrac{\hat{\kappa}}{\sqrt{\lvert\hat{\kappa}\rvert^2+1}}, \quad \hat{\kappa}=\dfrac{n+\frac{g}{h_2}h_3}{h_1\sqrt{\frac{1}{\sigma^2}+\frac{g^2}{{h_2}^2}}}$
	Coh. AMF $\hat{r}=\dfrac{\sigma}{h_1}\left[n+\dfrac{g}{h_2}h_3\right]$
Coh. CFAR MSD $cos=\dfrac{t}{\sqrt{\lvert t\rvert^2+1}}, \quad t=\dfrac{n}{g}$	Coh. CFAR ASD $\widehat{cos}=\dfrac{\hat{t}}{\sqrt{\lvert\hat{t}\rvert^2+1}}, \quad \hat{t}=\dfrac{1}{h_1}\left[n\dfrac{h_2}{g}+h_3\right]$
$n,\ g,\ h_1,\ h_2,\ h_3$ all $\perp\!\!\!\perp$ (independent)	
$n\sim CN[\frac{\mu}{\sigma}\sqrt{\underline{\psi}^\dagger\mathbf{R}^{-1}\underline{\psi}},1], \quad g^2\sim\chi^2_{N-1}[0]$	
${h_1}^2\sim\chi^2_{M-N+1}[0],\ {h_2}^2\sim\chi^2_{M-N+2}[0],\ h_3\sim CN[0,1]$	

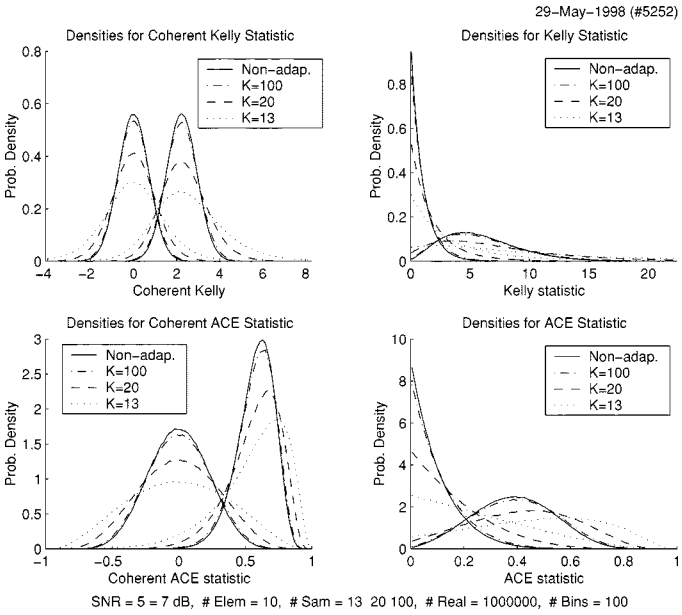

Fig. 3. *Densities* for the statistics $\sqrt{K}\,\hat{n}$, $K\widehat{\chi^2}$, $\widehat{\cos}$, and $\widehat{\cos^2}$. Densities are shown under both H_0 (symmetric about zero for the coherent statistics and weighted toward zero for the noncoherent statistics) and H_1 hypotheses. As the number of training samples K increases, the hypotheses become better separated. These densities were obtained from a Monte Carlo simulation using a million realizations of the *statistical decompositions* of Tables III and IV. Other parameters: dimension $N = 10$, SNR $= (\mu^2/\sigma^2)\underline{\psi}^\dagger\mathbf{R}^{-1}\underline{\psi} = 5$ (7 dB).

and decompositions for the ASD, AMF, and CFAR ASD statistics are recorded in the right-hand column. To illustrate the heuristic value of these decompositions, consider the multirank MSD statistic $\chi^2 = \|\underline{n}\|^2$ and the multirank AMF statistic $\widehat{r^2} = (\sigma^2/{h_1}^2)\|\underline{n}+(g/h_2)\underline{h}_3\|^2$. It is as if the statistic $\underline{n}$ undergoes the linear transformation $(\sigma/h_1)\,(\underline{n}+(g/h_2)\underline{h}_3)$ before its norm-squared is computed to obtain $\widehat{r^2}$. This linear transformation is of the form $a(\underline{n}+\underline{b})$, where a and $\underline{b}$ are independent noise terms. The influence of these additive and multiplicative noise terms decreases as the available training data increases. When $K\to\infty$, meaning $\mathbf{S}\to\mathbf{R}$ in distribution, then $(\sqrt{K}/h_1)\to 1$, and $(g/h_2)\underline{h}_3\to\underline{0}$ in distribution.

Similarly, consider the multirank CFAR MSD $F = (\|\underline{n}\|^2/g^2)$ and the multirank CFAR ASD $\hat{F} = (1/h_1^2)\|\underline{n}(h_2/g)+\underline{h}_3\|^2$. It is as if the statistic $(\underline{n}/g)$ undergoes the linear transformation $\hat{F} = (h_2/h_1)((\underline{n}/g)+(\underline{h}_3/h_2))$ before its norm-squared is computed. Again, the influence of the noise terms decreases as $K\to\infty$; then, $(h_2/h_1)\to 1$, and $(\underline{h}_3/h_2)\to\underline{0}$. The ASD has a similar interpretation.

