NOTICE

AVIS

The quality of this microform is heavily dependent upon the quality of the original thesis submitted for microfilming. Every effort has been made to ensure the highest quality of reproduction possible.

La qualité de cette microforme dépend grandement de la qualité de la thèse soumise au microfilmage. Nous avons tout fait pour assurer une qualité supérieure de reproduction.

If pages are missing, contact the university which granted the degree.

S'il manque des pages, veuillez communiquer avec l'université qui a conféré le grade.

Some pages may have indistinct print especially if the original pages were typed with a poor typewriter ribbon or if the university sent us an inferior photocopy.

La qualité d'impression de certaines pages peut laisser à désirer, surtout si les pages originales ont été dactylographiées à l'aide d'un ruban usé ou si l'université nous a fait parvenir une photocopie de qualité inférieure.

Reproduction in full or in part of this microform is governed by the Canadian Copyright Act, R.S.C. 1970, c. C-30, and subsequent amendments.

La reproduction, même partielle, de cette microforme est soumise à la Loi canadienne sur le droit d'auteur, SRC 1970, c. C-30, et ses amendements subséquents.

Canada

RESOURCE CONSTRAINED DETECTION: A NEW DETECTION CRITERION AND ITS APPLICATION TO RADAR SYSTEMS DESIGN AND ANALYSIS

by

Martie Meryle Goulding

B.A.Sc. (E.E.) University of British Columbia, 1987

M.A.Sc. (Eng.Sc.) Simon Fraser University, 1989

A THESIS SUBMITTED IN PARTIAL FULFILLMENT

OF THE REQUIREMENTS FOR THE DEGREE OF

DOCTOR OF PHILOSOPHY

in the School

of

Engineering Science

© Martie Meryle Goulding 1993

SIMON FRASER UNIVERSITY

February 1993

Canada

APPROVAL

Name: Martie Meryle Goulding

Degree: Doctor of Philosophy

Title of thesis: Resource Constrained Detection: A New Detection Criterion
                      and its Application to Radar Systems Design and Analysis

Examining Committee: Dr. Shahram Payandeh
                         Chair

Dr. John S. Bird
Senior Superivsor

Dr. James K. Cavers
Supervisor

Dr. Paul K.M. Ho
Supervisor

Dr. Shawn Stapletoh
Internal Examiner

Dr. Albert W. Bridgewater
External Examiner
Department of National Defence

Date Approved: FEB 4, 1993

ii

PARTIAL COPYRIGHT LICENSE

I hereby grant to Simon Fraser University the right to lend my thesis, project or extended essay (the title of which is shown below) to users of the Simon Fraser University Library, and to make partial or single copies only for such users or in response to a request from the library of any other university, or other educational institution, on its own behalf or for one of its users. I further agree that permission for multiple copying of this work for scholarly purposes may be granted by me or the Dean of Graduate Studies. It is understood that copying or publication of this work for financial gain shall not be allowed without my written permission.

Title of Thesis/Project/Extended Essay

"Resource Constrained Detection: A New Detection Criterion and its

Application to Radar Systems Design and Analysis"

_____

_____

Author:

_____
(signature)

Martie M. Goulding
_____
(name)

February 9, 1993
_____
(date)

Abstract

Real-world detection systems, such as radars, control resources that are limited in their processing ability. This thesis introduces a new detection criterion designed to operate under the limitations imposed by such resource constraints. In contrast to the Neyman-Pearson framework which constrains the false alarm rate, the rate-constraint criterion maximizes the detection probability of the test subject to a constraint on the threshold crossing (hit) rate. The resulting likelihood ratio test is practical since the hit rate is both controllable and observable, unlike the false alarm rate which is only controllable.

The single-stage rate-constraint criterion is extended to the multi-stage case where a number of individual detectors are cascaded; each test is slower but more capable than the previous. A new parameter, the SLOC function (the slope of the Receiver Operating Characteristic curve when plotted on log-log axes), is derived and seen to be critical to the performance of a detection system utilizing such a sequence of tests. Provided the SLOC numbers for the individual tests are properly ordered, the overall system detection performance under a wide range of criteria (Bayes, Neyman-Pearson, Maximum Information) is maximized when the rate-constraint criterion is used for all stages but the last. The resulting sequence of tests is again practical to control since each threshold depends only on local variables.

The properties of the SLOC function are investigated for a number of radar target, clutter, and processing models. Included are new results for the noncoherent integration of tone and Rayleigh fading targets in K-distributed clutter.

Techniques are presented for dealing with the variance of the hit rate. Analysis is given for two types of systems: one for actuators that require the same processing power to treat false alarms and targets, and the second for actuators that treat false alarms and targets differently. A test that chooses the targets to be investigated by ranking and choosing the largest instead of using an explicit threshold is also investigated.

For Dad, with loving memory,
for my child to come, with longing anticipation,
and for my one and only Sue.

Of making many books there is no end, and much study wearies the body.

Ecclesiastes 12:12

Acknowledgements

I would like to express my gratitude to the Radar Research Division at the Defence Research Establishment Ottawa and the National Sciences and Engineering Research Council for their financial support.

Thanks to all of the guys in the Underwater Reserach Lab who put up with me for so long, and especially my office mates, Neil (Will I ever finish my thesis?) Fried, William (Dr. Bill) McMullan and Paul (What should I do my thesis on?) Kraeutner. I would also like to acknowledge the great job done by our computer support staff, Chao Cheng and Ken Huckell, and by our Grad Secretary, Brigitte Rabold.

Thanks to my family (Sue, Mom, and Max) for your patience, understanding and support.

Finally, Dr. Bird (John), thank you for being generous with your encourgagement, patience, and ideas.

Contents

List of Tables

List of Figures

List of Symbols

A, a Amplitude of return

b Shape parameter for K-Distribution

$\mathcal{B}_n^C(p)$ Binomial probability, $= \begin{pmatrix} C \\ n \end{pmatrix} p^n (1-p)^{C-n}$

C Number of cells examined per service cycle

C_{ij} Cost of declaring H_i when H_j is true

$\mathcal{D}$ Class of decision functions

$E_\theta(x)$ Expected value of x given $\Theta = \theta$

$f_i(x)$ Probability density function of x under H_i

$F_i(x)$ Probability distribution function of x under H_i

g Global function to maximize for multi-stage system

$G_q(s)$ Moment generating function for random variable q

$G_0(s)$ Moment generating function under H_0

$G_1(s)$ Moment generating function under H_1

H_0 Target Absent Hypothesis (null hypotheses)

H_1 Target Present Hypothesis (alternative hypothesis)

I Identity Matrix

$I_0(x)$ Zeroth order modified Bessel function of first kind

$\mathcal{I}$ Shannon's Mutual Information function

k Multiplier for Normalization (CFAR) threshold

$I_0(x)$ Modified Bessel Function of zeroth order and first kind

$J_0(x)$ Ordinary Bessel Function of zeroth order

$K_\nu(x)$ Modified Bessel Function of νth order and third kind

L_Q	Time average queue length
$\ln(x)$	Natural logarithm
$\ell(x)$	Likelihood ratio $= \frac{f_1(x)}{f_0(x)}$
$\mathcal{L}^{-1}$	Inverse Laplace Transform
M	Number of pulses integrated
N_0	Number of empty cells interrogated per service cyle
N_1	Number of target cells interrogated per service cyle
N_{st}	Number of stages in a detection sequence
N_Q	Maximum queue length
p, q	Probability of success and failure for binomial trial (chapter 6)
pd	Probability of detection
pf	Probability of false alarm
Pc_k	Cumulative probability of detection for M/N detector
pc_k	Probability mass function for detection for M/N detector
pd_{eff}	Effective probability probability (including randomization effects)
pf_{eff}	Effective false alarm probability (including randomization effects)
pd_{ov}	Overall detection probability (including queue blockage)
pd_T	Total detection probability for a sequence of tests
pf_T	Total false alarm probability for a sequence of tests
$P_i(\bullet)$	Probability of ($\bullet$) given H_i
$\tilde{P}_i$	Steady-state probability of i cells in queue
P_b	Probability of cell being blocked out by a full queue
q	Test statistic for quadratic detector (chapter five)
$Q(A, t)$	Marcum's Q function
r	hit rate ($r = \pi_1 pd + \pi_0 pf$)
r_k	Return function for stage k ($=$ detection probability for stage k)
$\mathcal{R}(\delta)$	Bayes cost associated with decision rule δ
SDR	Specular-to-Diffuse Ratio (for Rician target)
SNR	Signal-to-Noise Ratio
S_k	Input state function for stage k
S	Number of servers
$\tilde{S}_k$	Output state function for stage k
W_Q	Mean queuing delay

t	Threshold for threshold test
X	Random Variable
x	Outcome of random variable X
$\overline{X}$	Random vector of observations
$\overline{x}$	Outcome of random vector $\overline{X}$
z	Magnitude of return
α	Probability of false alarm
β	Probability of miss
Γ	Borel set containing observations of X
Γ_i	Set of x indicating belief in H_i
$\Gamma_{pc}(z)$	Moment generating function for pc_k
γ	Randomizing constant for rate constraint and Neyman Pearson tests
$\delta(x)$	Decision rule, $=P(H_0\text{rejected}|X=x)$
η	SLOC number $(=\frac{\partial \ln pd}{\partial \ln pf})$
Θ	Set of parameters describing distributions of X
κ	Randomization number for test operating under rate constraint
Λ_i	Parameter space under H_i
λ	Threshold for Likelihood ratio test
λ_{NP}	Threshold for Neyman-Peason test
λ_{rc}	Threshold for rate constraint test
λ_S	Average carried load by system
$\mu(x)$	Measure of x
ν	Shape parameter for K-Distribution
ξ_i	A posteriori probability of declaring H_i to be true
π_i	A priori probability of H_i
π_0^k	Rate of empty cells into stage k
π_1^k	Rate of target cells into stage k
ϱ	Single pulse signal to noise ratio
ρ	Output rate test normalized to actuator's maximum processing rate
σ^2	Power of Rayleigh component in return (noise plus target component)
σ_d^2	Power of signal Rayleigh component

φ Phase for return

$\Phi(\omega)$ Hankel transform

ω Argument for Hankel transform

Chapter 1

Introduction

1.1 Why More Detection Theory?

The application of statistical hypothesis testing techniques to radar has a long history. Neyman and Pearson published their classic work in 1933 [1][1]. The Bayesian or cost function approach to decision theory was promoted in the 1950's by Good [2] and Savage [3]. Woodward's "Probability and Information Theory with Applications to Radar", published in 1953, brought together much of the theory to be applied to the problem of radar detection [4]. By 1960, Middleton [5] was able to objectively assess the applicability of decision theory to the radar detection problem. He listed three problems that continue to plague practical deployments, namely

1. the uncertainty in the choice of an optimality criterion,

2. the arbitrariness of cost assignments, and

3. the lack of a priori information.

The same detection criteria that Middleton considered in 1960 are still offered as solutions to radar detection today, and they continue to have the same shortcomings. Let us consider the latter two problems first.

Detection theory has its roots in statistical decision theory; this is appropriate since detections are in fact decisions, with real consequences. In decision theory a cost function is

---

[1]The list of references is found at the end of the thesis.

often specified to characterize the consequences of correct and incorrect decisions. However, it is not possible to specify meaningful costs for the radar problem. Moreover, the Bayesian techniques also require a priori information which is seldom available (for example the a priori probability of a target being present in a given cell, and the distribution of the radar return given the presence or absence of a target in the cell.)

The Mini-max approach eliminates the requirement for a priori information by finding the detection rule that minimizes the maximum cost; however, meaningful cost functions are again required. Information theory provides a less arbitrary cost function but both the costs and solution are functions of the priors.

Due to the inapplicability of cost function approaches, Neyman-Pearson theory is most widely applied to sensor systems. A design false alarm rate is chosen, and the resulting detection probability is accepted. By adopting the Neyman-Pearson approach, the problem of choosing an optimality criterion is replaced by the problem of choosing an acceptable false alarm probability.

Despite its prominence, Neyman-Pearson theory has a number of problems in its application to Radar systems. First, the underlying noise distribution must be known in order to achieve the design false alarm rate. A control loop cannot be used because the false alarm rate, while controllable, is not observable. Generally, it is assumed that the form of the distribution is known to within a small number of parameters, which are then estimated and used to derive the appropriate test. The resulting Constant False Alarm Rate (CFAR) methods perform well in the environments they are designed for. Unfortunately, in non-homogeneous clutter environments, the interference distribution is not known. Suboptimal non-parametric approaches are then often employed.

The major deficiency of the Neyman-Pearson framework is that the false alarm rate is only indirectly related to the goals of the detection system. Thus we have a trade-off: in order to achieve a tractable theory that doesn't presuppose unavailable information, the solution is distanced from practical measures.

To appreciate the ramifications of neglecting the context of the test, consider the problem of setting the appropriate design false alarm rate for a very simplistic system consisting of a gun control radar and a gun designed to protect a ship. Assume the gun is capable of firing one round per second and is assigned the task of destroying all incoming targets. We would like to design a system that has the highest probability of intercepting the targets. A

Neyman-Pearson strategy to accomplish this would set the radar threshold to yield a given false alarm probability and shoot at any cell that crosses that threshold. This strategy results in a firing rate that is acceptable provided no targets are present and the false alarm rate is less than the maximum firing rate. As targets appear, the firing rate increases until, if the number of targets is large enough, targets are missed because the gun is busy. This is an overload condition that puts the ship in jeopardy. To reduce the potential for overload the radar false alarm rate is lowered. But this reduces the radar's ability to detect small targets even when the number of targets is small and there is no danger of an overload; so again the ship is unnecessarily vulnerable.

As a second example, consider a Track-While-Scan (TWS) radar system. A TWS system consists of two stages: a detection stage and a tracking computer that uses the results of the detector to determine tracks for the targets. With the Neyman-Pearson approach, we would set the first stage to allow a specific number of false alarms per radar sweep. As with the gun control radar, if the first stage is not selective enough, the tracking computer will be overloaded. If the first stage is too selective, the probability of detection for smaller targets is smaller than necessary.

The solution for the TWS example is given by Trunk in [6]: "The (false alarm probability) of the detector should be as high as possible without saturating the tracking computer." The computer is a resource that should be fully utilized, but not overloaded, by the data out of the detection stage. Thus we have a resource-constrained detection system where the resource is the processing rate of the tracking computer. Similarly, it is clear in the first example that the degree of protection is determined by the firing rate of the gun and therefore the gun should be fully utilized. The radar should simply direct the gun to the radar cells most likely to contain targets, subject to the rate of the gun. This is a resource-constrained detection system where the resource is the gun's firing rate.

In either example, the threshold for the radar is easily set to keep the gun or tracker busy but not overloaded. There need be no appeal to probability distributions to calculate false alarm rates. If no targets are present the threshold drops to keep the gun firing or the computer busy, albeit with false alarms. But the radar is very sensitive to small targets and therefore the system is optimally protected. As the number of radar cells crossing the threshold increases, so does the threshold and therefore the cells most likely to contain targets are passed to the gun or tracker.

While it seems that the resource-constrained criterion is wasteful it does provide the greater degree of protection. After all, is that not the reason for the system in the first place? If the criterion is held in disfavour then there is a hidden resource that is actually driving the design. If that is the case, this hidden resource should be brought into the open and the control criterion built around it.

Whatever the case may be, real world detection systems are usually governed by a variation on the resource-constraint principle. Therefore it is beneficial to develop a detection theory around this principle to guide future designs and to assess the performance of the present ones.

The concept of matching the performance of a detection stage to the follow-on resources dependent on the detector is relatively new [7], with perhaps the first example being the integrated detection and tracking system of Kurniawan et al [8]. Despite the fact that the philosophy has been expressed previously, this dissertation provides the general framework and analysis required to justify the concepts that were hitherto heuristic in nature.

Before narrowing our scope to man-made detection systems, we present a brief overview of general information-processing systems. The material is interesting because it provides a broader context for the work of this thesis.

1.2   General Principles of Information Processing, and Their Application to Detection Theory

Two relevant studies are Resnikoff's examination of biological detection systems [9] and Conant's study of general information-processing theory [10]. Despite their very different approaches and interests, the two studies both list the following principles[2].

The first principle is the *invariance of information-processing structures and measures under appropriate group actions.* This concept emphasizes the relativistic nature of measurements with regard to information content. A simple example of information invariance is the independence of information content on the zero point or units of measurement. In a detection context, the principle points to the need for test statistics that are relevant to the task at hand. To quote Conant: "Arrange for the sensors of the system to respond only to those aspects of the environment which are potentially relevant." In chapter two we present

---

[2]Here we use Resnikoff's names for the concepts.

ways that detection systems determine the statistics to indicate relevance.

The second principle is the use of *hierarchical structures* in information-processing structures. The idea here is to use faster, lower resolution systems to cue slower, higher resolution systems. For example, in the human vision system, the peripheral vision system has comparatively low resolution but can scan a large area rapidly. In contrast, the foveal system has very high resolution but is incapable of quickly scanning the entire field of view. The peripheral system cues the foveal system to objects of interest. Hierarchical or cueing systems are prevalent in natural and man-made detection systems. A few technical applications are combined radar/lidar systems [11], medical diagnoses [12], and mine-hunting sonar systems [13]. Chapter four of this thesis is devoted to the application of resource-constrained theory to multiple-stage detection systems using cueing.

The third principle noted by Resnikoff is that information-processing systems tend to *extremize the quantity of information relative to some processing cost constraint.* Detection systems tend to be *resource limited* in that their performance will increase with increased processing resources [14]. "The resources are such things as processing effort, the various forms of memory capacity and communications channels... (Resources) are always limited") [14]. The constraints imposed by the limited resources necessitate the need for the *Selective Omission of Information* [9] or information blocking [10]. In detection systems consisting of multiple stages, this means that earlier stages pass on only that information relevant to subsequent stages. For example, in most cueing systems, the faster system passes on only the addresses of those cells considered most likely to contain targets. The actual value for the test statistic is not passed on, nor are the return values for any of the cells considered unlikely to be target candidates. Of course the information in the test statistic could be passed on, leading to increased performance at the cost of more resources [15].

In biological detection systems optimality is not the ultimate goal. The aim of this thesis is not to prove or claim optimality of resource constrained systems for all radar scenarios. While it is true that the principles derived herein are optimal over a general range of situations typical of radar, the more important claim is that the techniques do provide adequate performance. By adequate, we mean that a system operating under our principles will achieve performance the same or very close to that which would be achievable in theory, if the designer had perfect knowledge, *over the meaningful range of operation for the radar.* For example, we do not concern ourselves with possible losses that arise with

overall system false alarm probability rates lower than, say, 10^{-8} or so.

While the emphasis of this thesis is on the application of our new concepts to radar systems, the concepts are valid for any detection system operating in an environment where the presence of a target is relatively rare. For example, the theory could be applied to medical diagnoses or to the problem of finding the best applicant for an employment opportunity. We now turn our attention to a brief outline of the thesis.

1.3   Outline and Scope of Thesis

In chapter two we review classical detection theory and discuss the problems of applying existing criteria to practical radar scenarios. The review gives an opportunity to introduce the relevant concepts and notation required for the remainder of the thesis.

Rate constraint detection theory is introduced in chapter three. We prove that the optimal form for a detection test operating under a rate constraint is a likelihood ratio test with its threshold set to match the output hit-rate to the maximum processing rate of the following processing stage. A thorough analysis of a single-stage system controlling a resource constrained actuator is given, along with a comparison with a fixed-threshold (Neyman-Pearson) based system. Graphical and analytical arguments are given to show the superiority of the rate-constraint criterion. A simplified obstacle-avoidance sonar analysis gives an example of rate-constraint theory's applicability at the control and command level.

In chapter four, we proceed to multi-stage systems. A new parameter called the SLOC function is derived[3]. The region of optimality for the rate-constrained principle is determined by the SLOC functions for the individual detector stages. The SLOC function gives a relative measure of goodness for any two tests when the governing distributions are known. Unlike most of the work involved in the fusion of multiple tests, our results do not require the stages to perform synchronous tests on identically distributed returns. We use a dynamic programming argument to show that, provided the SLOC's for the two tests are appropriately ordered, the function for the detection performance of a multiple stage cueing system is a separable and monotonic function. Therefore, the problem is decomposable into a series of single-stage sub-problems, each of which may be solved using the theory of chapters two and three. We further show that the optimal operating strategy for any overall detection

---

[3]The SLOC (Slope of Log receiver Operating Characteristic Curve) function is the slope of the curve generated by plotting the detection probability as a function of the false alarm probability on a log-log scale.

criterion is to operate every stage but the last under the rate-constraint criterion, with the last stage threshold chosen according to the overall criterion. The resulting detection structure is robust, exhibiting graceful degradation when loss of a stage occurs.

In chapter five, the SLOC function is derived and examined for a number of radar target and clutter models. SLOC curves are given for Marcum, Swerling, Rician and Two-tone models in Gaussian noise, and for non-fluctuating and Rayleigh-fading targets in K-distributed clutter. New results are given for the noncoherent integration of non-fluctuating and Rayleigh-fading targets in a mixture of K-distributed and Gaussian noise using the new technique of [16]. The applicability of normalization or CFAR techniques in the resource-constraint framework is also briefly discussed. Finally, a practical example is given, whereby a large reduction in computational complexity is gained by using a stage performing non-coherent integration to cue a stage employing coherent integration.

A test operating under a rate constraint is required to have a fixed output rate; however, threshold tests will always have some variance in their output due to the stochastic nature of the noise. Chapter six discusses some of the ways that the variance in the output rate can be handled. The simplest technique is to introduce a queue between the detection stage and the subsequent actuator (or detection stage). We give two analyses of such systems. The first is for when targets and false alarms are treated identically by the follow-on resource; the second is for when targets require a different amount of resource than false alarms. Another way of dealing with the stochastic nature of the output rate is to rank the returns, and choose only the largest ones to pass on (the ratio of which cells are passed on to the total number of cells tested being equal to the processing ratio between the stage and the subsequent resource.) If a test always yields the same number of hits, regardless of the input, does the test yield any information? Yes, if a time history can be given of the test output. If a given cell is consistently the largest among its peers, it is more likely to be a target [17]. Chapter six includes a thorough analysis of binary integrators or "M/N" detectors. The chapter concludes with an involved example of a resource limited network of trackers, each using an M/N rule for track initiation.

Conclusions are given in chapter seven.

Throughout the thesis the emphasis is on analysis and the interpretation of results. In fact, no simulation results are presented. Graphical explanations are used, whenever possible, to supplement the analysis and lend insight. Hundreds of pages of graphs could

have been included; instead, we have tried to choose a representative sample to illuminate the pertinent points.

1.4 Conclusions

Resource constraint theory recognizes the fact that detections are decisions - decisions that have consequences for subsequent processing. The resulting criterion openly admits and addresses the ignorance that we must deal with when designing practical detection systems. By bringing into the open the implicit constraints and assumptions behind detection systems, the theory allows us to better understand, evaluate and ultimately design them. Optimal or adequate performance is achievable by accounting for the abilities of the systems that are dependent on the detection decisions. The theory is applicable at several system levels: overall control (in the obstacle-avoidance scenario of chapter three), data processing (in the tracking processor of chapter six), and signal processing (in the noncoherent/coherent processing system of chapter four.)

The rate-constraint is well-defined, since it is a function of the processing rates, which are design parameters. Furthermore, the hit-rate is both controllable and observable. The design of a system operating under resource-constraint principles requires no arbitrary cost functions, or a priori statistics. Unlike non-parametric methods, a design value for the false alarm rate is not required.

The criterion achieves optimal detection probability for a single-stage controlling a limited resource. The SLOC parameter is a useful design tool which naturally arises out of multi-stage rate-constraint theory. It gives guidance on how best to combine a chain of tests for optimal performance. There is no need for the tests to be synchronous, or identically distributed. The resulting cueing model is prevalent in both natural and man-made designs.

It is interesting to note that in the latest editions two of the most respected books in radar, the sections dealing with false alarm control actually use a hit-rate based argument and do not mention the Neyman-Pearson lemma [6],[18]. There is clearly a need for a new framework for detection theory that addresses the lack of a priori information common in most radar scenarios.

One warning before commencing: some of the concepts of this thesis may appear, at least at first glance, to be self-evident or even trivial. Beware! There are a few surprises

left in detection theory. To quote a recent paper studying counterexamples in distributed detection: "The obvious conclusion is that these are treacherous waters, and something may hold very often, without being true. It may even be supported by a plausible argument, but remain false" [19].

Chapter 2

Classical Detection Theory

2.1 Introduction

In this chapter, classical detection theory is briefly reviewed, and some of the inherent difficulties of using it for practical radar problems are discussed. The material is well documented in many detection texts (for example [20], [21]); our goal here is to introduce notation and to review concepts that are important later in the thesis.

As noted in chapter 1, Middleton in 1960 listed three problems in the application of statistical detection theory to practical sensor deployments: (1) the apparent arbitrariness of the cost assignments; (2) the usual inadequacy of the a priori information; and (3) the selection of the criterion of optimality itself [5]. In reviewing the classical theory we see that the same problems enumerated in 1960 continue to plague practical detection systems today.

2.1.1 Chapter Outline

In section 2.2 the fundamental concepts of binary hypothesis testing as applied to detection theory are discussed. From a discussion of the Neyman-Pearson lemma, and Birdsall's insight, it is shown that optimal tests under a number of criteria have the same form: a likelihood ratio test with a possibly different threshold. Various classical criteria and their choices for the best threshold are discussed in section 2.3. However, it is also seen that the a priori information required to determine the thresholds is not typically available in the radar problem. Section 2.4 discusses some of the problems associated with the practical

implementation of classical detection theory. A couple of classical radar detection problems are investigated in section 2.5, and finally section 2.6 reviews the subject of the chapter. The appendix to the chapter (section 2.7) presents a proof of the Neyman-Pearson lemma.

2.2   The model

We begin by assuming that there exist two mutually exclusive and exhaustive states of nature that we wish to distinguish between. Associated with the two states are two hypotheses: H_0, the *null* (or noise alone) hypothesis and H_1, the *alternate* (or target plus noise) hypothesis.

In order to distinguish between H_0 and H_1, we are given an observation x, of a random variable $X \in \Gamma$, or in general a vector of observations, $\overline{x} \in \Gamma^N$, where Γ^N is the Borel set (or power set, depending on whether x is a continuous or discrete variable) containing values for the random variable $\overline{X}$. Usually $\overline{x}$ is reduced to a single number x called a *test statistic*. If no detection performance is lost in reducing $\overline{x}$ to x, then x is called a *sufficient statistic*; however, non-sufficient statistics are often used in radar applications. Sufficient statistics are further discussed in section 2.4.3; until then we suppress the vector notation for x.

Assume that X is a random variable drawn from a distribution parameterized by a set Θ, where $\Theta \in \Lambda = \Lambda_0 \cup \Lambda_1$ and

$$\Lambda_i = \{\Theta : H_i\} \qquad\qquad \text{i=0,1} \qquad\qquad (2.1)$$

For the radar problem, usually $\Lambda_0 = \{0\}$ and we say the null hypothesis is *simple*. In general, Λ_1 contains more than one point and we say the alternate hypothesis is *composite*. For now we consider only simple hypotheses, returning to the problem of composite alternate hypotheses in section 2.4.

Let the family of probability distribution functions generating X be denoted by $F_{\Lambda_i}(x)$, and their associated densities (or mass functions) by $f_{\Lambda_i}(x) = \frac{dF_{\Lambda_i}(x)}{d\mu(x)}$ where $\mu(x)$ is the measure of X (Lebesgue or counting depending on whether X is a continuous or discrete random variable[1]). By defining the density functions in term of Radon-Nikodym derivatives (and subsequent integrals by Lebesgue-Stieljes integrals), continuous and discrete random variables can be treated with the same formulation [20].

---

[1] For the radar examples of this thesis, X will always be a continuous random variable, and the Lebesgue integral may be interpreted as an ordinary Riemann integral. However, we use the general notation because X may be a discrete random variable. For example, in lidar systems, X is often a photon count.

2.2.1 Decision Functions

Given the observation $x \in \Gamma$, the goal is to partition the observation space into two regions $\Gamma = \Gamma_0 \cup \Gamma_1$ where Γ_0 is called the acceptance region and Γ_1 the rejection region. If $x \in \Gamma_0$, the null hypothesis H_0 is accepted, and if $x \in \Gamma_1$, the null hypothesis H_0 is rejected. A decision function, $\delta(x) : X \rightarrow [0,1]$ is defined as

$$\delta(x) = \begin{cases} 1 & \text{if } x \in \Gamma_1 \\ 0 & \text{if } x \in \Gamma_0 \end{cases} \tag{2.2}$$

Equivalently, $\delta(x)$ is the probability of rejecting H_0 given the observation x, i.e.

$$\delta(x) = P(\text{rejecting } H_0 | X = x)$$

It may seem odd to express $\delta(x)$ as a probability when it only takes on values of 0 or 1; however, a third value for $\delta(x)$, say $0 < \gamma < 1$, is often assigned to the boundary of Γ_0 and Γ_1 if x is a discrete variable, in order to allow arbitrary values of false alarm probability. Decision functions that have values of $\delta(x)$ that are not zero or one are called *randomized* rules; we will use them in the Neyman-Pearson and Rate Constraint Lemmas. If X is a continuous variable, the assignment of probability to any point of Γ, in particular the boundary of Γ_0 and Γ_1, is superfluous since the point is of measure zero.

2.2.2 Errors

Given the decision function, $\delta(x)$, there are two types of errors one can make. The first is called a type I error, or false alarm; the probability of a false alarm for a test with a simple null hypothesis is defined by

$$\alpha = pf = E\{\delta(x) | H_0\} = \int_\Gamma f_{\Lambda_0}(x)\delta(x)\, d\mu(x) = \int_{\Gamma_1} f_{\Lambda_0}(x)\, d\mu(x) \tag{2.3}$$

where $E\{\bullet | H_0\}$ is the expected value of $\{\bullet\}$ given H_0. The false alarm probability, $pf = \alpha$ is also called the size of the test $\delta(x)$. Note that (2.3) is valid for both discrete and continuous x provided $\mu(x)$ is the appropriate (i.e. counting or Lebesgue) measure. If H_0 is composite, pf is defined by

$$pf = \sup_{\Theta \in \Lambda_0} E\{\delta(x)\} \tag{2.4}$$

In this thesis we assume that H_0 is simple since this is almost always the case for radar detection problems.

The other type of error we can make is called a Type II error or a miss, with a probability given by

$$\beta = 1 - pd \tag{2.5}$$

where pd is the detection probability or power of the test, given by

$$pd = E\{\delta(x)|H_1\} = \int_\Gamma f_{\Lambda_1}(x)\delta(x)\,d\mu(x) = \int_{\Gamma_1} f_{\Lambda_1}(x)\,d\mu(x) \tag{2.6}$$

To simplify the notation for now, since we are dealing only with simple hypotheses, we will suppress the Λ and write $f_{\Lambda_i}(x) = f_i(x)$.

2.2.3 Likelihood Ratio Tests: The Neyman-Pearson Lemma

There is a trade-off between the values of α and β; we can make either arbitrarily close to zero while raising the other to one (by setting $\Gamma_0 = \Gamma$ or $\Gamma_1 = \Gamma$). The sum of the error probabilities is given by

$$\alpha + \beta = \int_\Gamma [\delta(x)f_0(x) + (1 - \delta(x))f_1(x)]\,d\mu(x) \tag{2.7}$$

For each x, the integrand is a weighted average of $f_0(x)$ and $f_1(x)$, so therefore

$$\min\{f_0(x), f_1(x)\} \le \delta(x)f_0(x) + (1 - \delta(x))f_1(x) \le \max\{f_0(x), f_1(x)\} \tag{2.8}$$

Hence

$$\alpha + \beta \ge \int_\Gamma \min\{f_0(x), f_1(x)\}\,d\mu(x) \tag{2.9}$$

Two natural questions to ask are what is the optimal form of $\delta(x)$ and how does changing α affect β? The Neyman-Pearson lemma [1] answers both by minimizing β for a fixed value of α. Consider

$$\lambda\alpha + \beta \ge \int_\Gamma \min(\lambda f_0(x), f_1(x))\,d\mu(x) \tag{2.10}$$

where λ is a Lagrange multiplier. The equality holds (and therefore β is minimized, or equivalently pd is maximized) when $\delta(x)$ is a likelihood ratio test of the form

$$\delta(x) = \begin{cases} 1 & \text{if } l(x) > \lambda \\ \gamma(x) & \text{if } l(x) = \lambda \\ 0 & \text{if } l(x) < \lambda \end{cases} \tag{2.11}$$

where $l(x) = \frac{f_1(x)}{f_0(x)} \geq 0$ is called the likelihood ratio and $\lambda \geq 0$ is a threshold. As mentioned above, $\gamma(x)$ is a randomizing function [20] that may be ignored when the support of Γ is continuous since the probability that $l(x) = \lambda$ is zero.

In the appendix at the end of the chapter, the following lemma is proven. The proof is based on that found in [20]; we include it for comparison with the proof of the Rate Constraint lemma that is presented in chapter 3.

Lemma 1 (Neyman-Pearson) *For a hypothesis test between H_0 and H_1 (both simple) we have the following.*

1. *Optimality: Let δ be any decision rule satisfying $pf(\delta) \leq \alpha$ and let δ' be any decision rule of the form*

$$\delta'(x) = \begin{cases} 1 & \text{if } \ell(x) > \lambda \\ \gamma(x) & \text{if } \ell(x) = \lambda \\ 0 & \text{if } \ell(x) < \lambda \end{cases} \tag{2.12}$$

where $\lambda \geq 0$ and $0 \leq \gamma(x) \leq 1$ are such that $pf(\delta') = \alpha$. Then $pd(\delta') \geq pd(\delta)$.

2. *Existence: For every $\alpha \in (0,1)$, there is a decision rule δ_{NP} of the form of (2.12) with $\gamma(x) = \gamma_0$, a constant, for which*

$$pf(\delta_{NP}) = \alpha.$$

3. *Uniqueness: Let δ'' be any $\alpha-$level decision rule for H_0 versus H_1. Then δ'' must be of form (2.12) except possibly on a set of measure zero under H_0 and H_1.*

The Neyman-Pearson lemma says that the probability of detection for a simple hypothesis test is maximized for any given probability of false alarm when a likelihood ratio test is used. A unique test of the form (2.12) can always be found that will yield any desired value of pf.

2.2.4 Receiver Operating Characteristic Curves and Their Properties

Armed with the Neyman-Pearson lemma, we can re-write the probability of false alarm from (2.3) as

$$pf = \int_\lambda^\infty f_0(\ell) \, d\ell \tag{2.13}$$
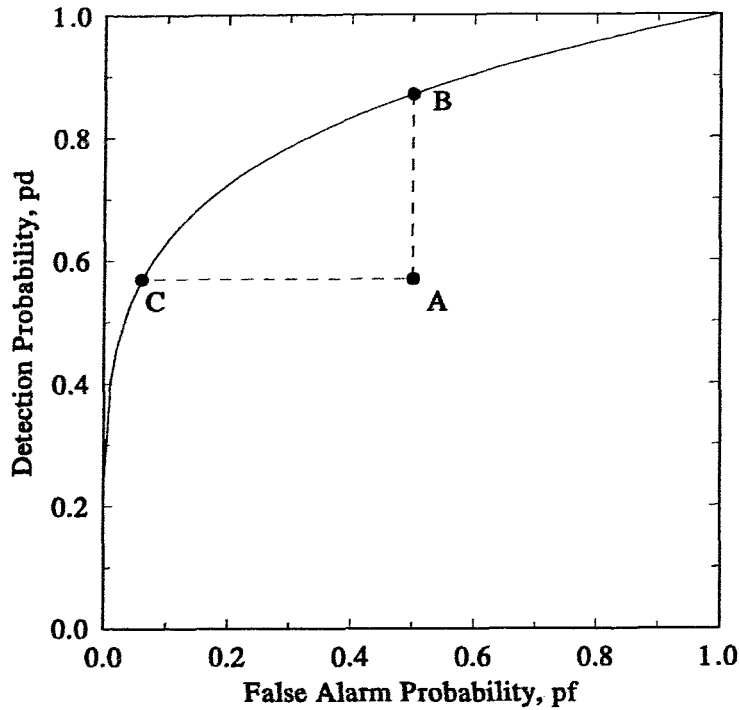
Figure 2.1: Receiver Operating Characteristic (ROC) Curve.

and the probability of detection (2.6) as

$$pd = \int_{\lambda}^{\infty} f_1(\ell)\, d\ell \qquad (2.14)$$

where $f_i(\ell)$ is the probability density function of the likelihood ratio given H_i is true. It is not always convenient to calculate pf and pd through (2.13) and (2.14) since the density of the likelihood ratio under the two hypotheses is not simple to determine. Fortunately, as explained in section 2.4.2, the detection and false alarm probabilities can be often expressed in terms of $f_0(x)$ and $f_1(x)$.

