Space-Time Detection Theory

William L. Melvin
Georgia Tech Research Institute
7220 Richardson Road, Smyrna, GA 30080, USA

bill.melvin@gtri.gatech.edu

Summary

Radar systems must detect targets in the presence of clutter and jamming (colored noise) signals. By exploiting signal diversity, space-time adaptive processing (STAP) improves radar detection performance in colored noise-limited environments. This set of lecture notes describes fundamental aspects of space-time detection theory in Gaussian noise. We first describe basic detection theory; key aspects of the discussion include the formation of a sufficient statistic, threshold setting, matched filtering, and the notion of a whitening filter. We then examine the role space-time signal diversity plays in enhancing target detection performance in colored noise environments. Next, we define commonly used performance metrics, including signal-to-interference-plus-noise ratio (SINR) loss factors. The latter part of the paper describes practical aspects of space-time detection: we consider various detector structures, estimation of unknown parameters, and the impact of clutter heterogeneity on detection performance.

1. Introduction

Target detection serves as the primary goal of radar. Moving target indication (MTI) is a common radar mission involving the detection of airborne or surface moving targets. The signal-to-noise ratio (SNR) – a characterization of the noise-limited performance of the radar against a target with radar cross section σ_T at range r – is approximated as

$$SNR(\phi,\theta)=\left(\frac{P_t G_t(\phi,\theta)}{4\pi r^2}\right)\left(\frac{\sigma_T}{4\pi r^2}\right)\left(\frac{A_e}{N_{in}\ F_n\ L_{rf}}\right), \tag{1}$$

where P_t is peak transmit power, $G_t(\phi,\theta)$ is antenna gain, A_e is the effective receive aperture area, N_{in} is the input noise power, F_n is the receiver noise figure and L_{rf} represents radio frequency (RF) system losses [1]. Assuming the noise is uncorrelated (white) and Gaussian, the probability of detection is a one-to-one function of both SNR and probability of false alarm. It is also important to point out that by maximizing SNR, the processor maximizes probability of detection. In light of (1), by increasing power-aperture $P_t A_e$, the radar designer ensures detection of targets with diminishing radar cross section at farther range. System constraints and cost limit the deployable power-aperture product.

In addition to noise-limited constraints, the aerospace radar system design must accommodate the impact of ground clutter and jamming on moving target detection. Clutter and jamming represent colored noise; unlike white noise, clutter and jamming exhibit a degree of correlation. Either individually, or collectively, we refer to clutter and jamming as interference. Interference increases the amount of ambiguity in the target decision process. In other words, as the interference increases, it becomes more difficult to decide whether a target is present in a given observation. Analogous to the white noise detection scenario, the probability of detection depends on both signal-to-interference-plus-noise ratio (SINR) and the specified false alarm rate. Since $SINR \le SNR$, interference always degrades detection performance in comparison with the noise-limited case.

Signal diversity, in the form of spatial and temporal degrees of freedom (DoF), greatly enhances radar detection in the presence of colored noise. Specifically, the appropriate application of space-time DoFs maximizes signal-to-interference-plus noise ratio (SINR) when the target competes with ground clutter and barrage jamming. Clutter exhibits correlation in both spatial and temporal dimensions, while jamming is predominantly correlated in angle for modest bandwidth. Space-time adaptive processing (STAP) involves

Paper presented at the RTO SET Lecture Series on "Military Application of Space-Time Adaptive Processing", held in Istanbul, Turkey, 16-17 September 2002; Wachtberg, Germany, 19-20 September 2002; Moscow, Russia, 23-24 September 2002, and published in RTO-EN-027.

adaptively (or dynamically) adjusting the two-dimensional filter response in an attempt at maximizing the filter's output SINR, and consequently, improving radar detection performance.

The objective of this set of lecture notes is to develop the basic theory of space-time detection. Six sections comprise this paper. In the next section, we begin by considering basic elements of hypothesis testing and detection theory. The important notions of matched filtering and whitening arise from this discussion. Subsequently, in Section 3 we describe the importance of signal diversity in advanced radar sensor design. Radar offers signal diversity in the following dimensions: spatial, slow-time (or Doppler), fast-time (or range), polarization and multi-scan. An appropriate selection of these available degrees of freedom (DoF) is necessary for effective performance enhancement. Herein, our interest lies in the space and slow-time dimensions. (In a companion paper [2], we describe the simultaneous application of space and fast-time DoFs to mitigate terrain scattered jamming.) After discussing signal diversity, we develop several typical STAP performance metrics in Section 4; the metrics enable comparison of competing STAP techniques and non-adaptive processing schemes. Section 5 investigates several practical STAP implementation structures. STAP is a data domain implementation of an optimum filter; practical implementation issues include estimating unknown quantities, alleviating computational burden and minimizing training data requirements. Another issue of practical concern is clutter heterogeneity. Clutter heterogeneity degrades STAP performance by exacerbating the estimation the interference covariance matrix, an essential STAP component. Thus, we devote the final section of the paper – Section 6 – to issues centering on covariance estimation errors resulting from clutter heterogeneity.

The following is a list of the more common variables used in this paper:

P_D = probability of detection;

P_{FA} = probability of false alarm;

$p_X(x)$ = probability density function of X;

η, η' = detection threshold;

N = number of slow-time pulses;

M = number of spatial channels;

ϕ, θ = azimuth and elevation angles;

f_d = Doppler frequency;

$\mathbf{s}_s(\phi,\theta)$ = spatial steering vector;

$\mathbf{v}_s(\phi,\theta)$ = surrogate spatial steering vector;

$\mathbf{s}_{s\text{-}t}(\phi,\theta,f_d)$ = space-time steering vector;

$\mathbf{v}_{s\text{-}t}(\phi,\theta,f_d)$ = surrogate space-time steering vector;

$\mathbf{x}_{k/s} \in C^{M\times 1}$ = spatial snapshot, kth range realization

$\mathbf{x}_k \in C^{NM\times 1}$ = space-time snapshot, kth range realization;

y_k = (scalar) filter output;

$\mathbf{w}_k$ = space-time weight vector;

$\mathbf{R}_k$ = space-time covariance matrix;

$\hat{\mathbf{R}}_k, \hat{\mathbf{w}}_k$ = estimated covariance matrix and estimated weight vector.

In general, a boldface, lowercase variable indicates a vector quantity; a boldface, upper case variable indicates a matrix; and, a variable with a caret is an estimate. Additionally, an optimum filter implies clairvoyant knowledge of the statistics of the signal environment. Primary data, or test cells, are those ranges the processor tests for target presence. Secondary data and training data are synonymous, and indicate those range cells used to estimate the unknown characteristics of the primary data. Superscripts "T" or "H" applied to a vector or matrix denote the transpose or hermitian (conjugate transpose) operations. The notation $\mathbf{a} \sim \mathrm{CN}(\boldsymbol{\mu}_a, \mathbf{R}_a)$ indicates that $\mathbf{a}$ is complex normal (Gaussian) with mean $\boldsymbol{\mu}_a$ and covariance matrix $\mathbf{R}_a$, and $\mathbf{b} \sim \mathrm{N}(\boldsymbol{\mu}_b, \mathbf{R}_b)$ implies $\mathbf{b}$ is normally distributed with mean $\boldsymbol{\mu}_b$ and covariance matrix $\mathbf{R}_b$.

2. Basic Detection Theory

The primary objective of a radar is to detect a target at a given range; in fact, radar is an acronym for "radio detection and ranging." This section of the paper describes the basics of radar detection. Additionally, we unify detection theory and predominant notions of adaptive filtering. Excellent sources of further information on radar detection include [3-6]. The whitening operation, matched filtering, and SINR are the key elements of this unification.

Clutter, jamming and receiver noise are random variables evolving in time. For instance, we can never precisely predict the clutter voltage. Rather, we can describe the statistical distribution of clutter, jamming and thermal receiver noise. Similarly, the target amplitude is a random variable; the famous Swerling models are usually used to describe target amplitude fluctuation [1]. Due to the random nature of the radar signal environment, ambiguity as to whether or not a target is present is always a concern. The goal of detection theory is to provide a rational procedure for confidently determining, in a statistical sense, which of two possible models – target present or target absent – generated a given observation. Hypothesis testing is the cornerstone of detection theory.

Prior to our hypothesis testing discussion, we briefly provide some germane background information. The probability density function (pdf), given as $p_X(x)$, describes the relative frequency behavior of random variable X. The joint pdf between X and Y is $p_{X,Y}(x,y)$. If we consider n trials, and then observe specific occurrences of $X=x$ a total of n_X times, and the simultaneous occurrence of $X=x$ and $Y=y$ a total of n_{XY} times, then $p_X(x) \approx n_X/n$ and $p_{X,Y}(x,y) \approx n_{XY}/n$; similarly, we find $p_Y(y) \approx n_Y/n$ for random variable Y. In hypothesis testing, the conditional probability density is most significant. Thus, we ponder the probability associated with observing Y given that we have already observed certain behavior in X. We may then define the conditional probability density as $p_{Y|X}(y|x) = p_{X,Y}(x,y)/p_X(x) \approx (n_{XY}/n)/(n_X/n)$. If X and Y are statistically independent, then $p_{X,Y}(x,y) = p_X(x)p_Y(y)$, since $p_{Y|X}(y|x) = p_Y(y)$ as a result of the fact that knowledge of X has no bearing on our estimating the probability that $Y=y$.