The rank-1 versions of Table III, together with the decompositions for all of the rank-1 coherent detectors, have been obtained in [3] and [4]. The rank-1 versions that apply to coherent problems are summarized in Table IV. All of the observations just made about ASDs being equivalent to noise-corrupted versions of their MSD counterparts apply here as well.

It should be noted that these decompositions are not merely "stochastic representations" in the sense of being "equal in distribution to" (or "$\stackrel{d}{=}$"). They are identical decompositions on a realization by realization basis, i.e., for a given realization of $\underline{y}$, the adaptive statistics are expressed in terms of the corresponding realizations of n, g, etc.

In addition to the intuitive insight into the structure of the adaptive detectors that these decompositions bring, they also help in quantitative analysis in several ways. First, they make it possible to directly calculate the moments, such as mean and variance, of the adaptive detectors, without the need to find analytical expressions for their densities or characteristic functions. To see how the complexity of the density or characteristic functions can be bypassed, consider the decomposition for the coherent AMF $\hat{r}$ in Table IV. It is written as sums and products of n, g, $(1/h_1)$, $(1/h_2)$, and h_3. It is possible to write the moments of $\hat{r}$ *exactly* in terms of the moments of these five random variables, which are distributed as normal, Rayleigh, or the reciprocal of a Rayleigh. In this way, we can analyze how the output SNR of the adaptive statistics improves as the number of available training samples increases. This is discussed in full detail in [22] and [23].

A second advantage comes in performing Monte Carlo simulations. With the statistical decompositions, the generation of random realizations of an adaptive statistic can be achieved much more efficiently. If the gamma random variables are generated by summing normals, only $2(2K-N+4)$ normal random variables need to be generated for a realization of a rank-1 detector in Table IV, compared with $2N\times(2K+1)$ if the training and test data were generated explicitly. This is a significant reduction: about a factor of N for $K\gg N$. The results of Monte Carlo simulations performed in this manner are shown in Figs. 3–5. Fig. 3 shows how the densities under H_0 and H_1 separate as the available training data increases. Densities are shown for the rank-1 case $(p=1)$ of the ASD and CFAR ASD statistics, which in this figure and all subsequent figures are labelled Kelly and ACE. Fig. 4 shows the corresponding improvement in the receiver operating characteristics (ROCs) as the training data increases. Fig. 5 shows the same plots, but they are grouped to compare the detection performance of the ASD statistics $\hat{n}$ and $\widehat{\chi^2}$, and the

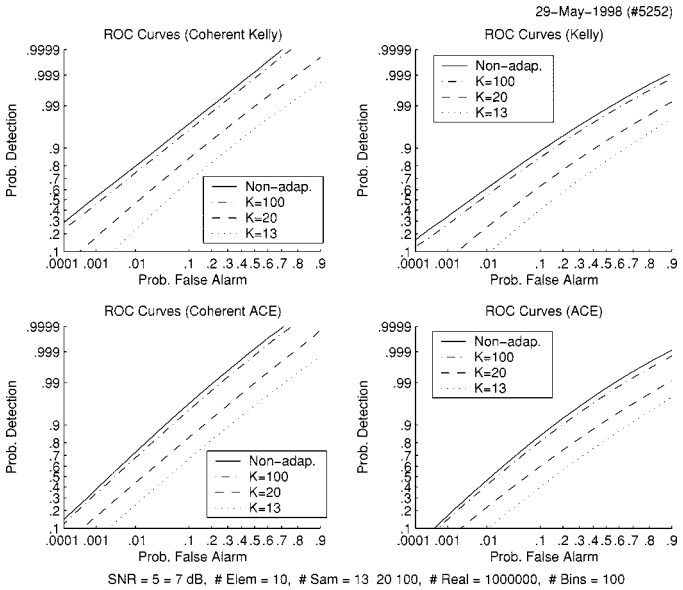

Fig. 4. *Receiver operating characteristics* (ROCs) for the $\hat{n}$, $\widehat{\chi^2}$, $\widehat{\cos}$, and $\hat{\beta}$ statistics. As the number of training samples K increases, the hypotheses become better separated. These curves are obtained from the same Monte Carlo simulation discussed in the caption to Fig. 3; they are plotted as though on "normal probability paper."

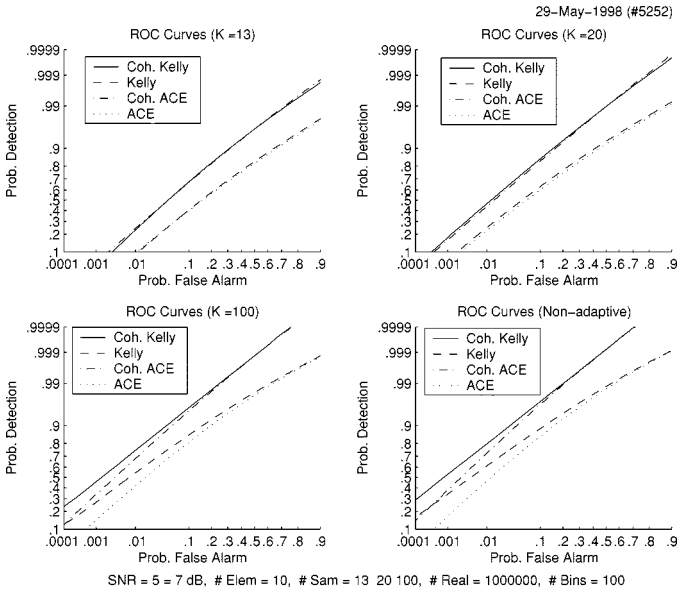

Fig. 5. ROCs when the number of training vectors K is 13, 20, 100, and infinite. These are the *same* curves as those in Fig. 4 but plotted to compare the behavior of the $\hat{n}$, $\widehat{\chi^2}$, $\widehat{\cos}$ and $\sim\hat{\beta}$ statistics.