As presented in (2.13) and (2.14), pd is a function of pf through the parameter (threshold) λ. Figure 2.1 is a typical plot of pd versus pf, called a Receiver Operating Characteristic (ROC) curve. There are a number of points to note regarding the ROC curve.

First, the curve must lie above the chance line $pd = pf$ at all points. Consider what

$pd < pf$ implies; we could choose a new $\delta'(x) = 1 - \delta(x)$ which lies above the chance line and therefore is optimal. But by the Neyman-Pearson criterion, $\delta(x)$ is optimal; therefore the ROC curve lies above the chance line. The situation $pf = pd$ also makes no sense; in that case the test gives no information since we could achieve the same performance by flipping an unbiased coin.

Next consider the slope of the ROC curve.

$$\frac{\partial pd}{\partial pf} = \frac{\partial pd}{\partial \mu(x)} \left(\frac{\partial pf}{\partial \mu(x)} \right)^{-1} = \frac{f_1(\lambda)}{f_0(\lambda)} = \lambda \tag{2.15}$$

Thus, the slope of the ROC curve at any point is the threshold, λ, required to operate at that point. This relationship is not surprising because in the Lagrange Multiplier formulation of (2.10), λ is the sensitivity coefficient and gives the effect on the objective function, pd, from changing the constraint, pf [22].

As the threshold, λ, is increased from 0 to ∞, both pd and pf fall from 1 to 0, while the slope $\frac{dpd}{dpf}$ must also increase from 0 to ∞. This implies that the ROC curve must be concave as shown in the figure. The ROC curve is of interest not only to radar detection experts [20], but also to Psychophysicists [12],[23].

Birdsall's North by Northwest Insight   We have seen that the likelihood ratio test is optimal under the Neyman-Pearson criterion. An insight credited to Birdsall by Scharf [24] is that for *any sensible detection criterion*, the operating point must lie on the ROC curve, and therefore must be a likelihood ratio test. To be sensible, a detection criterion should emphasize high pd and low pf.

To understand Birdsall's insight, consider point A in figure 2.1. From point A, we can increase pd without increasing pf by moving up to point B; we can also decrease pf without decreasing pd by moving to point C. Any point along the ROC curve between B and C will have better performance (i.e. higher pd and lower pf) than point A.

Birdsall's insight is a very useful tool; it tells us that any sensible detection criterion must lead to a likelihood ratio test. Given the ROC curve for the detection scenario, we know the optimal performance will be somewhere on the ROC curve. Different detection criteria choose the exact optimal operating point by trading off increases in pf with those in pd. We now consider a couple of different detection criteria.

2.3 Cost Function Approaches

It is important to recognize that detection systems make decisions - decisions that will have consequences. It is senseless to think of a detection system working in isolation - if no action results from the decisions, then why does the system exist in the first place[2]?

The cost function method of detection theory seeks to associate a cost structure with the consequences of false alarms and misses, and then to find the decision rule that minimizes the expected cost. Let $\mathcal{A}$ be the set of all possible actions associated with the partitioning of Γ. Then the cost function is $\mathcal{C} : \mathcal{A}x\Theta \rightarrow [0, \infty]$. In this thesis, the action set is always binary, $\mathcal{A} = \{0, 1\}$, and for now we assume that both the null and alternate hypotheses are simple, so that $\Theta = \{0, 1\}$. Then $\mathcal{C}$ can be expressed as C_{ij} where C_{ij} is the cost incurred by deciding that H_i is true when in fact H_j is.

2.3.1 Bayes

In the Bayesian framework, it is assumed that the designer is able to subjectively assign a fixed cost to decisions. The Bayesian criterion seeks to minimize the expected cost; therefore a priori probabilities of the occurrence of the two hypothesis occurring are required. Let $\pi_j = P(\Theta \in \Lambda_j) = P(H_j)$. (Note that $\pi_0 + \pi_1 = 1$.) Then the Bayes risk of the decision rule, $\mathcal{R}(\delta)$, is given by

$$\mathcal{R}(\delta) = \sum_{j=0}^{1} \pi_j \sum_{i=0}^{1} C_{ij} P_j(\Gamma_i) \qquad (2.16)$$

where $P_j(\Gamma_i)$ is the probability th t $X \in \Gamma_i$ given that H_j is true. If we assume that $f_j(x) = \frac{dF_j(x)}{d\mu(x)}$ is the density of X and recognize that $P_j(\Gamma_0) = 1 - P_j(\Gamma_1)$, we can write

$$\mathcal{R}(\delta) = \sum_{j=0}^{1} \pi_j C_{0j} + \int_{\Gamma_1} \left[\sum_{j=0}^{1} \pi_j (C_{1j} - C_{0j}) f_j(x) \right] d\mu(x) \qquad (2.17)$$

Since the first term in (2.17) is independent of δ, $\mathcal{R}(\delta)$ is minimized by choosing

$$\Gamma_1 = \left\{ x \in \Gamma : \sum_{j=0}^{1} \pi_j (C_{1j} - C_{0j}) f_j(x) \leq 0 \right\} \qquad (2.18)$$

It is straightforward to show that (2.18) is equivalent to choosing [20]

$$\Gamma_1 = \{ x \in \Gamma : \ell(x) \geq \lambda \} \qquad (2.19)$$

---

[2]To quote James 2:17, "Faith without deeds is useless".
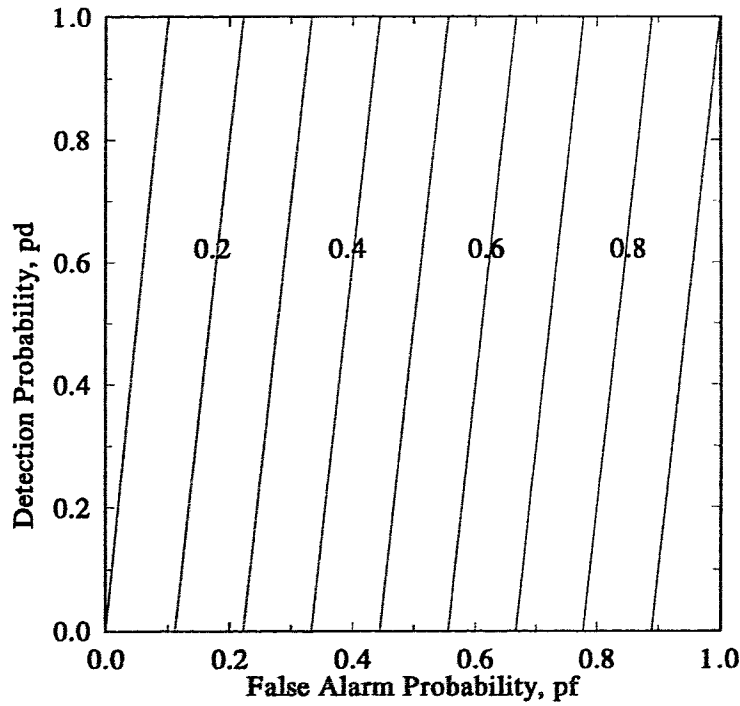
Figure 2.2: Iso-Cost Curves on ROC axes for Bayesian Test with $C_{01} = C_{10} = 1$, $C_{00} = C_{11} = 0$, and $\pi_1 = 0.1$. The numbers on the curves are the cost incurred by a test with the pf and pd at that point.

where $\ell(x)$ is the likelihood ratio and

$$\lambda = \frac{\pi_0(C_{10} - C_{00})}{\pi_1(C_{01} - C_{11})} \tag{2.20}$$

is the threshold.

As expected from Birdsall's insight, the optimal Bayes test is a likelihood ratio test. If the cost of a correct decision is zero (i.e. $C_{ii} = 0$), and the cost of each type of error is equal (i.e. $C_{01} = C_{10}$), the total a posteriori error is minimized (and the total probability of a correct decision is maximized). Such a cost assignment is useful in communications systems and is employed in a Maximum A Posteriori (M.A.P.) receiver [25]. In the radar scenario, however, the cost of a false alarm is expected to be much smaller than the cost of a miss.

In figure 2.2, the contour lines of constant Bayes risk as given by (2.16) are plotted. As expected, the contour lines are linear, with slope given by (2.20). For the figure, the M.A.P. cost assignment was used, along with a value of $\pi_1 = 0.1$; therefore the lines have slope = 10. Since $\pi_0 > \pi_1$ and $C_{10} = C_{01}$, such a cost assignment emphasizes the cost of false alarms more than misses. In order to graphically determine the optimal operating point for a given $f_0(x)$ and $f_1(x)$, the ROC curve as illustrated in figure 2.1 must be overlayed on figure 2.2; the operating point lies where the ROC curve is tangent to one of the contour lines.

While Bayes rules are useful for making decisions when priors and cost functions can be supplied, neither are available in a typical radar scenario. Information theory provides a less subjective cost structure, while the mini-max formulation eliminates the requirement for priors.

2.3.2 Maximum Information (or Minimum Equivocation)

One way to eliminate the subjectivity of Bayesian cost functions is to invoke Information Theoretic concepts. The mutual information expressed in the decision from a binary test is given by [5]

$$\mathcal{I} = \sum_{i=0}^{1} \pi_i \sum_{j=0}^{1} P_j(\Gamma_i) \log\left(\frac{P_j(\Gamma_i)}{\sum_{k=0}^{1} P_k(\Gamma_i)}\right) \tag{2.21}$$

This is equivalent to a Bayesian cost function with

$$C_{ij} = \log\left(\frac{P_j(\Gamma_i)}{\sum_{k=0}^{1} P_k(\Gamma_i)}\right) \tag{2.22}$$

Now let

$$\xi_i = \sum_{k=0}^{1} P_k(\Gamma_i)$$

be the a posteriori probability of H_i being decided, then we can re-write (2.21) as

$$\mathcal{I} = \sum_{i=0}^{1} \pi_i \sum_{j=0}^{1} P_j(\Gamma_i)[\log(P_j(\Gamma_i)) - \log \xi_i] \tag{2.23}$$

In order to find the optimal threshold, we must express (2.23) as a function of λ, which is accomplished by taking the derivative of $\mathcal{I}$ with respect to $pf = P_0(\Gamma_1)$ and using (2.15),

$$\frac{\partial \mathcal{I}}{\partial pf} = -\pi_0 \log\left(\frac{1-pf}{pf}\right) + [\pi_0 + \lambda\pi_1]\log \xi_0 - \lambda\pi_1 \log\left(\frac{1-pd}{pd}\right) - [\pi_0 + \lambda\pi_1]\log \xi_1 \tag{2.24}$$
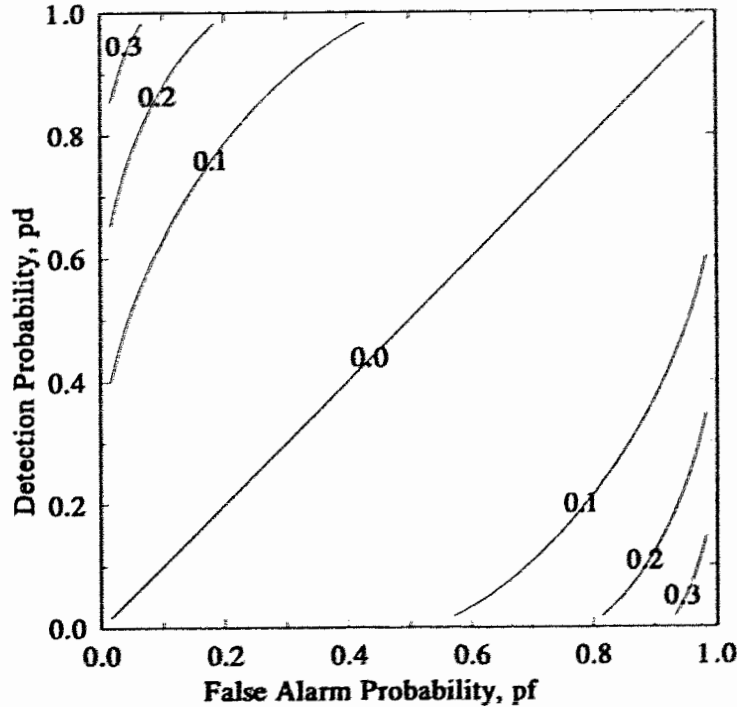
Figure 2.3: Constant mutual information contours on ROC axes for $\pi_1 = 0.1$. The numbers on the lines are the value for the mutual information.

To maximize $\mathcal{I}$, we set (2.24) = 0 and solve for λ, yielding [26]

$$\lambda = \frac{-\pi_0[\log(\xi_0/\xi_1) - \log((1 - pf)/pf)}{\pi_1[\log(\xi_0/\xi_1) - \log((1 - pd)/pd)} \qquad (2.25)$$

In figure 2.3 the contour lines of constant mutual information are plotted for $\pi_1 = 0.1$. The contours are curvilinear, and symmetric: $\mathcal{I}(pf, pd) = \mathcal{I}(1 - pf, 1 - pd)$. However, since we are interested only in the points above the chance line $pf = pd$, the symmetry is of no practical consequence.

Information theory has been extremely successful for its intended application, communications systems. While attempts have been made to incorporate information theoretic concepts into detection theory, the dependence of the threshold on the π_i values precludes its use since the priors are almost always unknown in practical radar scenarios. Books from

as early as 1953 [4] were attempting to utilize information theory in the radar context. Attempts have also been made to describe clinical diagnostic tests in terms of the mutual information function [27]. Despite the interest and attempts of many, it is still not clear how the cost functions of (2.22) relate to any real-world performance parameters in a radar system. Kanaya and Nakagawa [28] recently wrote an interesting paper investigating the connection between the mutual information function and Bayesian decision making. Unfortunately the relationship is through an asymptotic detection performance figure which is valid only as the number of samples goes to ∞. Other information theoretic measures such as the discrimination or divergence [29] have also been applied to detection theory; however, they also give asymptotic relationships that are not relevant in radar where we don't have a large number of independent identically distributed looks at a given cell.

2.3.3 Mini-Max

The mini-max criterion seeks to find

$$\min_{\delta} \max_{0 \le \pi_1 \le 1} \mathcal{R}(\delta) \qquad (2.26)$$

It can be shown [20] that the mini-max criterion also leads to a likelihood-ratio test; however, the test is seldom if ever useful in radar contexts, because of the difficulty in assigning meaningful cost functions. The criterion is pessimistic, and may lead to penalties in performance by basing the threshold on priors π_i that seldom occur.

2.3.4 The Neyman-Pearson Approach to Detection

We have seen that the cost function approaches require more information than is typically available in the radar problem. Given the problems of cost function approaches, the radar community has traditionally adopted to use the Neyman-Pearson philosophy.

For radar, the consequences of a miss are typically much higher than those of a false alarm. False alarms are nuisances that lead to waste of resources but misses are system failures that can cause loss of life. The Neyman-Pearson criterion asks, "How many false alarms can we afford to allow?" Then the appropriate threshold is chosen to allow that number of false alarms. and the resulting detection probability is hoped to be acceptable.

The Neyman-Pearson framework makes sense, and systems supposedly working under its philosophy have worked successfully for decades. In fact, Skolnik [30] wrote in 1962,

"The Neyman-Pearson criterion is well suited for radar application and is often used in practice, whether knowingly or not." However, a closer look reveals that it is the *hit-rate* (i.e. threshold-crossing rate), not the false alarm rate, that is actually the parameter of interest. For example, in the latest edition of two influential books in radar [6], [18], the sections discussing false alarm control actually discuss hit-rates.

Why is the hit-rate more useful? By definition, the false alarm rate is unobservable; if we knew a threshold crossing was a false alarm, we wouldn't call it a target in the first place. Furthermore, the choice of a design false alarm rate is purely subjective. As explained in the gun-control and TWS examples of section 1.1, the optimal value depends on the radar context - the number of targets etc. In chapter 3, we will introduce a new detection criterion that is based on the hit-rate rather than the false alarm rate.

We have seen that an optimal test (in the sense of Neyman-Pearson, Bayes, Maximum Information or Minimax) must be a likelihood ratio test. In general, a likelihood ratio test may be difficult to synthesize. In the following section we examine some of the ways that radar systems practically approximate likelihood ratio tests, and deal with composite alternate hypotheses.

2.4   Practical Tests

2.4.1   Composite Alternate Hypotheses: Uniformly Most Powerful Tests

Thus far, we have assumed that both the null and alternate hypotheses are simple. We now consider the most common scenario in radar detection problems: a simple null and composite alternate hypothesis.

Recall that a composite alternate hypothesis means that under H_1, the observation X has a density $f_\Theta(x)$ where $\Theta \in \Lambda_1$ has more than one point. Perhaps it is obvious that the simplest thing to do is to define a density for Θ and average over $f_\Theta(x)$:

$$f_1(x) = E_{\Theta \in \Lambda_1}\{f_\Theta(x)\} = \int_{\theta \in \Lambda_1} f(\theta)f_\theta(x)\,d\theta \qquad (2.27)$$

The detection probability for a composite alternate hypothesis is then defined as

$$1 - \beta(\Theta) = E_{\Theta \in \Lambda_1}\{\delta(x)\} = \int_{x \in \Gamma_1} \int_{\theta \in \Lambda_1} f(\theta)f_\theta(x)\,d\theta\,dx \qquad (2.28)$$

Similarly the likelihood ratio (sometimes called a *generalized likelihood ratio* [31]) is expressed as

$$\ell(x) = \frac{E_{\theta \in \Lambda_1}\{f_\theta(x)\}}{f_0(x)} \tag{2.29}$$

We would like to design a single test in the spirit of the Neyman-Pearson lemma for every pair $(\theta \in \Lambda_0, \theta \in \Lambda_1)$. A test $\delta(x)$ for testing H_0 versus H_1 is called *Uniformly Most Powerful (UMP)* if, for all $(\theta \in \Lambda_0, \theta \in \Lambda_1)$ it has power (detection probability) greater than any other test with the same size (false alarm probability).

UMP tests are highly desirable, since they guarantee optimal performance over the entire range of Θ. In the following theorem, we will consider scalar random variables X with density functions $f_{\theta_i}(x)$ parameterized by scalar parameters. While this appears to be restrictive, most radar detection problems do indeed reduce to the comparison of a single scalar that has a density parameterized by a scalar parameter. The Karlin-Rubin theorem, which we now consider, is extremely important for such problems, since it allows seemingly complicated likelihood ratio tests to be made by simple threshold tests [24].

2.4.2 Threshold Tests: The Karlin-Rubin Theorem

Assume that X is a scalar random variable whose density function is a function of the scalar parameter θ. Further, assume that the likelihood ratio function

$$\ell(x) = \frac{f_{\theta_1}(x)}{f_{\theta_0}(x)}$$

is a nondecreasing function of x for all (θ_0, θ_1) where $\theta_0 < \theta_1$. Then $\ell(X)$ is a monotonically increasing function of X (we say X has a *monotone-likelihood ratio*). A monotone-likelihood ratio function means that H_1 is more likely as X increases.

Given a monotonic likelihood function, the likelihood ratio test of (2.12) is equivalent to the threshold test

$$\delta(x) = \begin{cases} 1 & \text{if } x > t \\ \gamma & \text{if } x = t \\ 0 & \text{if } x < t \end{cases} \tag{2.30}$$

The proof is straightforward: Assume that H_0 is simple, and let θ_0 be fixed (note that this is not necessary for the proof, but we are considering only simple null hypotheses anyway). Begin with a fixed $\theta_1 > \theta_0$. By the Neyman-Pearson lemma, the test of (2.12) is

optimal for the simple $H_0 : \theta = \theta_0$ versus the simple alternative $H_1 : \theta = \theta_1$. Since $\ell(x)$ is monotone, we can replace the likelihood ratio test of (2.12) with the threshold test of (2.30) with

$$E\{\delta(X|H_0)\} = P_0(X > t) + \gamma P_0(X = t) = \alpha$$

Since the test is independent of θ_1, it is uniformly most powerful among all tests with false alarm probability $\leq \alpha$. $\square$[3]

The Karlin-Rubin Theorem is very powerful since it means that very simple tests can be used and still be guaranteed to be UMP. Threshold tests are almost always used in radar installations and it is important to know how the resulting performance compares with the optimal.

The Karlin-Rubin Theorem also simplifies calculation of the detection and false alarm probabilities. From (2.30), we see that the decision regions can now be expressed in terms of x instead of $\ell(x)$. Specifically

$$\Gamma_1 = \{x \in \Gamma : \ell(x) > \lambda\} = \{x \in \Gamma : x > t\}$$

and

$$\Gamma_0 = \{x \in \Gamma : \ell(x) < \lambda\} = \{x \in \Gamma : x < t\}$$

Then the detection and false alarm probabilities can be expressed in terms of the densities of X:

$$pd = \int_t^\infty f_1(x)\,dx \tag{2.31}$$

and

$$pf = \int_t^\infty f_0(x)\,dx \tag{2.32}$$

Furthermore if X has a monotone likelihood ratio, (2.31) is still valid for composite alternate hypothesis if $f_1(x)$ is defined as in (2.27).

The Karlin-Rubin Theorem allows considerable simplification when the observation has a monotonic likelihood ratio function. The question to ask is what densities produce monotonic likelihood ratios?

---

[3]The end of proofs is indicated by the $\square$.

One-Parameter Exponential Family Consider a random vector $\overline{X}$ drawn from a family of distributions of the form

$$f_\theta(\overline{x}) = C(\theta)h(\overline{x})\exp[Q(\theta)T(\overline{x})] \tag{2.33}$$

(Such a family is called a one-parameter exponential family). If Q is a nondecreasing function of θ, then $\overline{X}$ has a monotonic likelihood ratio function. A random sample $\overline{X}$ drawn from this distribution will produce a scalar statistic, $T(\overline{x})$ with a one-parameter distribution and monotonic likelihood ratio function. $T(\overline{X})$ is called a *sufficient statistic* for $\overline{X}$, which by the Karlin-Rubin theorem can be optimally tested by a simple threshold test.

It is easy to show that binomial, Poisson, one-parameter Normal (i.e. either the mean or variance known), and one-parameter Gamma distributions all generate monotonic likelihood ratio functions.

We have seen how composite hypothesis tests can be handled, and how monotonic likelihood ratio tests yield very simple forms for the UMP test. In the next section we discuss the problem of forming the test statistic, $T(\overline{x})$ from a vector of observations.

2.4.3 Multiple Observations: Sufficient Statistics

In the previous section we saw how that a vector of observations, $\overline{X}$, could be reduced to a sufficient statistic $T(\overline{X})$ when $\overline{X}$ is drawn from a distribution from a one-parameter exponential family. In most radar installations, however, a simpler (non-optimal) function and not $T(\overline{x})$ is calculated and used as a test statistic.

For example, in section 2.5.1 we derive the likelihood ratio for a common radar target model: a constant amplitude random phase target in Gaussian noise. It is shown that the likelihood ratio function is

$$\ell(z) = \exp\left[-\tfrac{1}{2}A^2\right]I_0(Az) \tag{2.34}$$

where z is the amplitude of the return (target plus noise), A is the amplitude of the target return alone and $I_0(z)$ is the zeroth-order modified Bessel function of the first type.

Now consider the problem of making a detection decision based on N returns. Assuming that the returns are statistically independent, the overall likelihood ratio is the product of the individual likelihood ratios:

$$\ell(\overline{z}) = \prod_{i=1}^{N}\exp\left[-\tfrac{1}{2}A^2\right]I_0(Az_i) \tag{2.35}$$

where z_i is the amplitude of the ith observation. Since A and N are known[4], the likelihood ratio test is equivalent to

$$\delta(\bar{z}) = \begin{cases} 1 & \text{if } \prod_{i=1}^{N} I_0(Az_i) > t' \\ \gamma(z) & \text{if } \prod_{i=1}^{N} I_0(Az_i) = t' \\ 0 & \text{if } \prod_{i=1}^{N} I_0(Az_i) < t' \end{cases} \qquad (2.36)$$

where $t' = t\exp(\frac{1}{2}NA^2)$ and t is the original threshold. The logarithm being a monotonic function, we can re-write (2.36) as

$$\delta(\bar{z}) = \begin{cases} 1 & \text{if } \sum_{i=1}^{N} \ln I_0(Az_i) > t'' \\ \gamma(x) & \text{if } \sum_{i=1}^{N} \ln I_0(Az_i) = t'' \\ 0 & \text{if } \prod_{i=1}^{N} \ln I_0(Az_i) < t'' \end{cases} \qquad (2.37)$$

where $t'' = \ln t'$.

Unfortunately, it is difficult to implement the $\ln I_0(\bullet)$ function at radar processing speeds. Therefore the function is approximated with either a square-law or a linear detector. Clearly there is some loss incurred in using a sub-optimum combining scheme which doesn't use a sufficient statistic[5]. For linear and quadratic detectors, the loss is typically on the order of a few dB or less for common radar targets in Gaussian noise [16].

For the purposes of this thesis where we are concerned with the practical consequences of not knowing the distributions, only one point must be made regarding the use of non-sufficient statistics. In the next section and in chapter 5, we will generate ROC curves for a number of different radar scenarios. In all cases the calculations are for threshold tests applied to the test statistic derived from a quadratic combining rule. Given the preceding discussion, we are not guaranteed that threshold tests are justified since we are not forming the likelihood ratio for our statistic.

However, it will turn out that the ROC curves for all tests considered in this thesis are concave (such ROC curves are called *proper* [23]). It is easy to show that a concave ROC curve implies a monotonic likelihood ratio test for the underlying distributions. Consider

---

[4]For this thesis, we assume that N is a given constant. Wald [32] has developed a theory called Sequential Detection that requires a variable number of returns. Unfortunately, in addition to cost functions and priors, Wald's test requires that an acceptable value for both the detection probability and false alarm probability be chosen. We will not consider Sequential Detection further.

[5]Note that even if the $\ln I_0(\bullet)$ function could be evaluated at radar speeds, it would only be the form of the optimal combination rule for a constant amplitude target in Gaussian noise, and not for any other target and noise models.

expanding the ROC curve in a Taylor series for the likelihood ratio function about a point $\ell(pf_0)$. We can write

$$\ell(pf_0 + \Delta pf) \approx \ell(pf_0) + \frac{\partial \ell}{\partial pf}\Delta pf \qquad (2.38)$$

but (2.15) shows that the slope of the ROC curve is equal to the likelihood ratio function evaluated at that point:

$$\ell = \frac{\partial pd}{\partial pf} \qquad (2.39)$$

Therefore

$$\ell(pf_0 + \Delta pf) \approx \ell(pf_0) + \frac{\partial^2 pd}{\partial pf^2}\Delta pf \qquad (2.40)$$

By definition of concavity $\frac{\partial^2 pd}{\partial pf^2} \leq 0$, hence

$$\ell(pf_0 + \Delta pf) \leq \ell(pf_0) \qquad (2.41)$$

ℓ is therefore a decreasing function of pf and pd, or equivalently an increasing function of X. Thus a concave ROC curve is sufficient to guarantee a monotonic likelihood ratio function (and to justify use of a threshold test, from the Karlin-Rubin theorem). This equivalence is important for some examples in chapter five, where we are unable to calculate the densities for some of the models, but are able to calculate the detection and false alarm probabilities. (The seeming anomaly of being able to determine the integral of a function without being able to calculate the function itself is a result of the calculation technique used.)

Of course, it is impossible to rigorously prove the concavity of a ROC curve by using numerical techniques which can calculate only a limited number of points on the curve, and to finite precision. However, such numerical techniques do generate some evidence to support use of a threshold test.

Once the detector of a radar system is fixed, so are the densities of the return under the two hypotheses - we may not know what they are, but provided they produce a concave ROC curve the equivalence of a threshold test to an optimum likelihood ratio test is guaranteed. Furthermore, if a scenario is envisioned where the test statistic does not generate a concave ROC curve, there is not much that we, as a radar designer operating under ignorance, can do. The point is that we are going to use a quadratic (or linear) detector, and a threshold test; if the test statistic is not optimally tested by a threshold test, then none of the classical detection criteria nor the criterion introduced in this thesis apply since we are not using a likelihood ratio test.

In the next section a couple of examples are offered to illustrate the ideas introduced thus far.

2.5 Examples

In order to clarify some of the concepts introduced thus far, we now give two classical examples. The first involves the detection of a constant amplitude random phase target in Gaussian noise; such a target is known as a Marcum or Swerling 0 target [21]. The second example involves a Rayleigh-fading target in Gaussian noise.

2.5.1 Nonfluctuating Target in Gaussian Noise

Consider the detection problem for a coherent radar. We assume that the two hypotheses are described as

$$H_0 : \overline{X} = (X_I, X_Q) = (N_I, N_Q)$$

and

$$H_1 : \overline{X} = (N_I + A\cos\Psi, N_Q + A\sin\Psi)$$

where A is an unknown positive constant indicating the amplitude of the radar return due to a target, Ψ is the phase of the target return, N_I and N_Q are respectively, the in-phase and quadrature components of the noise. (Both A and Ψ are random variables.)

The distributions of $\overline{X}$ under the two hypotheses are functions of the amplitude (0 or A) and phase Ψ of the target return. Let the parameter space Θ be given by

$$\Theta = \{\Theta_1, \Theta_2\}$$

where $\Theta_1 = [0, \infty)$ and $\Theta_2 = [0, 2\pi]$. We assume that Θ_2 is uniformly distributed over $[0, 2\pi]$. Then we can write

$$\Lambda_0 = \{0\} \times [0, 2\pi] \qquad \text{and} \qquad \Lambda_1 = [0, \infty) \times [0, 2\pi] \qquad (2.42)$$

If the noise components n_I and n_Q are normally distributed with zero mean and unity variance, then

$$f_\Theta(\overline{x}) = \tfrac{1}{2\pi} \exp[-s(\overline{x}, \theta)] \qquad (2.43)$$

where

$$s(\overline{x}, \theta) = \tfrac{1}{2}[(x_I - \theta_1 \cos\theta_2)^2 + (x_Q - \theta_1 \sin\theta_2)^2] \qquad (2.44)$$

Then we can write the density under H_0 as

$$f_0(\overline{x}) = f_\Theta(\overline{x}|\theta \in \Lambda_0) = \tfrac{1}{2\pi} \exp\left[-\tfrac{1}{2}(x_I^2 + x_Q^2)\right] \qquad (2.45)$$

If we let $z = \sqrt{x_I^2 + x_Q^2} = |\overline{x}|$, and $\varphi = \tan^{-1}(x_Q/x_I)$ we can re-write (2.45) as

$$f_0(z, \varphi) = \frac{z}{2\pi} \exp\left[-\tfrac{1}{2}z^2\right] \qquad (2.46)$$

Integrating with respect to φ over $[0, 2\pi]$ is trivial, and leaves the Rayleigh distribution for z:

$$f_0(z) = z \exp\left[-\tfrac{1}{2}z^2\right] \qquad (2.47)$$

Note that $f_0(z)$ is independent of Θ; H_0 is simple.

The distribution of $\overline{x}$ under H_1 is a function of Θ. In order to remove this dependence, we assume θ_2 is uniformly distributed over $[0, 2\pi]$ and integrate:

$$f_1(\overline{x}) = f_\Theta(\overline{x}|\theta \in \Lambda_1) = \frac{1}{2\pi} \int_0^{2\pi} f_\Theta(x|\theta_1 = A)\, d\theta_2 \qquad (2.48)$$

Expanding and simplifying,

$$f_1(\overline{x}) = \frac{z}{2\pi} \exp\left(-\tfrac{1}{2}(A^2 + z^2)\right) \frac{1}{2\pi} \int_0^{2\pi} \exp[A(x_I \cos\theta_2 + x_Q \sin\theta_2]\, d\theta_2 \qquad (2.49)$$

From the definition of z and φ, $x_I = z\cos\varphi$ and $x_Q = z\sin\varphi$. Rewriting (2.49) in terms of z and φ, and remembering the trigonometric identity

$$\cos\varphi \cos\theta_2 + \sin\varphi \sin\theta_2 = \cos(\theta_2 - \varphi)$$

we get

$$f_1(z, \varphi) = \frac{z}{2\pi} \exp\left(-\tfrac{1}{2}(A^2 + z^2)\right) I_0(Az) \qquad (2.50)$$

where $I_0(z)$ is the zeroth-order modified Bessel function of the first kind and is given by the integral of (2.49). Integrating (2.50) with respect to φ over $(0, 2\pi)$ is trivial, and yields a Rician distribution for z:

$$f_1(z) = z \exp\left(-\tfrac{1}{2}(A^2 + z^2)\right) I_0(Az) \qquad (2.51)$$

By combining (2.47) and (2.51), the likelihood ratio can be written as

$$\ell(z) = \exp\left[-\tfrac{1}{2}A^2\right] I_0(Az) \qquad (2.52)$$

Note that likelihood ratio is a function of a scalar, $z = |\bar{x}|$. Since $I_0(Az)$ is a monotonically increasing function of z, so is $\ell(z)$. Therefore, from the Karlin-Rubin theorem, we can replace the likelihood ratio test by an equivalent threshold test. Since the rejection region is the same for all values of $\theta_1 = A$, the threshold test is UMP even though H_1 is composite.

In order to evaluate the probabilities of detection and false alarm, we integrate the density functions for z. The probability of false alarm is given by

$$pf = \int_t^\infty f_0(z)\,dz = \exp\left[-\tfrac{1}{2}t^2\right] \qquad (2.53)$$

We can solve (2.53) for the threshold as a function of pf:

$$t = \sqrt{-2\ln pf} \qquad (2.54)$$

$f_1(z)$ must be integrated numerically to determine pd:

$$pd = \int_t^\infty z\exp\left[-\tfrac{1}{2}(A^2 + z^2)\right] I_0(Az) = Q(A,t) \qquad (2.55)$$

where $Q(A,t)$ is called the Marcum Q function [33]. In chapter 5 we show how moment generating functions can be used to calculate $Q(A,t)$.

In figure 2.4 the ROC curves for Marcum targets with SNR = 0,5 and 10 dB above the (unit variance) noise are plotted. (The SNR is given by $\tfrac{1}{2}A^2$.) As expected, all lines are concave and above the chance line. As the SNR is increased to 10 dB, the curve gets closer to the northwest $(pf = 0, pd = 1)$ corner. For SNR = 10 dB, pd appears to remain very close to 1 for all but lowest pf; this is due to the nature of the target. For a Marcum (constant) target, pd is expected to stay large until the threshold exceeds the amplitude of the return; then pd decreases very quickly. We will see different behaviour with the next example.