Radar detection is a binary hypothesis testing application. The Neyman-Pearson criteria (NPC) are most commonly used to implement this hypothesis testing procedure since no *a priori* probabilities are necessary in the decision mechanism. In radar, the NPC aids in deciding between the two hypotheses: H_0 and H_1. H_0 is known as the null hypothesis and corresponds to the case of target absence, while H_1 is the alternative hypothesis and indicates target presence. The NPC nomenclature includes the probability of detection, P_D, and the probability of false alarm, P_{FA}:

$$\begin{aligned} P_D &= \Pr[\text{choose } H_1 \,|\, H_1 \text{ is true}] = \int_{T_1} p_{X|H_1}(x|H_1)dx; \\ P_{FA} &= \Pr[\text{choose } H_1 \,|\, H_0 \text{ is true}] = \int_{T_1} p_{X|H_0}(x|H_0)dx, \end{aligned} \tag{2}$$

where T_1 defines the "target present" decision region. The goal of the NPC is then to maximize P_D while maintaining $P_{FA} \le \gamma$, where γ is the maximum false alarm rate the radar system and associated data processor can tolerate.

Figure 1 shows the relationship between P_D, P_{FA} and the threshold setting η. Observe a fundamental truth from this figure: lowering the threshold to increase P_D necessarily increases P_{FA}, and viceversa. The regions of overlap among the two pdf's signifies the ambiguity in the decision-making process. The NPC detector provides an optimum mechanism for determining target presence or absence in light of this ambiguity.

Applying the NPC in an optimization procedure (see [3-4] for details) leads to a decision mechanism known as the likelihood ratio test (LRT), which is given by

$$\Lambda(x) = \frac{p_{X|H_1}(x \mid H_1)}{p_{Y|H_0}(x \mid H_0)} \underset{H_0}{\overset{H_1}{\gtrless}} \eta. \tag{3}$$

The LRT is the detector! It compares the ratio of the two likelihood functions, $p_{X|H_m}(x \mid H_m)_{m=0,1}$, to the threshold setting; if the value exceeds the threshold, the processor declares target presence, while if the ratio is less than the threshold setting, the processor assumes the null hypothesis.

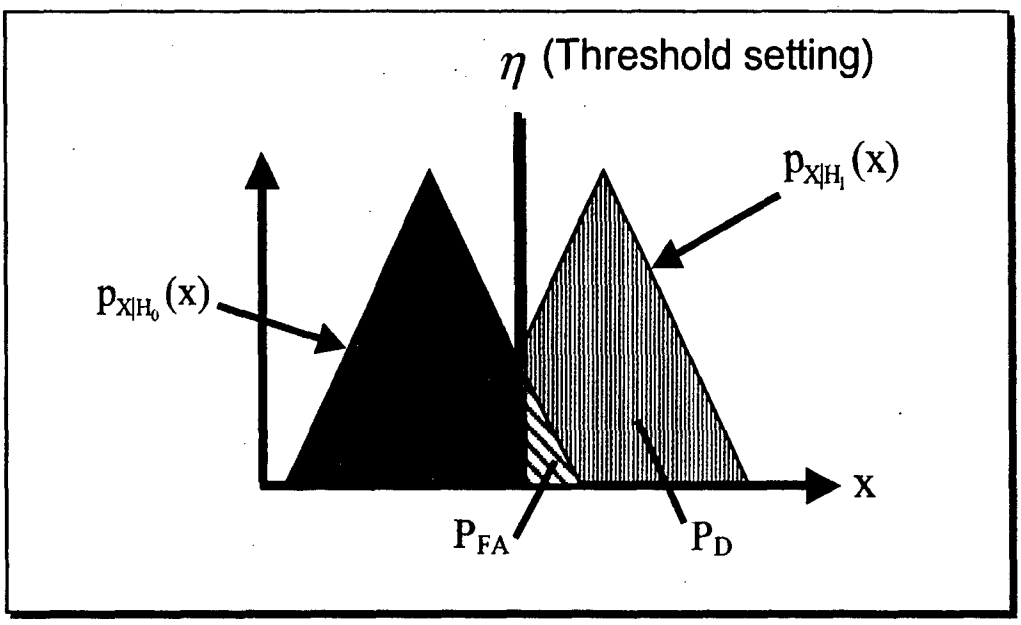

Figure 1. Relationship between P_D, P_{FA} and threshold setting, η.

It is commonplace to apply a monotonic operation to simplify the LRT. For example, the natural logarithm is often applied to reduce expressions involving exponential functions. A sufficient statistic results after manipulating the LRT into a canonical form involving a function of the observed data. We denote the sufficient statistic as

$$\Psi(x) \underset{H_0}{\overset{H_1}{\gtrless}} \eta'. \tag{4}$$

We obtain the modified threshold in (4) by deriving the pdf of $\Psi(x)$ under the null hypothesis condition and then solving

$$P_F = \int_{\eta'}^{\infty} p_{\Psi(x)|H_0}(\Psi(x) \mid H_0) dx = \gamma \tag{5}$$

for η'. As one might expect, we ascertain detection performance by calculating the pdf of $\Psi(x)$ under the alternative hypothesis case.

Application of the LRT is best understood through example. For this reason, let's consider a colored Gaussian noise (CGN) scenario. Our hypotheses are given as

$$H_i : \mathbf{x} = \mathbf{s}_i + \mathbf{n} \tag{6}$$

for $i = 0,1$, where $\mathbf{x}$ is the vector observation, $\mathbf{s}_1$ is the target signal, $\mathbf{s}_0 = \mathbf{0}$, and $\mathbf{n}$ is the interference vector. Also, suppose $\mathbf{x}$ represents a coherent dwell of N pulses, spaced by the pulse repetition interval, T_p, i.e.,

$$\mathbf{x} = \left[x(0) \quad x(T_p) \quad \cdots \quad x((N-1)T_p) \right]^T. \tag{7}$$

The joint pdf of (7) is

$$p_{\mathbf{x}|H_i}(\mathbf{x}|H_i) = p_{\mathbf{n}}(\mathbf{x} - \mathbf{s}_i). \tag{8}$$

Assuming that $\mathbf{n}$ is distributed according to a complex Gaussian distribution with zero mean and covariance matrix $\mathbf{R}_{\mathbf{n}} = \mathrm{E}[\mathbf{n}\mathbf{n}^{\mathrm{H}}]$, then

$$p_{\mathbf{n}}(\mathbf{n}) = \frac{1}{|\pi \mathbf{R}_{\mathbf{n}}|} \exp\left(-\mathbf{n}^{\mathrm{H}} \mathbf{R}_{\mathbf{n}}^{-1} \mathbf{n}\right). \tag{9}$$

Forming the LRT using (8)-(9), then taking the natural log and simplifying yields

$$\mathrm{Re}\left(\mathbf{s}_1^H \mathbf{R}_{\mathbf{n}}^{-1} \mathbf{x}\right) - \frac{\mathbf{s}_1^H \mathbf{R}_{\mathbf{n}}^{-1} \mathbf{s}_1}{2} \underset{H_0}{\overset{H_1}{\gtrless}} \frac{1}{2} \ln(\eta). \tag{10}$$

The first term on the left-hand side of (10) is a whitening filter, $\mathbf{R}_{\mathbf{n}}^{-1/2}$, while $\mathbf{R}_{\mathbf{n}}^{-1/2}\mathbf{s}_1$ is a linearly transformed matched filter. The matched filter maximizes SNR (in the noise-limited case); hence, in the presence of colored-noise, the detector first whitens (decorrelates) the colored-noise, and then applies a matched filter as in the noise-limited case. The whitening filter and matched filter will be discussed in further detail in subsequent sections.

It then follows from (10) that a sufficient statistic is

$$\Psi(\mathbf{x}) = \mathrm{Re}\left(\mathbf{s}_1^H \mathbf{R}_{\mathbf{n}}^{-1} \mathbf{x}\right) \underset{H_0}{\overset{H_1}{\gtrless}} \eta'; \quad \Psi(\mathbf{x}) \sim N\left(\mathrm{Re}\left(\mathbf{s}_i^H \mathbf{R}_{\mathbf{n}}^{-1} \mathbf{s}_1\right), \mathbf{s}_1^H \mathbf{R}_{\mathbf{n}}^{-1} \mathbf{s}_1\right). \tag{11}$$

The threshold and performance of the detector in (11) are then given by

$$P_{FA} = Q\left(\frac{\ln(\eta) + \mathbf{s}_1^H \mathbf{R}_{\mathbf{n}}^{-1} \mathbf{s}_1}{2\sqrt{\mathbf{s}_1^H \mathbf{R}_{\mathbf{n}}^{-1} \mathbf{s}_1}} \right); \quad P_D = Q\left(\frac{\ln(\eta) - \mathbf{s}_1^H \mathbf{R}_{\mathbf{n}}^{-1} \mathbf{s}_1}{2\sqrt{\mathbf{s}_1^H \mathbf{R}_{\mathbf{n}}^{-1} \mathbf{s}_1}} \right), \tag{12}$$

where $Q(x) = \frac{1}{\sqrt{2\pi}} \int_x^\infty e^{-\alpha^2/2} d\alpha$ is the right-tail distribution (complementary cumulative distribution) of a Gaussian variate with zero mean and variance of unity. Numerical tables provide values of $Q(x)$ and $Q^{-1}(x)$ [3]. We then use the expression for P_{FA} and $Q^{-1}(x)$ to calculate the threshold η.