CFAR ASD statistics $\widehat{\cos}$ and $\widehat{\cos^2}$, for various amounts of training data. As a general rule, the detectors that correspond to increased knowledge of system parameters (phase $e^{j\alpha}$ or σ^2) have better performance. However, as seen in Fig. 5, the performance loss of the CFAR ASD detectors compared with the ASD detectors becomes negligible for small training sets (even becoming a slight performance gain in places), where the CFAR ASD becomes a better approximation of the ASD.

A third advantage of the statistical decompositions is that they simplify the derivation of analytical expressions for the density functions of the ASDs, a topic that we discuss in detail in the next section.

V. Obtaining Distributions from the Statistical Decompositions

In this section, we use the statistical decompositions of Table III to find analytical expressions for the distributions of the ASD, AMF, and CFAR ASD statistics. Those less interested in some of the details may wish to skip to (68).

A. Multirank ASD, AMF, and CFAR ASD

We will start with the ASD $\widehat{\chi^2}$. Recall from Table III that the equivalent decomposition writes $\widehat{\chi^2}$ as $\widehat{\chi^2} = (\widehat{\kappa^2}/\widehat{\kappa^2}+1)$. The first step here is to separate out the mean of the vector $\underline{n}$. We denote by $\underline{\text{VSNR}}$ the mean of $\underline{n}$: $\underline{\text{VSNR}} = E[\underline{n}] = (\mu/\sigma)(\mathbf{\Phi}^\dagger\mathbf{\Phi})^{(1/2)}\underline{\theta}$. Then, from Table III, $\widehat{\kappa^2}$ may be written

$$\widehat{\kappa^2} \stackrel{d}{=} \frac{\left\|\underline{n}_0 + \dfrac{g}{h_2}\underline{h}_3 + \underline{\text{VSNR}}\right\|^2}{h_1^2\left(\dfrac{1}{\sigma^2} + \dfrac{g^2}{{h_2}^2}\right)} \tag{63}$$

where $\underline{n}_0$ and $\underline{h}_3$ are distributed as $CN[0, \mathbf{I}_p]$, and $\stackrel{d}{=}$ denotes "equal in distribution to." The sum $\underline{n}_0 + (g/h_2)\underline{h}_3$ is distributed the same as $\underline{m}\sqrt{1+(g^2/h_2^2)}$, where $\underline{m} \sim CN[0, \mathbf{I}_p]$ (this is true, irrespective of how h_2 and g are distributed, assuming they are independent of $\underline{n}_0$ and $\underline{h}_3$). Therefore, we can write $\widehat{\kappa^2}$ as

$$\begin{aligned}\widehat{\kappa^2} &\stackrel{d}{=} \frac{\left\|\underline{m}\sqrt{1+\dfrac{g^2}{h_2^2}} + \underline{\text{VSNR}}\right\|^2}{h_1^2\left(\dfrac{1}{\sigma^2}+\dfrac{g^2}{h_2^2}\right)} \\ &= \frac{\left\|\underline{m} + \dfrac{1}{\sqrt{1+\dfrac{g^2}{h_2^2}}}\underline{\text{VSNR}}\right\|^2}{h_1^2} \cdot \frac{1+\dfrac{g^2}{h_2^2}}{\dfrac{1}{\sigma^2}+\dfrac{g^2}{h_2^2}}.\end{aligned}$$

At this point, we can follow a line of reasoning similar to that of Kelly, *et al.* [5], [10], [24] and find the distribution conditioned on the ratio

$$b = \frac{1}{1+\dfrac{g^2}{{h_2}^2}}. \tag{64}$$

In the multidimensional case, g^2 and h_2^2 have different degrees of freedom, and b is distributed as $b \sim \beta_{K-N+p+1,\,N-p}$ [interestingly, in the rank-1 case ($p=1$), this has the same distribution as the Reed *et al.* normalized output SNR [5], [20]].