2.5.2 Rayleigh-Fading Target in Gaussian Noise

We now consider returns with non-constant amplitude. Note that nothing has changed under H_0, hence $f_0(\bar{x})$ can be expressed as $f_0(z)$ as previously. We will show that the likelihood ratio is monotonic and therefore the likelihood test can be replaced with a threshold test; hence (2.53) and (2.54) are still valid.
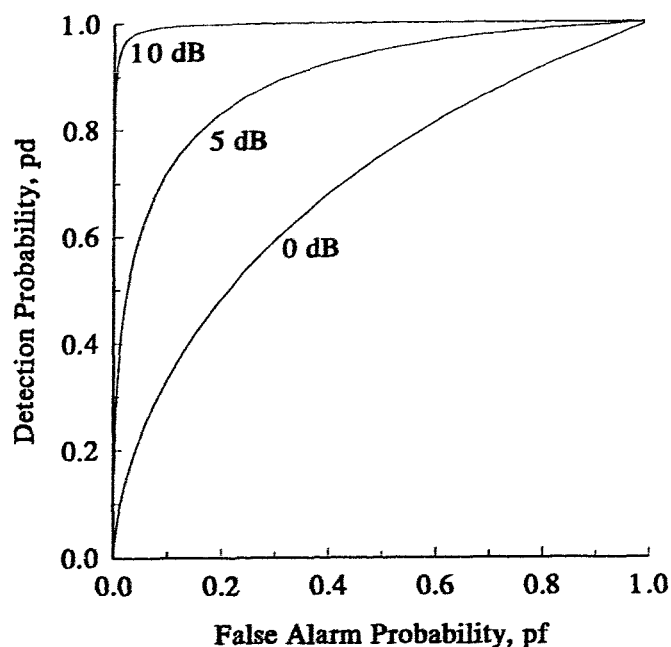
Figure 2.4: ROC Curves for Marcum targets with SNR = 0,5 and 10dB.

In order to deal with Λ_1 containing more than one point (i.e. H_1 is composite), we integrate $f_1(\overline{x})$ as before:

$$f_1(\overline{x}) = \int_{\Lambda_1} f_\Theta(\overline{x}) \overline{dx} = \int_0^{2\pi} f(\theta_2) \int_0^\infty f(\theta_1) f_\Theta(\overline{x}) d\theta_1 d\theta_2 \qquad (2.56)$$

where $f(\theta_1)$ and $f(\theta_2)$ are the density functions for the magnitude and phase of the return and $f_\Theta(\overline{x})$ is the density function of $\overline{x}$ given that the return has magnitude θ_1 and phase θ_2 (again it is assumed that Θ_2 is uniformly distributed over $[0, 2\pi]$). From the previous section, the likelihood function can be written

$$l(z) = \int_0^\infty \exp(-\tfrac{1}{2}\theta_1^2) I_0(\theta_1 z) f_{\theta_1}(z) \, dz \qquad (2.57)$$

Since $I_0(x)$ is monotonically increasing, and $f_{\theta_1}(z)$ is strictly non-negative, (2.57) shows that the likelihood ratio function is a monotonic function of $z = |\overline{x}|$. Therefore as claimed above, pf is the same as for a constant amplitude return.

In order to proceed further, we require a probability density function for the amplitude return. To simplify the notation, let $\theta_1 = A$ be the amplitude of the return. Note that A is now a random variable and not a constant as in the previous section. The most common assumption for a non-constant A assumes that the return is comprised by a large number of specular returns. If this is the case, we can invoke the central limit theorem. The in-phase and quadrature components are then described by a Gaussian distribution, and A is Rayleigh distributed:

$$f(a) = \frac{a}{\sigma^2} \exp\left(-\frac{a^2}{2\sigma^2}\right) \tag{2.58}$$

where σ^2 is the mode for a. We can also show that σ^2 is the single pulse signal-to-noise ratio (SNR), which we denote by ϱ:

$$\varrho = \frac{1}{2}E\{A^2\} \tag{2.59}$$

$$= \frac{1}{2}\int_0^\infty \frac{a^3}{\sigma^2}\exp\left(-\frac{a^2}{2\sigma^2}\right)\,da \tag{2.60}$$

$$= \sigma^2 \tag{2.61}$$

From (2.31) the detection probability is

$$pd = \int_t^\infty z\exp(-\tfrac{1}{2}z^2)\int_0^\infty \frac{a}{\sigma^2}\exp\left(-\frac{(1+\sigma^2)a^2}{2\sigma^2}\right)I_0(az)\,da\,dz \tag{2.62}$$

Now consider the substitution

$$y = \frac{a}{\sigma}\sqrt{1+\sigma^2}$$

Then the second integral in (2.62) can be re-written

$$\int_0^\infty \frac{y}{1+\sigma^2}\exp(-\tfrac{1}{2}y^2)I_0\left(\frac{y\sigma z}{\sqrt{1+\sigma^2}}\right)\,dy \tag{2.63}$$

$$= \frac{1}{1+\sigma^2}\exp\left(\frac{z^2\sigma^2}{2(1+\sigma^2)}\right)\int_0^\infty y\exp\left(-\frac{1}{2}\left[y^2 + \frac{z^2\sigma^2}{1+\sigma^2}\right]\right)I_0\left(\frac{\sigma yz}{\sqrt{1+\sigma^2}}\right)\,dy \tag{2.64}$$

By comparing (2.64) with (2.50), we see that the integral within (2.64) is the integral of a Rician density with $z = y$ and $A = \sigma z/\sqrt{1+\sigma^2}$, and is therefore equal to one. Thus re-writing (2.62)

$$pd = \int_t^\infty \frac{1}{1+\sigma^2}\exp\left(\frac{z^2\sigma^2}{2(1+\sigma^2)}\right)z\exp(-\tfrac{1}{2}z^2)\,dz \tag{2.65}$$
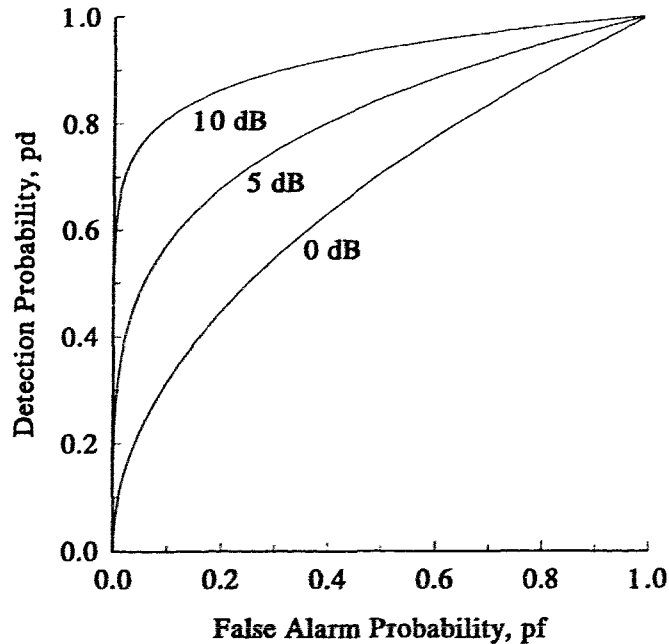
Figure 2.5: ROC Curves for Rayleigh-fading targets with SNR = 0,5 and 10dB.

$$= \exp\left(-\frac{t^2}{2(1+\sigma^2)}\right) \tag{2.66}$$

which can be re-written as a function of pf by using (2.54):

$$pd = pf^{\frac{1}{1+\sigma^2}} \tag{2.67}$$

which is a very well known result [21].

The ROC curves for a Rayleigh-fading target with SNR = 0,5 and 10 dB are given in figure 2.5. Again the curves are concave and above the chance line, with increased performance for larger SNR. Comparing with figure 2.4, it is seen that the Rayleigh target appears to fall off much quicker at larger pf than the Marcum target. This is because of the effect mentioned previously; the Marcum target is composed of a single tone plus noise and so pd remains high when the threshold is lower than the amplitude of the tone.

2.6 Conclusions

This chapter has given a brief introduction to classical detection theory, and its application to radar detection theory. We have seen from the Neyman-Pearson lemma and Birdsall's insight that a likelihood ratio test is always optimal for any sensible optimality criteria. When the observations are governed by densities that produce monotonic likelihood ratio functions, the Karlin-Rubin theorem shows that a threshold test is equivalent to the optimal likelihood ratio test. In most radar scenarios, a suboptimal statistic and not the optimal likelihood ratio is used. However, the threshold test applied to the suboptimal statistic is equivalent to a likelihood ratio test *for that test statistic* provided that the resulting ROC curve is concave.

Once the mechanics of producing the test statistic are set, the question is how to set the threshold. Table 2.1 compares the various criteria discussed in chapter 2. Cost function approaches such as Bayesian and Maximum Information tests require more a priori information than is commonly available for most radar scenarios. The Neyman-Pearson philosophy attempts to define an acceptable false alarm rate and then to accept the resulting detection performance. However, the false alarm rate is unobservable and cannot be controlled; therefore the density of the return under H_0 is required. Furthermore, there is no way to object₁vely choose the appropriate design value for the false alarm probability.

In the next chapter we consider a new detection criterion that has, in fact, been employed in many radar systems, but never analyzed in a formal framework.

Table 2.1: Comparison of Detection Criteria

Optimality Criteria	Optimal Threshold, λ	Implementation Considerations
Bayes	$\frac{\pi_0(C_{10}-C_{00})}{\pi_1(C_{01}-C_{11})}$	C_{01}, C_{10}, C_{00} and C_{11} are subjective and arbitrary. π_0 and π_1 must be known.
Ideal Observer (or Maximum a Posterior)	$\frac{\pi_1}{\pi_0}$	See above
Maximum Information	$\frac{\pi_0 \ln(\frac{1-pf}{pf}(\frac{\xi_1}{\xi_0}))}{\pi_1 \ln(\frac{pd}{1-pd}(\frac{\xi_0}{\xi_1}))}$	π_0 and π_1 are unknown. pf and pd require probability distributions under H_0 and H_1.
Neyman-Pearson	λ is such that pf is constant.	pf is unobservable unless sure target is absent. Takes a long time to estimate if small or we need the distribution of the test statistic under H_0 to calculate
Rate Constraint (discussed in chapter 3)	λ is such that $\xi_1 =$ constant.	ξ_1 is observable (doesn't require distributions - although may use distributions if available).

2.7 Appendix: Proof of Neyman-Pearson Test

The form of proof is based on [20]. It is included to allow comparison with the proof of the rate constraint criterion found at the end of chapter 3.

Proof of Optimality For δ and δ' defined as in equation (2.12), we always have

$$(\delta'(x) - \delta(x))(\ell(x) - \lambda) \geq 0 \tag{2.68}$$

Multiplying by $f_0(x)$ and integrating, we have

$$\int_\Gamma (\delta'(x) - \delta(x))(f_1(x) - \lambda f_0(x))\, d\mu(x) \geq 0 \tag{2.69}$$

and

$$\int_\Gamma \delta'(x) f_1(x)\, d\mu(x) - \int_\Gamma \delta(x) f_1(x)\, d\mu(x) \geq \lambda \int_\Gamma \delta'(x) f_0(x)\, d\mu(x) - \lambda \int_\Gamma \delta(x) f_0(x)\, d\mu(x) \tag{2.70}$$

which yields

$$pd(\delta') - pd(\delta) \geq \lambda[pf(\delta') - pf(\delta)] = \lambda[\alpha - pf(\delta)] \geq 0 \tag{2.71}$$

Therefore

$$pd(\delta') \geq pd(\delta)_\square \tag{2.72}$$

Proof of Existence Let λ_0 be the smallest number such that

$$P_0(\ell(x) > \lambda_0) \leq \alpha$$

where by $P_i(\bullet)$ we mean $P(\bullet|H_i)$, with $i = 0, 1$; that is the probability of ($\bullet$) occurring given that hypothesis i is true.

Then if

$$P_0(\ell(x) > \lambda_0) < \alpha$$

let

$$\gamma_0 = \frac{\alpha - P_0(\ell(x) > \lambda_0)}{P_0(\ell(x) = \lambda_0)} \tag{2.73}$$

Then defining the Neyman-Pearson rule δ_{NP} to be of the form of (2.12) with $\lambda = \lambda_0$ and $\gamma(x) = \gamma_0$, the false alarm probability of δ_{NP} is given by

$$pf(\delta_{NP}) = P_0(\ell(x) > \lambda) + \gamma_0 P_0(\ell(x) = \lambda) = \alpha_\square$$

Proof of Uniqueness Let δ' be an α-level test of the form given in (2.12) and let δ'' be any other α-level with the same pd. Since $pd(\delta') = pd(\delta'')$ equation (2.71) shows that $pf(\delta'') = \alpha$, which in turn (from (2.71)) implies that $pf(\delta') = pf(\delta'')$. Therefore working from (2.71) back to (2.68) we have

$$\int_\Gamma (\delta'(x) - \delta''(x))(\ell(x) - \lambda)\, d\mu(x) = 0 \tag{2.74}$$

Since the integrand must be positive or zero, $\delta''(x)$ must be of the same form as $\delta'(x)$ except possibly when $\ell(x) = \lambda$. Therefore δ'' can differ from δ' only in the randomizing function $\gamma(x)$. □

Chapter 3

Rate-Constraint Detection Criterion

3.1 Introduction

In chapter 2 various classical detection criteria were discussed and found to be inappropriate for radar target detection where a priori information is missing. While the Neyman-Pearson criterion is the one most often cited in radar detection literature, it is impractical or impossible to design, or even to describe the performance of, a statistical detection system in the Neyman-Pearson framework. There are two major problems in using the Neyman-Pearson criterion in practical detection systems. First, the false alarm rate is unobservable; therefore the distribution of the returns under H_0 must be fully specified since it is impossible to control an unobservable. Secondly and more importantly, assuming that we could achieve any false alarm rate desired the question remains: What is the appropriate rate to choose? While values such as 10^{-6} are typically given, there is no reason to pick this value over any other.

For example, consider the the gun-control radar discussed in section 1.1. If the false alarm rate is set too low, the ship is not optimally protected. If the false alarm rate is set too high, the gun is easily overloaded and not able to handle all targets. Clearly the optimal value for the false alarm rate is a function of the environment (number of targets) that the radar is working in. The optimal strategy is to set the threshold such that the resource is always fully utilized.

Nearly all detection systems control resources that are constrained by some combination of limits in processing power, memory constraints, and communications channels [14]. This chapter investigates the best way to make full use of the resources. In his study on information flow through general systems, Conant noted that for optimal performance we should "match components to tasks so that each component is operated at capacity, ... work it as hard as you can." The importance of matching the hit-rate to the tracker resources has also been noted in discussion of optimal radar tracking systems [6]. It is important to understand exactly how such a strategy performs.

3.1.1 Chapter Outline

This chapter examines the rate-constraint criterion, a new detection criterion that maximizes the effective probability of detection given that the resource utilizing the detection decisions can handle only a given number of threshold crossings per unit time. In section 3.2 the form of the optimal test is derived - not surprisingly it is a likelihood ratio test. The threshold for the test is such that the number of cells crossing the threshold is matched to the processing rate of the follow-on resource. The rate-constraint test is formally compared with a Neyman-Pearson test operating under a rate-constraint in section 3.2. In section 3.3 rate-constraint theory is applied to the detec⁺ion of Marcum and Rayleigh targets. The application of rate-constraint theory to a systems-level control system for an obstacle avoidance sonar is discussed in section 3.4. Conclusions are discussed in section 3.5. Finally, the chapter appendix (section 3.6) includes a proof of the rate constraint lemma.

3.2 Single Stage plus Resource

A simple two-stage detection system is shown in figure 3.1. It consists of a first stage that investigates the environment and makes decisions at a rate r_r, the radar cell rate, and an actuator acting on the detection decisions of the first stage. We assume that the actuator services each cell that produces a threshold crossing in the first stage and ignores the rest. We furthermore assume that the actuator is capable of handling only some fixed number, r_a, threshold crossings per unit time, and that the performance of the actuator is not a function of the processing load presented to it by the first stage (provided there are less than r_a crossings per unit time.)
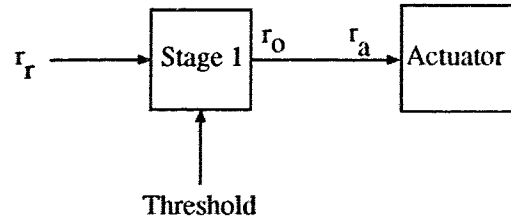
Figure 3.1: A Detector Cueing a Rate-Constrained Actuator

The question is what test to use in the first stage so that the probability of a target being serviced by the actuator is maximized. If the first stage were removed, the cells passed to the actuator would be chosen randomly and the resulting probability of detection would be

$$pd = \frac{r_a}{r_r} \tag{3.1}$$

The purpose of the first stage is to match the radar rate to the actuator rate in such a way that the probability of detection is a maximum, hopefully greater than (3.1).

First Stage Detection Criterion The goal of the first stage is to screen the radar cells in such a way that detection performance is optimized. In other words, the first stage must choose those cells most likely to contain targets to pass on to the actuator. Another way to view the problem is that the actuator is a resource that must be utilized in the most efficient fashion to detect targets in the radar space.

The model is quite general, and could describe a large number of detection problems. For example, the first stage could be a scout for a hockey team, an obstacle avoidance sonar, or a surveillance radar. Corresponding actuators could be a hockey coach conducting a number of skill tests, a thruster, or a jet interceptor. The actuator could even be another detection stage such as an automated tracking computer. In chapter 4 we will consider multi-stage detection systems, where the actuator is a series of detection stages that operate more slowly than the first stage, but give better detection performance.

The primary purpose of the first stage is to reduce the radar cell rate r_r to the actuator cell rate r_a. The output (hit or threshold crossing) rate r_o of the first stage can be described in terms of the probability of detection and false alarm of the first stage as

$$r_o = [\pi_1 pd + \pi_0 pf]r_r \tag{3.2}$$
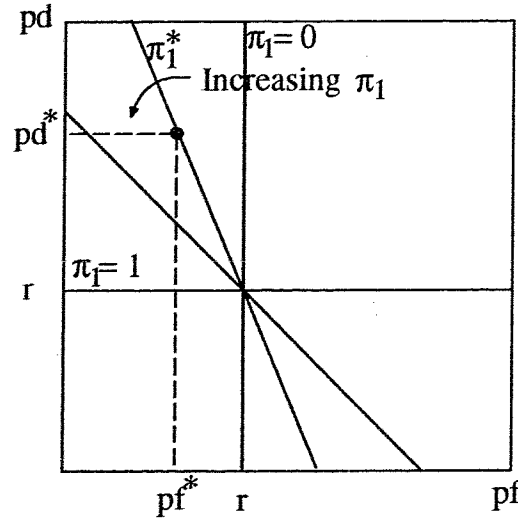
Figure 3.2: Constraint Boundaries for Various Values of π_1.

We require $r_o \leq r_a$ or

$$\pi_1 pd + \pi_0 pf \leq r \tag{3.3}$$

where $r = r_a/r_r$ is the normalized actuator input rate. We want the first stage to maximize pd subject to constraint (3.3).

Before discussing the rate-constraint detection criterion, it is helpful to describe briefly the constraint in terms of the standard ROC (receiver operating characteristic) diagram. Figure 3.2 illustrates the linear constraint boundaries (equation (3.3) with equality) for various values of π_1. For $\pi_1 = 0$ we have the vertical line intersecting the pf axis at $pf = r$ and for $\pi_1 = 1$, the horizontal line intersecting the pd axis at $pd = r$. For $0 < \pi_1 < 1$ the boundaries are a series of straight lines all passing through the point $pf = pd = r$ as shown. Any test developed must have a rate that is less than or equal to r and therefore pd must be less than or equal to the boundary value for a given pf and π_1. For example, given values of $pf = pf^*$ and $\pi_1 = \pi_1^*$ we have $pd \leq pd^*$. These constraint boundaries help to describe an operating point for the test criterion.

Operation of a Neyman-Pearson Test Under a Rate-Constraint Our objective here is to investigate the Neyman-Pearson criterion in light of the rate-constraint forced on the test by the actuator and expressed in (3.3). We consider Neyman-Pearson tests for two reasons. First, the Neyman-Pearson formulation is the one most often discussed in

literature; therefore it provides the benchmark to compare our criterion with. Secondly, the Neyman-Pearson test is representative of any test that applies a fixed threshold to a likelihood ratio test. Therefore, our consideration of Neyman-Pearson tests will yield insight into how Bayesian, Maximum Mutual Information and other tests will work.

Recall that a Neyman-Pearson test is a likelihood ratio test of the form of (2.12). If λ_{NP}, the threshold for the Neyman-Pearson test, is set such that $pf(\lambda_{NP}) = pf_{NP}$ then the hit-rate for the test is given by

$$\pi_1 pd_{NP} + \pi_0 pf_{NP} = r_{NP} \tag{3.4}$$

As long as $r_{NP} \leq r$, (3.3) is satisfied. If π_1 and pd are such that $r_{NP} > r$ then the hit-rate must be randomized to bring it down to r. This simply means that a fraction of the hits are thrown away and not passed to the actuator. Denoting the fraction of hits not thrown away as κ, we have

$$\kappa = \frac{r}{r_{NP}} = \frac{r}{\pi_1 pd_{NP} + \pi_0 pf_{NP}} \tag{3.5}$$

and the effective probabilities of detection and false alarm are

$$pd'_{NP} = \kappa pd_{NP} \tag{3.6}$$

and

$$pf'_{NP} = \kappa pf_{NP} \tag{3.7}$$

Solving (3.7) for κ and substituting it into (3.6) we have

$$pd'_{NP} = \frac{pd_{NP}}{pf_{NP}} pf'_{NP} \tag{3.8}$$

In summary, there are two expressions for the probability of detection under the Neyman-Pearson criterion. Which one is valid depends upon the fraction κ.

$$pd'_{NP} = \begin{cases} pd_{NP} & \kappa \geq 1 \\ \frac{pd_{NP}}{pf_{NP}} pf'_{NP} & \kappa < 1 \end{cases} \tag{3.9}$$

where pf'_{NP} is given by (3.7). Figure 3.3 illustrates these relationships on the ROC diagram.

As long as the target population, described by π_1, is such that $\kappa \geq 1$, the Neyman-Pearson test operates at point A. If the target population increases, however, pf and pd are forced to drop to maintain the rate at r. This drop is described by (3.7) for $\kappa < 1$ and is
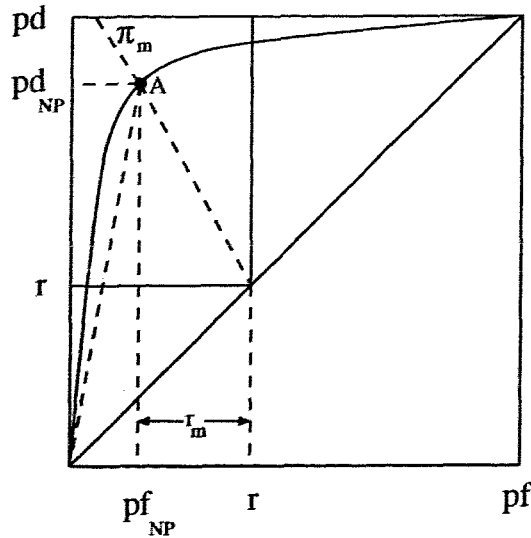
Figure 3.3: A ROC diagram illustrating the performance of a Neyman-Pearson test modified to operate under a Rate-Constraint

represented by the straight line from (pf_{NP}, pd_{NP}) to $(0,0)$ on the ROC diagram. Therefore if the actuator is overloaded $(\kappa < 1)$, random selection is necessary and performance degrades until finally for $\pi_1 = 1$, $pd'_{NP} = r$.

To guard against entering the overload region a rate margin r_m can be introduced. The rate margin is defined as the difference between pf_{NP} and r:

$$r_m = r - pf_{NP} \tag{3.10}$$

If r_m is positive the test will admit a certain target population π_m before going into overload. If r_m is negative the test is in overload for all π_1.

The behavior of the Neyman-Pearson criterion described above is familiar to designers of practical systems that work under a hit-rate constraint. We now discuss the optimum rate-constraint test.

3.2.1 The Rate-Constraint Criterion

The rate-constraint criterion maximizes *pd* subject to the normalized hit-rate being less than or equal to a constant *r*. A rate-constrained detection rule is defined as a decision rule having this property. The following lemma is proved in the appendix at the end of the chapter.

Lemma 2 (Rate-Constraint) *For a hypothesis test between H_0 and H_1 we have the following:*

1. *Optimality: Let δ be any decision rule satisfying $\pi_1 pd(\delta) + \pi_0 pf(\delta) = r_1 \leq r$ and let δ' be any decision rule of the form*

$$\delta'(x) = \begin{cases} 1 & \text{if } \ell(x) > \lambda \\ \gamma(x) & \text{if } \ell(x) = \lambda \\ 0 & \text{if } \ell(x) < \lambda \end{cases} \quad (3.11)$$

   where $\lambda \geq 0$ and $0 \leq \gamma(x) \leq 1$ are such that $\pi_1 pd(\delta') + \pi_0 pf(\delta') = r$. Then $pd(\delta') \geq pd(\delta)$.

2. *Existence: For every $r \in (0,1)$, there is a decision rule δ_r of the form of (3.11) with $\gamma(x) = \gamma_0$, a constant, for which*

$$\pi_1 pd(\delta_r) + \pi_0 pf(\delta_r) = r.$$

3. *Uniqueness: Let δ'' be any r rate-constrained decision rule for H_0 versus H_1. Then δ'' must be of form (3.11) except possibly on a set of measure zero under H_0 and H_1.*

Therefore a rate-constrained decision rule is a likelihood ratio test, as expected. In fact if the a priori probability of a target being present, π_1, is zero, the rate-constrained decision rule and the Neyman-Pearson decision rule are identical. However, for non-zero target densities the tests are different with the rate-constrained decision rule yielding the greater probability of detection.

This does not mean that the Neyman-Pearson test is not optimum for its stated constraint ($pf \leq \alpha$) but rather that it is not optimum under the rate-constraint. In fact to operate under the rate-constraint, the Neyman-Pearson test has to be randomized by the factor κ discussed earlier resulting in the following test.

$$\delta(x) = \begin{cases} \begin{cases} 1 & \text{if } \ell(x) > \lambda_{NP} \\ \gamma(x) & \text{if } \ell(x) = \lambda_{NP} \\ 0 & \text{if } \ell(x) < \lambda_{NP} \end{cases} & \text{if } \pi_1 pd_{NP} + \pi_0 pf_{NP} \leq r \\[3em] \begin{cases} \kappa & \text{if } \ell(x) > \lambda_{NP} \\ \kappa\gamma(x) & \text{if } \ell(x) = \lambda_{NP} \\ 0 & \text{if } \ell(x) < \lambda_{NP} \end{cases} & \text{if } \pi_1 pd_{NP} + \pi_0 pf_{NP} > r \end{cases} \quad (3.12)$$

Equation (3.12) describes a decision rule that satisfies the rate-constraint but it is not a rate-constrained decision rule as defined in (3.11) and therefore necessarily has a probability of detection less than the optimum.

The rate-constraint decision rule does not have some of the nice mathematical properties of the Neyman-Pearson decision rule because the rule changes for different targets and target concentrations; however, it does have one major advantage. Unlike the Neyman-Pearson rule, the constraint variable for the rate-constraint rule is observable. The constraint variable for the Neyman-Pearson rule is the probability of false alarm which is impossible to observe unless one is sure that there are no targets present. On the other hand the hit-rate is observable and therefore can be employed to control the detection process through a feedback loop.

The performance of the rate-constrained rule is illustrated on a ROC diagram in figure 3.4. If we assume that a control mechanism is in place to adjust the hit rate then the system operating point will move along the *pd/pf* curve until it intersects with the constraint line that corresponds to the target density. The operating point will remain stationary until the target density or the statistics change.

For example, consider operating point A of figure 3.4. If, say the signal-to-noise ratio drops from ϱ_A to ϱ_B while the target density remains constant, the operating point will move along the rate-constraint line to B, with an accompanying decrease in *pd* and increase in *pf*. Note that pd_B is higher than would have been obtained $(pd_{B'})$ if the false alarm were maintained at pf_A. On the other hand, if the operating point were A and the target population increased while the SNR remained constant at ϱ_A, the operating point would move to C. At C both *pd* and *pf* are reduced from their values at A to compensate for the higher density of targets so the following stage will not be overloaded. Conversely, if the population concentration decreases, *pd* and *pf* increase thereby increasing the detection sensitivity of the stage.

We define the *detection context* of a target as the set of all parameters which influence the position of the operating point. In mixed target and variable clutter situations these parameters may be difficult to identify but the operating point is still easily determined by observing and controlling the hit-rate. In these complex situations it may not be easy to translate the operating point into a performance number such as the probability of detection but at least the system is controllable.
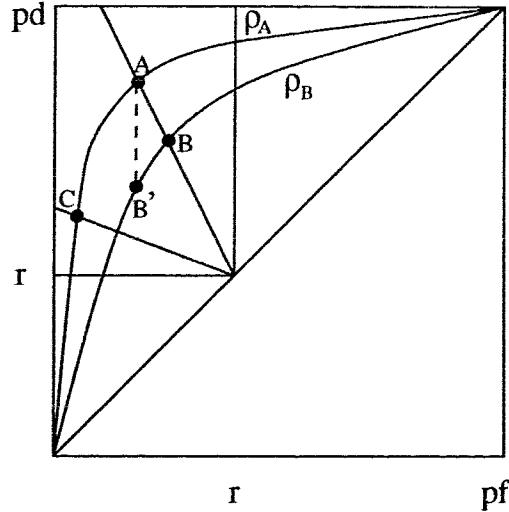
Figure 3.4: ROC diagram showing two pd versus pf curves ($\varrho_A > \varrho_B$) and linear constraint boundaries

Comparison of Rate-Constraint and Neyman-Pearson Criteria Under a Rate-Constraint

As was discussed in the previous section the modified Neyman-Pearson test (modified to perform under a rate-constraint) does not perform as well as the rate-constraint test. Specifically,

$$pd_{RC} \geq pd_{NP} \tag{3.13}$$

The equality holds when $pf_{RC} = pf_{NP}$ for the given detection context which occurs say at point A in figure 3.5. If π_1 decreases then the operating point for the RC test moves up the pd/pf curve, say to the point where $\pi_1 = \pi_B$. The detection probability pd_B is necessarily larger than pd_{NP} from lemma 2. It is also evident from observing that at the new operating point we have another Neyman-Pearson test with a larger pf than pf_A and therefore $pd_B > pd_A$.

If the operating point is at A and π_1 increases, then the rate-constraint operating point will move to, say, point C consistent with the constraint curve corresponding to π_C. Now this increase in π_1 results is an overload condition for the Neyman-Pearson test and so the operating point for this test must fall to point C' as described earlier. The operating point C' is at the intersection of the straight line from (pf_{NP}, pd_{NP}) to $(0,0)$ and the constraint line for $\pi_1 = \pi_C$. We already know that $pd_C \geq pd_{C'}$ because of lemma 2; furthermore, by
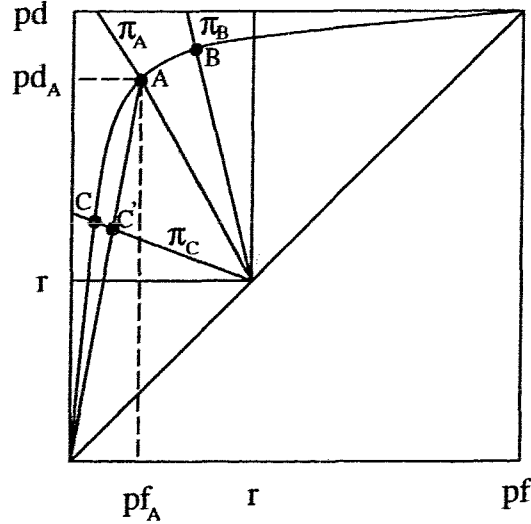
Figure 3.5: Comparison of modified Neyman-Pearson and rate-constrained tests.

observing the ROC curve we also see that it is true if

$$\frac{pd_C}{pf_C} \geq \frac{pd_{NP}}{pf_{NP}} \qquad (3.14)$$

for all possible points C on the pd/pf curve. Since (3.14) is both a necessary and sufficient condition for $pd_C \geq pd_{C'}$, we conclude that it must be true for all ROC curves developed using the likelihood ratio. In fact, (3.14) is merely another way of stating the concavity of the ROC curve.

One of the properties of the rate-constraint test is that it approaches the Neyman-Pearson test as the detectability of the target goes to zero (for example, if the signal-to-noise ratio of the target goes to zero). This property is implied by the rate-constraint which for $pd \rightarrow 0$ becomes $\pi_0 pf \leq r$ or $pf \leq r/\pi_0$, which is the Neyman-Pearson criterion. Consider figure 3.6. For the large signal-to-noise ratio (curve A) there is a significant difference between pd_{NP} and pd_{RC}. For curve B, however, which describes a low signal-to-noise ratio curve, there is very little difference because the pd/pf curve for the rate-constraint criterion approaches a straight line. Therefore the two tests are equivalent when the target detectability is small.

In summary, we have described the behavior of a Neyman-Pearson and rate-constraint test under a rate-constraint. The Neyman-Pearson test has to be modified to meet the constraint and the resulting test is suboptimal. The rate-constraint test is optimum and
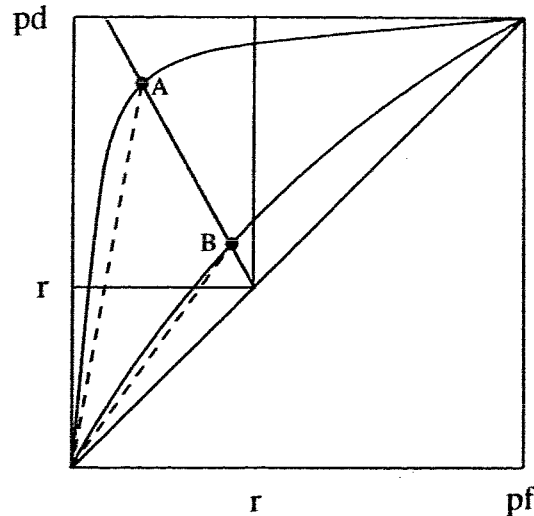
Figure 3.6: Comparison of modified Neyman-Pearson and rate-constrained tests for high (A) and low (B) SNR.

therefore has a greater probability of detection than the modified Neyman-Pearson test. One of the main reasons for considering the rate-constraint test is the fact that most practical systems employ some sort of hit-rate constraint to control the detection process. It is comforting to know that such a strategy has a desirable result.

In the next two sections we investigate the optimality of rate-constraint tests. Section 3.3 compares the performance of a rate-constraint test with a fixed-threshold test (Neyman-Pearson, for example) when detecting Marcum and Rayleigh-fading targets. Then in section 3.4 an example involving an obstacle avoidance sonar is given.

3.3 Examples

We now consider the problem of detecting the Marcum and Rayleigh-fading targets described in section 2.5 with the resource limited system of figure 3.1. By comparing the rate-constraint test with Neyman-Pearson tests, we can better understand the optimality of the former.