Figure 2 shows an upper and lower bound on $Q(x)$. The important observation from (12) and Figure 2 concerns the role of the factor $\mathbf{s}_1^H \mathbf{R}_{\mathbf{n}}^{-1} \mathbf{s}_1$; this term represents the SINR. As $\mathbf{s}_1^H \mathbf{R}_{\mathbf{n}}^{-1} \mathbf{s}_1$ increases, the argument in (12) decreases for the P_D calculation and so detection performance improves. In effect, the argument in (12) for P_{FA} increases, and so the false alarm rate decreases below the specification unless the threshold is increased. An increase in SINR serves to separate the pdf's for null and hypothesis conditions, making it easier to distinguish target absence from target presence. Figure 3 clarifies this point by comparing the histograms for null and alternative hypothesis with three different values of signal-to-noise ratio. We plot the amplitude of the complex observations; in the null hypothesis case this yields a Rayleigh distribution, while

for the alternative hypothesis we have the well-known Rician distribution [6]. Adequate SNR or SINR is essential to acceptable detection performance.

The appropriate exploitation of signal diversity can be used in the Gaussian case to maximize SINR, hence yielding maximal detection performance for any specified false alarm rate. Perhaps the simplest example of the benefits of signal diversity involves sidelobe noise jamming. For a single channel system, the jamming enters the sidelobes of the antenna and masks target detection over all Doppler. However, since the noise jammer is spatially correlated, spatial sampling of the jammer waveform using multiple spatial channels provides information on the jammer direction of arrival. This information can then be used to construct a whitening filter to mitigate all signals coming from the jamming angle. Effectively, the detector implements a contrained matched filtering operation, where the constraint places a filter null in the jammer direction. The consequent increase in SINR leads to greatly improved detection performance, as our prior discussion indicates.

Ground clutter exhibits a two-dimensional correlation in angle and slow-time. The seminal result on space-time detection theory for radar application is given by Brennan and Reed in [5]. Using the hypothesis testing theory of the prior discussion, we have the mathematical framework to develop Brennan and Reed's result. Herein, we elect to synopsize the key aspects of this development.

Consider the hypotheses

$$H_0:\ \mathbf{x}_{k/H_0} = \mathbf{x}_{k/I} + \mathbf{x}_{k/N};\ H_1:\ \mathbf{x}_{k/H_1} = \mathbf{s}_T + \mathbf{x}_{k/H_0}, \tag{13}$$

where $\mathbf{x}_{k/H_m} \in C^{NMx1}$ is the space-time snapshot for N pulses and M spatial channels for the kth range realization and hypothesis H_m. Additionally, $\mathbf{x}_{k/H_0} \sim CN(\mathbf{0}, \mathbf{R}_k)$, where $\mathbf{R}_k$ is the null-hypothesis covariance matrix, $\mathbf{s}_T \in C^{NMx1}$ is the space-time signal vector, and $\mathbf{x}_{k/I} \in C^{NMx1}$ and $\mathbf{x}_{k/N} \in C^{NMx1}$ are space-time interference and noise snapshots. Each space-time snapshot is organized as

$$\mathbf{x}_k = \left[\mathbf{x}_{k/s}^T(0) \quad \mathbf{x}_{k/s}^T(T_p) \quad \cdots \quad \mathbf{x}_{k/s}^T((N-1)T_p)\right]^T \in C^{MNx1} \tag{14}$$

where $\mathbf{x}_{k/s}(n) \in C^{Mx1}$ is the spatial snapshot for the n^{th} pulse. The detector involves forming a function of the filter output, $y_k = \mathbf{w}_k^H \mathbf{x}_k$, where $\mathbf{w}_k \in C^{NMx1}$ is the space-time weight vector. The pdf for the two hypothesis of (13) are

$$p\left(y_{k/H_i}\right) = \frac{1}{\sqrt{2\pi\sigma_y^2}} \exp\left(\frac{-\left|y_{k/H_i} - \mu_{y/H_i}\right|^2}{2\sigma_y^2}\right), \quad i = 0,1; \tag{15}$$

with the mean and interference-plus-noise variance given by

$$\begin{aligned} \mu_y &= E[y_k] = \begin{cases} 0 & \text{for } H_0 \\ \mathbf{w}_k^H \mathbf{s}_T & \text{for } H_1 \end{cases}; \\ \sigma_y^2 &= E\left[\left(y_k - \mu_y\right)\left(y_k - \mu_y\right)^*\right] = \mathbf{w}_k^H \mathbf{R}_k \mathbf{w}_k. \end{aligned} \tag{16}$$

Upon forming the LRT and algebraically manipulating, we find a sufficient statistic taking the simple form

$$|y_k| \underset{H_0}{\overset{H_1}{\gtrless}} \eta'. \tag{17}$$

The pfd's of (17) for null and alternative hypotheses are Rayleigh and Rician, respectively [5]. Armed with the distributions, we can then determine the threshold setting and detection performance. The threshold is calculated from

$$P_{FA} = \Pr\left[|y_k| > \eta' \,|\, \mathbf{x}_{\mathbf{k/H_0}}\right] = \exp\left(\frac{-(\eta')^2}{2\sigma_y^2}\right). \tag{18}$$

Detection performance is given by

$$P_D = Q_M(\alpha,\beta) = \int_{\beta}^{\infty} v \exp\left(\frac{-\left(v^2+\alpha^2\right)}{2}\right) I_0(\alpha v)\, dv; \text{ where,}$$
$$v = \frac{r}{\sigma_y},\ r = |y_k|,\ \alpha^2 = \frac{a^2}{\sigma_y^2} = \mathrm{SINR}_{\mathrm{out}},\ \mathrm{a} = \left|\mathbf{w}_{\mathbf{k}}^H \mathbf{s}_{\mathbf{T}}\right|, \text{ and } \beta = \frac{\eta'}{\sigma_y}. \tag{19}$$

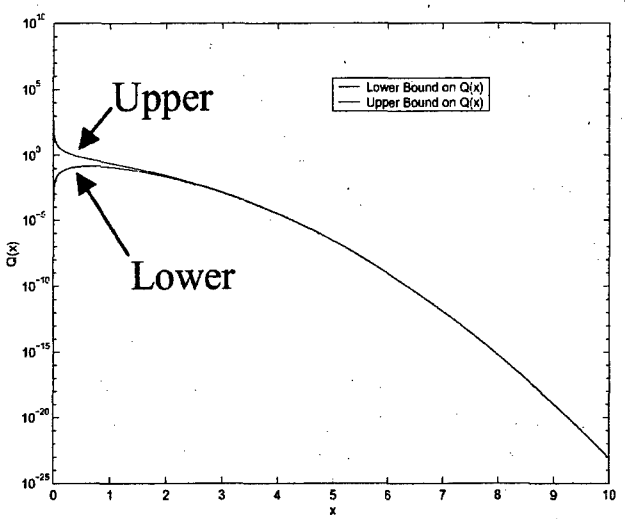

Figure 2. Upper and lower bounds on the complementary cumulative distribution for a Gaussian variate.

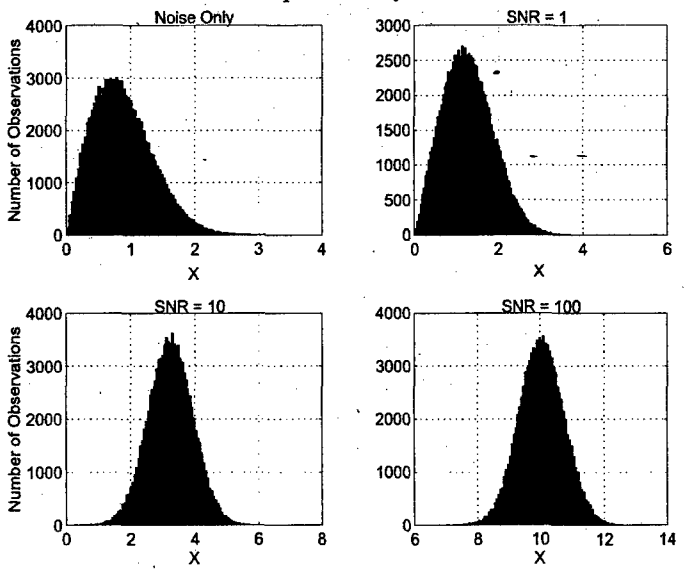

Figure 3. Comparison of noise-only histogram with signal-plus-noise histograms for varying SNR.

$I_0(\cdot)$ is the modified Bessel function of the first kind and order zero. $Q_M(\alpha,\beta)$ is known as Marcum's Q and is a monotonically increasing function of the square-root of SINR and monotonically decreasing function of normalized threshold setting, β; numerical techniques are required to solve for P_D in (19).

The receiver operating characteristic (ROC) characteristic of the space-time detector in (17) is shown in Figure 4. The different curves shown in the figure correspond to different false alarm rates. Observe that as the false alarm rate increases – implying a decrease in the threshold – the detection probability increases.