With this identification, the factor on the right can be rewritten with the following algebra:

$$\frac{\frac{1}{\sigma^2}+\frac{g^2}{h_2^2}}{1+\frac{g^2}{h_2^2}} = \frac{\frac{h_2^2}{\sigma^2}+g^2}{h_2^2+g^2}$$
$$= \frac{1}{\sigma^2}\left(\frac{h_2^2}{h_2^2+g^2}\right)+\frac{g^2}{h_2^2+g^2} = \frac{b}{\sigma^2}+(1-b)$$
$$= 1+\left(\frac{1}{\sigma^2}-1\right)b \tag{65}$$

and $\widehat{\kappa^2}$ can be rewritten as

$$\widehat{\kappa^2} \stackrel{d}{=} \frac{\left\|\underline{m}+\sqrt{b}\,\underline{\text{VSNR}}\right\|^2}{h_1^2}\cdot\frac{1}{1+\left(\frac{1}{\sigma^2}-1\right)b}. \tag{66}$$

Conditioned on b, the left-hand factor of $\widehat{\kappa^2}$ is an F-distributed random variable with $(p, K-N+1)$ degrees of freedom scaled by $(p/K-N+1)$

$$\widehat{\kappa^2}|b \stackrel{d}{=} \frac{\|\underline{\zeta}\|^2}{h_1^2}\cdot\frac{1}{1+\left(\frac{1}{\sigma^2}-1\right)b} \tag{67}$$

where $\underline{\zeta} \sim CN[\underline{\text{VSNR}}\cdot\sqrt{b}, \mathbf{I}_p]$ is a complex normal random vector whose mean is conditioned on b. We denote the distribution function of the scaled noncentral F by $F_{[\nu_n,\nu_d,a]}(\eta)$; we define this to be the probability that the ratio of a complex-chi-square with ν_n degrees of freedom and a noncentrality parameter a, divided by a complex chi-square with ν_d degrees of freedom is less than η. The conditional distribution of $\widehat{\kappa^2}$ is then given by

$$\Pr[\widehat{\kappa^2} \le \eta|b] = F_{[p,\,K-N+1,\,\text{SNR}\cdot b]}\left(\eta\left(1+\left(\frac{1}{\sigma^2}-1\right)b\right)\right) \tag{68}$$

where $\text{SNR} = \|\underline{\text{VSNR}}\|^2 = (\mu^2/\sigma^2)\underline{\theta}^\dagger(\mathbf{\Phi}^\dagger\mathbf{\Phi})\underline{\theta} = (\mu^2/\sigma^2)\underline{\phi}^\dagger\underline{\phi}$. Again, an expression for the final distribution can be obtained by integrating this distribution over the beta density of b.

Using the same procedure for the AMF $\widehat{r^2}$ yields, from Table III

$$\widehat{r^2} \stackrel{d}{=} \frac{\sigma^2\left(1+\frac{g^2}{h_2^2}\right)}{h_1^2}\left\|\underline{m}+\frac{1}{\sqrt{1+\frac{g^2}{h_2^2}}}\underline{\text{VSNR}}\right\|^2. \tag{69}$$

Conditioning on b, we obtain the following statistical equivalent for $\widehat{r^2}$:

$$\widehat{r^2}|b \stackrel{d}{=} \frac{\|\underline{\zeta}\|^2}{h_1^2}\cdot\frac{\sigma^2}{b} \tag{70}$$

where $\underline{\zeta}$ is defined as above. Then, the distribution of $\widehat{r^2}$ is given by

$$\Pr[\widehat{r^2} \le \eta|b] = F_{[p,\,K-N+1,\,\text{SNR}\cdot b]}\left(\eta\,\frac{b}{\sigma^2}\right). \tag{71}$$

Finally, we can go through the same procedure for the $\hat{F}$-version of the CFAR ASD, $\hat{\beta} = (\hat{F}/\hat{F}+1)$. From Table III

$$\hat{F} \stackrel{d}{=} \frac{\left(\frac{h_2^2}{g^2}+1\right)}{h_1^2}\left\|\underline{m}+\frac{1}{\sqrt{1+\frac{g^2}{h_2^2}}}\underline{\text{VSNR}}\right\|^2. \tag{72}$$

Conditioned on b, the statistical equivalent of $\hat{F}$ is

$$\hat{F}|b \stackrel{d}{=} \frac{\|\underline{\zeta}\|^2}{h_1^2}\cdot\frac{1}{1-b}. \tag{73}$$

Then, the distribution of $\hat{F}$ is given by

$$\Pr[\hat{F} \le \eta|b] = F_{[p,\,K-N+1,\,\text{SNR}\cdot b]}(\eta(1-b)) \tag{74}$$

which notably does not depend on σ. The probability of detection (PD) for one of these statistics is given by one minus the distribution function evaluated at the threshold. The probability of false alarm (PFA) is the PD when $\text{SNR} = 0$.

B. Coherent ASD, AMF, and CFAR ASD

Using the decompositions in Table IV, we can apply the same techniques to find distributions for the coherent versions of the ASD, AMF, and CFAR ASD. Conditioned on the beta parameter b and σ, they are all related to the noncentral t distribution

$$\Pr[\text{Re}(\hat{\kappa}) \le \eta|b] = T_{[K-N+1,\,\sqrt{\text{SNR}\cdot b}]}\cdot\left(\eta\sqrt{1+\left(\frac{1}{\sigma^2}-1\right)b}\right) \tag{75}$$

$$\Pr[\text{Re}(\hat{r}) \le \eta|b,\,\sigma] = T_{[K-N+1,\,\sqrt{\text{SNR}\cdot b}]}\left(\eta\,\frac{\sqrt{b}}{\sigma}\right) \tag{76}$$

$$\Pr[\text{Re}(\hat{t}) \le \eta|b] = T_{[K-N+1,\,\sqrt{\text{SNR}\cdot b}]}\left(\eta\sqrt{1-b}\right). \tag{77}$$

Here, we denote the distribution function of the scaled noncentral t by $T_{[\nu_d,\,\sqrt{a}]}(\eta)$; this we define to be the probability that the ratio of the real part of a complex normal with mean $\sqrt{a}$, divided by the square root of a complex chi-square with ν_d degrees of freedom, is less than η. Note that the distribution of $\text{Re}(\hat{t})$ does not depend on σ.