In figure 3.7, the probability of detection as given by (3.9) is plotted as a function of the threshold for pd given by (2.55) for a Marcum target, and (2.66) for a Rayleigh-fading target. The probability of a target, $\pi_1 = 10^{-6}$ while the actuator processing rate is $r_a = 10^{-4}$ and
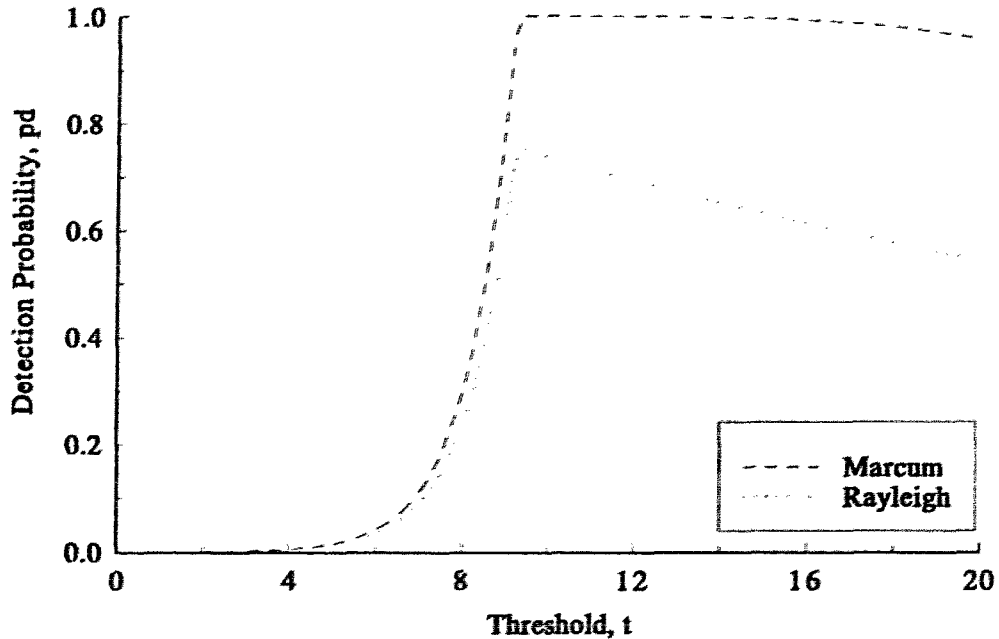
Figure 3.7: Detection Probability as a function of threshold for Marcum and Rayleigh targets. Single Pulse SNR = 15 dB, $r_a = 10^{-4}$, $\pi_1 = 10^{-6}$.

the SNR is 15 dB in both cases. The optimal thresholds are easily seen on the graph. In both cases, the optimal thresholds occur when the rate out of the first stage is 10^{-4}. If the threshold is made lower then its optimal value, randomization occurs and the effective detection probability falls off severely, eventually reaching that of (3.1) when the threshold is zero. If the threshold is raised from its optimal value, the detection probability falls off because the test is more selective than it needs to be.

Figure 3.7 shows the sensitivity of a test, operating under a rate-constraint, to the threshold. Clearly overload is a condition to be avoided. Any test operating under a fixed threshold (such as the Neyman-Pearson test) would likely have to set a threshold much higher than the optimal in order to guarantee against overload under all operating contexts.

The deleterious effect of using a fixed threshold in a dynamic environment is more easily
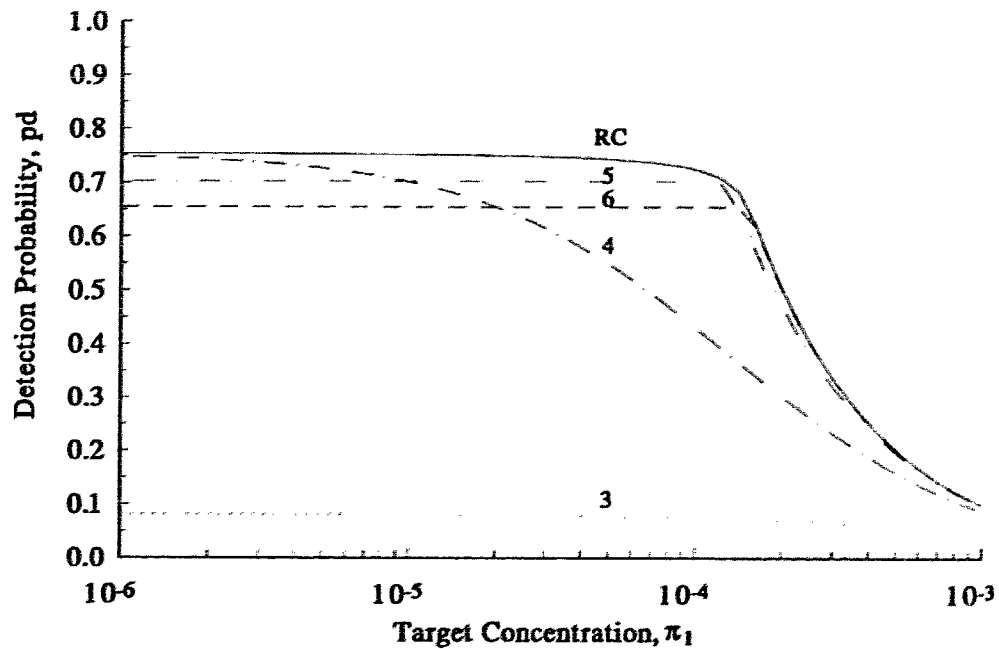
Figure 3.8: Detection probabilities for rate-constraint (curve RC) and various Neyman-Pearson rules (curves 3-6). Rayleigh target with SNR = 15 dB, $r_a = 10^{-4}$. For Neyman-Pearson rules the curve number is the negative exponent of false alarm probability (i.e. curve 3 is $pf = 10^{-3}$).

seen from figure 3.8. The effective detection probability is plotted for Neyman-Pearson rules with $pf = 10^{-6} - 10^{-3}$ and for the optimal rate constraint test with $r_a = 10^{-4}$. The target considered is a Rayleigh-fading target with SNR = 15 dB.

The various Neyman-Pearson tests become optimal at different values of π_1 (where their thresholds are the same as that for the optimal rate-constraint for that value of π_1). However, no single value of false alarm is best over the entire range of π_1. Note also that the rate-constraint test has poor detection performance as π_1 exceeds r_a: this simply means that the actuator isn't fast enough to even handle the targets, let alone the false alarms.

In the next section, we will consider a more involved example of a system that can be modeled by figure 3.1.

3.4 Obstacle Avoidance Sonar Application

In section 3.2 we saw that the rate-constraint criterion maximized the detection probability given a rate-constraint and, in addition, can be implemented in practical situations. In this section we apply the rate-constraint detection criterion to the problem of obstacle avoidance for an autonomous underwater vehicle (AUV). As expected, the rate-constraint criterion performs optimally. The example is presented for two reasons. First it illustrates the rate-constraint in a slightly different context than the radar examples presented in the rest of the thesis. Second it gives an example of rate-constraint theory applied at the mission control level.

We assume that the vehicle must navigate through a field of obstacles with average concentration π_1. In other words, π_1 is the probability of a particular sonar cell containing a target. We also assume that the sonar starts to collect information on a particular cell (i.e. starts building the test statistic) as soon as the cell comes within range of the sonar. When the cell is a certain distance away, called the decision distance, the sonar must decide to avoid the cell or not.

If the vehicle encounters k cells on its way through the field then the probability of mission success (i.e. probability of not bumping into an obstacle for k independent cells) is

$$p(success) = (\pi_0)^k \qquad (3.15)$$

If a sonar with an obstacle detection probability of pd is added to the vehicle the probability of success increases because the only obstacles that will stop the mission are those that are missed by the sonar. Since the probability of not detecting an obstacle is $(1 - pd)$, we have

$$p(success) = (1 - \pi_1(1 - pd))^k \qquad (3.16)$$

Therefore the effect of the sonar is to reduce the target concentration by a factor of $(1 - pd)$, subsequently increasing the mission success probability. In fact, as pd increases to one the mission success probability also increases to one.

There is a price to be paid, however, for the higher mission success probability. Every time an obstacle is detected by the sonar, the vehicle must manoeuvre around it, thus delaying its arrival at the assigned destination. We assume that the vehicle is capable of a top speed of v_o m/s and that it must travel at an average velocity of v_a m/s $(0 < v_a \leq v_o)$

in order to meet the mission requirement. Therefore there are two ways that the mission can fail: impacting an obstacle or arriving late. The first is mitigated by a higher pd but this aggravates the second. The second is mitigated by maintaining v_a regardless of the obstacles which, of course, increases the likelihood of hitting an obstacle. Hence we must risk the vehicle to maintain v_a and our task is to minimize this risk.

The detection performance of the sonar is related to the mission objective by first obtaining a relationship between the vehicle's velocity and the threshold crossing rate of the sonar. If the vehicle is traveling at the required average velocity, v_a, then the sonar cell rate (the rate new cells are presented to the sonar) is

$$r_s = \frac{v_a}{\Delta R} \qquad (3.17)$$

where ΔR is the range extent of a cell. The average velocity of the vehicle may be expressed in terms of the fraction of time the vehicle spends in manoeuvre mode and transit mode assuming that it travels at v_o when in transit mode. Therefore

$$v_a = v_o(1 - f) + \frac{\Delta R}{t_o}f \qquad (3.18)$$

where f is the fraction of time in manoeuvre mode and t_o is the time it takes to manoeuvre around an object. In other words, the vehicle travels at a velocity of v_o when it is transiting and at an effective velocity of $\frac{\Delta R}{t_o}$ when it is manoeuvring around an object.

Let r_a be the average threshold crossing rate. Then the average time spent manoeuvring out of a total time t is $r_a t t_o$, and therefore $f = r_a t_o$. But $r_a = r r_s$ where r is the normalized rate; therefore from (3.17), $r_a = r\frac{v_a}{\Delta R}$, and the fraction of time spent manoeuvring is

$$f = \frac{r v_a t_o}{\Delta R} \qquad (3.19)$$

Substituting (3.18) into (3.19) and solving for r we obtain

$$r = \frac{\Delta R(v_o - v_a)}{v_a(v_o t_o - \Delta R)} \qquad (3.20)$$

Therefore, given a sonar resolution cell size ΔR, a maximum velocity v_o, a manoeuvre time t_o and a mission specified average velocity v_a, we can determine a required threshold crossing rate. If this rate is exceeded the vehicle will spend too much time in manoeuvre mode and will not meet the mission deadline. If the rate is less than r the vehicle will arrive ahead of schedule but will be exposed to a greater than necessary risk of hitting an obstacle.

Then we have a rate-constrained detection problem with constraint

$$\pi_1 pd + \pi_0 pf \leq r = \frac{r_a}{r_s} = \frac{\Delta R(v_o - v_a)}{v_a(v_o t_o - \Delta R)} \qquad (3.21)$$

From section 3.2, we know that the rate-constraint criterion will be optimal. To be definite in our results and to provide an example, let us assume that we are detecting the Rayleigh-fading target of section 2.5.2. Recall that from (2.67),

$$pd = pf^{\frac{1}{1+\varrho}} \qquad (3.22)$$

where ϱ is the signal-to-noise ratio. It should be noted that this model typically is used to describe the test statistic that results from the a single sonar return, or from coherent integration of a number of sonar returns over time. A more accurate model for the problem is one that employs noncoherent integration from pulse-to-pulse as the cell of interest moves closer to the vehicle (such a model is discussed in chapter 5.) This latter model, however, does not have a neat closed-form expression for pd as a function of pf and therefore would not be efficient for showing the trends we wish to show. Since the purpose of this section is to illustrate principles rather than model particular signal statistics we have chosen to use the mathematically simpler model here.

For the rate-constraint criterion, we have

$$r = \pi_1 pd_{rc} + \pi_0 pd_{rc}^{1+\varrho} \qquad (3.23)$$

where pd_{rc} is the detection probability resulting from the rate-constraint criterion. Therefore to determine the maximum pd attainable, namely pd_{rc} that satisfies the mission velocity constraint we must solve the above transcendental equation.

In this section we again compare the performance of the rate-constraint test with that of the Neyman-Pearson test. For the Neyman-Pearson criterion we have

$$r_{NP} = \pi_1 pd_{NP} + \pi_0 pd_{NP}^{1+\varrho} \qquad (3.24)$$

where r_{NP} is the threshold crossing rate. Remembering that the threshold λ is chosen to provide a specific false alarm probability pf_{NP}, pd_{NP} is fixed through equation (3.24). Therefore as long as $r_{NP} \leq r$ we maintain pd at pd_{NP}, but if the target concentration increases so that $r_{NP} > r$, pd_{NP} must be reduced by the factor κ (as discussed in section

3.2) to maintain our desired velocity v_a. Hence for the Neyman-Pearson criterion we have an effective probability of detection $pd_{NP,eff}$ as described by equation (3.9). Since we have

$$\frac{pd_{NP}}{pf_{NP}}pf_{NP,eff} = \frac{pd_{NP}}{pd_{NP}^{1+\varrho}}\frac{rpd_{NP}^{1+\varrho}}{(\pi_1 pd_{NP} + \pi_0 pd_{NP}^{1+\varrho})} = \frac{r}{\pi_1 + \pi_0 pd_{NP}^{\varrho}} \tag{3.25}$$

equation (3.25) becomes

$$pd_{NP,eff} = \begin{cases} pd_{NP} & \text{if } r_{NP} \leq r \\ \frac{r}{\pi_1 + \pi_0 pd_{NP}^{\varrho}} & \text{if } r_{NP} > r \end{cases} \tag{3.26}$$

By comparing equations (3.23) and (3.26) we are able to verify that $pd_{rc} \geq pd_{NP,eff}$. First consider $r_{NP} \leq r$; then we have

$$\pi_1 pd_{NP} + \pi_0 pd_{NP}^{1+\varrho} \leq \pi_1 pd_{rc} + \pi_0 pd_{rc}^{1+\varrho} \tag{3.27}$$

for which we require $pd_{NP,eff} = pd_{NP} \leq pd_{rc}$.

Second, for $r_{NP} > r$ we have

$$\pi_1 pd_{NP} + \pi_0 pd_{NP}^{1+\varrho} > \pi_1 pd_{rc} + \pi_0 pd_{rc}^{1+\varrho} \tag{3.28}$$

for which we require $pd_{NP} > pd_{rc}$, but for this condition we have

$$pd_{NP,eff} = \frac{r}{\pi_1 + \pi_0 pd_{NP}^{1+\varrho}} \tag{3.29}$$

and by rearranging equation (3.23) we have

$$pd_{rc} = \frac{r}{\pi_1 + \pi_0 pd_{rc}^{1+\varrho}} \tag{3.30}$$

which when compared with (3.29) shows that $pd_{NP,eff} < pd_{rc}$. Hence $pd_{rc} \geq pd_{NP,eff}$ for both conditions as expected.

In figure 3.9, we illustrate the detection performance of the rate-constraint criterion as a function of the target concentration for two signal-to-noise ratios and average velocities. The maximum velocity was set at 3m/s, the sonar resolution at 10cm and the time required to manoeuvre around an obstacle at 5s. For a required average velocity of 0.5m/s when compared with 2.5ms, we have a relatively large margin between the maximum velocity and the average, and therefore more time can be spent in manoeuvring. Hence, the vehicle can successfully navigate a denser target field as indicated by the breakpoint in the detection
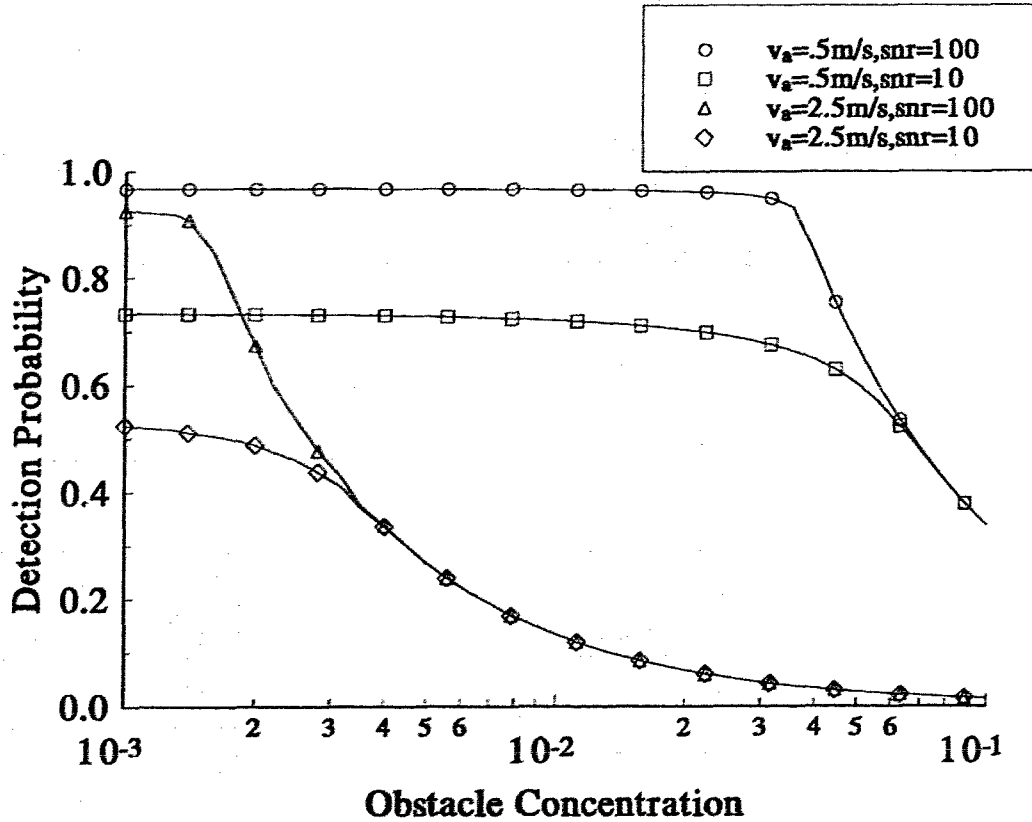
Figure 3.9: Detection performance of rate-constraint criterion under various SNR and average velocities. v_o =3 m/s, $\Delta R = 10$cm.

probability curve being farther to the right. The breakpoint is located where the target concentration π_1 equals the threshold crossing rate r. For example, with v_a =0.5m/s the rate given by plugging v_a, v_0, ΔR, and t_0 into (3.20) is $r = 0.0333$. Examination of figure 3.9 indicates that the breakpoint does indeed occur at $\pi_1 = .0333$.

For the lower signal-to-noise ratio the detection probability is lower, as expected. These detection curves approach those of the higher signal-to-noise ratio as the concentration increases and both curves tend to $pd = r$.

Figure 3.10 compares the high signal-to-noise ratio performance of the rate-constraint criterion with that of the Neyman-Pearson criterion. Curves for three probabilities of false alarm are shown. It is evident that none of these curves is higher than that for the rate constraint criterion. Before the breakpoint the Neyman-Pearson curves will increase with
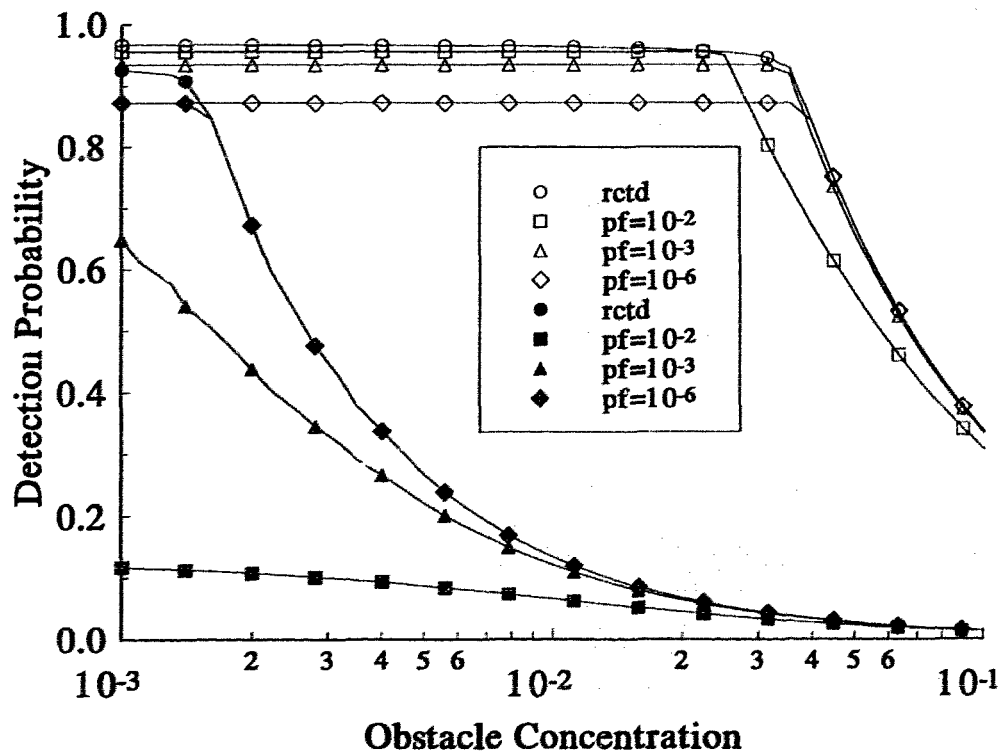
Figure 3.10: High SNR comparison between rate-constraint and Neyman-Pearson criterion. The filled symbols correspond to $v_a = 2.5$m/s the hollow symbols to $v_a = 0.5$ m/s.

decreasing pf if $r_{NP} > r$ but otherwise will decrease as shown in the figure. After the breakpoint, low pf's allow the Neyman-Pearson performance to approach the rate-constraint performance but never surpass it.

These figures illustrate the expected performance and optimality of the rate constraint criterion over a wide range of target concentrations. Two points are of significance. The first is that even though we don't know the operating point in the sense of where it is on the performance curve we can still set it simply by adjusting the threshold to obtain the desired velocity, v_a. Once the threshold is set we are assured of maximizing the probability of mission success, even though we don't know what the maximum value is.

The second point is that the resource-constraint philosophy gives us a method by which we can relate our detection criterion to the goals of the organism or vehicle. This means

that the statistical sensor can be integrated more intimately into the command and control structure of the system.

3.5 Conclusions

In chapter 3 we have investigated a new detection criterion designed to achieve maximum detection performance with a resource-constrained -ctuator. Since almost all actuators following a detection test do in fact have finite resources, the model is quite general.

The resulting test is a likelihood ratio test (or an equivalent threshold test when the Karlin Rubin theorem holds), with its threshold set such that the threshold crossing rate matches the processing rate of the actuator. A test built around rate-constraint principles is practical; the hit rate is observable and controllable, and the processing rate of the actuator is a known design parameter.

Note that a rate-constraint test emphasizes maximizing pd regardless of the consequences of pf. It may be dangerous to operate an entire system under such a philosophy; for example the gun-control example of section 1.1 could end up shooting down friendly aircraft if further tests such as Identify Friend or Foe (IFF) are not used to identify the potential targets. Typically more than one test is used in a detection system before a final decision is made. In chapter 4 we show how the rate-constraint criterion extends naturally to systems consisting of a sequence of tests.

3.6   Appendix: Proof of Rate-Constraint Test Properties

In this appendix, we prove the claims made in Lemma 2 concerning the optimality, existence and uniqueness of the rate-constraint criterion. The proof parallels that given in section 2.7 for the Neyman-Pearson test.

Proof of Optimality   For δ and δ' defined as in equation (3.11), we always have

$$(\delta'(x) - \delta(x))(\ell(x) - \lambda) \geq 0 \tag{3.31}$$

Multiplying both sides by $f_0(x)$ and integrating over Γ, we have

$$\int_\Gamma (\delta'(x) - \delta(x))(f_1(x) - \lambda f_0(x))\, d\mu(x) \geq 0 \tag{3.32}$$

and

$$\int_\Gamma \delta'(x) f_1(x)\, d\mu(x) - \int_\Gamma \delta(x) f_1(x)\, d\mu(x) \geq \lambda \int_\Gamma \delta'(x) f_0(x)\, d\mu(x) - \lambda \int_\Gamma \delta(x) f_0(x)\, d\mu(x) \tag{3.33}$$

which yields

$$pd(\delta') - pd(\delta) \geq \lambda[pf(\delta') - pf(\delta)] \tag{3.34}$$

From the constraint we have

$$pf(\delta') = \frac{r - \pi_1 pd(\delta')}{\pi_0} \tag{3.35}$$

and

$$pf(\delta) = \frac{r_1 - \pi_1 pd(\delta)}{\pi_0} \tag{3.36}$$

Inserting (3.35) and (3.36) into (3.34), we have

$$pd(\delta') - pd(\delta) \geq \frac{\lambda(r - r_1)}{\pi_0 + \lambda\pi_1} \tag{3.37}$$

but

$$\frac{\lambda(r - r_1)}{\pi_0 + \lambda\pi_1} \geq 0 \tag{3.38}$$

because $r \geq r_1$, so we have the desired result,

$$pd(\delta') - pd(\delta) \geq 0_\square \tag{3.39}$$

Proof of Existence Let λ_0 be the smallest number such that

$$\pi_0 P_0(\ell(x) > \lambda_0) + \pi_1 P_1(\ell(x) > \lambda_0) \leq r$$

Then if

$$\pi_0 P_0(\ell(x) > \lambda_0) + \pi_1 P_1(\ell(x) > \lambda_0) < r$$

let

$$\gamma_0 = \frac{r - \pi_0 P_0(\ell(x) > \lambda_0) - \pi_1 P_1(\ell(x) > \lambda_0)}{\pi_0 P_0(\ell(x) = \lambda_0) + \pi_1 P_1(\ell(x) = \lambda_0)} \tag{3.40}$$

Then defining the rate-constrained decision rule δ_r to be of the form of (3.11) with $\lambda = \lambda_0$ and $\gamma(x) = \gamma_0$ we have the rate of δ_r, $r(\delta_r)$ given by

$$r(\delta_r) = \pi_0 P_0(\ell(x) > \lambda_0) + \pi_1 P_1(\ell(x) > \lambda_0) +$$

$$\gamma_0(\pi_0 P_0(\ell(x) = \lambda_0) + \pi_1 P_1(\ell(x) = \lambda_0)) \tag{3.41}$$

which, after substituting (3.40) for γ_0 gives us the desired result,

$$r(\delta_r) = r \quad \square \tag{3.42}$$

Proof of Uniqueness Let δ' be a rate-constrained test of the form given in equation (3.11) and let δ'' be any other rate-constrained test with the same *pd*. Since $pd(\delta') = pd(\delta'')$ equation (3.37) shows that $r = r_1$, which in turn (from (3.35) and (3.36)) implies that $pf(\delta') = pf(\delta'')$. Therefore from (3.33) and then (3.31) we have

$$\int_\Gamma (\delta'(x) - \delta''(x))(\ell(x) - \lambda) \, d\mu(x) = 0 \tag{3.43}$$

Since the integrand must be positive or zero, $\delta''(x)$ must be of the same form as $\delta'(x)$ except possibly when $\ell(x) = \lambda$. Therefore δ'' can differ from δ' only in the randomizing function $\gamma(x)$. $\square$

Chapter 4

Multi-Stage Detection Systems

4.1 Introduction

In most detection systems, a given cell must pass several tests before being declared a target. For example, in Nathanson's chapter on false alarm control for radar, seven stages of processing and decision making are listed [18]. As a given cell passes through the various stages (Pulse compression, Moving Target Indicator, CFAR, binary integration, clutter map, digital track extraction and tracking), the likelihood of that cell containing a target becomes larger. The later stages in the detection chain interrogate less cells, but require more processing power to investigate those cells.

Considering the problem of medical diagnoses based on several tests, Metz writes:

> Diagnostic tests are rarely used alone. Instead the results of several diagnostic tests are usually combined with clinical background information to decide the disease state of the patient or to decide that additional diagnostic tests should be performed. In order to choose the best sequence of diagnostic tests, that is, to optimize diagnostic strategy, one must recognize that (pd) and (pf) for each diagnostic test usually can be changed together by changing the decision threshold for the test... Full optimization of diagnostic strategy involves choosing not only the best sequence of tests, but also the best operating point on the ROC curve for each test [12].

In this chapter we investigate a system where faster, coarser sensors use their detection

ability to cue the attention of slower, higher resolution sensors. Such models have been described before. For example, Lacoss [34] described a system where acoustic sensors were used to cue television cameras to track aircraft and Hovanessian [11] has suggested the concept of using low resolution search radars to cue lidars. Schweizer et al, describing an architecture for mine-hunting using autonomous underwater vehicles wrote:

> When lower resolution systems are used as cuers for high resolution systems, collections of large volumes of data may be significantly avoided, thus also reducing field analysis time [13].

While these examples have been described, this thesis provides the general framework and analysis to justify their use.

There is currently much interest in *sensor fusion*: the problem of optimally combining information gathered by a number of sensors or tests. (For example, [35],[36].) A difficulty with many papers discussing fusion is their implicit assumption that all of the sensors process data at the same rate and resolution. For example, in the serial fusion system of [35], every stage in the network processes every cell in the radar space. The disparity in cell production rates between different types of sensors can make such systems impractical.

In an appropriately titled (Fusion or confusion: knowledge or nonsense?) review of fusion technology, Rothman and Denton noted:

> Another problem which is infrequently discussed in the theoretical literature, but is commonly encountered in implementation, is the lack of synchronous updates between sensors... in advanced systems and in virtually all distributed systems, it is not possible to synchronize the sensing systems [37].

Cued systems provide the same robustness and graceful degradation features that many fusion structures do. They don't require the same level of co-ordination, and can be practically implemented. There are no requirements for the sensors to be identical, or to provide statistics that are identically distributed.

This chapter presents the framework and analysis required to model and understand the performance of a cued detection system. The rules that govern the choice of operating point and the conditions of optimality will be explained.

4.1.1   Chapter Outline

In section 4.2 a two-stage cueing system is shown to be identical in performance to a specific type of fusion network. The optimal form of the local tests is a likelihood ratio test. The conditions for optimal detection performance of the overall system are derived with no resource constraints. A new parameter, the SLOC number is seen to be important in determining the optimal operating points for the individual tests. The SLOC function is the slope of the ROC curve when plotted on log-log axes.

In section 4.3 we introduce rate constraints into the second stage. Again the SLOC number is central in determining the optimal operating strategy. We show that a comparison of the SLOC functions for the two stages determines the region of operation where the rate constraint criterion is optimal for the first stage.

In section 4.4 a dynamic programming argument is used to show that the results for the two-stage system extend generally to systems with an arbitrary number of stages, provided the SLOC functions for the stages are properly ordered. We further show that the overall system detection performance is optimized, under any sensible detection criterion, when the rate constraint is applied at every stage but the last.

Finally the chapter's conclusions are given in section 4.5.

4.2   Unconstrained Two Stage Systems

We begin by examining the relationship between our model and that of a 'conventional' fusion model. In figure 4.1 we have a number of sensors independently interrogating the environment. The sensors make individual decisions and report them to a central fusion center which then makes the global decisions. Thomopoulos et al have shown that the optimal local decision rules are likelihood ratio tests [36].

Consider the 'And' fusion rule, where a target is declared if and only if all of the individual tests report a threshold crossing. If any one of the tests fails to report a hit, the fusion center doesn't need to examine the rest of the local tests. Thus the system of figure 4.1 performs the same test as that of 4.2.

Each sensor of figure 4.2 passes on the addresses of those cells which have exceeded the local threshold to be examined by subsequent sensors. Therefore the sensors later in the chain need to examine far less data than those earlier in the chain. At first glance it
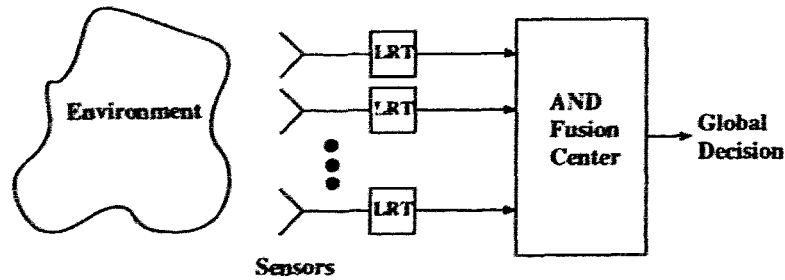
Figure 4.1: Sensor fusion using 'And' rule and Likelihood Ratio Tests (LRT)
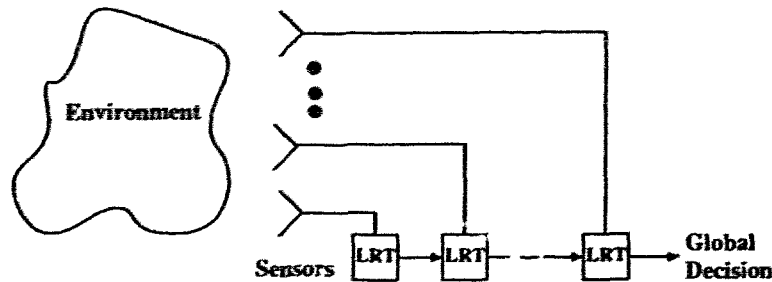
Figure 4.2: Alternate form of 'And' fusion. A given sensor-LRT stage only examines those cells that have exceeded the thresholds in all of the previous stages.

may appear that the model of figure 4.2 requires too much co-ordination between sensors; however, the system is actually quite robust. If any given sensor (except for the first) becomes unserviceable, the overall system can be gracefully degraded by removing that stage from the overall system and linking the two on either side. Of course the first stage is crucial, since no other stages are capable of processing at the radar rate.

4.2.1 Optimal Solution For an Unconstrained Two Stage System

Consider a system without any constraints on the processing power of the second stage. Since the optimal form for the individual tests is a likelihood ratio test, we can formulate the Neyman-Pearson problem as

$$\max_{\lambda} \prod_{i=1}^{N} pd_i(\lambda_i) = pd_T \qquad (4.1)$$

subject to

$$\prod_{i=1}^{N} pf_i(\lambda_i) = pf_T$$

where pd_i, pf_i and λ_i are, respectively, the detection probability, the false alarm probability and the threshold for sensor i, $\overline{\lambda}$ is the vector of the $\lambda_i's$, and pf_T is the desired false alarm probability out of the fusion sensor.

Why are we using the Neyman-Pearson formulation? From Birdsall's insight, any optimal test must lie on the ROC curve for the overall system. The system ROC curve is generated by solving (4.1) for pf_T varying from 0 to 1.

Consider the case $N = 2$. To find the optimal λ_i, we use the Lagrange Multiplier optimization technique with objective function

$$J = pd_1 pd_2 - \zeta(pf_1 pf_2 - pf_T) \tag{4.2}$$

where ζ is the Lagrange Multiplier, and we have suppressed the functional dependence of the pd_i and pf_i on λ_i. Taking the partial derivatives of J and equating them to zero yields

$$\frac{\partial J}{\partial \lambda_1} = pd'_1 pd_2 - \zeta pf'_1 pf_2 = 0 \tag{4.3}$$

$$\frac{\partial J}{\partial \lambda_2} = pd_1 pd'_2 - \zeta pf_1 pf'_2 = 0 \tag{4.4}$$

and

$$\frac{\partial J}{\partial \zeta} = pf_1 pf_2 - pf_T = 0 \tag{4.5}$$

Solving (4.3) and (4.4) for ζ,

$$\zeta = \frac{pd'_1 pd_2}{pf'_1 pf_2} = \frac{pd_1 pd'_2}{pf_1 pf'_2} \tag{4.6}$$

Multiplying by $\frac{pf_1 pf_2}{pd_1 pd_2}$, we find that the condition for optimal pd_T becomes

$$\frac{pd'_1 pf_1}{pf'_1 pd_1} = \frac{pd'_2 pf_2}{pf'_2 pd_2} \tag{4.7}$$

where $pd'_i = \frac{dpd_i}{d\lambda_i} = f_1(x_i)$ is the probability density function of x_i given the existence of a target. Similarly $pf'_i = f_0(x_i)$. Since $\lambda_i = \frac{f_1(x_i)}{f_0(x_i)}$ is the threshold for the likelihood test applied to test i, (4.7) can be re-written as

$$\lambda_1 \frac{pf_1}{pd_1} = \lambda_2 \frac{pf_2}{pd_2} \tag{4.8}$$

Still another form is obtained by recognizing that $\frac{g'(x)}{g(x)} = \frac{d\ln g(x)}{dx}$. Then (4.8) can be expressed as

$$\frac{\partial \ln pd_1}{\partial \ln pf_1} = \frac{\partial \ln pd_2}{\partial \ln pf_2} \tag{4.9}$$

Equations (4.7) through (4.9) have obvious generalizations for $N > 2$. The derivatives of (4.9) are important for our purposes. Let

$$\eta_i = \frac{\partial \ln pd_i}{\partial \ln pf_i} \qquad (4.10)$$

Equations (4.8) and (4.9) suggest two methods of calculating η_i graphically. From (4.8), η_i at any given operating point on a ROC curve is the product of the threshold (i.e. the slope of the ROC curve) and the secant line to the point. Equation (4.9) yields the preferred method: plot pd versus pf on a log-log scale and the slope of the resulting curve is η_i. We call such plots *Log-ROC* curves, and η_i the *SLOC* (i.e. the Slope of the Log receiver Operating Characteristic curve). Clearly, the SLOC number is an important relational parameter for determining the optimal operating point of a test.