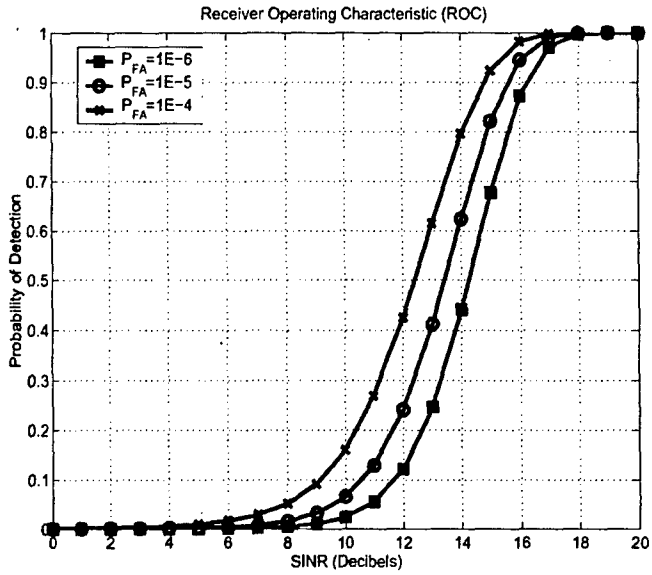

Figure 4. Receive operating characteristic for space-time detector.

3. Signal Diversity

Pulsed radar sensors can potentially measure angle (azimuth, elevation or cone angle), Doppler (slow-time), range (fast-time), polarization, and the multi-scan behavior of target and interference behavior. Different classes of interference exhibit varying correlation over the available sensor measurement spaces. Table 1 describes different interference types and their respective correlation properties.

Table 1. Correlation Characteristics of Some Interference Types

Interference Type	Correlation Characteristics	Comments
Ground Clutter	Azimuth (or cone angle) and Doppler, polarization response may differ from clutter.	Space and slow-time adaptive processing (STAP) can mitigate clutter, polarization may aid the detection of very slow targets.
Narrowband Noise Jamming	Angle (azimuth and elevation).	Spatial null suppresses jamming signal.
Wideband Noise Jamming	Angle and fast-time.	Variation with wavelength must be compensated with fast-time processing, also helps with channel mismatch.
Hot Clutter	Angle and fast-time.	Jammer waveform often correlated in fast-time.
Cold and Hot Clutter	Angle, slow-time and fast-time.	Requires 3-D STAP.

The ability to separate interference and target characteristics in the chosen measurement spaces enables the processor to enhance the target while mitigating the interference. Inherently, the processor takes advantage of the differences in correlation properties to maximize SINR in the multidimensional detection space, thereby maximizing detection performance. Figure 5 provides a notional view of the impact of signal diversity on detection performance. In this figure, X, Y and Z are Fourier transforms of measurement domains x, y and z; hence, Figure 5 is a three-dimensional power spectral density (PSD). The inverse Fourier transform of the PSD yields the correlation function. Notice the separability of the target from the interference in the three-dimensional space. However, projecting the data into any two-dimensional data space, we observe the masking of the target by the interference. When the target shares the same space as the interference, filtering

has little or no impact on SINR. In the three-dimensional space, the detection processor can construct a filter to maximize the gain on the target while nulling the interference.

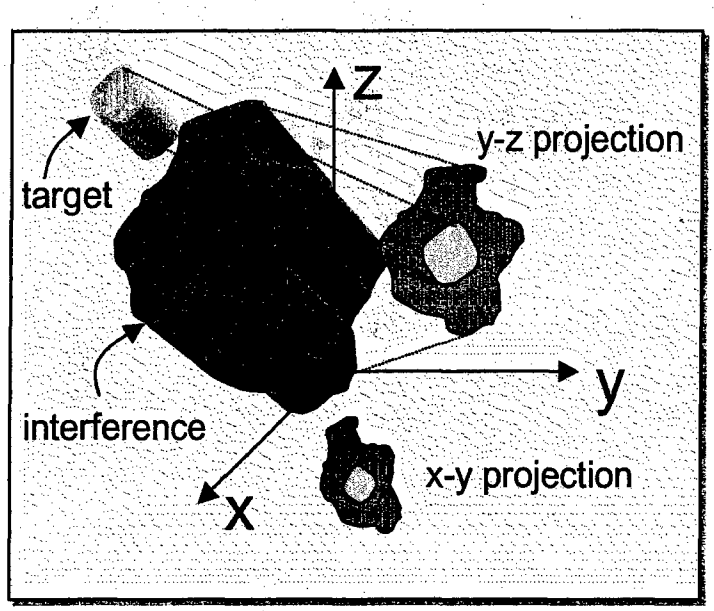

Figure 5. Notional view of the importance of signal diversity on target detection. In three dimensions, the target and interference are separable, and hence more detectable, while in two dimensions the interference masks the target.

The impact of ground clutter on detection performance serves as our primary focus. For this reason, the remainder of the paper considers the space (azimuth or cone angle) and slow-time (Doppler) dimensions. Figure 6 shows the PSD for simulated ground clutter; observe the coupling in angle and Doppler marked by those two-dimensional frequencies with larger clutter power values. The dashed line denotes the transmit direction. The processor searches for targets along this line at all Doppler. Mainlobe clutter – located near zero Doppler and zero angle – impedes the detection of slow-moving targets. Endo-clutter detection refers to detection of low radial velocity (low Doppler) targets existing in the diffraction-limited clutter regions. Additionally, sidelobe clutter – present at those Doppler frequencies away from zero Doppler – masks the detection of faster moving, exo-clutter targets. Figure 7 shows the corresponding space-time filter response. The filter places a null on the clutter ridge. This matched two-dimensional response leads to maximal SINR.

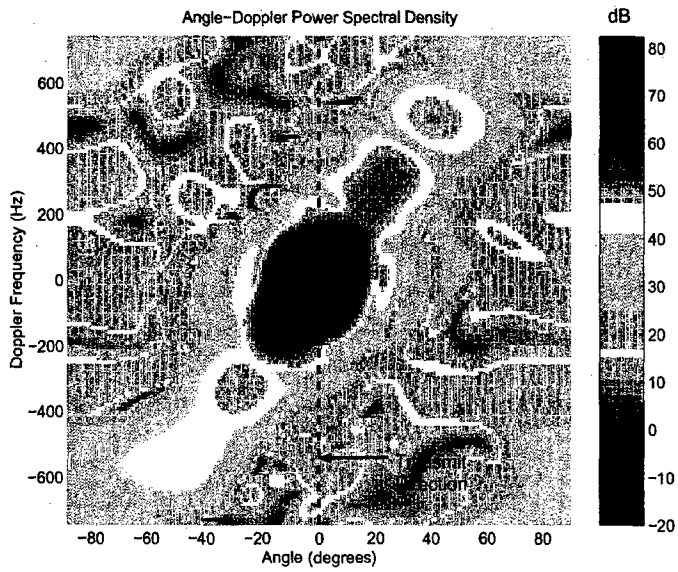

Figure 6. Power spectral density for ground clutter.

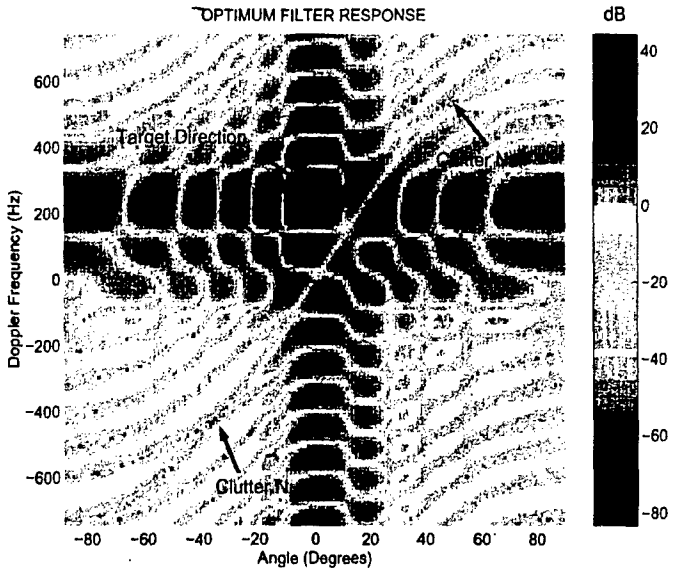

Figure 7. Space-time optimal filter response.

Signal diversity is an essential component in enhancing the detection of slow moving and/or weak targets competing with ground clutter returns. Choosing other measurement spaces, with perhaps the exception of polarization, will not help improve performance. As can be seen from Figures 6-7, it would be very difficult, if not impossible, to detect a target at zero angle and zero Doppler. Yet, since STAP is a member of the class of super-resolution algorithms, the processor can detect targets very close to mainbeam clutter.

4. Performance Metrics

We briefly consider several of the more useful performance metrics used to characterize STAP performance. We begin by describing the difference between adaptive and optimal processors.

STAP is a data domain implementation of an optimal filter. In developing the space-time detector in (13)-(19), we pointed out that by maximizing SINR, the processor maximizes P_D. The results in these equations are valid for any weight vector $\mathbf{w}_k$. We are interested in the weight vector leading to maximal SINR, which is given as

$$\mathbf{w}_k = \mu \mathbf{R}_k^{-1} \mathbf{s}_{s\text{-}t}(\phi, \theta, f_d), \tag{20}$$

where μ is an arbitrary constant and $\mathbf{s}_{s\text{-}t}(\phi, \theta, f_d)$ is the target space-time steering vector. The space-time steering vector is

$$\mathbf{s}_{s\text{-}t}(\phi, \theta, f_d) = \mathbf{s}_t(f_d) \otimes \mathbf{s}_s(\phi, \theta), \tag{21}$$

with $\mathbf{s}_t(f_d) \in C^{Nx1}$ representing the temporal (Doppler) steering vector pointing to Doppler frequency f_d, and $\mathbf{s}_s(\phi, \theta) \in C^{Mx1}$ denoting the spatial steering vector for a signal at azimuth ϕ and elevation θ. The space-time steering vector is the normalized response of the radar sensor to a target with a specified angle of arrival and Doppler frequency, i.e., $\mathbf{s}_T = \alpha \mathbf{s}_{s\text{-}t}(\phi, \theta, f_d)$ for complex amplitude α.