In summary, the statistical decompositions of Table IV may be used to get statistically equivalent random variables conditioned on the beta-distributed random variable b. Each of these conditional random variables is a linear transformation of either a t-distributed or an F-distributed random variable. This means that their distributions may be obtained by integrating a noncentral t- or F-distribution against a beta density to get integral representations for the distributions of the coherent and multirank

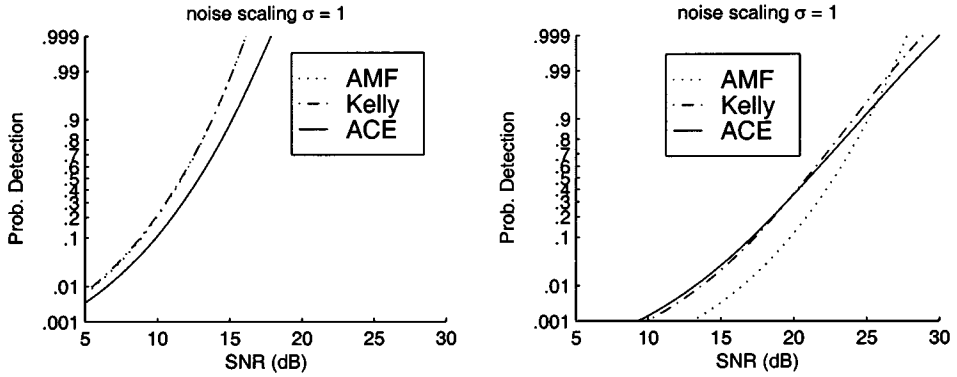

Fig. 6. *Detection probabilities* for (left) $K = 100$ and (right) $K = 25$; PFA $= 10^{-5}$, and $N = 20$.

versions of the ASD, AMF, and CFAR ASD statistics. The distribution results for the coherent ASD, AMF, and CFAR ASD statistics are original with this paper. The distributions for the rank-1 noncoherent detectors (in [5], [10], and [24]) arise when the noise scaling σ^2 is set to 1, and the signal subspace rank p is set equal to 1, in the expressions obtained here for the general multirank case.

VI. Numerical Performance Comparisons

Now, let us compare the detection performance of the ASD, AMF, and CFAR ASD statistics. We will consider the rank-1 case of these statistics ($p = 1$), referring to them in this case as the Kelly, AMF, and ACE to be consistent with the standard nomenclature of the radar literature. In the rank-1 case, the ASD (Kelly), AMF, and CFAR ASD (ACE) and associated statistical decompositions (which are obtained by setting $p = 1$ in Table III, or by magnitude squaring the expressions in Table IV) take on the form

$$\widehat{\chi^2} = \frac{|\underline{\psi}^\dagger \mathbf{S}^{-1}\underline{y}|^2}{(\underline{\psi}^\dagger \mathbf{S}^{-1}\underline{\psi})(K + \underline{y}^\dagger \mathbf{S}^{-1}\underline{y})} = \frac{\widehat{\kappa^2}}{\widehat{\kappa^2}+1}$$

$$\widehat{\kappa^2} = \frac{\left|n + \dfrac{g}{h_2}\, h_3\right|^2}{h_1^2\left(\dfrac{1}{\sigma^2} + \dfrac{g^2}{h_2^2}\right)} \tag{78}$$

$$\widehat{r^2} = \frac{|\underline{\psi}^\dagger \mathbf{S}^{-1}\underline{y}|^2}{\underline{\psi}^\dagger \mathbf{S}^{-1}\underline{\psi}} = \frac{\sigma^2}{h_1^2}\left|n + \frac{g}{h_2}\, h_3\right|^2 \tag{79}$$

$$\widehat{\cos^2} = \frac{|\underline{\psi}^\dagger \mathbf{S}^{-1}\underline{y}|^2}{(\underline{\psi}^\dagger \mathbf{S}^{-1}\underline{\psi})(\underline{y}^\dagger \mathbf{S}^{-1}\underline{y})} = \frac{\hat{F}}{\hat{F}+1}$$

$$\hat{F} = \frac{1}{h_1^2}\left|n\frac{h_2}{g} + h_3\right|^2. \tag{80}$$

We use expressions for the distribution functions obtained in Section V conditioned on b and σ. We also make use of a useful finite-sum expression for the noncentral-F distribution, which was found by Kelly [25]. Instead of numerical integration, we approximate the integral of the distribution, over the densities of b, by averaging realizations of distribution given by 1000 realizations of b. (Based on examination of these realizations, we estimate the uncertainty to be about $(1/100)$ the height of the plots; the qualitative behavior discussed here is still seen with far fewer realizations.)