Condition (4.9) can be understood by examining the effect of changing pf_1 and pf_2 from their optimal values by a small amount while maintaining the overall false alarm probability at pf_T.

To see the change in the overall detection probability, consider

$$\Delta \ln(pd_1pd_2) \quad = \quad \Delta \ln pd_1 + \Delta \ln pd_2 \qquad (4.11)$$

$$= \quad \frac{\partial \ln pd_1}{\partial \ln pf_1} d \ln pf_1 + \frac{\partial \ln pd_2}{\partial \ln pf_2} d \ln pf_2 \qquad (4.12)$$

but $d \ln pf_1 = -d \ln pf_2$ to maintain the false alarm probability at pf_T. Then

$$\Delta \ln(pd_1pd_2) = (\eta_1 - \eta_2) d \ln pf_1 \qquad (4.13)$$

Now consider what happens if (4.9) is not true. Then the difference term in (4.13) is non-zero, and by making $d \ln pf_1$ the same sign as the difference, $\Delta \ln(pd_1pd_2)$ can be made positive. But a positive $\Delta \ln(pd_1pd_2)$ would imply that the starting point was a non-optimal solution; therefore (4.9) must hold at the optimal operating point.

We now consider a couple of examples to better understand the meaning of (4.9).

4.2.2 Examples for Unconstrained Two-Stage Systems

Let us begin with the Marcum target discussed in section 2.5.1. Plugging the results for the likelihood ratio (2.52), and probabilities of false alarm (2.53) and detection (2.55) into
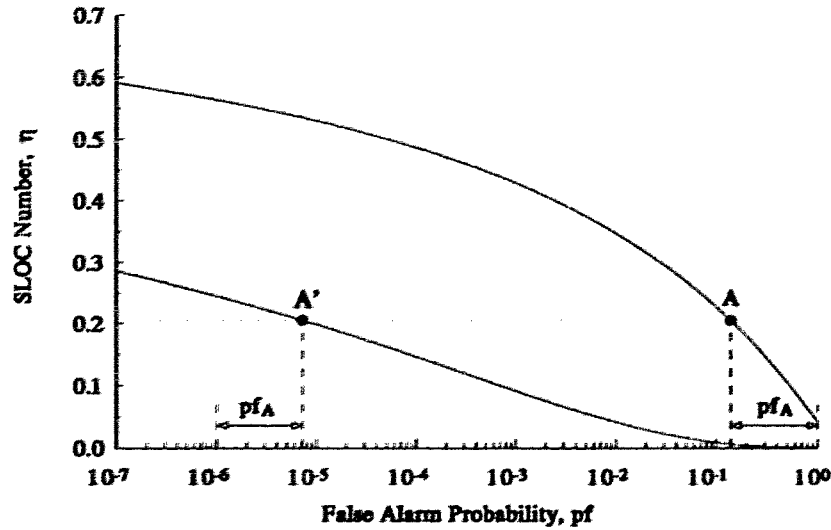
Figure 4.3: SLOC curves for two Marcum targets. The upper curve has SNR = 5dB, the lower 10dB.

(4.8) the resulting SLOC function is given by[1]

$$\eta_{Marcum}(\lambda) = \frac{\exp\left[-\frac{1}{2}A^2\right] I_0(A\lambda)}{Q(A,\lambda)} \qquad (4.14)$$

where λ is the threshold, $Q(A,\lambda)$ is the Marcum Q function defined in (2.55), and $I_0(x)$ is the zeroth-order modified Bessel function of the first kind.

In figure 4.3 the SLOC functions are plotted for a marcum target with SNR = 5 dB (the upper curve) and 10 dB (the lower curve) [2]. The optimal operating points for the two tests (A and A') are given for an overall false alarm probability $pf_T = pf_A pf_{A'} = 10^{-6}$. Since the false alarm probability is plotted on a logarithmic scale, the distance along the axes (as shown in the figure) must add to pf_T.

It may seem surprising to see that the lower SNR test has an operating point other than

---

[1]We drop the subscripts on λ and η for now, since we are only considering a single stage here.

[2]Note the two-stage system is detecting the same target; however the individual sensors or tests are operating at different signal-to-noise ratios, possibly due to longer observation times allowed by different cell processing rates.)
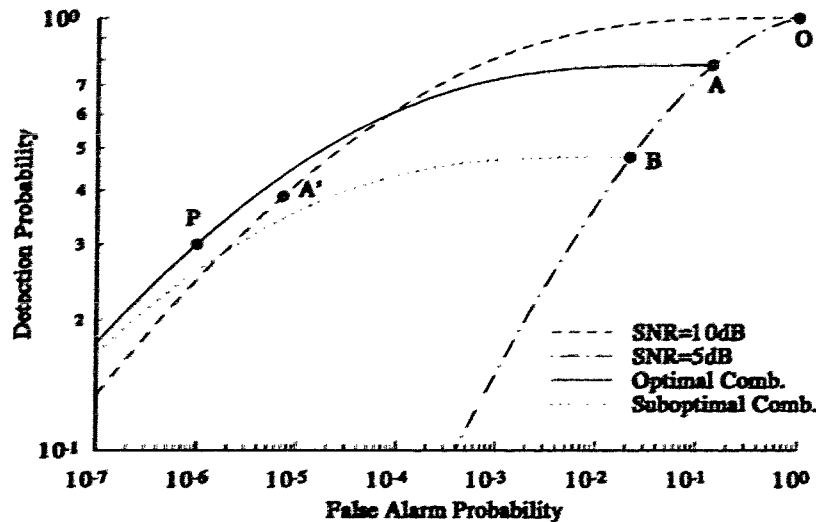
Figure 4.4: Log-ROC curves for figure 4.3

$pd = pf = 1$. This means that - for the Marcum model at least - it is possible for a test with a higher SNR to gain performance by using a lower SNR test to provide some selectivity.

Figure 4.4 illustrates the optimality on a log-ROC diagram. To understand figure 4.4, begin with the curves marked "SNR = 5dB" and "SNR = 10dB". Since Birdsall's insight also holds on log-ROC curves, it is no surprise that the curve for the 10dB test is always above that for the 5dB test, except for the point O, where $pf = pd = 1$.

To generate the curve marked "Optimal Comb.", the curve for the 10dB target is merely slid down the 5dB curve; this moves point O to point A (the same point A as for figure 4.3), and point A' is translated to point P. The figure shows that for a false alarm probability of 10^{-6}, the resulting detection probability (point P) is indeed larger than the detection probability for the 10dB test alone. In fact, if the 5 dB test is operated anywhere between point O and A, the resulting detection probability for $pf = 10^{-6}$ will lie between point P and the resulting pd for the 10dB test alone. Furthermore, if the operating point for the 5dB test is moved down, say to point B, the resulting combined detection probability will also lie below point P for $pf = 10^{-6}$.
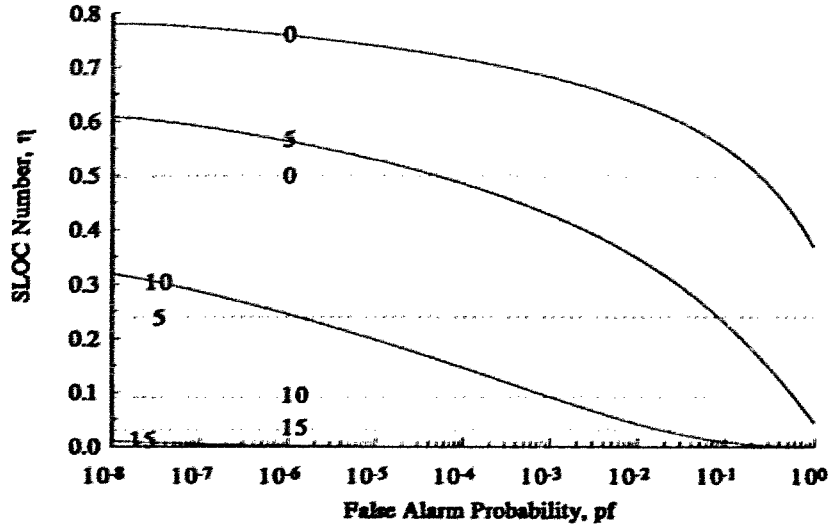
Figure 4.5: SLOC functions for Marcum (solid lines) and Rayleigh (dashed lines) targets. The number on the curves indicates the signal-to-noise ratio in dB.

We have just shown that for a Marcum target, a test with a large SNR could improve performance by co-operating with a test with a lower SNR. This is not so for a Rayleigh target. The easiest way to evaluate the SLOC for Rayleigh targets is to use (2.67) in (4.9). Then, for a Rayleigh target with SNR $= \varrho$

$$\eta_{Rayleigh} = \frac{1}{1+\varrho} \tag{4.15}$$

The Rayleigh SLOC function is a constant independent of the operating point and decreasing with increased SNR. Therefore if two different SNR stages are cascaded together as in figure 3.1, condition (4.9) can never be met. The optimum solution is found by considering the boundary conditions, whereby it is seen that the optimal detection probability is found by setting the threshold of the lower SNR test to zero (effectively removing that test altogether).

The constant SLOC of the Rayleigh target appears to be the exception, rather than the rule, as shown by the SLOC functions presented in chapter 5. However, it is interesting to note that studies on sensor fusion cited in the literature often utilize Rayleigh targets (due to the simple functional form of pd); one questions the generality of some of the results in

the light of the Rayleigh model's behaviour described here.

In figure 4.5 SLOC functions are plotted for Rayleigh and Marcum targets for various signal-to-noise ratios. Note that for a given pf and target model, the SLOC number is a positive, decreasing function of the SNR. Furthermore, for the Rayleigh and Marcum targets, the SLOC is a nonincreasing function of pf; however, this is not a universal trait of SLOC curves. (Some exceptions are given in chapter 5.)

In this section there have been no constraints applied to the processing rate of either test. We now turn our attention to the practical scenario where the second test is slower, but more capable than the first.

4.3 Effect of Rate Constraints

In section 4.2 we saw that the optimal solution for an unconstrained two-stage system is derived by finding the operating points that match the SLOC values. Of course, it would be impossible to operate under such a criterion in practice since the SLOC values are strongly dependent on the signal-to-noise ratios which are unknown.

We now consider the effect of introducing a rate constraint on the test with the higher SNR. Such a model is extremely common; tests that take more time to operate tend to perform better.

Let us begin by examining the deleterious effects of setting the threshold too low, so that randomization occurs, as discussed in section 3.2. Recall that randomization increases the effective false alarm probability while decreasing the effective detection probability. Therefore we expect the effective SLOC to be negative when randomization occurs. When randomization occurs, the effective probabilities of detection and false alarm are related:

$$\pi_1 pd_{eff} + \pi_0 pf_{eff} = r \tag{4.16}$$

Solving for pd_{eff} as a function of pf_{eff}:

$$pd_{eff} = \frac{1}{\pi_1}[r - \pi_0 pf_{eff}] \tag{4.17}$$

Then the effective SLOC in randomization is

$$\eta_{Rand} = \frac{d \ln pd_{eff}}{d \ln pf_{eff}} \tag{4.18}$$

$$= \left(\frac{d\ln pd_{eff}}{dpd_{eff}} \middle/ \frac{d\ln pf_{eff}}{dpf_{eff}} \right) \frac{dpd_{eff}}{dpf_{eff}} \tag{4.19}$$

$$= \frac{pf_{eff}}{pd_{eff}} \left(-\frac{\pi_0}{\pi_1} \right) \tag{4.20}$$

$$= -\frac{\pi_0 pf}{\pi_1 pd} \tag{4.21}$$

where the last equality results from comparing (3.7) and (3.9). The negative SLOC means that a test should never operate in randomization; the test can produce both higher pd and lower pf by increasing its threshold.

Now let us consider again (4.13), reprinted here for convenience:

$$\Delta \ln(pd_1 pd_2) = (\eta_1 - \eta_2) d\ln pf_1 \tag{4.22}$$

First, note that if test one is in randomization, $\eta_1 < 0$, and therefore an increase in pf_1 will indeed decrease the overall pd. We have shown that we never want to operate any stage such that it forces randomization. In our subsequent discussion, we will use the SLOC as defined in (4.10) and not worry about the effective SLOC since we never want to operate in randomization.

Now, since the first test has a lower SNR, we expect that it will have the larger SLOC. We say a two-stage system is *uniformly properly ordered* if

$$\eta_1 \geq \eta_2 \qquad \forall (pf_1, pf_2) \tag{4.23}$$

For example, a system consisting of two stages each detecting a Rayleigh target is a properly ordered sequence if the first test has a lower SNR than the second. (We consider a less strict condition on ordering in section 4.3.1)

If a system is uniformly properly ordered, (4.23) in (4.22) shows that $\Delta \ln(pd_1 pd_2)$ is an increasing function of $d\ln pf_1$. Equivalently, pd_T is a monotonically increasing function of pd_1. If there is no rate constraint, the optimal pd_T for a given pf_T is then found for a uniformly properly ordered detection sequence by setting the first threshold to zero (so that $pd_1 = pf_1 = 1$).

However, when the second stage is limited by resource constraints, the problem becomes

$$pd_T = \max_{\delta_1, \delta_2 \in \mathcal{D}} pd_1(\delta_1) pd_2(\delta_2) \tag{4.24}$$

where $\mathcal{D} = \{\delta_1, \delta_2 : \pi_1 pd_1(\delta_1) + \pi_0 pf_1(\delta_1) \leq r_2, pf_1(\delta_1) pf_2(\delta_2) = pf_T\}$

That is, we want to find the maximum pd_T for a given overall pf_T, subject to the rate out of the first stage not exceeding the input rate of the next stage. If the rate out of the first stage is too high, randomization will bring it down to the maximum rate as described in section 3.2.

Note that $pf_2(\delta_2)$ is in fact a function of $pf_1(\delta_1)$ through the constraint on pf_T. Therefore once δ_1 (and therefore $pf_1(\delta_1)$ and $pd_1(\delta_1)$) is set, so is $pf_2(\delta_2)$ and therefore $pd_2(\delta_2)$. Furthermore note that δ_1 is constrained only by the processing rate of the second stage and not by δ_2 at all. Therefore, provided (4.23) holds, we can re-write (4.24) as

$$\max_{\delta_1 \in \mathcal{D}_1} [pd_1(\delta_1) \max_{\delta_2 \in \mathcal{D}_2} pd_2(\delta_2)] \tag{4.25}$$

where $\mathcal{D}_1 = \{\delta_1 : \pi_1 pd_1(\delta_1) + \pi_0 pf_1(\delta_1) \le r_2\}$

and $\mathcal{D}_2 = \left\{\delta_2 : pf_2(\delta_2) = \dfrac{pf_T}{pf_1(\delta_1)}\right\}$

The decoupling of (4.24) into two single-stage maximization problems in (4.25) may appear familiar to those readers with knowledge of Dynamic Programming.

From chapter three, we already know how to find δ_1: it is merely a likelihood ratio test with threshold set to match the output rate to the input rate of the second stage. From chapter two, δ_2 must be the Neyman-Pearson test that gives the desired false alarm rate.

In summary, we have seen that for a uniformly properly ordered test, the overall detection probability for a given false alarm probability is maximized by using a test operating under rate-constraint principles in the first stage and then setting the second stage's threshold to choose the appropriate operating point on the overall ROC curve.

In fact, this argument can be extended to show that under any detection criterion (Bayes, etc.) for the entire system, the first stage should operate under the rate constraint criterion if the system is uniformly properly ordered. Consider the shape of the overall ROC curve. We have argued that in order to lie on the overall ROC curve, the first stage is determined by the rate constraint. If the operating point is at (pd_1, pf_1), then the ROC curve has a second derivative

$$\frac{\partial^2(pd_1 pd_2)}{\partial(pf_1 pf_2)^2} = \frac{pd_1}{pf_1} \frac{\partial^2 pd_2}{\partial pf_2^2} \le 0 \tag{4.26}$$

Thus the overall ROC curve is concave, and Birdsall's insight holds. Therefore, a detection system operating under any sensible criterion emphasizing high pd and low pf is optimized by matching the rate out of the first stage to the processing rate of the second stage. The

threshold for the second stage is then determined by the criterion for the overall system. This somewhat surprising result is only strictly true for properly ordered detection sequences. In the next section we will consider test sequences that are properly ordered only in a specific region of operation.

4.3.1 Compression Factors: Locally Properly Ordered Sequences

In the above we considered only properly ordered sequences. In general, it will be impossible to claim a sequence of tests are uniformly properly ordered (unless the second test is detecting a Rayleigh-fading target in Gaussian noise, so that its SLOC is constant and smaller than the minimum SLOC for the first stage.) Instead, we will have to be satisfied with claiming that a sequence is properly ordered over a restricted range of operating conditions. Then we say that the sequence is *locally properly ordered.*

For example, consider again figure 4.3. Say that the processing rate of the second stage is such that the first test can operate at or to the left of point A. (That is, the false alarm rate associated with the rate constraint is $pf_1 \approx .14$, and the SLOC is .205.) Then the sequence will be locally properly ordered for any tests requiring the second stage to operate with a SLOC less than .205, or equivalently for $pf_2 > 7.2(10)^{-6}$. Put another way, if the rate constraint on the second stage limits the first stage to operate at point A or slower, the rate-constraint criterion is guaranteed to be optimal for the first stage for overall false alarm probability greater or equal to 10^{-6}. We call the smallest value of pf_2 for a given operating point for the first test the *compression factor* for the sequence. From figure 4.3, we see that a slower second stage would give a larger compression factor; larger in the sense that the rate constraint is guaranteed optimal over a larger range of overall false alarm probabilities. For example, if the second stage has a processing rate 10^{-2} that of the first (assuming no targets), then the compression factor for the second test is larger than 10^{-7}, and the rate constraint criterion is optimal for tests with pf_T greater than at least 10^{-9}.

The SLOC curves also allow us to compare different tests with different statistics. From figure 4.5, we see that a 5 dB Rayleigh target yields a constant SLOC value of 0.24, and that a 10 dB Marcum target for the second stage would yield additional compression of 10^{-6} before the rate constraint is not optimal.

The SLOC curves provide insight into the merit of various tests. They do require known and tractable statistics to be useful (like ROC curves.) However, if experimental ROC curves

can be generated, then the associated SLOCS can be graphically determined. Even with rough SLOC estimates, intelligent guesses of suitable compression factors for the tests can be made. Furthermore, the individual test stages need not be identically distributed. For example, if a microwave radar is cueing a lidar, it is possible that a given target might be Rayleigh distributed when viewed by the radar, but appear as a Marcum target to the lidar. By examining the SLOC curves, we are still able to understand how the two subsystems will interact.

What if a test sequence is not properly ordered? Then from section 4.2, the optimal detection would be found by matching the SLOC numbers of the two stages; however, that is clearly not practical. There will be some loss in operating the first stage under the rate-constraint criterion, but at least the criterion can be practically applied.

In the next section, we will extend the two-stage rate-constraint concepts to consider multi-stage systems. By use of dynamic programming arguments, we show that the results of this section hold for multi-stage systems. For any overall detection criterion and a properly ordered system, the rate-constraint criterion is optimal for all stages but last.

4.4   Extension to More Stages

Often detection sequences consist of more than two stages. For example, radar systems often use several processing and decision-making stages [18]. Similarly, clinical diagnosticians may use several tests before declaring a positive or negative diagnoses [12].

In figure 4.6 the model for an N-stage detection system is given. In keeping with Dynamic Programming models, each stage has an input state S_k, an output state $\tilde{S}_k$, a decision function δ_k, and a return function r_k [38][3].

Let the input state into stage k be

$$S_k = \begin{cases} (\pi_0, \pi_1) & k = 1 \\ \left(\pi_0 \prod_{i=1}^{k-1} pf_i, \pi_1 \prod_{i=1}^{k-1} pd_i\right) = (\pi_0^k, \pi_1^k) & 1 < k \leq N \end{cases} \qquad (4.27)$$

where $pf_i = pf(\delta_i)$ and $pd_i = pd(\delta_i)$, and δ_i is a likelihood ratio test (which is determined by choosing λ_i, or by the t_i for the equivalent threshold test). Then the output state of

---

[3]Note that we have reversed the ordering of the indices for the stages from the traditional in Dynamic Programming.

stage k is given by

$$\tilde{S}_k = (\pi_0^k p f_k(\delta_k), \pi_1^k p d_k(\delta_k)) \tag{4.28}$$

The state functions are necessary to determine the appropriate rate-constraints at each stage.

The return function out of stage k is given by

$$r_k(S_k, \delta_k) = \max_{\delta_k \in \mathcal{D}_k} p d_k(\delta_k) \tag{4.29}$$

where $\mathcal{D}_k$ is the set of all decision rules meeting the rate constraint:

$$\mathcal{D}_k = \begin{cases} \{\delta_k : \pi_0^k p f_k(\delta_k) + \pi_1^k p d_k(\delta_k) \leq r_{k+1}\} & 1 \leq k < N \\ \{\delta_k : \frac{\pi_0^k}{\pi_0} p f_k(\delta_k) = p f_T\} & k = N \end{cases} \tag{4.30}$$

Note that the definition of $\mathcal{D}_k$ doesn't preclude the use of tests that cause randomization due to overload; it merely means that the effects of such randomization must be included in the optimization. Since randomization always causes the SLOC to become negative (forcing a southeast movement in the ROC curve) we know that randomization will never be present in an optimal system.

The definition for $\mathcal{D}_N$ is for the Neyman-Pearson formulation. As with the two-stage system of section 4.3, the final stage can operate under any criterion subject to Birdsall's Insight; the first $N - 1$ thresholds will not change.

Thus we see that the return at each stage is a function of the input state (through the constraint) and the decision function, δ_k. Note that each of the constraints can be expressed in terms of local variables (i.e. variables for that stage only.) This locality suggests that local control in a practical system is feasible.

The goal is to find

$$g = \max_{\delta_1, \dots, \delta_N} \prod_{k=1}^{N} r_k(S_k, \delta_k) \tag{4.31}$$

Being a product of the r_i, g is separable: [38]

$$g[p d_1(S_1, \delta_1), p d_2(S_2, \delta_2), \dots, p d_N(S_N, \delta_N)] = g_1[p d_1(S_1, \delta_1), g_2[p d_2(S_2, \delta_2), \dots, p d_N(S_N, \delta_N)]] \tag{4.32}$$

Now if the system is properly ordered (at least in the region of interest) for the various tests, then g is a monotonically non-decreasing function of g_1. Then maximizing $p d_T$ is equivalent to maximizing $p d_1(S_1, \delta_1)$.
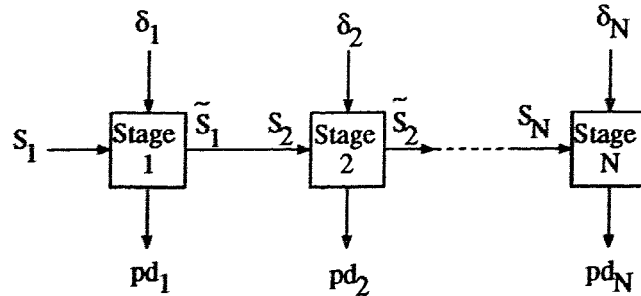
Figure 4.6: A Multi-stage Detection System.

Given that g is separable and that g is a monotonically nondecreasing function of g_1, we say that the problem is decomposable [39]. This means that it is possible to solve the N stage problem one stage at a time. Let

$$h_1(S_1) = \max_{\delta_1,\ldots,\delta_N} [pd_1(S_1,\delta_1)pd_2(S_2,\delta_2)\cdots pd_N(S_N,\delta_N)] \tag{4.33}$$

where $S_{k+1} = \tilde{S}_k = b_k(S_k,\delta_k)$. Given the monotonicity of g on g_1, we can write

$$h_1(S_1) = \max_{\delta_1}[pd_1(S_1,\delta_1) \max_{\delta_2,\ldots,\delta_N} pd_2(S_2,\delta_2)\cdots pd_N(S_N,\delta_N)] \tag{4.34}$$

and

$$h_2(S_2) = \max_{\delta_2,\ldots,\delta_N} [pd_2(S_2,\delta_2)\cdots pd_N(S_N,\delta_N)] \tag{4.35}$$

Therefore,

$$g_1 = \max_{\delta_1} pd_1(S_1,\delta_1)h_2(S_2) \tag{4.36}$$

but $S_2 = b_1(S_1,\delta_1)$ is a function of S_1 and δ_1. Therefore

$$g_1 = \max_{\delta_1} H_1(S_1,\delta_1) \tag{4.37}$$

Equation (4.37) states what we expect: to maximize the overall detection probability, we must first maximize pd_1 subject to its rate constraint. Clearly the argument is recursive, and *provided the sequence is properly ordered*, the optimal threshold for each stage but the last is determined by the rate-constraint at that local stage and not the overall detection criterion. The argument on the concavity for the overall system given in section 4.3 again holds, and therefore the rate-constraint thresholds for the first $N-1$ stages will be optimal under any overall detection criterion. The threshold for the final stage must be chosen in order to optimize under whatever criterion the entire system is working under.

An interesting point to note is that the detection performance of any N-stage system will always be improved, if another stage is added such that it makes a new properly ordered $(N+1)$-stage system. We know this, for if the new stage didn't improve the detection, its threshold would be zero, effectively removing the stage. Therefore a given detection sequence can be improved if intermediate stages are added, at the expense of increased latency in decision time due to the added stage.

4.5 Conclusions

In chapter four we have investigated a cueing structure whereby faster tests cue slower, but better tests. The structure is equivalent to a Fusion system operating under an "AND" rule, and provides the same graceful degradation as many fusion systems. However, there is no requirement for the individual tests to provide synchronous outputs.

The SLOC number is an important relational parameter that provides the key to the optimal detection strategy for a cued system. For unconstrained detection sequences, the optimal operating point for the overall detection probability is found by matching the SLOC number of the individual detection stages. A test with a lower SNR can improve the detection performance of a higher SNR test, except for the case of Rayleigh targets in Gaussian noise (which have constant SLOC functions). The SLOC number allows comparison of any two detectors, provided the Receiver Operating Characteristic (ROC) curves can be determined, experimentally or analytically. The tests needn't be identically distributed.

When a resource constrained detection sequence is properly ordered, so that each stage is faster and has a larger SLOC function than the stage succeeding it, the optimal strategy is to operate the earlier stage under the rate-constraint criterion. This strategy is optimal for any overall detection criterion that is sensible in terms of Birdsall's insight. A local control structure is possible for a properly ordered cueing system; a given stage need not worry about previous or downstream processors other than the succeeding one.

The SLOC function provides important information for the understanding of how multi-stage detection systems work. Chapter five is devoted to a study of the SLOC functions for a number of practical radar scenarios.

Chapter 5

Applicability to the Radar Problem

5.1 Introduction

In chapter four the SLOC parameter emerged as a central theme to the understanding of how cascaded detection systems work. We showed that the rate-constraint criterion should be applied at every given stage in a detection chain, provided that the SLOC function of that stage is larger than that of subsequent stages. Thus to understand the role of rate-constraint theory in radar systems, we must understand the SLOC functions for radar problems. In this chapter the log-ROC and SLOC curves for several classical and nonclassical radar models are presented. During the course of the discussion, we feel compelled to comment on some of the apparent confusion regarding certain models.

There are a couple of points to note regarding the content of this chapter. First, it must be stressed that that in practical deployments, the statistics required for the calculations made in this chapter are usually unavailable. Furthermore, the models are just that, models; they don't even necessarily provide bounds on the behaviours that may be observed in real radar systems. Thus our intent is to investigate the validity of the theory in terms of existing radar models, and not to prove the optimality of rate-constraint in all circumstances.

Second, it will become apparent to the reader that, although the chapter does include a number of graphs, the coverage is by no means exhaustive. The results presented are intended to be representative, and to yield understanding into applying rate-constraint

principles. Most of the graphs are relatively straightforward to generate. We leave it to the interested reader to calculate any required.

5.1.1 Chapter Outline

In section 5.2, the classical target models of Marcum and Swerling are briefly reviewed. The Rician target model is shown to be a more valid representation for a target consisting of a large scatterer plus a large number of smaller scatterers than the Swerling III and IV models normally used. It is also shown that the behaviour of the Rician model is intermediate to that of the Marcum and Swerling I/II models. Another target model, consisting of two specular returns is considered. The moment generating function is used as the tool to calculate the detection and false alarm probabilities in this chapter. Section 5.2.1 gives the moment generating functions for the models discussed.

In section 5.2.2, the log-ROC and SLOC curves are given for the various target models, based on a single return. The integration of multiple returns to develop the test statistic is discussed in section 5.2.3, and numerous curves are given.

In section 5.3, we consider an application of rate-constraint systems to a signal processing problem. Specifically, we investigate a two-stage detection system where the first stage performs noncoherent integration, and then cues only those cells likely to contain targets to the second stage which then performs coherent integration.

The relatively new technique of [16] for calculating the detection probability for noncoherent integration is briefly reviewed in section 5.4 and then applied to investigate the effect of K-Distributed clutter on log-ROC's and SLOC's. New results for noncoherent integration of Marcum and Rayleigh targets in K-distributed clutter are given.

A brief discussion of the roles of normalization (Constant False Alarm Rate) techniques in the rate-constraint framework is given in section 5.5. The chapter's conclusions are given in section 5.6.

5.2   Target Models

The common radar target return models are a result of the pioneering work of Marcum [33] and Swerling [40]. As with all models discussed in this section, Marcum's and Swerling's

models assume that the interfering noise is Gaussian distributed[1]; we consider other types of clutter in section 5.4. The detector (combination rule for multiple returns) is assumed to use a square-law.

For some of the models in this chapter, no close-form solution exists. Others can be expressed only in terms of functions that are not common; therefore we will not concern ourselves with the explicit expressions. Instead, we give the Moment Generating Functions for the various models in a pair of tables.

Marcum's Model We saw Marcum's model in section 2.5.1 [33]; it consists of a constant amplitude random-phase return immersed in Gaussian noise. Marcum targets are sometimes called Swerling 0 targets; they result from returns from very large, stable objects such as a large sphere.

Swerling Models The Rayleigh target was discussed in section 2.5.2. When a series of Rayleigh-fading returns are added noncoherently, the resulting models are called Swerling I if the returns remain fully correlated from pulse-to-pulse, and Swerling II if they become fully de-correlated between pulses [40].

The Rayleigh model is valid when a number of independently identically distributed echoes contribute to the return. Then, by the Central Limit Theorem, the in-phase and quadrature components can be assumed to be Gaussian. While an infinite number of echoes is theoretically required to invoke the Central Limit Theorem, in practice, 6 or more echoes is sufficient [41]. The Rayleigh distribution for the amplitude (envelope) then results from the quadratic combination of the two Gaussian components.

Swerling also introduced another model, an approximation to the combination of a constant target plus a Rayleigh component, the so-called *One-Dominant* model. When a One-Dominant target is noncoherently integrated, a Swerling III model results if the returns are pulse-to-pulse correlated, and a Swerling IV if the returns are fully de-correlated. The model is based on a Chi-Squared density function with four degrees of freedom (note that the Rayleigh is also Chi-Squared, but with two degrees of freedom.) Swerling made no attempt

---

[1]More properly, the in-phase and quadrature returns are i.i.d. Gaussian distributed. In the literature, such noise is often called 'Rayleigh' noise, since the envelope is Rayleigh distributed; however, it is often the square of the envelope which is of interest. We will use the term Gaussian noise, since it is valid whether a linear-law or square-law combiner is used. We will, however, continue to call a target with Gaussian amplitude a Rayleigh-fading target, regardless of the combining law used.

to physically justify the One-Dominant model; he merely suggested that it might apply to aircraft. Since 1957, the One-Dominant model has appeared often in the literature, yet there it has no physical meaning. In 1967, Scholefield wrote regarding the One-Dominant model:

> Standard radar textbooks ... tend to be uninformative about the precise application of (the One-Dominant Model), and experience shows that misunderstandings arise [42].

The problem is that there are only two degrees of freedom in a radar return. The One-Dominant model requires a receiver to "resolve the input signal into four orthogonal components, which it clearly cannot do"[42].

Unfortunately, contemporary radar books continue to be as vague as the texts mentioned by Scholefield in 1967. In fact, of several books written since 1987 that discuss Swerling's models [7],[6],[43], [44], only Nathanson [18] mentions the difficulties of the One-Dominant model.

In this thesis, we present only limited results for the One-Dominant model. Our reason for presenting any at all is that the One-Dominant model has become a yard-stick, albeit a contrived one.

Rician Targets The proper model for a return resulting from a single dominant scatterer and Gaussian noise is a Rician target[2]. The Rician model, unlike the One-Dominant, allows us to vary the specular-to-diffuse ratio (SDR) in the return. As the SDR varies from $-\infty$ to ∞ (dB), the Rician model goes from the Marcum to the Rayleigh model. Therefore, the Rician model produces a family of intermediate distributions between the Marcum and Rayleigh and includes both.

Two-Tone Targets We have seen models based on a single dominant echo, an infinite number of independent and identically distributed (i.i.d.) echoes, and a mixture of the two. A logical question to ask is whether an intermediate number of i.i.d. echoes provides an intermediate result? The answer is not always. Jao and Elbaum have shown that a model based on two equal-amplitude random-phase returns (glints) provides an extreme case (in

---

[2]The return from the target is Rician distributed only if a linear-law combiner is used; however the target is still called a Rician target for a square-law.

the sense of required SNR for a given detection performance) [41]. Thus, by considering a target with two equal glints we can bracket the behaviour expected from radar models composed of any number of glints in Gaussian noise. We do not consider more than two glints, because as the number of glints increases, the target becomes closer to being Rayleigh distributed (provided the glints have equal amplitude.)

5.2.1 Moment Generating Functions

Recall that the SLOC function as given by (4.10) and (4.8) is

$$\eta = \lambda \frac{pf}{pd} = \frac{f_1(t)}{f_0(t)} \frac{pf(t)}{pd(t)} \tag{5.1}$$

In order to calculate the densities and probabilities of (5.1), we use moment generating functions. Specifically, the density of the return under H_1 is

$$f_1(t) = \mathcal{L}^{-1}\{G_1(s)\} \tag{5.2}$$

where $G_1(s)$ is the Laplace transform or the moment generating function for the density of the return given a target, and $\mathcal{L}^{-1}$ is the inverse Laplace Transform. The detection probability is found by integrating $f_1(t)$, which is accomplished by dividing $G_1(s)$ by s.

$$pd(t) = \mathcal{L}^{-1}\left\{\frac{G_1(s)}{s}\right\} \tag{5.3}$$

(The expressions for $f_0(t)$ and $pf(t)$ are similar.)

The moment generating functions for the targets discussed above are given in table 5.1 and are derived from similar tables of characteristic functions in [45] and [16]. To evaluate the inverse Laplace transforms in (5.2) and (5.3), we used the IMSL routine 'DINLAP', which employs a numerical method described by [46].

In the table, M is the number of returns noncoherently combined (with square-law combining), and ϱ is the single-pulse SNR.