We now consider a short proof on the optimality of (20). Observe that

$$SINR = \frac{E\left[y_s y_s^*\right]}{E\left[y_{I+N} y_{I+N}^*\right]} = \frac{P_s}{P_{I+N}} = E\left[|\alpha|^2\right] \frac{\mathbf{w}_k^H \mathbf{s}_{s\text{-}t} \mathbf{s}_{s\text{-}t}^H \mathbf{w}_k}{\mathbf{w}_k^H \mathbf{R}_k \mathbf{w}_k}, \tag{22}$$

where y_s is the signal-only filter output, y_{I+N} is the null hypothesis filter output, and P_s and P_{I+N} represent output signal power and null hypothesis output power. Additionally, we have removed the space-time steering vector's dependence on angle and Doppler for notational convenience. Next, let $\mathbf{w}_k = \mathbf{A}\mathbf{z}$ and $\tilde{\mathbf{s}}=\mathbf{A}^H\mathbf{s}_{s\text{-}t}$, where $\mathbf{A} = \mathbf{R}_k^{-1/2}$. Substituting these expressions into (22) yields

$$P_{I+N} = \mathbf{w}_k^H \mathbf{R}_k \mathbf{w}_k = \mathbf{z}^H \mathbf{A}^H \mathbf{R}_k \mathbf{A}\mathbf{z} = \mathbf{z}^H\mathbf{z};\ P_s = E\left[|\alpha|^2\right]\left|\mathbf{z}^H\tilde{\mathbf{s}}\right|^2. \tag{23}$$

Inserting (23) into (22) gives

$$\frac{P_s}{P_{I+N}} = \frac{E\left[|\alpha|^2\right]\left|\mathbf{z}^H\tilde{\mathbf{s}}\right|^2}{\mathbf{z}^H\mathbf{z}}. \tag{24}$$

Using the Schwarz Inequality, we find

$$\left|\mathbf{z}^H\tilde{\mathbf{s}}\right|^2 = \left|\sum_m [\mathbf{z}]_m [\tilde{\mathbf{s}}]_m\right|^2 \le \mathbf{z}^H\mathbf{z}\tilde{\mathbf{s}}^H\tilde{\mathbf{s}}. \tag{25}$$

Notice that an optimal value for $\mathbf{z}$ is $\mathbf{z}_{opt} = \mu\tilde{\mathbf{s}}$; with this selection, the SINR is maximal. Hence,

$$\mathbf{w}_{k/opt} = \mu\mathbf{A}\mathbf{A}^H\mathbf{s}_{s\text{-}t} = \mu\mathbf{R}_k^{-1}\mathbf{s}_{s\text{-}t}, \tag{26}$$

which is the result we desire.

Several comments are in order. First of all, $\mathbf{A} = \mathbf{R}_k^{-1/2}$ is a whitening filter since $E[\mathbf{A}\mathbf{x}_k\mathbf{x}_k^H\mathbf{A}^H] = \mathbf{I}_{NM}$ (i.e., the whitening filter decorrelates the colored noise input, hence making it look like white noise). In the white noise case, $\mathbf{s}_{s\text{-}t}$ is the matched filter; it has a bandpass response with angle and Doppler and maximizes SNR. We may thus interpret the optimal weight vector as

$$\mathbf{w}_{k/opt} = \underbrace{\left(\mathbf{R}_k^{-1/2}\right)}_{\substack{\text{Whitening}\\ \text{Filter}}}\underbrace{\left(\mathbf{R}_k^{-1/2}\mathbf{s}_{s\text{-}t}\right)}_{\substack{\text{Warped}\\ \text{Matched Filter}}}. \tag{27}$$

The warped matched filter accommodates the linear transformation applied to the space-time signal vector during the whitening stage. Finally, keep in mind that SINR is a function of both angle and Doppler.

In the STAP case, both the covariance matrix and target space-time steering vector are unknown. For this reason, the processor must estimate both quantities. To handle the unknown steering vector, the processor sweeps across a series of pre-calculated space-time steering vectors, $\mathbf{v}_{s\text{-}t}(\phi,\theta,f_d)$, covering several angle bins about the transmit direction and the whole Doppler space. Performance degradation resulting from steering vector mismatch is generally slight. Covariance estimation is more complicated. Commonly, we use the estimate

$$\hat{\mathbf{R}}_k = \frac{1}{K}\sum_{m=1}^{K}\mathbf{x}_m\mathbf{x}_m^H, \tag{28}$$

where $\mathbf{x}_m$ are training (secondary data) taken from ranges adjacent to the test (primary) cell [7]. The adaptive weight vector is

$$\hat{\mathbf{w}}_k = \beta\hat{\mathbf{R}}_k^{-1}\mathbf{v}_{s\text{-}t}(\phi,\theta,f_d), \tag{29}$$

where β is a constant typically set to $\beta = 1$ or $\beta = \sqrt{\mathbf{v}_{s\text{-}t}^H(\phi,\theta,f_d)\hat{\mathbf{R}}_k^{-1}\mathbf{v}_{s\text{-}t}(\phi,\theta,f_d)}$.

If the secondary data are multivariate Gaussian and independent and identically distributed (iid) with respect to the null-hypothesis of the primary data, (28) is a maximum likelihood estimate. Selecting the number K of secondary data to obtain a suitable estimate via the calculation of (28) is addressed in [7]. Specifically, Reed, Mallett and Brennan have shown that

$$E\left[(\text{SINR}\,|\,\hat{\mathbf{w}}_k)/(\text{SINR}_{\text{Optimum}})\right] = (K+2-NM)/(K+1). \tag{30}$$

The famous Reed-Mallett-Brennan (RMB) Rule then states that for the adaptive processor to attain an SINR loss of -3 dB in comparison with the optimal processor requires $K = 2NM - 3$. The RMB Rule assumes the secondary data are iid.

Since a one-to-one correspondence exists between SINR and P_D in the Gaussian case, SINR loss factors are very convenient and commonly used metrics of STAP performance [8]. We may write the output SINR as

$$\begin{gathered}\text{SINR}(\phi,\theta,f_d) = \text{SNR}(\phi,\theta)\times L_{s,1}(\phi,\theta,f_d)\times L_{s,2}(\phi,\theta,f_d); \\ 0 \le L_{s,1}(\phi,\theta,f_d),\ L_{s,2}(\phi,\theta,f_d) \le 1.\end{gathered} \tag{31}$$

$\text{SNR}(\phi,\theta)$ follows from (1). $L_{s,1}(\phi,\theta,f_d)$ and $L_{s,2}(\phi,\theta,f_d)$ are SINR loss factors: $L_{s,1}(\phi,\theta,f_d)$ characterizes the loss in signal-to-noise ratio due to colored noise, while $L_{s,2}(\phi,\theta,f_d)$ measures the loss in SINR resulting from the error between optimal (known) and adaptive (estimated) weight vectors. $L_{s,1}(\phi,\theta,f_d)$ captures the impact of both the system design and signal processing algorithm selection on performance in interference-limited environments. Definitions of both SINR loss factors are

$$L_{s,1}(\phi,\theta,f_d) = \frac{\text{SINR}(\phi,\theta,f_d)|_{\mathbf{w}_k}}{\text{SNR}(\phi,\theta)};\ L_{s,2}(\phi,\theta,f_d) = \frac{\text{SINR}(\phi,\theta,f_d)|_{\hat{\mathbf{w}}_k}}{\text{SINR}(\phi,\theta,f_d)|_{\mathbf{w}_k}} = \frac{\text{Adaptive Output SINR}}{\text{Optimal Output SINR}}. \tag{32}$$

Substituting (32) into (31) yields the identity $\text{SINR}(\phi,\theta,f_d) = \text{SINR}(\phi,\theta,f_d)|_{\hat{\mathbf{w}}_k}$.

Figure 8 shows the $L_{s,1}$ term for the optimal space-time processor (STP) and the conventional digital beamformer (DBF) followed by Hanning-weighted Doppler processing (DOP). This example corresponds to the same interference environment shown in Figure 6, except the dwell has been doubled. The poor performance of the conventional approach is a consequence of the short dwell and small aperture used in this example. However, the example does highlight the tremendous performance potential of STAP (in an iid environment, the STAP SINR loss curve appears similar to the optimal STP $L_{s,1}$ curve of Figure 8, with a downward shift of a few decibels). The 0 dB line in Figure 8 corresponds to noise-limited performance. From the SINR loss curves, we derive another very important STAP metric: the minimum detectable velocity (MDV). The MDV follows from a specification of the maximum tolerable SINR loss if the system is to meet detection performance requirements. If, for example, 5 dB is the tolerable loss, the radar system MDV in our example is $\pm(\lambda/2)\times(60\text{ Hz}) = \pm 9$ m/s for the optimal STP, where $\lambda = 0.3$ m is the RF wavelength.