The resulting detection curves for the Kelly, AMF, and ACE are shown in Fig. 6 for σ^2 equal to unity and $N = 20$. They are more easily interpreted by recognizing that the probability of detection of the nonadaptive cosine detector (CFAR MSD) is always upperbounded by that of the MSD; this is the price paid for estimating the noise scaling σ. The performance of the CFAR MSD only approaches that of the MSD when N is large compared with the SNR, resulting in a relatively good estimate of the noise scaling, as given by (13) $\sqrt{N-1}\hat{\sigma} = \sqrt{\underline{z}^\dagger \mathbf{P}^\perp_{\underline{\phi}}\underline{z}}$ (see [1]; this can also be verified, in terms of the variance and expectation of n and t under H_1 and H_0, using the analysis approach of [22]).

When $K = 100$ in Fig. 6, the training data support is relatively high, and consequently, the adaptive detectors behave close to their nonadaptive counterparts. Consistent with this observation, we can see that the AMF does well against the ACE at high SNRs, with the difference becoming negligible at low SNRs.

When $K = 25$ in Fig. 6, the training data support is relatively low. Here, the adaptive detectors rely on poor sample covariance estimates $\mathbf{S}$. ACE is not only invariant to scaling of the measurement $\underline{y}$ but is separately invariant to global scaling of the training data set $\{\underline{x}_i\}$ (and, thus, to scaling of $\mathbf{S}$). For this reason, we expect it to be more robust under conditions of small sample support. For $K = 25$, the ACE begins to take advantage of its scale invariance to $\mathbf{S}$ and overtakes the AMF at low SNR.

By comparing (78) with (79) and (80), one can see that the Kelly GLRT approaches the AMF for very high values of the sample support K and more closely approximates ACE for very low values of sample support. In Fig. 6, this can be observed in how close the Kelly curve is to that of the AMF for $K = 100$ and to the ACE for $K = 25$. In the very regime of small K in which the ACE performs relatively well, the Kelly begins to behave more like the ACE, which would be expected by considering the normalization term in (78) in the small K limit.

A. False Alarm Stability

These comparisons have been made under the idealized condition of $\sigma = 1$, which the Kelly and AMF assume but the ACE does not. When σ, which is the true relative scaling of the measurement, deviates from unity, the probability of false alarm

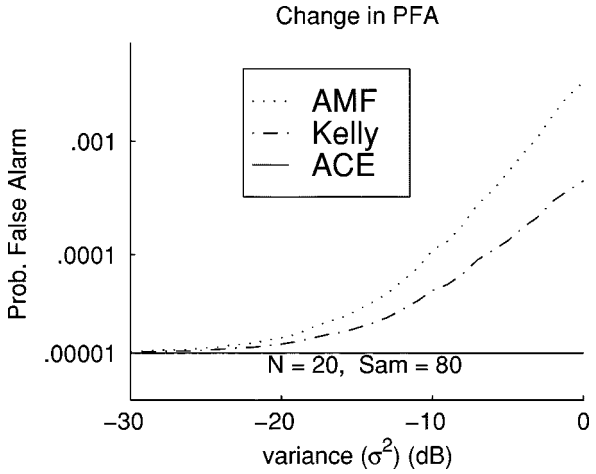

Fig. 7. *Change in false alarm probability for a constant threshold* as the variance of the noise scaling increases; $K = 80$, and $N = 20$.

(PFA) of the Kelly and AMF will be affected, but the PFA of the ACE will not be affected. In Fig. 7, we illustrate this effect by introducing a randomly fluctuating noise scaling. Under H_0, this is equivalent to the random amplitude scaling of the compound-Gaussian noise model of [12]. We chose a simple distribution for σ^2: complex chi-squared but normalized to have unit mean and with a variance equal to 1 divided by the number of degrees of freedom. Because ACE is invariant to scaling of $\underline{y}$, then under H_0, its PFA is completely insensitive to fluctuations in the noise scaling.

Conversely, while the Kelly and AMF may do well against the ACE in terms of plots of PD versus PFA, their thresholds cannot be set to achieve a desired PFA without requiring the user to have some knowledge of the scaling distribution. The ACE does not require this information; for a set threshold, its PFA does not depend on the statistics of the noise scaling, whereas the PFA of the AMF and Kelly may vary considerably, as illustrated in Fig. 7.

In summary, ACE sacrifices a small amount of SNR performance (about 2 dB in Fig. 6, for a PFA of 10^{-5}; results are similar at other PFAs), for low SNRs (less than N), or for small sample support. In exchange, it has scaling invariance and true CFAR performance against scale fluctuations in the test data.

VII. Conclusion

In this paper, we have offered a unified treatment of two classes of generalized likelihood ratio tests: the MSDs, which use a known noise covariance structure, and the ASDs, which use training data to estimate an unknown noise covariance structure. Both matched and adaptive subspace detectors may be further classified according to whether the test signal is completely specified (coherent) or parameterized (subspace) and according to whether the noise level is known or unknown. In the adaptive case, the unknown noise-level problem translates to an unknown scaling between the noise in the training data and test data; it is assumed that the training data accurately represents the noise *structure* but may not accurately represent the noise *level*.