For the Rician model, the signal-to-noise ratio is given by

$$\varrho = \tfrac{1}{2}a^2 + \sigma_d^2 \tag{5.4}$$

while the specular-to-diffuse ratio (SDR) is given by

$$\text{SDR} = \frac{\tfrac{1}{2}a^2}{\sigma_d^2} \tag{5.5}$$

Table 5.1: Moment Generating Functions for Common Target Models

Swerling Case	Moment Generating Function
Noise	$G_0(s) = \frac{1}{(1+s)^M}$
0	$G_1(s) = \frac{\exp\left(-M\varrho\left(\frac{s}{1+s}\right)\right)}{(1+s)^M}$
I	$G_1(s) = \frac{1}{(1+s)^{M-1}(1+s(1+M\varrho))}$
II	$G_1(s) = \frac{1}{(1+s(1+\varrho))^M}$
III	$G_1(s) = \frac{1}{(1+s)^{M-2}(1+s(1+\frac{1}{2}M\varrho))^2}$
IV	$G_1(s) = \frac{(1+s)^M}{(1+s(1+\frac{1}{2}\varrho))^{2M}}$
Single Pulse Rician	$G_1(s) = \frac{\exp\left(\frac{-a^2 s}{2(1+\sigma^2 s)}\right)}{1+\sigma^2 s}$
Two-Tone Target	$G_1(s) = \frac{\exp\left(-\frac{M(a_1^2+a_2^2)s}{2(1+s)}\right)\left[I_0\left(\frac{a_1 a_2 s}{1+s}\right)\right]^M}{(1+s)^M}$

where a is the amplitude of the specular (constant) return, σ_d^2 is the power in the diffuse (Gaussian) component of the return, and the noise power is normalized to unity. In $G_1(s)$ for the Rician case, the variance of the Gaussian components in the signal, $\sigma^2 = 1 + \sigma_d^2$ is the sum of the powers in the noise and the diffuse return. Note that in table 5.1, $G_1(s)$ for the Rician model is valid for $M = 1$ only. The integration of multiple returns for Rician targets is discussed in detail in section 5.2.3 and table 5.2.

The signal-to-noise ratio for the two-tone target is given by

$$\varrho = \tfrac{1}{2}(a_1^2 + a_2^2) \tag{5.6}$$

where a_1 and a_2 are the amplitudes of the two specular components (also called glints). It is known that the two-tone target has its most extreme behaviour (in the sense of required SNR)) when $a_1 = a_2$ [41]. Therefore, in this thesis, we consider only equal strength glints.

5.2.2 ROCS and SLOCS for Single Returns

In figure 5.1 we have plotted the log-ROC and SLOC curves for a single return with $\varrho = 12$ dB for Marcum, Swerling I and III, and Rician targets. For a single return Swerling I and II are equivalent, corresponding to a Rayleigh-fading target. Similarly the Swerling III and IV models are equivalent for a single return, corresponding to a Chi-squared distribution with four degrees of freedom.

From figure 5.1 we can conclude that the log-ROC and SLOC curves for a single return vary considerably for the various target models. As expected from (4.15), the Rayleigh target (Swerling I/II) has a constant SLOC. This means that any two stages operating under rate-constraint principles will be guaranteed to have optimum performance for a Rayleigh target if the second stage has a higher SNR than the first.

While the Rayleigh target has a flat SLOC, the Marcum target has the steepest. This is due to the nature of the Marcum target. For high false alarm probability, the threshold is set lower than the amplitude of the target tone and so the detection probability is high and the SLOC is low. As the false alarm decreases, the threshold rises to the point where it becomes comparable and then larger than the tone's amplitude, resulting in a sharp decrease in the detection probability and a large value for the SLOC.

The Swerling III/IV and Rician (SDR = 0 dB) targets give similar results, with log-ROC's and SLOC's somewhat between the Marcum and Rayleigh cases. In one of the few
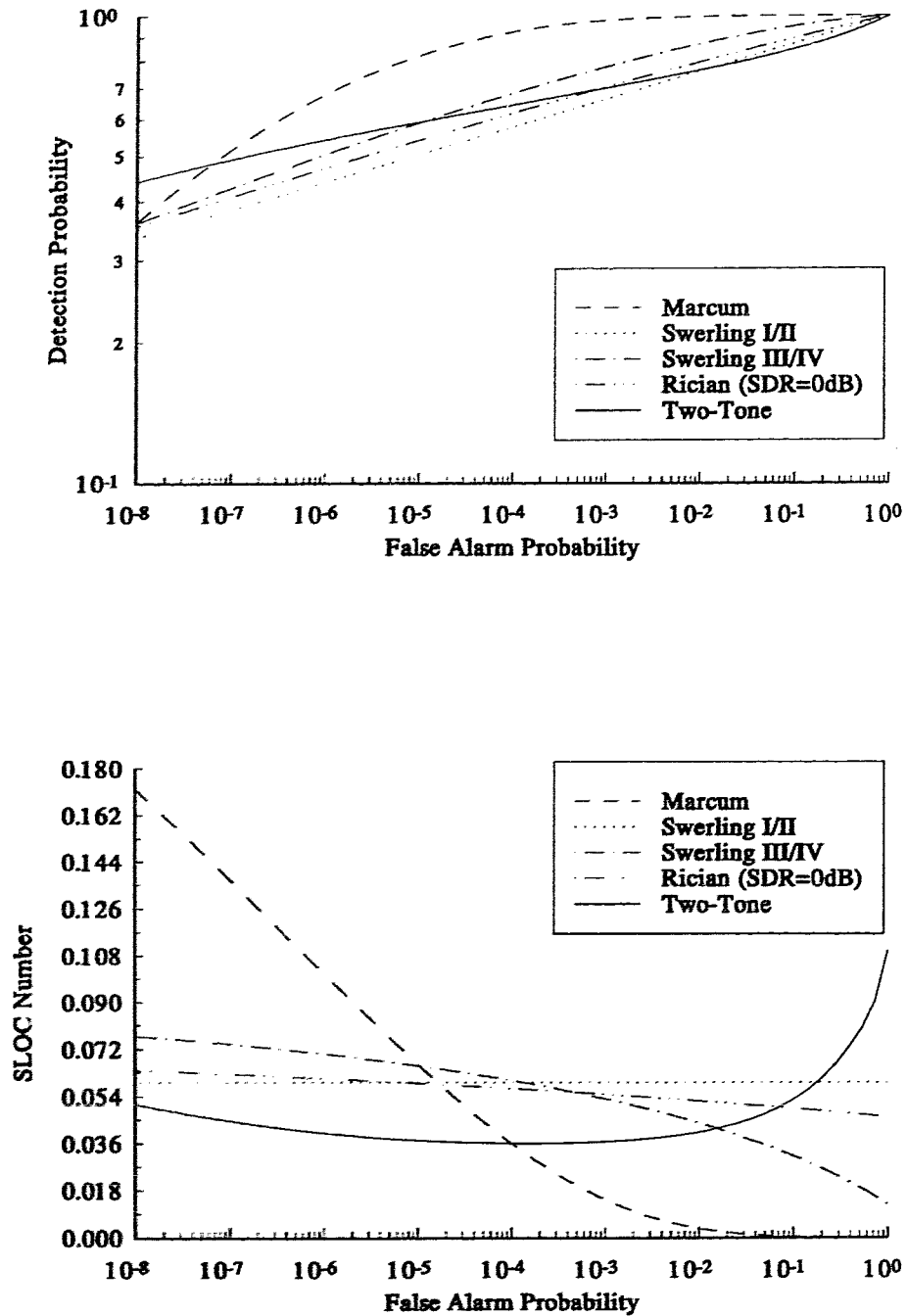
Figure 5.1: Log-ROC and SLOC curves for Swerling Class and Rician Targets. Curves are for a single return with $\varrho = 12$ dB.

papers discussing the Rician target, Scholefield showed that a Rician target and a Swerling III target with the same SNR "agree reasonably well except at high and low signal-to-noise ratios" [42]. Shapiro has also presented some results for the detection of Rician targets, in the context of lidar [47].

In figure 5.2, we illustrate the family of log-ROC's and SLOC's for single-pulse returns from Rician targets with the SDR as a parameter (Again $\rho = 12$ dB). As the SDR ranges from $-\infty$ to ∞ we generate a family of curves ranging from the Rayleigh to the Marcum target. Thus by considering the Marcum and Rayleigh models, we can generate the range of behaviours expected from the family of Rician targets.

Returning again to figure 5.1, we see that, unlike the Rician, the two-tone model is clearly not an intermediate between the Swerling I/II and Marcum models. Over all but the highest values of pf, the SLOC for the two-tone is the smallest of the families considered. Furthermore, the two-tone is our first example of a model that gives a non-monotonic SLOC; further examples are given later in the chapter.

The density functions for the two-tone model do not converge well due to the exponential increase of the modified Bessel function. (The extra s in the denominator makes the pd calculation possible.) Therefore, the SLOC for the two-tone model was numerically evaluated by fitting a cubic spline to the log-ROC data. The cubic spline was used since it guarantees continuity of the SLOC at the data points [48].

5.2.3 Effect of Integration

Thus far we have considered only single pulses ($M = 1$). We now consider the effect of integrating multiple pulses. Beginning with the Rayleigh target, we see that there are three different scenarios for integration depending on the level of correlation between consecutive pulses:

1. Coherent integration - this assumes pulse-to-pulse correlation and results in another Rayleigh model with ρ increased by a factor of M.

2. Noncoherent integration with pulse-to-pulse correlation - this is a Swerling I model.

3. Noncoherent integration with pulse-to-pulse decorrelation - this is a Swerling II model[3].

---

[3]Of course, we could add a fourth: coherent integration with the pulse-to-pulse decorrelation, but this would yield a Rayleigh target with no change.
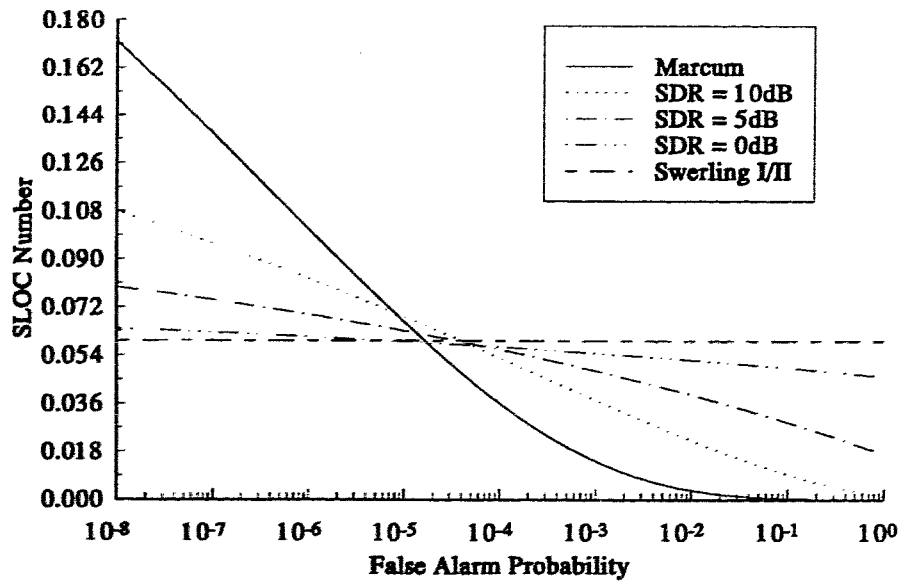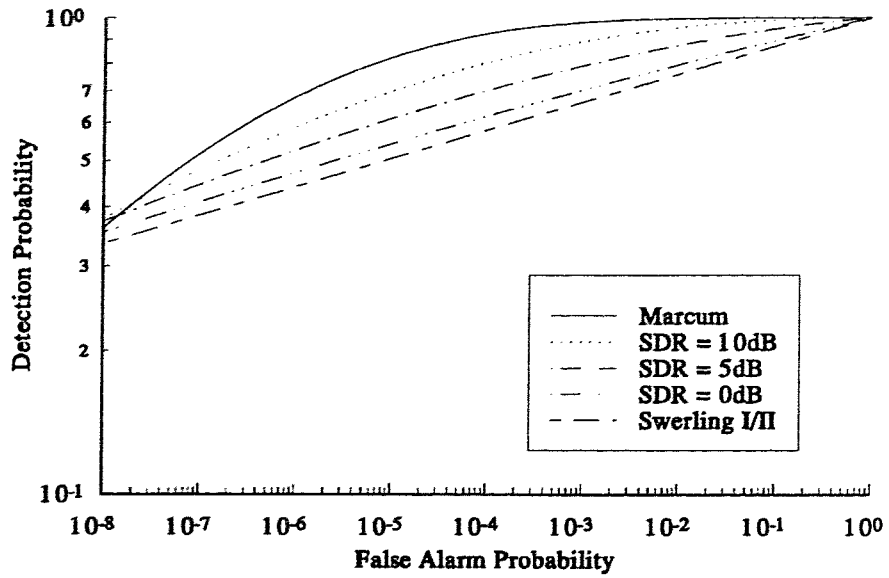
Figure 5.2: Log-ROC and SLOC curves for Rician targets with Specular to Diffuse Ratio (SDR) as a parameter. Curves are for a single return with $\varrho = 12$ dB.

Figure 5.3 shows the log-ROC and SLOC curves for the three types of integration of 16 looks at a Rayleigh target with $\rho = 0$ dB and 3 dB. For all integration techniques, the ROC curve is higher and the SLOC curve lower for the 3 dB than for the 0 dB target. The next thing to notice is that the Swerling II target actually achieves higher detection probability than coherent integration for false alarm probabilities greater than 10^{-6}. This well-known effect results from the probability of small values for the single-look SNR that will remain throughout the 16 returns for the coherent case, but will be averaged out by large returns in the uncorrelated (Swerling II) case. Thus at first glance, it may appear that one would always want to attempt to de-correlate the returns (for example by using frequency diversity in the radar waveform); however, coherent integration offers other gains such as the ability to use frequency discrimination (MTI or Pulse Doppler) techniques. Such methods depend on the radar returns remaining correlated pulse-to-pulse for several returns, and provide gains not evident in figure 5.3. Note that the Swerling I SLOC curve is at all points above that for the coherent integration. This suggests that a system could use a stage utilizing noncoherent integration to cue a stage utilizing coherent integration; we will return to this idea in section 5.3.

In figure 5.4, we have plotted the log-ROC and SLOC curves for coherent and noncoherent integration of 16 returns from a Marcum target with $\rho = 0$ dB and 3 dB. Note again the 3 dB target has a higher pd and a lower SLOC curve than the 0 dB target. The threshold effect for a tone target is evident for the coherent integration (which is now a tone with amplitude multiplied by 16). Note that, by definition, the Marcum target will have the same amplitude pulse-to-pulse and so there is only one type of noncoherent integration. From the SLOC curve of figure 5.4, we can see that for a Marcum target (at least with sufficiently high enough SNR), the SLOC for noncoherent integration will be greater than the SLOC for coherent integration.

The log-ROC and SLOC curves for coherent and noncoherent integration of 16 returns from a Two-Tone target with equal glints for $\rho = -3$ dB and 0 dB is given in figure 5.5. The SNR comparison is made between 0 and -3 dB because of convergence problems for 16 coherent integrations of a 3 dB target; as expected the higher SNR target has the larger pd and smaller SLOC. Again the SLOC for the noncoherent integration is larger than that for coherent integration when the false alarm probability is not greater than 10^{-1}.

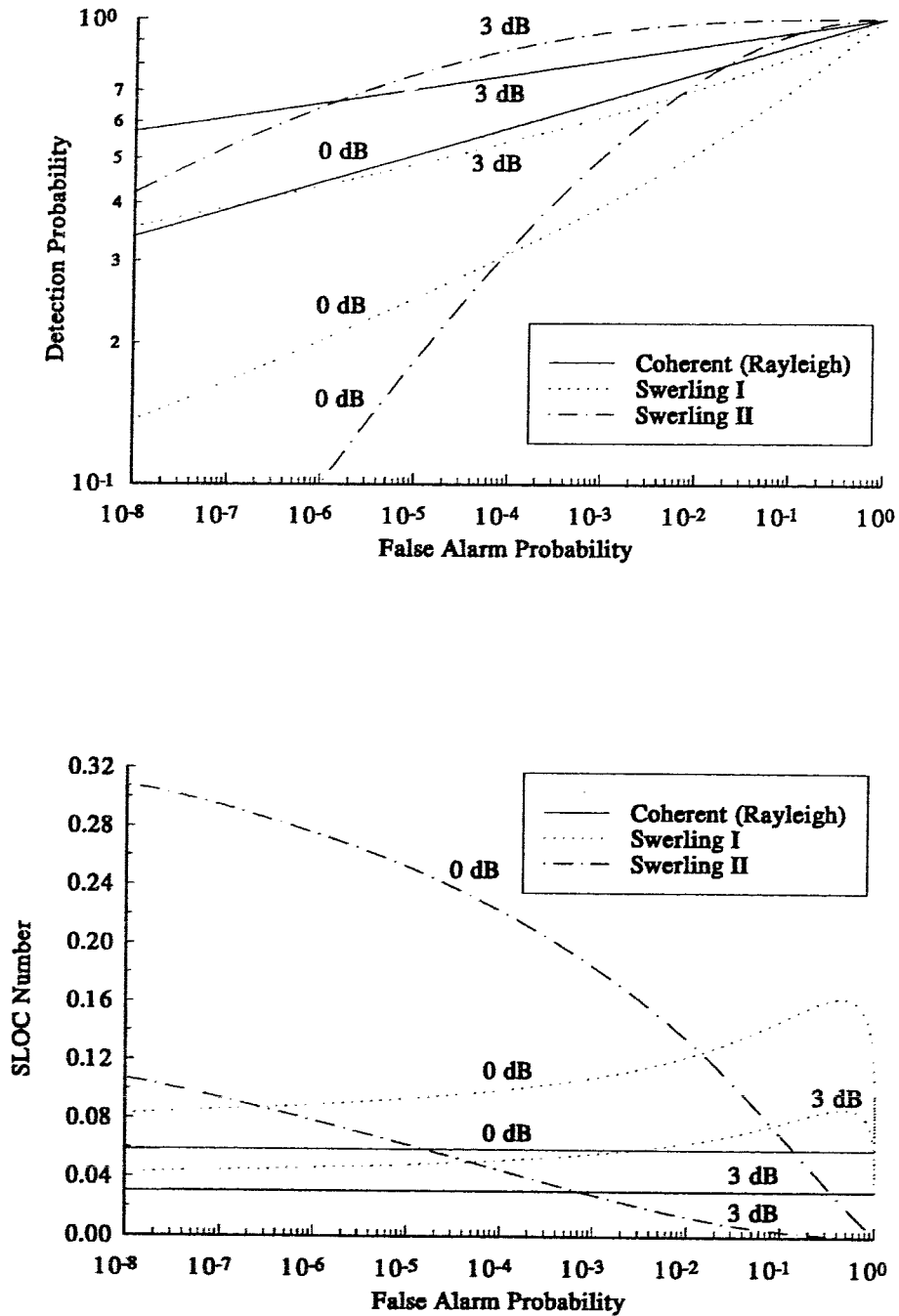Figure 5.3: Log-ROC and SLOC curves for integration of 16 returns from a Rayleigh fading target with single-pulse SNR = 0 dB and 3 dB
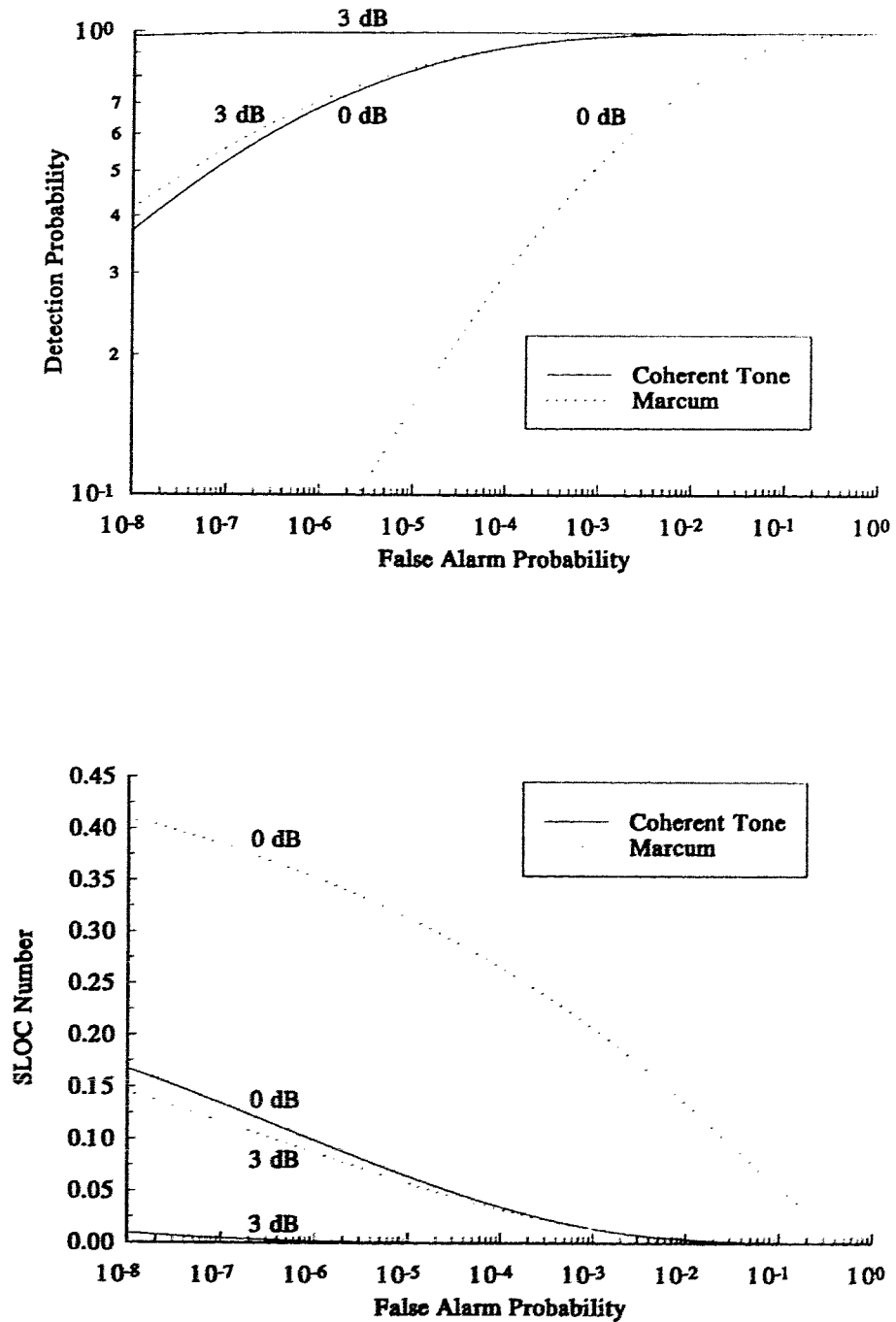
Figure 5.4: Log-ROC and SLOC curves for integration of 16 returns from a constant (Marcum) target with single-pulse SNR = 0 dB and 3 dB
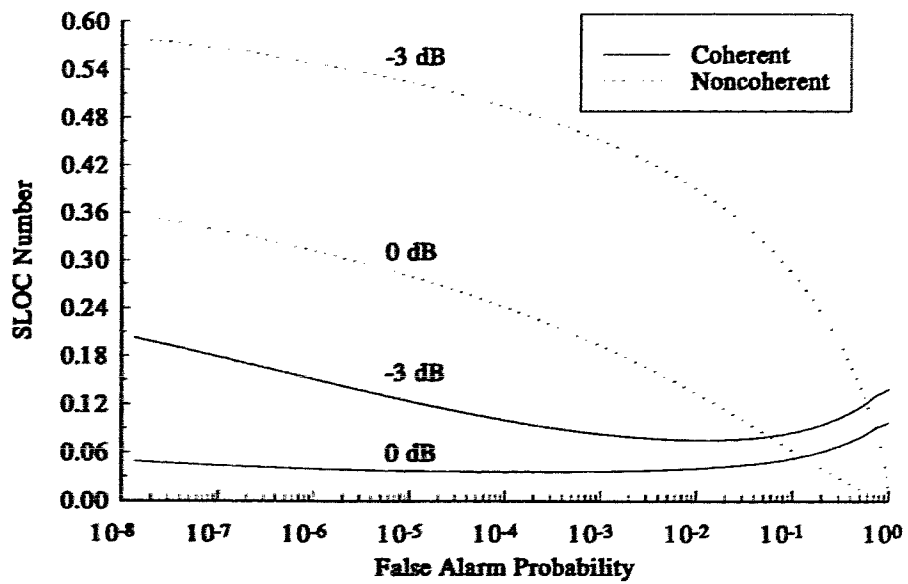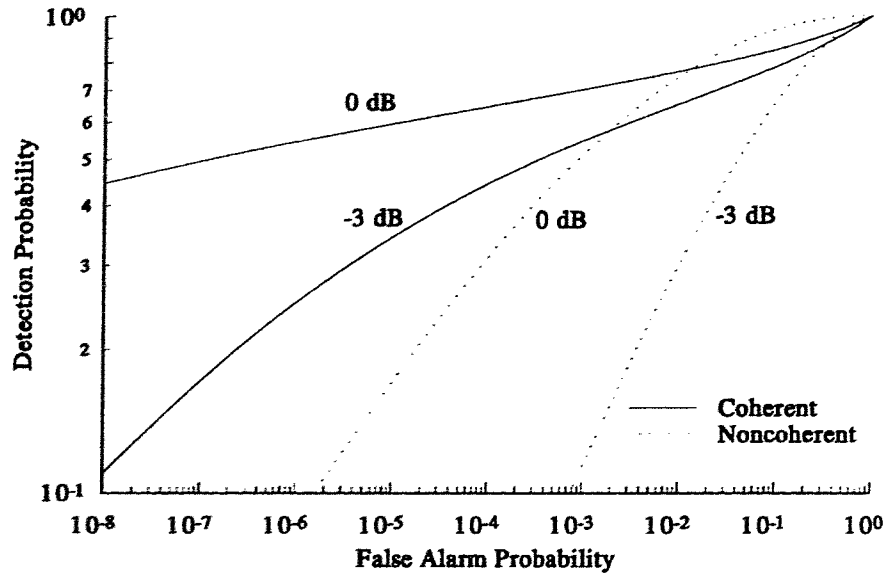
Figure 5.5: Log-ROC and SLOC curves for integration of 16 returns from a Two-Tone targets with equal glints with single-pulse SNR = -3 dB and 0 dB.

Rician Targets   For Rician targets, there are 4 integration models possible, namely:

1. Coherent integration, with diffuse component pulse-to-pulse correlated. The results will range from a single tone to a Rayleigh target (both with ϱ increased by M) as the SDR goes from ∞ to $-\infty$. See figure 5.6(a).

2. Coherent integration, with diffuse component decorrelated pulse-to-pulse. The results will range from a single tone with ϱ multiplied by M to a Rayleigh target with ϱ unchanged (i.e. no gain in SNR) as the SDR goes from ∞ to $-\infty$. See figure 5.6(b).

3. Noncoherent integration with diffuse component correlated from pulse-to-pulse. The results will range from a Marcum model to a Swerling I model as the SDR goes from ∞ to $-\infty$. See figure 5.7(a).

4. Noncoherent integration with diffuse component decorrelated from pulse-to-pulse. The results will range from a Marcum model to a Swerling II model as the SDR goes from ∞ to $-\infty$. See figure 5.7(b).

Note again the flexibility that the Rician model gives over the Swerling III/IV; for the latter we could define coherent integration only by increasing ϱ by M - however there is no physical significance. Also note that the performance curves for the Rician model are bounded by those for the Marcum and Swerling I/II models. Therefore, we need consider only the Marcum and Swerling models. This justifies our use of only a single value for ϱ when considering Rician targets; we are still guaranteed that the pd curve will become higher and the SLOC lower when ϱ increases.

For all of the target models considered, the SLOC has been lower for increased SNR. From a rate-constraint perspective, this is good news: slower systems are expected to have higher SNR, and we want the slower systems to have higher SLOC functions.

Furthermore, all the models considered have at least some region where noncoherent integration produces a larger SLOC than coherent integration. In the next section, we consider a two-stage detection scheme where the first stage implements noncoherent integration, and cues the second stage, which uses coherent integration.

Figure 5.6: Log-ROC curves for coherent integration of Rician targets with diffuse component (a) fully correlated and (b) fully decorrelated pulse-to-pulse. Curves are for a 16 returns with $\varrho = 0$ dB.
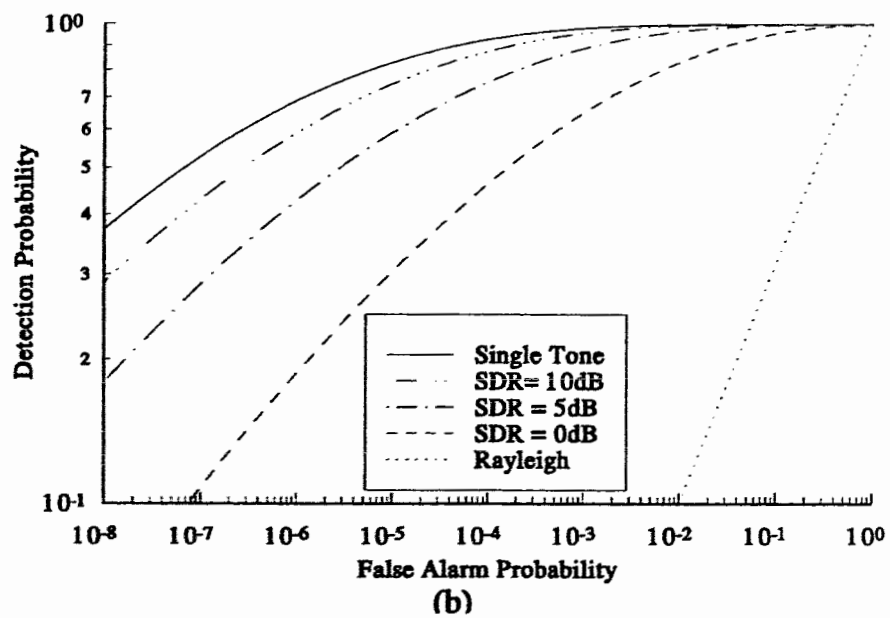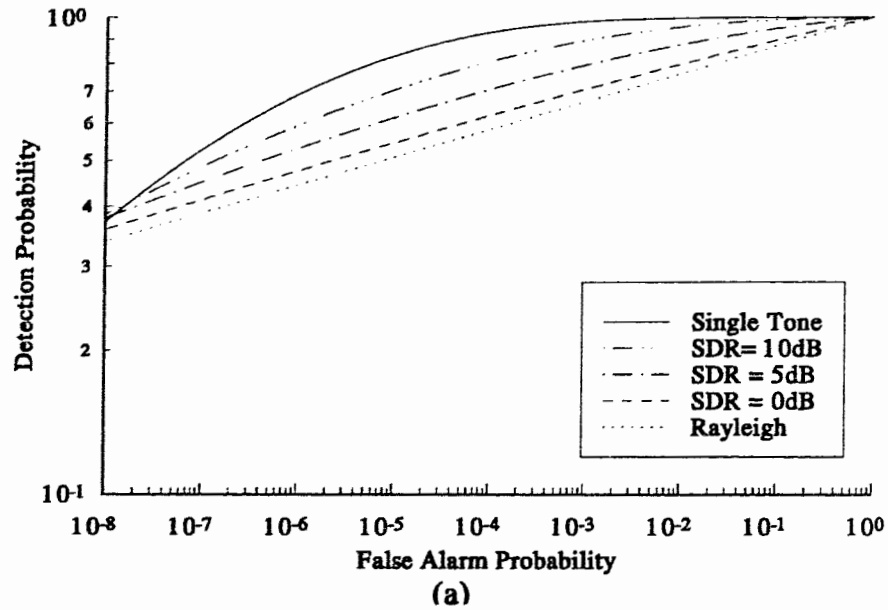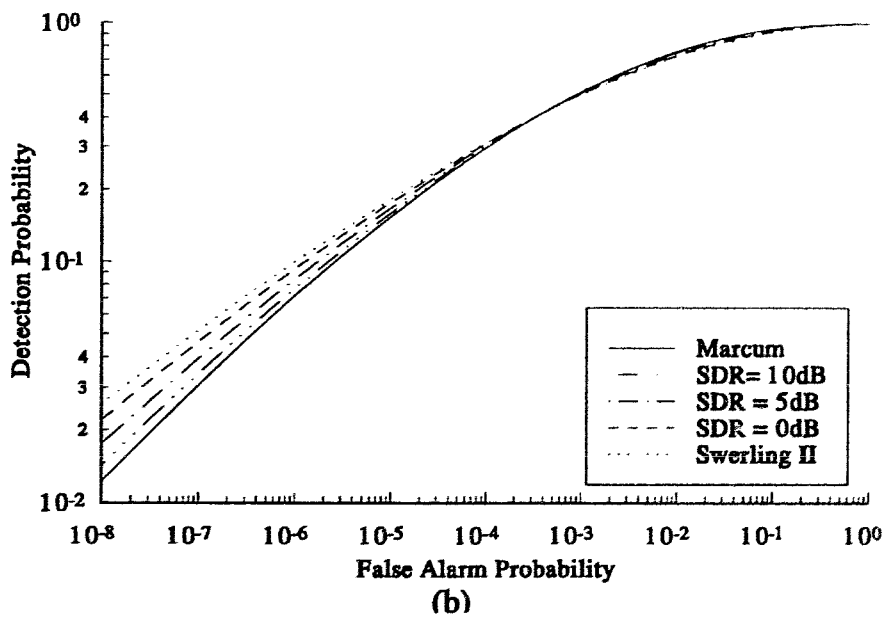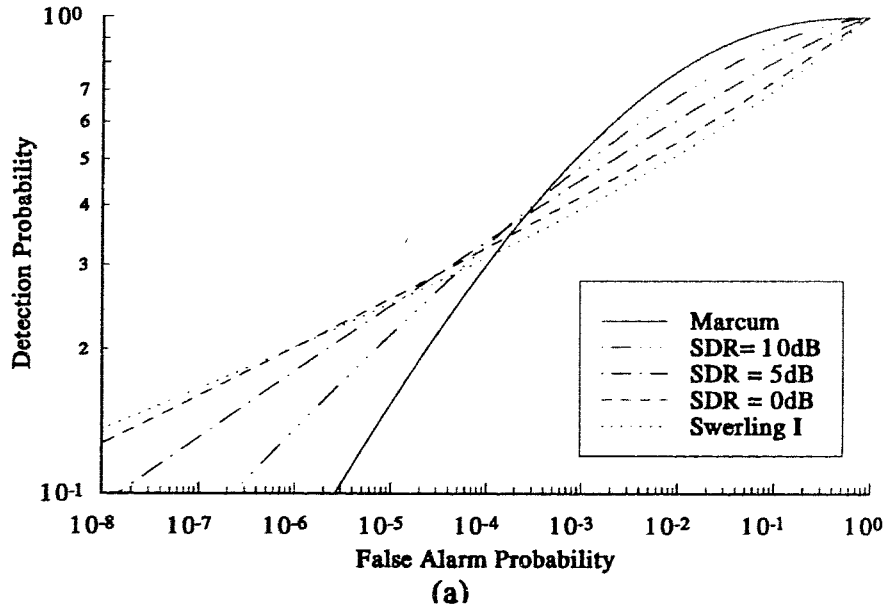
Figure 5.7: Log-ROC curves for noncoherent integration of Rician targets with diffuse component (a) fully correlated and (b) fully decorrelated pulse-to-pulse. Curves are for a 16 returns with $\varrho = 0$ dB. Note the change of scale for the detection probability in (b).