The optimal and adaptive filter responses are also useful metrics. Given a weight vector, we compute the filter output response by sweeping across all space-time steering vectors of interest, viz.

$$H_{\text{optimal}}(\phi,\theta,f_d) = \left|\mathbf{w}_k^H\mathbf{s}_{s\text{-}t}(\phi,\theta,f_d)\right|^2;\ H_{\text{adaptive}}(\phi,\theta,f_d) = \left|\hat{\mathbf{w}}_k^H\mathbf{s}_{s\text{-}t}(\phi,\theta,f_d)\right|^2. \tag{33}$$

An example of the optimal response is given in Figure 7; the filter is tuned to broadside and 250 Hz. Doppler.

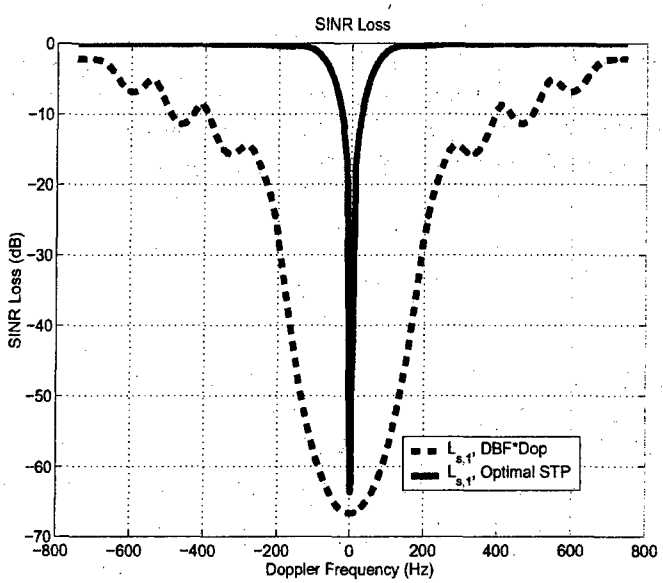

Figure 8. Example SINR loss curves.

5. Space-Time Detector Structures

In this section we begin by briefly considering some normalizations applied in STAP-type detectors. Then, we discuss the concept of interference rank, examine several reduced-rank or reduced-dimension processing schemes, and consider the role of diagonal loading in STAP implementation.

Recall, in (20) and (29) the optimal or adaptive weight vector includes arbitrary scalars μ and β, respectively. While a particular selection of either scalar does not affect SINR, certain selections are convenient in subsequent processing. A commonly used detector and normalization is given by

$$\left|\mathbf{w}_k^H \mathbf{x}_k\right|^2 = \frac{\left|\mathbf{s}_{s\text{-}t}^H \mathbf{R}_k^{-1} \mathbf{x}_k\right|^2}{\mathbf{s}_{s\text{-}t}^H \mathbf{R}_k^{-1} \mathbf{s}_{s\text{-}t}} \underset{H_0}{\overset{H_1}{\gtrless}} \eta_1 \tag{34}$$

where $\mu = 1/\sqrt{\mathbf{s}_{s\text{-}t}^H \mathbf{R}_k^{-1} \mathbf{s}_{s\text{-}t}}$ in this case [9]. Let $\tilde{\mathbf{w}}_k = \mathbf{R}_k^{-1} \mathbf{s}_{s\text{-}t}(\phi, \theta, f_d)$, such that $\mathbf{w}_k = \mu \tilde{\mathbf{w}}_k$. Then, $1/\mu^2$ equals the output noise power of the filter with weight vector $\tilde{\mathbf{w}}_k$. In the unknown covariance case, we substitute the form in (28) into (34), yielding the modified sample matrix inversion (MSMI) detection statistic. In this case, (34) becomes

$$\left|\hat{\tilde{\mathbf{w}}}_k^H \mathbf{x}_k\right|^2 \underset{H_0}{\overset{H_1}{\gtrless}} \eta_1 \left(\mathbf{s}_{s\text{-}t}^H \hat{\mathbf{R}}_k^{-1} \mathbf{s}_{s\text{-}t}\right); \quad \hat{\tilde{\mathbf{w}}}_k = \hat{\mathbf{R}}_k^{-1} \mathbf{s}_{s\text{-}t}, \tag{35}$$

where we further calculate η_1 for the single channel, single pulse case with the noise variance set to unity. The term $1/\beta^2 = \mathbf{s}_{s\text{-}t}^H \hat{\mathbf{R}}_k^{-1} \mathbf{s}_{s\text{-}t}$ modifies the threshold to account for varying interference power levels (and integration gain). As shown in [9], this normalization leads to a constant false alarm rate (CFAR) property. As the noise power increases, the threshold increases to maintain CFAR, and vice versa.

Kelly formulated the generalized likelihood ratio test (GLRT) of the form

$$\frac{\left|\mathbf{s}_{s\text{-}t}^H \hat{\mathbf{R}}_k^{-1} \mathbf{x}_k\right|^2}{\mathbf{s}_{s\text{-}t}^H \hat{\mathbf{R}}_k^{-1} \mathbf{s}_{s\text{-}t} \left(1 + \frac{1}{K} \mathbf{x}_k^H \hat{\mathbf{R}}_k^{-1} \mathbf{x}_k\right)} \underset{H_0}{\overset{H_1}{\gtrless}} K\eta_2, \tag{36}$$

where K is the number of secondary data samples [10]. This detector also possesses the CFAR property. It differs in form from (34) as a result of different modeling assumptions: in the GLRT, the detector assumes the covariance matrix is unknown. The derivation of (36) generally follows the same LRT development described in the initial sections of this report. Observe the detector can be written

$$\left|\hat{\mathbf{w}}_k^H \mathbf{x}_k\right|^2 \underset{H_0}{\overset{H_1}{\gtrless}} \left(\eta_1\ \mathbf{s}_{s\text{-}t}^H \hat{\mathbf{R}}_k^{-1} \mathbf{s}_{s\text{-}t}\right)\left[\frac{K\eta_2}{\eta_1}\left(1+\frac{1}{K}\mathbf{x}_k^H \hat{\mathbf{R}}_k^{-1}\mathbf{x}_k\right)\right]. \tag{37}$$

A data dependent threshold modifies the MSMI threshold to account for the finite number of training data. As K becomes large, the test statistic approaches the MSMI result.

Figure 9 provides a block diagram of the overall process. Essentially, a STAP filter suppresses interference. The scalar filter output is multiplied by a data-dependent "threshold multiplier" – either $1/\hat{\alpha}^2$ or $1/\hat{\alpha}^2$ multiplied by the factor in brackets on the far right-hand side of (37) – and then compared to a fixed threshold.

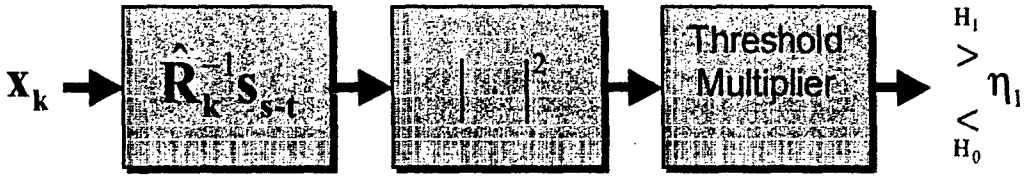

Figure 9. Embedded CFAR in STAP processing.

Other normalizations are possible. For instance, the minimum variance beamformer chooses $\mu_{MV} = 1/\mathbf{s}_{s\text{-}t}^H \mathbf{R}_k^{-1} \mathbf{s}_{s\text{-}t}$ [11]. This normalization accordingly affects the interpretation of the data.

STAP is effectively a catch-all for a class of linear filters involving SINR maximization of the input data [8]. We can broadly characterize the class of STAP methods into two sub-groups: reduced-dimension (RD) STAP and reduced-rank (RR) STAP. RD-STAP involves data-independent transformations and bin selection, whilst RR-STAP is a data-dependent approach. We'll consider these two sub-groups in turn. The ultimate goals of either RD-STAP or RR-STAP include reducing computational burden and reducing required sample support.

Figure 10 is an overview of RD-STAP methods. The processor can transform the space-time data via the data independent transformation

$$\breve{\mathbf{x}}_k = \mathbf{T}^H \mathbf{x}_k \in C^{PQx1}, \tag{38}$$

where $\mathbf{T}$ is the transformation matrix and ideally $PQ << NM$. Common transformations include beamforming and Doppler processing steps. By analogy to our prior discussion, the required optimal RD weight vector yielding maximum SINR in the reduced space is

$$\breve{\mathbf{w}}_k = \left(\mathbf{T}^H \mathbf{R}_k \mathbf{T}\right)^{-1} \mathbf{T}^H \mathbf{s}_{s\text{-}t}. \tag{39}$$

In practice, the RD-STAP weight vector is given by

$$\hat{\breve{\mathbf{w}}}_k = \hat{\breve{\mathbf{R}}}_k^{-1} \breve{\mathbf{v}};\ \hat{\breve{\mathbf{R}}}_k = \frac{1}{K'}\sum_{m=1}^{K'} \breve{\mathbf{x}}_m \breve{\mathbf{x}}_m^H;\ \breve{\mathbf{v}} = \mathbf{T}\ \mathbf{v}_{s\text{-}t}, \tag{40}$$

where $K' << K$ indicates a reduction in required secondary data. Also, RD-STAP computational burden is $O(P^3Q^3)$, which typically is far less than for the space-time case with computational load $O(N^3M^3)$.