Maximizing the likelihood ratios over this additional scaling parameter produces the cosine-based CFAR MSDs and the CFAR ASDs, which are invariant with respect to arbitrary scaling of the test data. In addition, the CFAR MSDs are CFAR with respect to the entire class of elliptically contoured distributions, which include compound-Gaussian distributions. It is interesting that the CFAR ASDs, which include the ACE statistic, have the *same form* as the CFAR MSDs, with the sample covariance replacing the known covariance. This is *not* true of the ASDs for known noise scaling, such as the Kelly GLRT [5], which does not take the form of the matched subspace detector.

The CFAR ASDs suffer some performance loss under the idealized scenario of homogeneity between training-data and test-data noise statistics. However, their invariance to test-data scaling makes them CFAR with respect to variation in the noise level between training and test data; other researchers have shown them to have robustness to more complicated inhomogeneities, such as changes in the power of clutter discretes [24].

We have presented a unified description of the statistical behavior of the class of ASDs, including those parameterized by multidimensional signal subspaces. We have shown that they each have an identical *statistical decomposition*, which is a simplified function of the same set of five statistically independent random variables. These same random variables appear in all such representations; they include the nonadaptive matched filter and the t/cosine statistic, plus three perturbing variables attributable to training data. In addition to their heuristic value, these representations provide some computational advantages. We have used them here for more efficient Monte Carlo simulations and to present in detail a simplified derivation of analytical expressions for the probability distributions of the ASDs.

VIII. Postscript

This paper traces its heritage to the collaboration of LLS and D. W. Lytle, who in [11] applied the theory of invariance in hypothesis testing to the problem of CFAR signal detection. These ideas were then generalized to incorporate multidimensional detectors in the collaborations of LLS and M. J. Dunn [26], resulting in the treatment of MSDs given in [1]. The collaboration of LLS and B. J. Friedlander led to the GLRT interpretation of [2]. The work of Kelly [5], a major contribution to adaptive detection, was followed by the work of Chen and Reed [9] and Robey *et al.* [10]. These papers are the natural predecessors of this paper.

About the time of [11], R. L. Spooner [27] and G. Vezzozi and B. Picinbono [28] derived CFAR detectors for spherically invariant noise. These papers are predecessors of the work on adaptive detection for spherically invariant noise by Conte *et al.* [12], [13], who suggest the rank-1 version of the noncoherent CFAR ASD derived in [6]. Conte *et al.* [12], [13] slightly predate the rank-1 version of the CFAR ASD presented in [7], [19], and [29]. However, as we show in this paper, the detector of [12], [13], and [19] is just one of a large class of adaptive detectors one can derive from a maximum likelihood principle, beginning with the MSDs of [1], [2], and [26]. In fact, it was not until the publication of [6] that we had a convincing derivation for the CFAR ASD, based on asymptotic arguments in [12] and [13] and based on heuristic arguments in [19].

ACKNOWLEDGMENT

The authors acknowledge C. Richmond for helpful discussions, particularly on obtaining analytical expressions for density functions. They also acknowledge helpful discussions with D. W. Tufts and I. S. Reed, which encouraged them to clarify the significance of arbitrary scaling between test and training data.