Table 5.2: Moment Generating Functions for Integrating Rician Targets

Integration Case	Moment Generating Function
1	$G_1(s) = \dfrac{\exp\left(\dfrac{-Ma^2s}{2[1+s(1+M\sigma_d^2)]}\right)}{1+s(1+M\sigma_d^2)}$
2	$G_1(s) = \dfrac{\exp\left(\dfrac{-Ma^2s}{2[1+s(1+\sigma_d^2)]}\right)}{1+s(1+\sigma_d^2)}$
3	$G_1(s) = \dfrac{\exp\left(\dfrac{-Ma^2s}{2[1+s(1+M\sigma_d^2)]}\right)}{(1+s)^{M-1}[1+s(1+M\sigma_d^2)]}$
4	$G_1(s) = \dfrac{\exp\left(\dfrac{-Ma^2s}{2[1+s(1+\sigma_d^2)]}\right)}{[1+s(1+\sigma_d^2)]^M}$

5.3 Application: Combined Noncoherent Coherent System

We now present an example of the application of rate-constraint theory to signal processing. Specifically, we will examine how the amount of processing required to derive Doppler information on targets can be decreased by orders of magnitude with only moderate increases in required SNR to achieve equivalent detection performance.

Consider the detection system of figure 5.8. The first stage uses noncoherent integration of M returns to determine the cells that are most likely to contain targets. The second stage then coherently combines another M (independent) returns, and makes the final detection decision. Such a system is very efficient in the number of Fast Fourier Transforms (FFT's) required.

One subtle point should be noted regarding the false alarm probability specification. When coherent integration of M returns is used, the total number of chances for a false alarm increases by a factor of M, since each return yields M frequency bins. Therefore, the specification for overall false alarm probability may also have to be changed when comparing the detection performance between a fully coherent and fully noncoherent system. For example, if the actuator following the coherent system treats each Doppler cell as individual
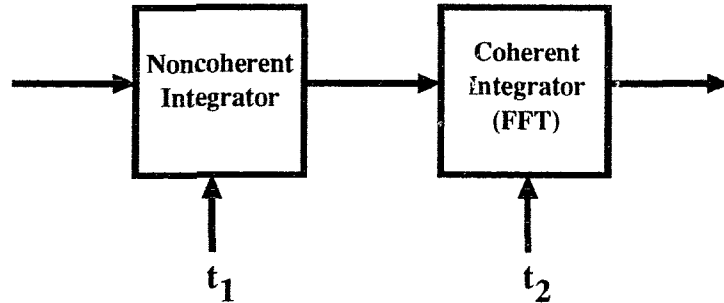
Figure 5.8: Combined Noncoherent/Coherent System. Only those cells crossing the first threshold are coherently integrated by the second stage.

(as in an automated tracking system), the designed false alarm probability must be divided by M; however, if the actuator is, for example, a jet interceptor that must interrogate a given range cell (regardless of the Doppler resolution of that cell), no such division is required.

To investigate such a combined detection system, we compare its detection performance with a 'conventional' system using same number of pulses, coherently combining each burst and noncoherently combining the two (using quadratic combination). The system of figure 5.8 utilizes two independent *bursts*, each containing a number of coherent pulses. The optimal combination system would coherently integrate the pulses within each burst, and then optimally combine the results from the two bursts; however, we will make our comparison against a system using quadratic combining.

To compare between the combined and conventional systems, we calculate the required SNR (assumed to be the same for each burst) to get $pd_T = 0.8$ for $pf_T = 10^{-6}$ when the a priori probability of a target, $\pi_1 = 10^{-4}$. In figure 5.9, the difference in required SNR between the optimal and combined systems is plotted as a function of the rate out of the noncoherent stage for the Rician family of targets. As expected, the curves are bracketed by the Rayleigh (equivalent to SDR= $-\infty$) and the Marcum (SDR $= \infty$) curves. Note that for the Rayleigh target, the processing loss is monotonically increasing in r; this is due to the constant SLOC of the coherent stage, which means that the system is globally properly ordered and will always get better performance as the second stage is able to process more data. For the Marcum target, the processing loss is less than for the Rayleigh, but is not monotonically decreasing for large values of r. The non-monotonicity results from the tests not being properly ordered, so operating under the rate-constraint is not optimal. However,
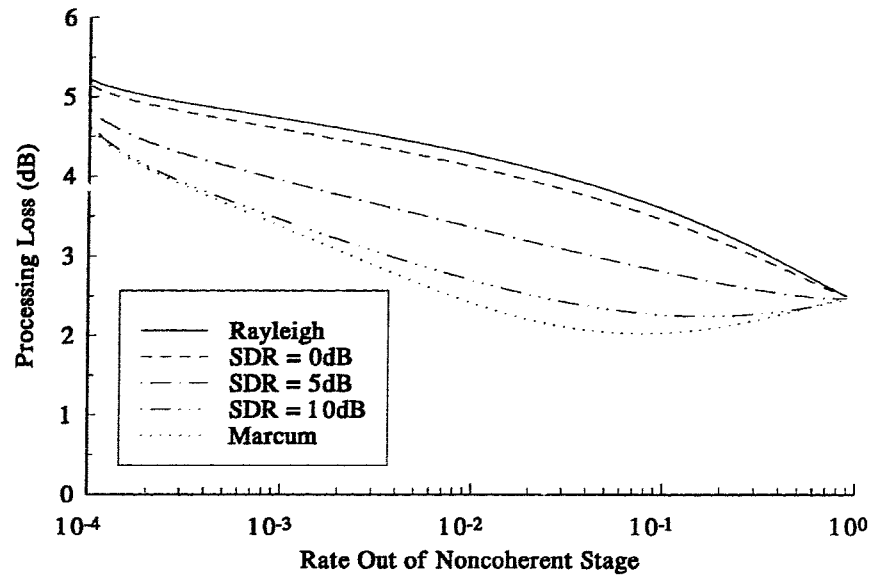
Figure 5.9: Processing loss for detection of Rician targets: $pd = 0.8, pf = 10^{-6}$, $\pi_1 = 10^{-4}$, $M = 16$

as shown by figure 5.9, the total loss is not going to be more than a couple of dB.

Note for $r = 1$, the curves converge to a value of ≈ 2.45 dB. This value gives the gain obtained by the noncoherent integration of the two bursts in the conventional system; for $r = 1$ the combined scheme just uses the results of the second burst. (The actual value of 2.45 dB is a function of the detection and false alarm probabilities used.)

In figure 5.10, we consider the effect of changing the number of cells integrated per burst. Since Rician is bounded by the two, we consider only Marcum and Rayleigh fading targets. The losses become smaller as M decreases.

As figures 5.9 and 5.10 indicate, it is possible to buy considerable simplifications in complexity with moderate prices in signal-to-noise ratio. For example, based on figure 5.9, when detecting a Rician target based on 16 returns, the number of FFT's required may be reduced by a factor of 100 with processing losses of less than 5 dB. If the target has a significant glint (i.e. SDR ≥ 10 dB), the loss drops to less than 3 dB.
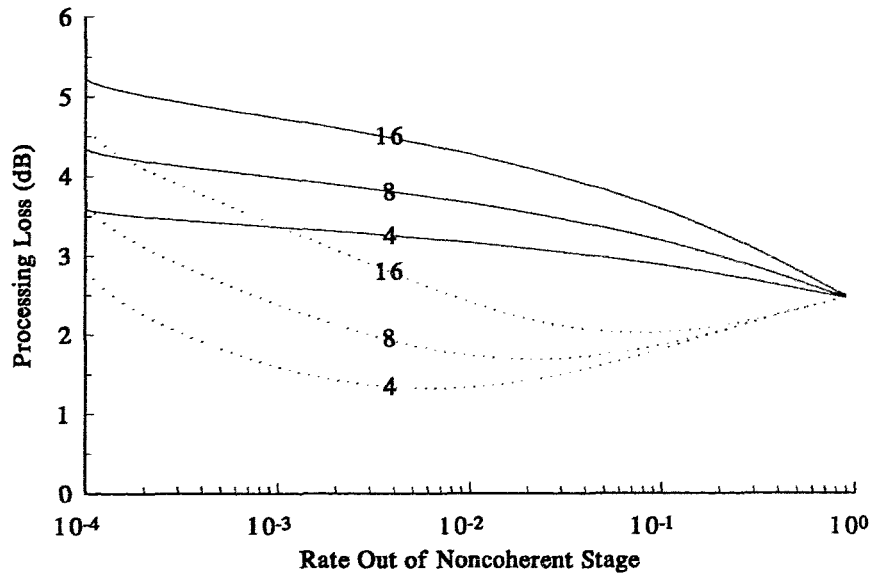
Figure 5.10: Processing loss for detection of Rayleigh and Marcum targets for various values of M: $pd = 0.8, pf = 10^{-6}, \pi_1 = 10^{-4}$

Thus far, we have considered the SLOC characteristic of different target models in Gaussian distributed noise. In the next section, we briefly investigate the effect of radar clutter which has heavier tails than the Gaussian distribution.

5.4 Effect of (K-Distributed) Radar Clutter

We have considered only radar returns for models based on Gaussian noise. In this section we investigate the log-ROC and SLOC properties of radar targets in non-Gaussian clutter.

The emphasis of this thesis is not on how to calculate probability integrals; however, we will briefly outline a new technique developed by Bird in [16]. While a thorough understanding of the method is not required, we do present enough detail to provide a context for the results to follow.

5.4.1 The Method

The technique of [16] treats both linear and quadratic detectors; however we consider only quadratic combining here. The two perform very similarly, with the linear providing the practical advantage of having a smaller dynamic range, but requiring a larger collapsing loss. (The collapsing loss is the additional signal required to maintain the same detection and false alarm probabilities when unwanted noise samples are integrated with the desired signal (plus noise) samples [6].)

The method works for circular symmetric distributions. Since the phase of the noise and target return is symmetric over $[0, 2\pi]$, the two-dimensional characteristic function of the joint distribution of the in-phase and quadrature components simplifies to a function of the single radial variable, z. Then we have the following transform pair

$$\Phi(\omega) = \int_0^\infty f_Z(z) J_0(\omega z) dz \tag{5.7}$$

and

$$f_Z(z) = z \int_0^\infty \Phi(\omega) \omega J_0(z\omega) d\omega \tag{5.8}$$

where $J_0(x)$ is the ordinary Bessel function of order zero, $f_Z(z)$ is the density function of the radial variable Z, and $\Phi(\omega)$ is the coherent characteristic function of $f(i, q)$. The method's appeal arises from the ease in which $\Phi(\omega)$ may be determined for a number of interesting target and noise scenarios.

For a square-law detector, the test statistic is $q = z^2$, so that (5.8) becomes

$$f_Q(q) = \tfrac{1}{2} \int_0^\infty \Phi(\omega) \omega J_0(\sqrt{q}\omega) d\omega \tag{5.9}$$

Now, in order to calculate the distribution for the noncoherent integration of M independent returns, we require the noncoherent moment generating function of q, given by

$$G_q(s) = \int_0^\infty f_Q(q) \exp(-sq) dq \tag{5.10}$$

Substituting (5.9) into (5.10), and exchanging the order of integration,

$$G_q(s) = \tfrac{1}{2} \int_0^\infty \Phi(\omega) \omega \left[\int_0^\infty J_0(\sqrt{q}\omega) \exp(-sq) dq \right] d\omega \tag{5.11}$$

but the integral in the square brackets is the Laplace transform of $J_0(\sqrt{q}\omega)$ which is $\exp(-\omega^2/4s)/s$ so therefore

$$G_q(s) = \tfrac{1}{2} \int_0^\infty \Phi(\omega) \frac{\omega}{s} \exp\left(-\frac{\omega^2}{4s} \right) d\omega \tag{5.12}$$

Then to calculate the detection probability,

$$pd(t) = 1 - \mathcal{L}^{-1} \left\{ \frac{1}{s} G_q(s)^M \right\} \tag{5.13}$$

where the s^{-1} factor indicates the integration required to calculate the cumulative density function, and the M exponent results from the noncoherent integration of M independent samples. (Note the false alarm probability is similarly calculated with $G_q(s)$ replaced by the appropriate moment generating function under H_0.)

Note that the technique may be used to generate the previous results for Marcum, Rayleigh and Rician targets (however, there is no appropriate model for a Swerling III/IV target in this formalism). Moreover, the technique can be used to create targets composed of any mixture of tones and Gaussian noise.

5.4.2 K-Clutter

It has long been recognized that for some radar scenarios (for example, shallow grazing angles) the noise and/or clutter is not well-fit by a Gaussian distribution [49],[50]. The log-normal and Weibull distributions have been used, with varying degrees of success. The Weibull distribution, while convenient, isn't motivated by a physical justification.

First introduced in [50], the K-distribution has been shown to provide better fits to experimental clutter data in [49]; in addition it generalizes to the Rayleigh and Weibull distributions[4].

The K-distribution is a compound model, consisting of a local mean level y which is fit well by a Chi distribution:

$$f(y) = \frac{2b^{2(\nu+1)}y^{2\nu+1}}{\Gamma(\nu+1)} \exp(-b^2 y^2) \tag{5.14}$$

and a 'speckle' component which is Rayleigh distributed with mean y:

$$f(x|y) = \frac{\pi x}{2y^2} \exp\left(-\frac{\pi x^2}{4y^2}\right) \tag{5.15}$$

Integrating to get $f(x)$ yields

$$f(x) = \int_0^\infty f(x|y)f(y)dy = \frac{2b}{\Gamma(\nu+1)} \left(\frac{bx}{2}\right)^{\nu+1} K_\nu(bx) \tag{5.16}$$

---

[4]Thus the Weibull distribution does have a physical interpretation: it results from a K-distribution with $\nu = -0.5$. The Weibull model is used for its mathematical convenience however, and the K-distribution produces the more general model.

where K_ν is the νth order modified Bessel function of the third kind, ν is a shape parameter and we have normalized x (divided by $\sqrt{\pi}$) in order to have the same notation as [51]. b is given by

$$b = 2\sqrt{\frac{\nu + 1}{E\{x^2\}}} \tag{5.17}$$

From the literature, sea clutter seems to be fit well by values of ν in the region $-0.75 < \nu < -0.5$ or so [51], while ground clutter may have values as low as $\nu = -0.9$ [49]. Note that $\nu = -0.5$ yields a Weibull distribution, while $\nu = \infty$ gives the Rayleigh distribution (smaller values of ν indicate more 'spikiness' in the distribution).

The coherent characteristic function for the K-distribution is given by [49]

$$\Phi(\omega) = \left(\frac{b^2}{b^2 + \omega^2}\right)^{\nu+1} \tag{5.18}$$

If we include Gaussian components (from thermal noise and a Rayleigh target), and a specular return of amplitude a, the resulting coherent characteristic function is

$$\Phi(\omega) = \left(\frac{b^2}{b^2 + \omega^2}\right)^{\nu+1} \exp\left(-\frac{\omega^2\sigma^2}{2}\right) J_0(a\omega) \tag{5.19}$$

where under H_0, $\sigma^2 = \sigma_n^2$, the power in the thermal noise component, while under H_1, the diffuse power in the target is included so that $\sigma^2 = \sigma_n^2 + \sigma_t^2$. In (5.19), a single glint of amplitude a is included; any number of glints with any amplitude may be included by replacing the Bessel function by a product of Bessel functions, each with its argument weighted by the appropriate amplitude. However, we will consider only single glint, and Rayleigh targets here.

From a study matching the K-distribution to experimentally measured clutter, it is noted that the clutter to thermal (Rayleigh) noise can be on the order of 23 dB or so [49]. In figure 5.11, we have plotted the log-ROC and SLOC curves for integration of 16 returns of a Rayleigh target in K-distributed clutter with a single-pulse clutter-to-noise ratio (CNR) of 23 dB and a signal-to-clutter ratio of 0 dB. The shape parameter for the K-distribution was set at $\nu = -0.85$, which corresponds to an experimentally determined value for land clutter described in [49], and at $\nu = -0.5$, which corresponds to a Weibull distribution. Of course, ν is a function the radar parameters (frequency, polarization, depression angle, etc.) as well as the area being investigated.
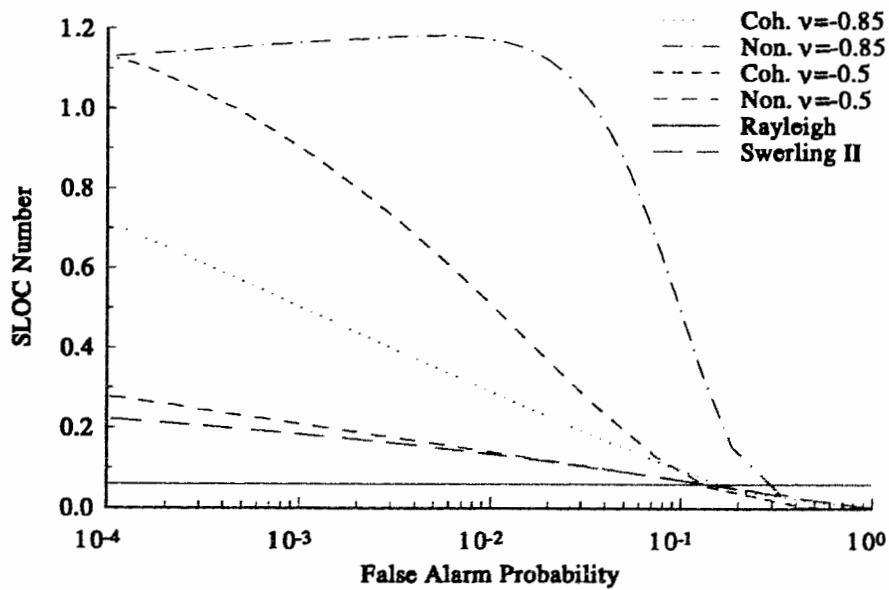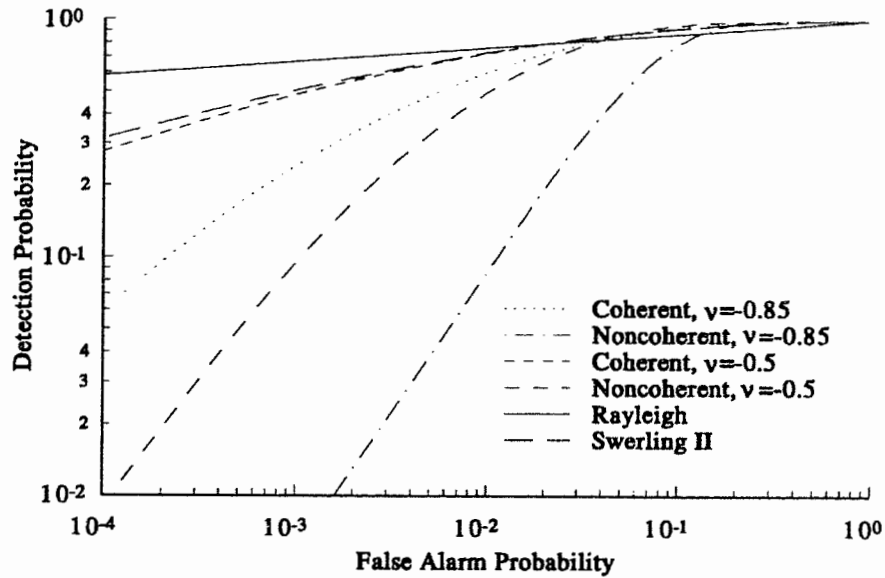
Figure 5.11: Log-ROC and SLOC curves for integrations of 16 returns of a Rayleigh target in a mixture of K-Distributed plus Gaussian noise (CNR = 23 dB, and SNR = 0 dB).

For comparison, the ROC and SLOC curves for coherent integration of a Rayleigh target and noncoherent integration of a Swerling II target (each with SNR = 0 dB and CNR = -∞ dB) are also given in figure 5.11. The effect of the heavy tails of the K-clutter is evident. The clutter results in lower detection probabilities for $pf < 10^{-2}$ or so. To obtain a given pf, the threshold must be raised significantly compared to the Gaussian noise case, resulting in a much lower detection probability. The heavier tails result in smaller densities at lower signal values, therefore pd is higher for high values of pf. Note that the same effects are evident in comparing the two clutter cases: for the Weibull clutter ($\nu = -0.5$), the tail is not as heavy as for $\nu = -0.85$ and so pd is lower at high pf and higher at low pf.

The SLOC functions of figure 5.11 were again evaluated by fitting a cubic spline to the log-ROC data. Note that the SLOC is higher for noncoherent integration than for coherent integration over some operating regions, so that the processing scheme of section 5.3 is still viable.

Similar curves are plotted in figure 5.12 for non-fading (Marcum) targets with the same SNR and CNR. The interesting curve is for the coherent integration of a target in clutter with heavy tails ($\nu = -0.85$). Note how the log-ROC literally crashes for false alarm probabilities smaller than $3(10)^{-3}$. The cause of the crash is that the threshold for $pf = 3(10)^{-3}$ just exceeds the value of the signal component. Since the clutter has such a heavy tail, the change in pf is not as extreme as the change in pd; hence the large slope. The large slope is reflected in a peak in the SLOC as indicated in the figure. For the smaller Weibull tail (with $\nu = -0.5$) the effect is less pronounced and occurs at a lower pf. The moral of figure 5.12 is clear: when dealing with a Marcum target, make sure the threshold is set higher than the amplitude of the return or else detection performance will be extremely poor.

Note that provided we consider only false alarm probabilities greater than that which would cause a crash, coherent integration again produces a smaller SLOC function than noncoherent.

In our analysis of K-clutter, we have assumed that the clutter is identically and independently distributed (i.i.d.) from cell-to-cell. In fact, we have implicitly assumed temporal and spatial stationarity for the statistics in all that we have done thus far. To combat the nonstationarity inherent in any radar problem, a number of normalization techniques called Constant False Alarm Rate (CFAR) methods have been developed. In the next section we will consider how such techniques may be incorporated into systems operating under
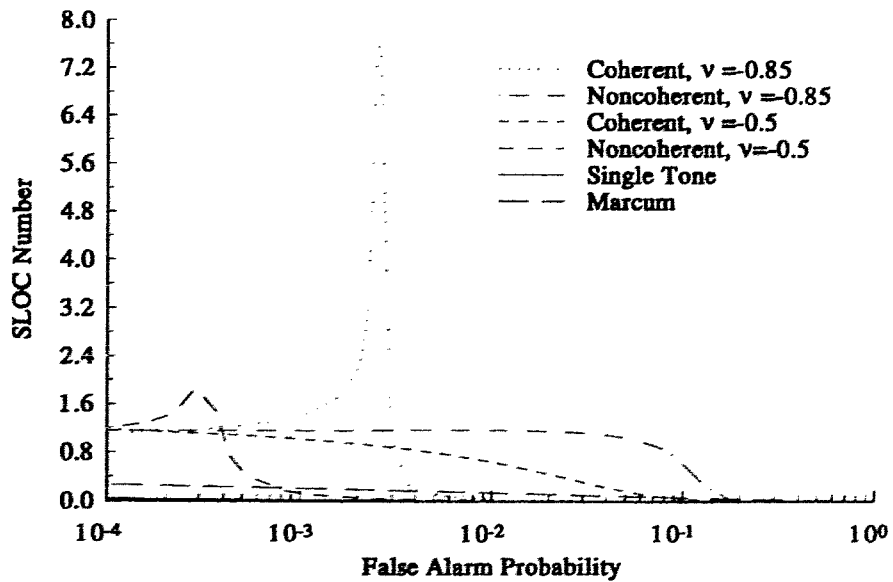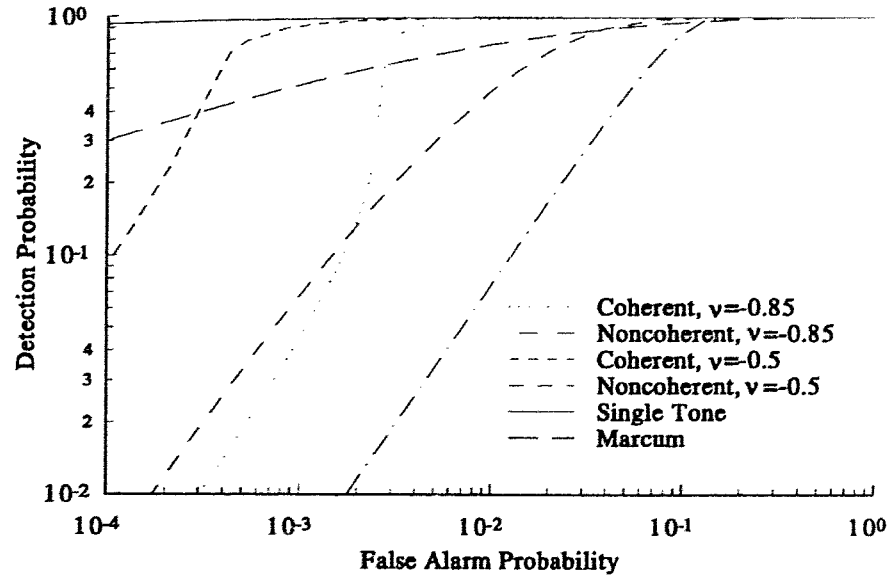
Figure 5.12: Log-ROC and SLOC curves for integration of 16 returns of a Marcum target in a mixture of K-Distributed plus Gaussian noise (CNR = 23 dB, and SNR = 0 dB).
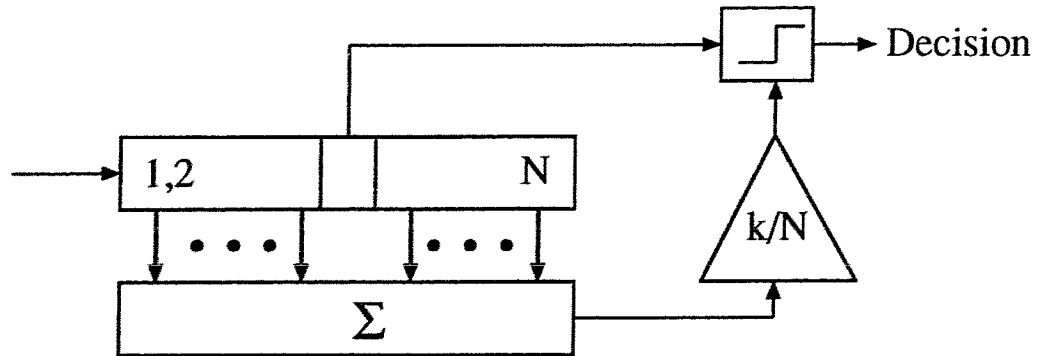
Figure 5.13: Cell Averaging Constant False Alarm Rate processor.

rate-constraint principles.

5.5 A Brief Note on Normalization Techniques

Thus far we have not considered the role of normalization (CFAR) techniques in Rate-Constraint Theory. Normalization methods are used to combat the global nonstationarity of interference (clutter or thermal noise). There are two different assumptions that can be made to motivate CFAR techniques: spatial homogeneity or temporal stationarity. Cell Averaging CFAR and Order Statistic CFAR work by detecting when a given cell is much larger than it neighbouring cells. Clutter map CFAR detects when a given cell exceeds the average of the previous returns for that individual cell.

Our coverage is intentionally brief; the intent is to show how normalization techniques can be used to combat nonstationarity in systems operating under a rate constraint. For further detail the interested reader is referred to an extremely readable review of CFAR by Farina and Studer [52].

Cell Averaging CFAR   The general algorithm for CA-CFAR processing is given in figure 5.13. The basic idea is to use the cells surrounding a given cell to estimate the background noise level. The estimate is then multiplied by a constant, k to generate the threshold that the return will be compared against. Note that the cells used to form the estimate may be neighbours in Doppler, azimuth, range or a combination.

It is well known that the probabilities of detection and false alarm for a CA-CFAR
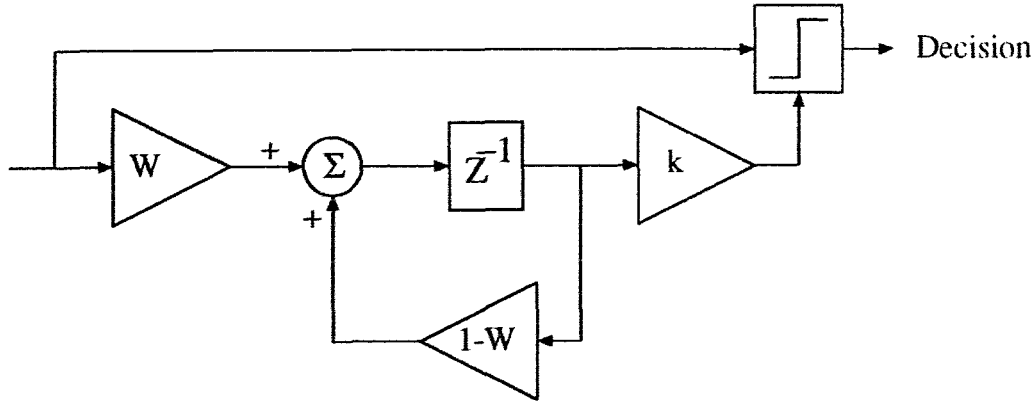
Figure 5.14: Clutter Map Constant False Alarm Rate processor.

system operating on a Rayleigh target in Gaussian noise are, respectively, given by [53], [43]

$$pd = \left(1 + \frac{k}{N(1+\varrho)}\right)^N \qquad (5.20)$$

and

$$pf = \left(1 + \frac{k}{N}\right)^N \qquad (5.21)$$

where N is the number of cells used in the estimator, k is the multiplier used to determine the threshold from the estimate, and ϱ is the signal-to-noise ratio.

Clutter Map CFAR   In figure 5.14, the processing scheme for clutter map CFAR is given. A recursive averager is used to estimate the mean value of the return for the individual cell. The estimate is again multiplied by a constant, k and then used for the threshold.

Nitzberg has shown that when detecting a Rayleigh target in Gaussian noise with a Clutter Map CFAR system such as in figure 5.14, the detection and false alarm probabilities are given by [54]

$$pd = \lim_{M \to \infty} \frac{1}{\prod_{m=0}^{M}[1 + k_D W(1 - W)^m]} \qquad (5.22)$$

and

$$pf = \lim_{M \to \infty} \frac{1}{\prod_{m=0}^{M}[1 + kW(1 - W)^m]} \qquad (5.23)$$

where k is again the CFAR multiplier and

$$k_D = \frac{k}{1 + \varrho} \qquad (5.24)$$
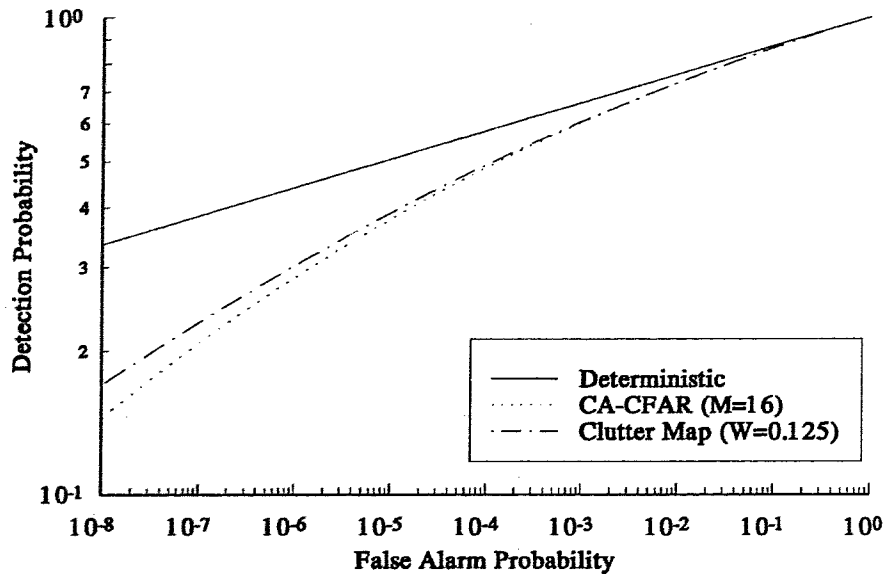
Figure 5.15: Log-ROC curve for Clutter Map and Cell Averaging CFAR systems detecting a Rayleigh target in Gaussian noise ($\varrho = 12dB$).

In figure 5.15, we have plotted the log-ROC curves for the two CFAR techniques, along with the curve for a deterministic threshold. Since the effective number of cells used in the estimate for the clutter map is [43]

$$\frac{2 - W}{W} = 15 \text{ for } W = .125 \tag{5.25}$$

and $N = 16$ for the CA-CFAR, it is not surprising that the two techniques give nearly identical performance for these parameters. Note that for large pf and pd, the log-ROC (and therefore SLOC) curves for the CFAR are very close to that for the Rayleigh case.

Order Statistics Another way to estimate the mean background interference level is to rank the returns, and multiply the mth largest by a constant. We introduce a similar technique in chapter 6. The basic idea is this: in rate-constraint we want those cells most likely to contain targets to be passed on to the next stage. Therefore, why not rank the returns and pass on the largest, thereby eliminating the need for an explicit threshold?

All of the CFAR techniques can be used in rate-constraint systems. While we don't require the constant false alarm rate, the normalization techniques ensure that large clutter regions won't mask small targets. The CFAR multiplier k can be controlled by a rate-constraint system to generate the desired rate.

5.6   Conclusions

In this chapter, we have presented graphs illustrating the log-ROC and SLOC curves for several radar models. For the models studied, a few general conclusions can be drawn:

1. The SLOC number tends to decrease with increased SNR; therefore a multi-stage system consisting of fast, low SNR stages cueing slower, higher SNR stages will have at least some operating region where the system is properly ordered.

2. The SLOC number tends to be higher for noncoherent integration than for coherent integration for the same number of returns of a given target type. A system of the type discussed in section 5.3 consisting of a stage utilizing noncoherent integration cueing a stage performing coherent integration is therefore feasible. For the target models considered, the loss incurred by operating under the rate constraint is small, even when the stages are operating in a region that is not locally properly ordered.

3. A Rician target is a more accurate representation of a target consisting of one single dominant and a number of smaller i.i.d. scatterers. The range of behaviours of a Rician target may be bracketed by considering Swerling I/II targets (for which the SDR $= -\infty$ dB) and Marcum targets (for which SDR $= \infty$ dB).

4. Radar clutter typically produces heavier tails than predicted by the Rayleigh distribution. The heavier tails result in larger SLOC values at lower values of pf.

5. Normalization techniques cause slight losses, and therefore increase the SLOC slightly when compared to 'crisp' thresholds.

Remember that in real radar deployments, the actual target type is not known - to us or to our detection scheme. Nevertheless, the radar models do allow us to investigate rate constraint theory in an established framework. It appears that rate-constraint systems are

in fact optimal over large operating regions for many common models; more importantly, rate-constraint tests can be practically implemented. The analysis also allows us to make more intelligent design decisions regarding required processing rates for multi-stage systems.

We saw that CFAR techniques were invented to deal with global nonstationarity, by assuming local stationarity. The nonstationarity suggests that a rate-constraint test will not be able to produce a constant output, but will have some variance about the desired value. In chapter 6 we investigate some of the consequences of such variances.

Chapter 6

Implementation Issues

6.1 Introduction

The basic strategy underlying resource-constraint detection is to utilize the available processing resources to the fullest extent. In chapters three and four we showed that the output rate from a detection stage should be matched to the maximum processing rate of the following stage. Like all of detection theory, our work thus far has been concerned only with first-order statistics. However, the hit-rate is a random variable with variances that must be accounted for. Another source of variance is due to the fact that the output rate, while observable, must be estimated.

A rule of thumb when estimating a binomial probability ξ_1 is to use at least $\frac{10}{\xi_1}$ cells [55]. The ξ_1 in question, the hit-rate, is often a small number for radar systems (say $\xi_1 \approx 10^{-2} - 10^{-4}$ [18]). Therefore a large number of cells are required to get a reasonable estimate. Nonstationarity may be a problem when such long observation periods are necessary. While the normalization techniques discussed in chapter five help, the output rate will vary somewhat about the desired operating point.

When the output rate has a mean ξ_1 and a non-zero variance, there will be times when the instantaneous rate is larger than ξ_1. Therefore, some of the hits will be lost to normalization unless allowances are made for the overloads.