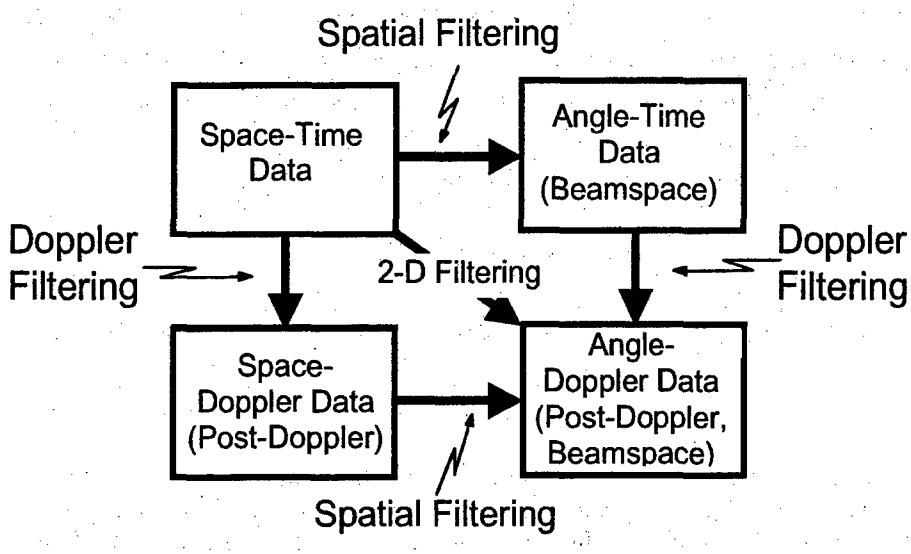

Figure 10. Overview of reduced-dimension STAP methods.

Examples of RD-STAP techniques are available in [8, 12-13]. The Factored Time-Space (FTS) algorithm is a post-Doppler method suitable for long coherent dwells and higher radial velocity targets. The FTS method essentially involves spatial notching of the clutter in a given Doppler filter. Since FTS provides no temporal adaptivity, it is not a true STAP algorithm. To enhance performance with only modest increase in required sample support and computational burden, DiPietro proposed the Extended Factored Algorithm (EFA) [12]. The EFA method involves adaptively combining several adjacent Doppler filters (typically three) and all spatial channels. The EFA method often exhibits performance very close to the theoretical joint-domain (JD) space-time bound. To provide diversity in spatial and temporal DoF, Wang developed the Joint Domain Localized (JDL) technique [13]. JDL is a post-Doppler, beamspace method. Basically, the processor forms multiple beams, then Doppler processes each beam, and finally selects a collection of adjacent angle-Doppler bins over which to adapt the filter response. JDL provides good performance with very low training data requirements and very modest computational burden. Three adjacent beams by three adjacent Doppler bins is a typical localized JDL processing regions [13]. Figure 11 shows the SINR loss for each of the RD-STAP methods for the same example shown in Figure 8; we compare performance against the upper bound given by the JD space-time filter. Each RD-STAP method uses secondary data support of twice the processor's DoF. Note, our discussion did not include an example of a joint pre-Doppler and angle method.

Reduced-rank STAP methods involve data-dependent transformations and selections [14-15]. Consider the following formulation of the optimal weight vector

$$\mathbf{w}_k = \frac{1}{\lambda_o}\left[\mathbf{s}_{s\text{-}t}(\phi,\theta,f_d) - \sum_{m=1}^{NM}\frac{\lambda_m - \lambda_o}{\lambda_m}\alpha_m \mathbf{q}_m\right] \tag{41}$$

where λ_m is an eigenvalue of $\mathbf{R}_k$ with eigenvector $\mathbf{q}_m$, $\lambda_o = \min(\lambda_m)$ and α_m is the projection of the m^{th} eigenvector onto the quiescent response given by $\mathbf{s}_{s\text{-}t}$. In the adaptive case, we substitute $\hat{\mathbf{R}}_k$ for $\mathbf{R}_k$. Thus, we can view STAP as a pattern synthesis problem: based on a sensing of the environment, the processor places notches in the quiescent pattern to mitigate interference. Notice from (41) that no subtraction occurs for eigenvalues at the noise floor. The principal components (PC) method involves truncating the summation in (41) to accommodate only the largest eigenvalue terms. A benefit of this approach is a reduction in training data support. However, computational burden remains high, since the processor must compute eigenvalues and eigenvectors. An alternative, yet similar, approach to the PC method is given in [15] and is known as the Principal Components Inverse (PCI) method. Specifically, the processor computes

$$\mathbf{x}_k' = \left(\mathbf{I}_{NM} - \mathbf{Q}_I \Omega_I \mathbf{Q}_I^H\right)\mathbf{x}_k = \mathbf{x}_k - \sum_{m=1}^{P} \frac{\lambda_{I/m}}{\lambda_m} \alpha_{x/m} \mathbf{q}_m \tag{42}$$

where $\mathbf{Q}_I$ is the collection of interference eigenvectors (corresponding to each column), Ω_I is the matrix enabling perfect subtraction of the designated interference term in $\mathbf{x}_k$, $\alpha_{x/m}$ is the projection of the m^{th} eigenvector onto the data snapshot, and $\lambda_{I/m}$ is a non-zero eigenvalue of the interference-only eigenvector. In practice, we cannot estimate the interference-only eigenvector, and so for strong, low-rank interference we can set $\lambda_{I/m} = \lambda_m$, or $\Omega_I = \mathbf{I}_{NM}(:,1:P)$ for the P dominant components. Since

$$\mathbf{Q}_I \mathbf{Q}_I^H + \mathbf{Q}_N \mathbf{Q}_N^H = \mathbf{I}_{NM}, \tag{43}$$

where $\mathbf{Q}_N$ is the collection of noise eigenvectors (corresponding to each column), we can express (42) as

$$\mathbf{x}_k' = \mathbf{Q}_N \mathbf{Q}_N^H \mathbf{x}_k. \tag{44}$$

Equation (44) indicates that PCI can operate by using the noise subspace information. This stands in contrast to the PC method using the dominant subspace. Using our familiar example found in Figure 8 and Figure 11, Figure 12 compares the SINR loss for the PC and PCI methods against the upper bound defined by the JD STAP solution. The adaptive solution involves a secondary data set of one times the joint domain processor's total DoF, i.e. $K = NM$. The interference subspace has dimension 40, and so we use all 40 principal components for the PC method and all 312 noise subspace components for the PCI approach. Armed with this information, observe that both methods very closely follow the optimal performance. In practice, choosing the correct number of components is challenging and an incorrect selection can lead to performance degradation.

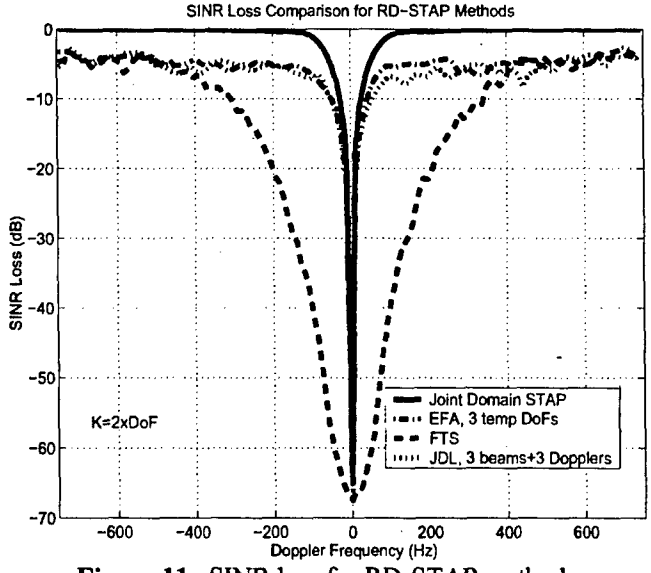

Figure 11. SINR loss for RD-STAP methods.

Oftentimes, the covariance estimate is ill-conditioned or exhibits a perturbed noise subspace. In such cases, diagonal loading can be used to tailor performance [16]. As the name implies, diagonal loading involves adding an identity matrix to the covariance estimate to stabilize the noise floor and condition the covariance matrix estimate for subsequent numerical processing. From (41) we find that if the noise eigenvalues are perturbed, the processor subtracts a noise eigenvector from the quiescent response. Noise eigenvectors exhibit random sidelobe behavior. Hence, the adaptive filter response for low secondary data set sizes shows elevated sidelobes. High sidelobes can be problematic, leading for instance to sidelobe target detection. Diagonal loading cures this problem.

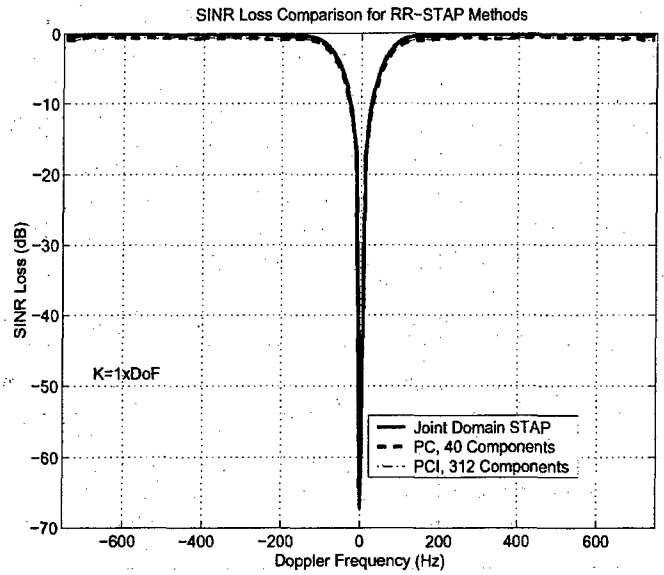

Figure 12. SINR loss for RR-STAP methods.