REFERENCES

[1] L. L. Scharf, *Statistical Signal Processing*. Reading, MA: Addison-Wesley, 1991, ch. 4.
[2] L. L. Scharf and B. Friedlander, "Matched subspace detectors," *IEEE Trans. Signal Processing*, vol. 42, pp. 2146–2157, Aug. 1994.
[3] S. Kraut, L. T. McWhorter, and L. L. Scharf, "A canonical representation for distributions of adaptive matched subspace detectors," in *Proc. 31st Asilomar Conf. Signals, Syst., Comput.*, Pacific Grove, CA, Nov. 1997.
[4] S. Kraut and L. L. Scharf, "The cosine GLRT: Comparison of this scale-invariant GLRT with the Kelly GLRT and the AMF," in *Proc. 7th Annu. Workshop Adaptive Sensor Array Process.*. Lexington, MA: Lincoln Lab., Mass. Inst. Technol., Mar. 1999.
[5] E. J. Kelly, "An adaptive detection algorithm," *IEEE Trans. Aerosp. Electron. Syst.*, vol. AES-22, pp. 115–127, Jan. 1986.
[6] S. Kraut and L. L. Scharf, "The CFAR adaptive subspace detector is a scale-invariant GLRT," *IEEE Trans. Signal Processing*, vol. 47, pp. 2538–2541, Sept. 1999.
[7] L. L. Scharf, "Adaptive matched subspace detectors and adaptive coherence," , submitted for publication.
[8] L. T. McWhorter, L. L. Scharf, and L. J. Griffiths, "Adaptive coherence estimation for radar signal processing," in *Proc. 30th Asilomar Conf. Signals, Syst., Comput.*, Pacific Grove, CA, Nov. 1996.
[9] W.-S. Chen and I. S. Reed, "A new CFAR detection test for radar," *Digital Signal Process.*, vol. 1, no. 4, pp. 198–214, 1991.
[10] F. C. Robey, D. R. Fuhrmann, E. J. Kelly, and R. A. Nitzberg, "A CFAR adaptive matched filter detector," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 28, pp. 208–216, Jan. 1992.
[11] L. L. Scharf and D. W. Lytle, "Signal detection in Gaussian noise of unknown level: An invariance application," *IEEE Trans. Inform. Theory*, vol. IT-17, pp. 404–411, July 1971.
[12] E. Conte, M. Lops, and G. Ricci, "Asymptotically optimum radar detection in compound-Gaussian clutter," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 31, pp. 617–625, Mar. 1995.
[13] ——, "Adaptive matched filter detection in spherically invariant noise," *IEEE Signal Processing Lett.*, vol. 3, pp. 248–250, Aug. 1996.
[14] R. J. Muirhead, *Aspects of Multivariate Statistical Theory*. New York: Wiley, 1982.
[15] C. D. Richmond, "Adaptive array signal processing and performance analysis in non-Gaussian environments," Ph.D. dissertation, Mass. Inst. Technol., Cambridge, 1996.
[16] E. J. Kelly, "Adaptive detection in nonstationary interference, Part III," Lincoln Lab., Mass. Inst. Technol., Lexington, Tech. Rep. 761, 1987.
[17] R. S. Raghavan, N. Pulsone, and D. J. McLaughlin, "Performance of the GLRT for adaptive vector subspace detection," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 32, pp. 1473–1487, July 1996.
[18] E. J. Kelly and K. M. Forsythe, "Adaptive detection and parameter estimation for multidimensional signal models," Lincoln Lab., Mass. Inst. Technol., Lexington, Tech. Rep. 848, 1989.
[19] L. L. Scharf and L. T. McWhorter, "Adaptive matched subspace detectors and adaptive coherence," in *Proc. 30th Asilomar Conf. Signals, Syst., Comput.*, Pacific Grove, CA, Nov. 1996.
[20] I. S. Reed, J. D. Mallett, and L. E. Brennan, "Rapid convergence rate in adaptive arrays," *IEEE Trans. Aerosp. Electron. Syst.*, vol. AES–10, pp. 853–863, Nov. 1974.
[21] J. Capon and N. R. Goodman, "Probability distributions for estimators of the frequency wavenumber spectrum," *Proc. IEEE*, vol. 58, pp. 1785–1786, 1970.
[22] S. Kraut and L. L. Scharf, "Performance evaluation of adaptive subspace detectors, based on stochastic representations," in *Proc. 9th Biennial Euro. Signal Process. Conf., EUSIPCO*, Rhodes, Greece, Sept. 1998.
[23] ——, "Performance convergence of the adaptive matched filter," in *Proc. 32nd Asilomar Conf. Signals, Syst., Comput.*, Pacific Grove, CA, Nov. 1998.
[24] C. D. Richmond, "Statistical performance analysis of the adaptive sidelobe blanker detection algorithm," in *Proc. 31st Asilomar Conf. Signals, Syst., Comput.*, Pacific Grove, CA, Nov. 1997.
[25] E. J. Kelly, "Finite-sum expressions for signal detection probabilities," Lincoln Lab., Mass. Inst. Technol., Cambridge, Tech. Rep. 566, 1981.
[26] M. J. Dunn, "Sufficiency and invariance principles applied to four detection problems," M.S. thesis, Univ. Colorado, Boulder, 1986.
[27] R. L. Spooner, "On the detection of a known signal in a non-Gaussian noise process," *J. Acoust. Soc. Amer.*, vol. 44, pp. 141–147, Jan. 1968.
[28] G. Vezzozi and B. Picinbono, "Detection d'un signal certain dans un bruit spheriquement invariant, structure et characteristiques des recepteurs," *Ann. Telecommun.*, vol. 27, pp. 95–110, 1972.
[29] L. L. Scharf, "Adaptive matched subspace detectors and adaptive coherence," Univ. Colorado, Boulder, Tech. Rep., June 1996.

Shawn Kraut (M'00) received the B.S. degree from the University of Arizona, Tucson, in 1993 and the Ph.D. degree from the University of Colorado, Boulder, in 1999.

He has authored or co-authored papers in adaptive detection theory, photorefractive physics, and optical signal processing for adaptive arrays. He is currently a research associate at Duke University, Durham, NC, where he is investigating nonlinear recursive Bayesian state estimation, and subspace selection for nonlinear parameter estimation.

Louis L. Scharf (F'86) received the Ph.D. degree in electrical engineering in 1969 from the University of Washington, Seattle.

From 1969 to 1971, he was with Honeywell's Marine Sytems Center, Seattle. From 1971 to 1982, he was with Colorado State University (CSU), Fort Collins. He was Professor and Chair of Electrical Engineering at the University of Rhode Island, Kingston, from 1982 to 1985. From 1985 to 2000, he was with the University of Colorado, Boulder. He is currently Professor of Electrical and Computer Engineering and Statistics at CSU.

Prof. Scharf is a past member of the IEEE ASSP AdCom. He was the Technical Program Chairman of the IEEE International Conference on Acoustics, Speech, and Signal Processing in 1980. In 1994, he served as a Distinguished Lecturer for the IEEE Signal Processing Society, in 1995, he received the Society's Technical Achievement Award, and in 2000, he received an IEEE Third Millenium Medal.

L. Todd McWhorter is currently at Mission Research Corporation, Fort Collins, CO, as a Member of the Applied Mathematics Group. His interests are, in general, statistical signal processing with particular interest in automatic target recognition, data compression, array processing, and wavelets.