6. Heterogeneous Clutter Effects

Adaptive processing requires secondary (training) data to estimate unknown quantities. For example, traditional, CFAR circuits [6], which involve scalar adaptive processing, typically select ten range bins on each side of the test and guard cells to estimate the ambient interference power. STAP similarly requires a training interval to estimate the null hypothesis covariance matrix. However, in contrast with the scalar CFAR case, the potentially high dimensionality of STAP requires training data selection over large regions. Suppose a STAP-based radar has $N = 32$ (thirty-two pulse dwell) and $M = 11$ (eleven spatial channels). The minimum training set is then on the order of $2NM = 704$ range bins. If the instantaneous bandwidth is 1 MHz, training data comes from a region extending over 10 km! Over the 10 km interval, the changing cultural features of the clutter environment lead to range-angle variation of ground clutter returns. This range-angle variation in then ground clutter return is known as clutter heterogeneity [17-18].

Common STAP training approaches include the following methods: block selection, windowed block selection and sliding window. In the block selection approach, the processor selects training data sequentially from a given region and then applies it to a distinct primary data region. The windowed block selection is a variant of the block selection technique; the training data is chosen sequentially about the primary data region, with each half of the training data coming from opposite sides of the primary data regions. In the sliding window approach, the processor attempts to localize training data within the vicinity of the primary data. Thus, the processor chooses training data symmetrically about the test data and adjacent guard cells. The sliding window approach is most computationally intense, but numerical techniques are available to alleviate the computational burden of both covariance estimation and covariance matrix inversion.

As a result of clutter heterogeneity, the precise yet unknown covariance matrix of a given range cell does not match other ranges. Hence, heterogeneous clutter environments are no longer iid. The non-iid environment gives rise to increased covariance estimation errors when using (28). Specifically, since $\mathbf{R}_p \neq \mathbf{R}_m$ in non-iid environments, where p and m indicate the training vector indices and $p \neq m$, it stands to reason that

$$E\left[\hat{R}_k\right] = E\left[\frac{1}{K}\sum_{m=1}^{K}\mathbf{x}_m\mathbf{x}_m^H\right] = \frac{1}{K}\sum_{m=1}^{K}E\left[\mathbf{x}_m\mathbf{x}_m^H\right] = \frac{1}{K}\sum_{m=1}^{K}\mathbf{R}_m \tag{45}$$

In other words, in heterogeneous clutter environments, the covariance matrix estimate tends to an average characteristic of the training data. For this reason, instantaneous performance can be poor.

Clutter heterogeneity manifests as a modification of certain components and the addition of additional terms. The angle and Doppler behavior of ground clutter is well defined and results from sensor geometry and

the platform velocity vector [8]. However, clutter amplitude and spectral width can vary, and target-like signals lead to extra components. Table 2 synopsizes the various classes of clutter heterogeneity and their impact on STAP performance.

The covariance estimation errors due to clutter heterogeneity lead to increased SINR loss, i.e. $L_{s,2}$ further approaches zero. Clutter heterogeneity is a challenge to STAP implementation, since adaptive and optimal performance does not necessarily converge with increasing number of training vectors. This realization is driving new STAP approaches holding the potential for robust performance in light of cluttter heterogeneity and other behavior (see the last entry in Table 2) leading to non-iid training data.

Table 2. Taxonomy of Clutter Heterogeneity

Heterogeneity Type	Cause	Impact on Adaptive Radar
Amplitude	Shadowing and obscuration, range-angle dependent change in clutter reflectivity, strong stationary discretes, sea spikes, urban centers, land-sea interfaces, *etc.*	Null depth depends on eigenvalue ratio – MLE "averaging" leads to underestimated eigenvalue magnitude, and consequently, uncancelled clutter and increased false alarm rate.
Spectral	Intrinsic clutter motion due to soft scatterers (trees, windblown fields, *etc.*), ocean waves; CNR-dependent spectral mismatch.	Null width set to mean spread – too narrow for some range cells, thereby leading to uncancelled clutter, seriously degraded MDV.
Moving Scatterers	Ground traffic, weather, insects and birds, air vehicles.	Mainlobe nulling, false sidelobe target declarations, distorted beam patterns, exhausts DoF.
Some Other Effects	Chaff, hot clutter, multi-bounce/ multipath, impact of platform geometry (*e.g.*, non-sidelooking or bistatic) on angle-Doppler behavior over range.	Combination of above effects.

7. Summary

In this set of lecture notes, we consider fundamental aspects of space-time adaptive detection theory. Section 2 describes the basics of hypothesis testing and relates this information to space-time detection. We introduce the likelihood ratio test in this section and through example demonstrate its use. Additionally, we describe the important whitening filter and matched filter concepts, relating them to the detection statistic. We then describe the application of signal diversity to advanced sensor design in Section 3. Signal diversity helps cull the weak target signal masked by interference. We provide an example showing the space-time characteristics of ground clutter and the advantage space-time degrees of freedom offer in enhancing detection performance. Section 4 first develops the maximum signal-to-interference-plus-noise ratio (SINR) filter, which maximizes the probability of detection in the Gaussian case. We then describe commonly used STAP performance metrics. We overview space-time detector structures in Section 5. In this section we first begin by examining STAP normalizations leading to a constant false alarm rate (CFAR) property. We then describe the philosophy behind reduced-dimension and reduced-rank STAP methods. Reduced-dimension STAP involves data independent transformations to characterize interference behavior in a smaller subspace. Advantages of reduced-dimension STAP include a requirement for smaller training data sets and minimal computational burden. Reduced-rank techniques involve data dependent transformations to derive a stochastic basis describing the interference properties. The principal component reduced-rank technique can achieve the upper bound on SINR performance when properly applied. We also introduce the principal components inverse method in this section. The paper concludes with a brief discussion of clutter heterogeneity. We define clutter heterogeneity and describe its impact on STAP implementation.

References

[1] M.I. Skolnik, Introduction to Radar Systems, 2nd Ed., McGraw Hill, New York, NY, 1980.

[2] W.L. Melvin, "Application of STAP in advanced sensor systems," *Proceedings of the Research and Technology Agency, North Atlantic Treaty Organization (RTA-NATO) Lecture Series 228 – Military Applications of Space-Time Adaptive Processing*, September 2002.
[3] D.H. Johnson and D.E. Dudgeon, Array Signal Processing: Concepts and Techniques, Prentice-Hall, Englewood Cliffs, NJ, 1993.
[4] J.V. DiFranco and W.L. Rubin, Radar Detection, Artech-House, Dedham, MA, 1980.
[5] L.E. Brennan and I.S. Reed, "Theory of adaptive radar," *IEEE Trans. AES*, Vol. 9, No. 2, March 1973, pp. 237-252.
[6] N. Levanon, Radar Principles, John Wiley & Sons, New York, 1988.
[7] I.S. Reed, J.D. Mallett, and L.E. Brennan, "Rapid convergence rate in adaptive arrays," *IEEE Trans. AES*, Vol. 10, No. 6, November 1974, pp. 853-863.
[8] R. Klemm, Space-Time Adaptive Processing: Principles and Applications, IEE Radar, Sonar, Navigation and Avionics 9, IEE Press, 1998.
[9] W.S. Chen and I.S. Reed, "A new CFAR detection test for radar," *Digital Signal Processing*, Vol. 1, Academic Press, 1991, pp. 198-214.
[10] E.J. Kelly, "An adaptive detection algorithm," *IEEE Trans. AES*, Vol. AES-22, No. 1, March 1986, pp. 115-127.
[11] S. Haykin, Adaptive Filter Theory, Third Ed., Prentice-Hall, Upper Saddle River, NJ, 1996.
[12] R.C. DiPietro, "Extended factored space-time processing for airborne radar," *Proc. 26th Asilomar Conf.*, Pacific Grove, CA, Oct. 1992, pp. 425-430.
[13] H. Wang and L. Cai, "On adaptive spatial-temporal processing for airborne surveillance radar systems," *IEEE Trans. AES*, Vol. 30, No. 3, July 1994, pp. 660-670.
[14] W.F. Gabriel, "Using spectral estimation techniques in adaptive processing antenna systems," *IEEE Trans. AP*, Vol. 34, No. 3, March 1986, pp. 291-300.
[15] D.W. Tufts, I. Kirsteins and R. Kumaresan, "Data-adaptive detection of a weak signal," IEEE Trans. AES, Vol. AES-19, No. 2, March 1983, pp. 313-316.
[16] B.D. Carlson, "Covariance matrix estimation errors and diagonal loading in adaptive arrays," IEEE Trans. AES, Vol. 24, No. 4, July 1988, pp. 397-401.
[17] W.L. Melvin, "Space-time adaptive radar performance in heterogeneous clutter," *IEEE Trans. AES*, Vol. 36, No. 2, April 2000, pp. 621-633.
[18] W.L. Melvin, J.R. Guerci, M.J. Callahan, M.C. Wicks, "Design of adaptive detection algorithms for surveillance radar," in *Proc. IEEE 2000 Int'l Radar Conf.*, Alexandria, VA, 7-12 May 2000, pp. 608-613.