In this chapter, we present ways to deal with the variance in the output rate from a test operating under a rate-constraint criterion. The first technique discussed is the introduction of a queue between the detection stage and the subsequent actuator. We examine queuing

structures to handle two different scenarios. The first is for a processor that treats all hits the same. For example, a clutter map might be applied to all cells that contain hits; the processing required for the map is the same whether the hit was caused by a target or a false alarm. The second situation is the more general; we consider actuators that treat false alarms and targets differently. For example, a target will presumably tie up the computing resources in a tracking computer longer than will a false alarm since the target will be continually monitored during its presence in the radar field, while the false alarm will require only a short number of radar looks to dispense with.

Many multi-stage detection systems can be characterized by a combination of a stage operating under the first scenario followed by a stage operating under the second. For example, a multi-stage detection system where all hits are processed by various forms of CFAR (clutter Map, binary integration etc.) is explained in [18]. Further in the detection chain, those cells that pass the CFAR processing are passed to a track initiator, which requires more processing for targets than for false alarms. In such systems, the earlier stages are characterized by smaller values for the target concentration and signal-to-noise ratio.

Another method of dealing with the stochastic nature of the output rate is to eliminate the variance altogether by ranking the returns and passing on only those cells most likely to contain targets. The load on the resource limited actuator can be controlled by varying the fraction of cells chosen during each processing interval. It might seem counter-productive to use a ranking scheme since the test will always provide exactly the same number of hits. Does such a test provide any information? The answer is yes, provided a time-history[1] is available for the cells that produce hits. If a given cell is consistently ranked as the largest in its neighbourhood, then it is more likely to contain a target. The idea of combining a ranking scheme with a test that declares a target present if a cell is chosen M or more times in N consecutive scans was first introduced in [17].

The structures in this chapter provide the ability to run systems at (or near) maximum capacity even in the face of ignorance of the underlying statistics.

---

[1]A spatial 'history' may be more appropriate in sonar applications; a hit is more likely to be a target if there are hits in adjacent cells.
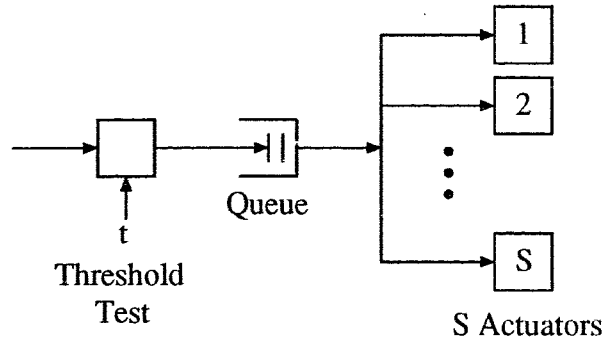
Figure 6.1: Queue for buffering output hits

6.1.1 Outline of Chapter

In section 6.2, we analyze the use of a simple queue between a detection stage and a bank of actuators. We assume that all hits, whether they be targets or false alarms, require a single service time. The trade-offs between the output rate of the test, the maximum queue length, and the mean delay in the queue are examined.

The method of ranking the returns and processing the largest is presented in section 6.3. We compare the results of ranking with a threshold test for detection of a Rayleigh target in Gaussian Noise. The loss attributed to ranking is not large for the resource-constraints considered.

The use of "M/N" Detectors is the subject of section 6.4. We include more detail than is required for the purposes of the thesis in order to present some new results for the mean and variance of the time required to confirm a target with a given confirmation rule.

In section 6.5, we extend the Markoff model introduced in section 6.2 to allow targets and false alarms to be treated differently. The results from the analysis stress the need for processors that are able to quickly dispense with false alarms. Section 6.6 then applies the analysis of section 6.5 to a network of track-initiators, each utilizing a M/N strategy. Finally, chapter conclusions are discussed in section 6.7.

6.2 Queues for Buffering

Consider figure 6.1 consisting of a queue with maximum length N_Q feeding a bank of S servers (actuators). Let $\tilde{P}_i$ be the steady-state probability that there are exactly i cells in

the system immediately after the beginning of a service cycle. It can be seen that $\tilde{P}_i$ is then equal to the probability of i cells in the queue immediately before the queue feeds S cells to the actuators[2]. Now let d_S be the probability that there are no more than S cells in the system at that time, that is

$$d_S = \sum_{i=0}^{S} \tilde{P}_i \tag{6.1}$$

Assume that the number of cells investigated per processing interval is C, of which, on average, $\xi_1 = \pi_1 pd + \pi_0 pf$ are hits. Then following [56],

$$
\begin{aligned}
\tilde{P}_0 &= d_S \mathcal{B}_0^C(\xi_1) \\
\tilde{P}_1 &= d_S \mathcal{B}_1^C(\xi_1) + \tilde{P}_{S+1} \mathcal{B}_0^C(\xi_1) \\
\tilde{P}_2 &= d_S \mathcal{B}_2^C(\xi_1) + \tilde{P}_{S+1} \mathcal{B}_1^C(\xi_1) + \tilde{P}_{S+2} \mathcal{B}_0^C(\xi_1) \\
&\vdots \qquad\qquad\qquad\vdots \\
\tilde{P}_k &= d_S \mathcal{B}_k^C(\xi_1) + \tilde{P}_{S+1} \mathcal{B}_{k-1}^C(\xi_1) + \cdots + \tilde{P}_{S+k} \mathcal{B}_0^C(\xi_1) \qquad\qquad \text{for } k \leq N_Q - S \\
&\vdots \qquad\qquad\qquad\vdots \\
\tilde{P}_k &= d_S \mathcal{B}_k^C(\xi_1) + \tilde{P}_{S+1} \mathcal{B}_{k-1}^C(\xi_1) + \cdots + \tilde{P}_{N_Q-1} \mathcal{B}_{k+1-(N_Q-S)}^C(\xi_1) \\
&\quad + \tilde{P}_{N_Q} \mathcal{B}_{k-(N_Q-S)}^C(\xi_1) \qquad\qquad\qquad\qquad \text{for } N_Q - S < k \leq N_Q - 1 \\
&\vdots \qquad\qquad\qquad\vdots \\
\tilde{P}_{N_Q} &= 1 - (\tilde{P}_0 + \tilde{P}_1 + \cdots + \tilde{P}_{N_Q-1}) \tag{6.2}
\end{aligned}
$$

where

$$\mathcal{B}_i^C(p) = \binom{C}{i} p^i (1-p)^{C-i}$$

denotes the binomial probability mass function expressing the probability of getting i successes out of C tries given that the probability of success for an individual trial is p. The final equation in (6.2) normalizes the solution.

Solving the linear system of equations of (6.2) for the $\tilde{P}_i$ yields the steady-state distribution for the queue population. The average load carried by the system is then

$$\lambda_S = \sum_{i=1}^{S-1} i \tilde{P}_i + S \sum_{i=S}^{N_Q} \tilde{P}_i \qquad \text{for } N_Q \geq S \tag{6.3}$$
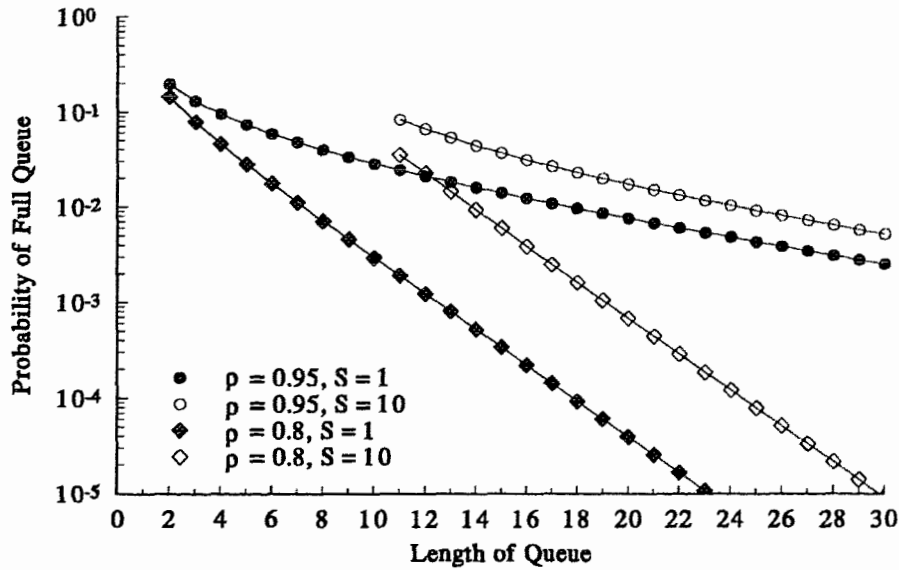
Figure 6.2: Probability of queue overload

The probability of a hit being rejected by the queue is given by

$$P_b = \frac{\text{offered load - carried load}}{\text{offered load}} \tag{6.4}$$

but the offered load is given by $\xi_1 C$, so that

$$P_b = 1 - \frac{\lambda_S}{\xi_1 C} \tag{6.5}$$

Figure 6.2 shows the probability of queue blockage, P_b as a function of the maximum queue length. Curves are given for two values of S and normalized load

$$\rho = \frac{\xi_1 C}{S}$$

As expected, the probability of a queue overload is higher for larger offered loads. Decreasing the number of actuators from 10 to 1 reduces the probability of queue overload because there

---

[2]Note that the analysis is equivalent for a single actuator that takes S cells and serves them before updating the queue; however, in that case we must remember to divide the processor rate by S.

are more cells examined per cycle for the lower value of S, and therefore a smaller variance in the output rate. The results of figure 6.2 are for $C = 1000$, but do not change appreciably with C, provided C is large enough. The lack of dependence on C is to be expected since a binomial distribution may be approximated by a Poisson for large values of C[3]. In fact, Chu's curves, derived from the Poisson approximation are identical to those of figure 6.2 [56].

Figure 6.2 shows that in order to reduce the risk of a cell being blocked out by a full queue, there are two options. Either the normalized load, ρ must be reduced or else the maximum length of the queue must be increased.

The former solution, reducing ρ, introduces a trade-off between the probability of the queue availability ($= 1 - P_b$) and the detection probability. If ρ is decreased, the output rate of the threshold test, ξ_1 and therefore the detection probability is also decreased. The probability of queue overload is, however, reduced. The trade-off between pd and P_b for a particular scenario (Rayleigh target with $\varrho = 20$ dB in Gaussian noise, $\pi_1 = 10^{-3}, N_Q = 10, C = 1000$ and $S = 1$) is shown in figure 6.3. Note the introduction of the queue has softened the effect of overloading the actuator when compared with the randomization shown in figure 3.7, for example.

The overall detection probability, given by

$$pd_{ov} = pd(1 - P_b)$$

achieves its maximum at a value of ρ that is a function of all of the detection parameters. For example figure 6.4 includes plots for several systems detecting a 10 dB Rayleigh target in Gaussian noise.

There are a number of general conclusions to be inferred from figure 6.4:

1. The exact value of ρ for the maximum overall detection probability is a function of the radar scenario. For the cases considered here, the location of the maximum ranges from $\rho \approx 0.8 \rightarrow 1$.

2. Increasing C, the number of cells, from 100 to 1000 while maintaining the expected number of targets, $\pi_1 C$, decreases the overall detection probability, pd_{ov}. The decrease results from the increased variance in the output rate, which increases the probability

---

[3]The Poisson approximation is independent of the number of cells.
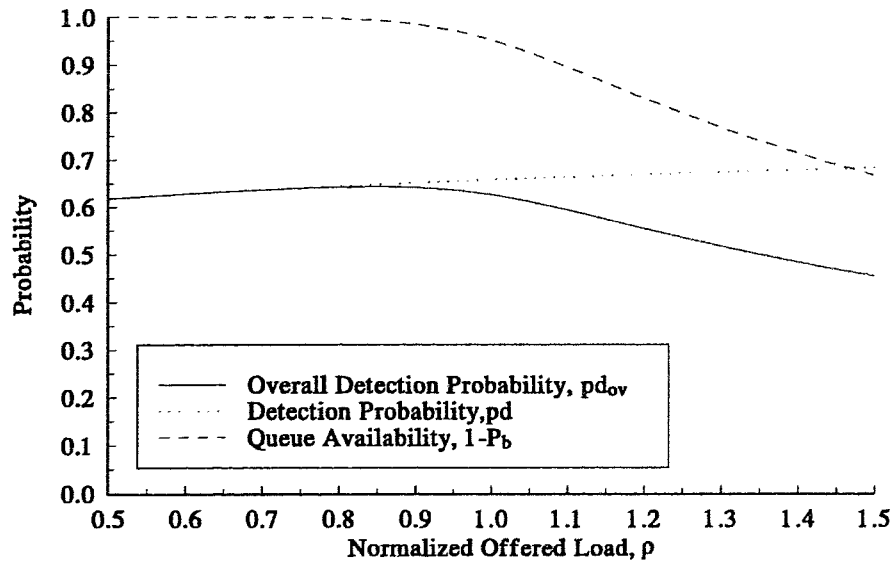
Figure 6.3: Trade-off Between Detection Probability and Queue Availability for $\pi_1 = 5(10)^{-5}, \varrho = 10$ dB, $N_Q = 10, C = 100, S = 1$.

of a hit being blocked by a full queue. Note that the value of ρ for maximum overall detection probability doesn't change; in fact the entire shape of the curve appears to be similar for the two values of C, with only a vertical shift caused by the increase in P_b.

3. Decreasing the number of places in the queue from 10 to 5 while maintaining all other parameters causes pd_{ov} to fall off at a lower value of ρ. This is because the queue blocking probability becomes more significant at lower offered loads since the queue is shorter. As a result, the maximum for the overall detection probability is found at a lower value of ρ when N_Q is decreased.

4. Increasing the number of servers from 1 to 5 produces the highest overall detection probability. The slight decrease in queue availability (for example see figure 6.3) is more than compensated for by the increase in pd due to the increased rate.
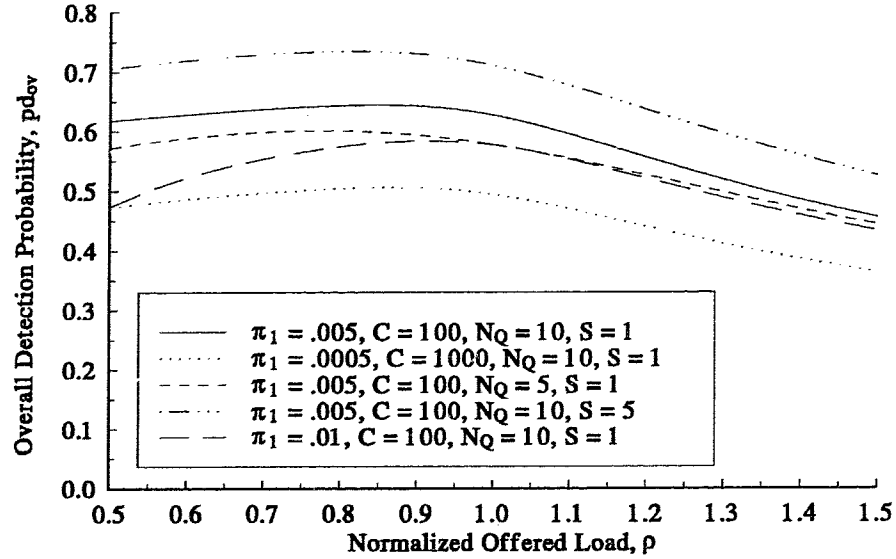
Figure 6.4: Overall Detection Probability for $\varrho = 10$ dB.

5. The largest change in shape for the overall detection probability results from doubling the target concentration from $\pi_1 = .005$ to $.01$. The increase in π_1 causes a sharper decrease-off in pd (as expected from the discussion in section 3.2.1). As a result, the value of ρ for maximum pd_{ov} is increased and in fact is the largest for any of the curves of figure 6.4.

It appears that the way to increase pd_{ov} is to increase the length of the queue. However, increased queue lengths lead to increased queuing delay. If a target remains in the queue too long, it may have moved by time the actuator gets around to serving it.

The time-average of the queue length is given by

$$L_Q = \sum_{i=S}^{N_Q} i\tilde{P}_i + \tfrac{1}{2}\xi_1 C \tag{6.6}$$

where the first term in (6.6) is the expected number of cells in the queue at the beginning of the service interval, and the second term is the time-average of the number of hits occurring
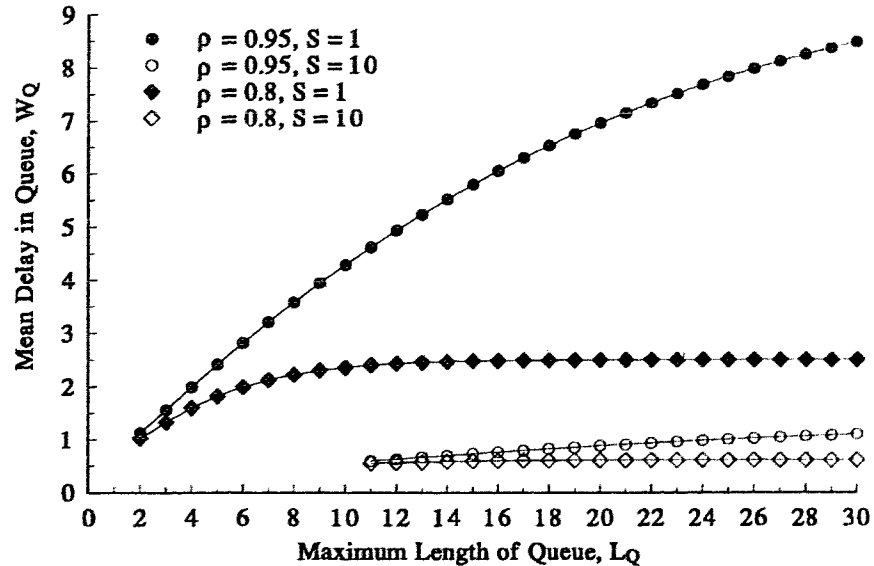
Figure 6.5: Mean delay for queue

during the service interval.

Then from Little's equation, the mean delay in the queue is given by [57]

$$W_Q = \frac{L_Q}{\lambda_S} \qquad (6.7)$$

In figure 6.5, W_Q is plotted as a function of L_Q for two values of S and ρ. As expected, the mean queue waiting time increases with N_Q. The effect is more prominent for larger values of ρ, when the queue is busier. For $S = 10$, the expected time in the queue is very short when compared to that for $S = 1$. Recall, however, that if we are comparing two systems with the same processing ability, the service times will be 10 times longer for the $S = 10$ system than for the $S = 1$ system.

In this section, we have considered some of the aspects of queuing to smooth out the variance in the hit-rate of a detection stage. We have assumed that all hits, whether they are caused by targets or false alarms, require exactly one processor cycle to service. In section 6.5 we extend the analysis to account for random service times, not necessarily the same for

both targets and false alarms. However, we now turn our attention to another technique of providing a fixed-rate output.

6.3 Rate-Constrained Detection by Ranking

In this section, we consider the detection performance of a test that operates without a threshold. Instead, the test picks the largest cells to pass onto the resource-constrained actuator. We assume that during each service interval the radar interrogates C cells, and passes on A for servicing.

In order to evaluate the performance of a system operating under a ranking system, we require the distribution of the cells, given their position in the ranking.

If there are a total of N_0 empty cells, the kth largest one has a density given by [58]

$$f_0^k(x) = k \binom{N_0}{k} (F_0(x))^{k-1} (1 - F_0(x))^{N_0-k} f_0(x) \tag{6.8}$$

where $f_0(x)$ and $F_0(x)$ are the probability density and cumulative distribution functions for the returns due to noise only. Similarly, the lth largest return out of a total of N_1 cells[4] containing targets has a density given by

$$f_1^l(x) = l \binom{N_1}{l} (F_1(x))^{l-1} (1 - F_1(x))^{N_1-l} f_1(x) \tag{6.9}$$

Now let

$$P_t(t) = P(\text{at least } t \text{ of the } N_1 \text{ targets are chosen}) \tag{6.10}$$

$$= P(l^{th} \text{ largest target} > k^{th} \text{ largest false alarm}) \tag{6.11}$$

$$= \int_{x_0=-\infty}^{\infty} f_0^k(x_0) \int_{x_1=-\infty}^{x_0} f_1^l(x_1) dx_1 dx_0 \tag{6.12}$$

where

$$l = N_1 - t + 1$$

and

$$k = N_0 + t - A$$

---

[4]Note $N_1 + N_0 = C$.

Then the probability of exactly t targets being chosen is given by

$$p_t(t) = \begin{cases} P_t(t) - P_t(t+1) & \text{for } t \leq A \\ P_t(A) & \text{for } t = A \end{cases} \tag{6.13}$$

The probability of detection, given that there are exactly N_1 targets is then

$$pd_{N_1} = \sum_{i=1}^{N_1} \frac{i}{N_1} p_t(i) \tag{6.14}$$

and to finally get the probability of detection, given an a priori target concentration, π_1, we average over all values of N_1:

$$pd = \sum_{N_1=1}^{C} \frac{\mathcal{B}_{N_1}^{C}(\pi_1) pd_{N_1}}{1 - \mathcal{B}_0^C(\pi_1)} \tag{6.15}$$

The denominator in (6.15) is included because in order to calculate pd, we assume the presence of at least one target.

We now become specific and consider the problem of detecting a Rayleigh target in Gaussian noise. The density and distribution under the noise-only hypothesis is

$$f_0(x) = x e^{-\frac{x^2}{2}} \tag{6.16}$$

and

$$F_0(x) = 1 - e^{-\frac{x^2}{2}} \tag{6.17}$$

Under the alternate hypothesis,

$$f_1(x) = \frac{x}{1+\varrho} e^{-\frac{x^2}{2(1+\varrho)}} \tag{6.18}$$

and

$$F_1(x) = 1 - e^{-\frac{x^2}{2(1+\varrho)}} \tag{6.19}$$

where ϱ is the signal-to-noise ratio. Of course, $x \geq 0$ for all of the above distributions and densities.

In figure 6.6, we have plotted pd as a function of $r = \frac{A}{C}$ for Rayleigh targets in Gaussian noise, along with the results from using a threshold. For each value of ϱ (indicated on the curves), there are two different curves drawn. The solid line is the log-ROC curve for a threshold detector implementing a rate-constrained rule with mean output rate $= r$. From
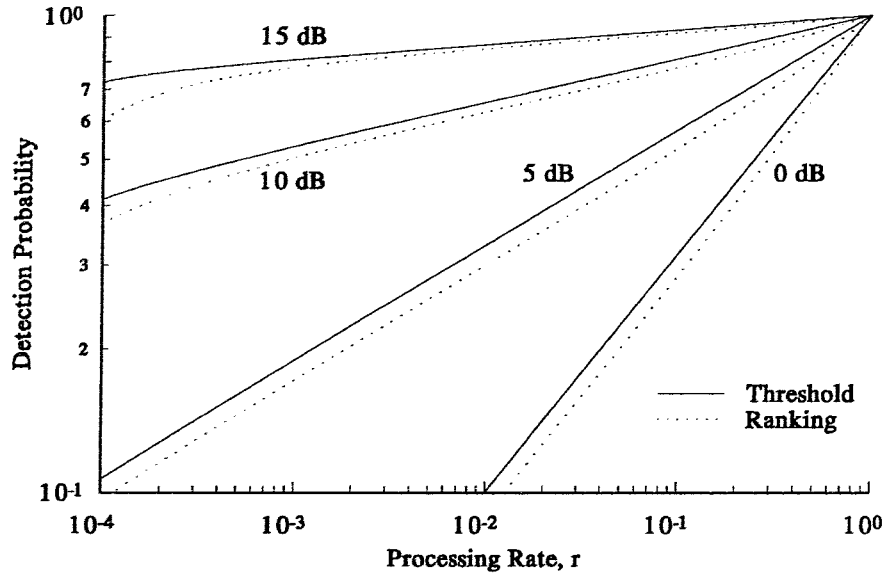
Figure 6.6: Comparison of Detection curves for ranking scheme and threshold detection for Rayleigh targets in Gaussian Noise. Target Concentration, $\pi_1 = 10^{-4}$.

the Rate-Constraint lemma, we know that the threshold must give the highest achievable pd for a given value of pf; hence ranking must result in performance loss. However, figure 6.6 shows that the loss is not very large, and decreases with increased ρ and increased r. The increased loss in pd with decreases in r results from the fact that the threshold curves fail to account for the need of a queue. If the queue blocking probability were included in the threshold curve, the two curves would be closer.

The ranking technique gives a rate-constrained output, with zero variance. No latency is introduced, since there is no queuing wait. However, it may seem that there is no significance to the hits, since the test always produces the same number of them. One technique to gain significance from the output of a rate-constrained test (if required) is to count the number of times a given cell is picked in consecutive scans. We now consider M/N detectors.

6.4 Coincidence Detectors

We have shown optimality of the rate-constraint criterion for detection systems with resource limited actuators. Consider now a system operating under the rate-constraint but without an actuator; the threshold crossings have no significance since there will always be the same number of hits out of the first stage, regardless of the actual distribution of targets in the detection field. While the hits denote the cells most likely to contain targets, further testing of some form is required to establish their significance. Such post-processing is not required under other detection strategies such as Neyman-Pearson; however, it often is included to improve detection performance (e.g. [59]).

In order to gain significance from the threshold crossings, we seek further evidence or support to confirm target existence. One form of support is coincidental crossings; if a cell is consistently ranked highly we expect it to be a target. Note that this form of support is temporal - we consider N consecutive scans and declare a target if a cell has M detections. However, we might also consider spatial support. For example, in an obstacle avoidance sonar, we expect an object to extend to more than one cell, so we might examine the cells adjacent to a potential target. In general, establishing any form of support requires resources in the form of processing time and therefore inflicts a rate-constraint. Gerlach and Andrews [60] recently have also discussed a similar trade-off between detector performance and calculation complexity.

Note that the combination of a ranking test with a M/N detector was first introduced in [17].

6.4.1 The M/N Detector

An M/N detector[5] is often used to initiate target tracks in a Track-While-Scan (TWS) radar system [61]. We will return to track initiators in section 6.6.

The following description of the operation of M/N rules is included for two reasons. First, before one can understand how M/N track initiators work under a rate-constraint (in section 6.6), it is important to understand how M/N detectors work in isolation. The second reason for the detail is to present a context for new results on the mean and standard

---

[5]The M/N detector has many names, including "Two-stage detector", "Sliding Window Detector" and "Binary Integrator"
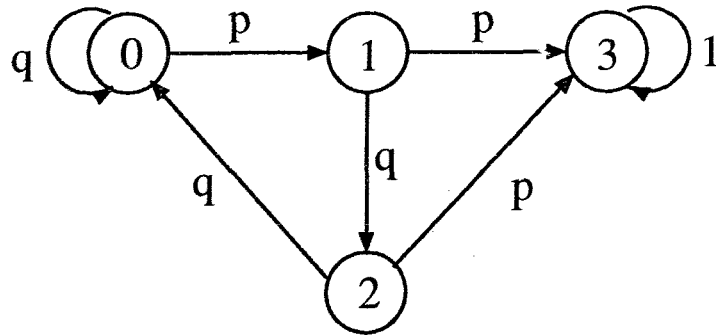
Figure 6.7: Markoff Chain For 2/3 Decision Rule

deviation of the time required to confirm a target. Previous approximate results given in [62] and quoted in [52] are extended.

The dynamics of an individual M/N detector may be analyzed using conventional Markoff chain techniques [61]. For example, consider the 2/3 rule as illustrated in figure 6.7. A target is confirmed if it is detected on two or more of three consecutive scans.

The equations governing the detectors are

$$
\begin{bmatrix} P(0) \\ P(1) \\ P(2) \\ P(3) \end{bmatrix}_k = \begin{pmatrix} q & 0 & q & 0 \\ p & 0 & 0 & 0 \\ 0 & q & 0 & 0 \\ 0 & p & p & 1 \end{pmatrix} \begin{bmatrix} P(0) \\ P(1) \\ P(2) \\ P(3) \end{bmatrix}_{k-1} \tag{6.20}
$$

$$
= \left(\begin{array}{c|c} A & 0 \\ \hline b & 1 \end{array} \right) \begin{bmatrix} \overline{P} \\ \hline P(3) \end{bmatrix}_{k-1} \tag{6.21}
$$

where p and q are the probability of a hit or no hit from the detector stage (i.e. p and q are either pd and $(1 - pd)$ or else pf and $(1 - pf)$ depending on whether the cell contains a target or a false alarm.) We have partitioned the state transition matrix into a square matrix A, row vector b, column vector of 0 and single entry 1 to simplify the analysis.

One of the parameters of interest for a detector is the mean delay or number of scans required to confirm a target, given a specific single look detection probability. The problem has been addressed by Castella in [62] for M/N rules, where $N \leq 4$. While Castella

solved the system of equations governing the tracker by using a recursive algorithm, classical Markoff chain methods provide a simpler, more accurate approach. We solve (6.21) for the probability of the track being in a confirmed state (i.e. in state 3):

$$Pc_k \;=\; b\overline{P_{k-1}} + Pc_{k-1} \tag{6.22}$$

$$=\; bA^{k-1}\overline{P_0} + Pc_k \tag{6.23}$$

where we have set $Pc_k = P(3)_k$, and $\overline{P_0} = \begin{pmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix}$ since the tracker begins in state 0 with

probability 1. Then the probability of the track being confirmed in exactly k cycles is given by

$$pc_k \;=\; Pc_k - Pc_{k-1} \tag{6.24}$$

$$=\; bA^{k-1}\overline{P_0} \tag{6.25}$$

Re-writing (6.23), the cumulative distribution function of (6.25) is given by

$$Pc_k \;=\; \sum_{l=1}^{k} bA^{l-1}\overline{P_0} \tag{6.26}$$

$$=\; b\left(\sum_{l=1}^{k} A^{l-1}\right)\overline{P_0} \tag{6.27}$$

$$=\; b(I - A^k)(I - A)^{-1}\overline{P_0} \tag{6.28}$$

Note that this assumes that the target statistics remain stationary for at least k cycles, which may be unrealistic for large k.

Figures 6.8 and 6.9 show the cumulative detection probability (CDP) as given by (6.28) for M/N detectors with various parameters.

In figure 6.8, the CDP is plotted for a 3/5 rule for various p. As expected, the CDP is an increasing function of p and k. If we are willing to wait long enough, (i.e. for k large enough), high values of the CDP are attainable even with values of the single look hit probability, p as low as 0.5. However, there are two problems with increasing k. First the CDP for false alarms as well as targets will increase; second and more importantly, the target statistics
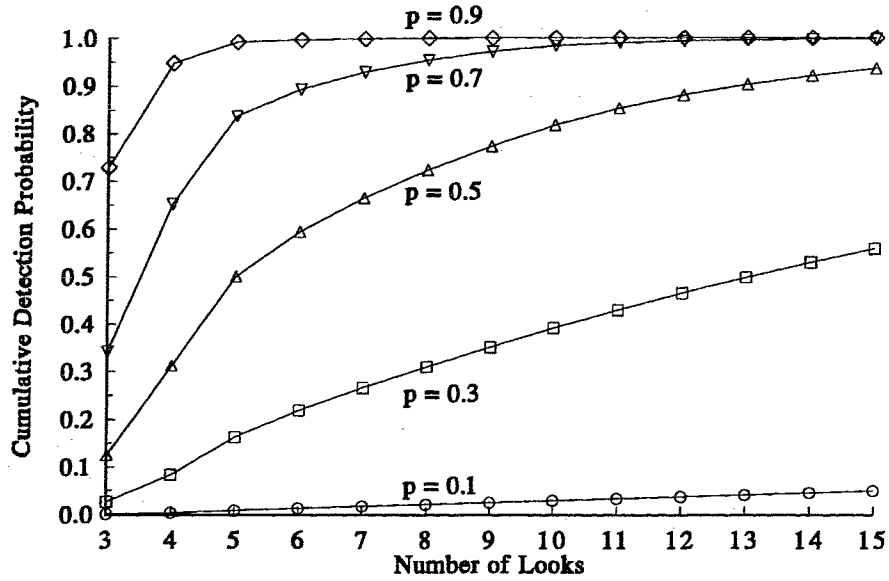
Figure 6.8: Cumulative Probability of detection for 3/5 Rule

may not remain stationary long enough for large k. (For example, the target may perform evasive maneuvers).

Figure 6.9 shows the effect of changing M with $N = 5$ and $p = 0.5$. As expected, the easier tests (i.e. with lower M) appear to perform better, although they also increase the CDP for false alarms.

We can find the mean and standard deviation of the number of looks required to confirm a target by taking moments of pc_k. The moment generating function of pc_k is

$$\Gamma_{pc}(z) = E\{z^k\} = \sum_{k=1}^{\infty} pc_k z^k = b \sum_{k=1}^{\infty} A^{k-1} z^k \overline{P_0}$$

$$= bz \sum_{k=0}^{\infty} (zA)^k \overline{P_0} = bz(I - zA)^{-1} \overline{P_0}$$

Then the mean number of looks required is given by

$$E\{k\} = \frac{d\Gamma(1)}{dz} = b(I - A)^{-2} \overline{P_0} \tag{6.29}$$
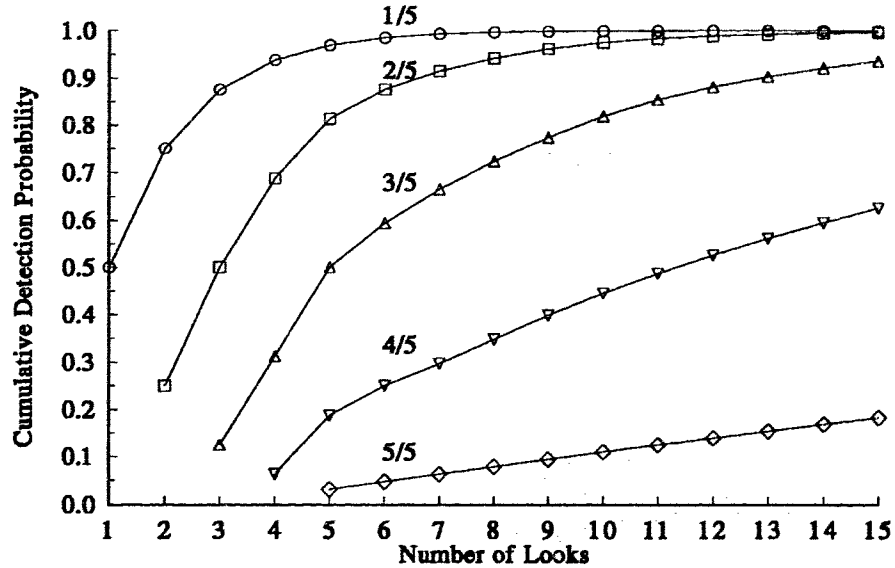
Figure 6.9: Cumulative Probability of detection for M/5 Rule with single look detection probability = 0.5

Now

$$\frac{d^2\Gamma(1)}{dz^2} = E\{k^2\} - E\{k\} = 2b(I - A)^{-3}A\overline{P_0}$$

which can be solved for $E\{k^2\}$, yielding

$$E\{k^2\} = b((I - A)^{-3}(I + A))\overline{P_0} \tag{6.30}$$

The variance is then given by $E\{k^2\} - (E\{k\})^2$.

This new method of solution gives the same results as Castella's; however its simplicity allows initiation rules with $N > 4$ to be easily analyzed given the state transition matrix governing the rule. We have developed a program to generate the state transition matrix for an arbitrary M/N rule and are therefore able analyze any detection strategy. Table 6.1 gives the mean and standard deviation of the number of returns required to confirm a target for various values of M, given specific values of p ($p = pd$ or pf).

As expected, the average delay increases as M increases, and as p decreases. However,

Table 6.1: Mean and Standard Deviation of number of looks required to confirm a target for various rules.

M/N	p=0.1	p=0.3	p=0.5	p=0.7	p=0.9
1/N	10,9.49	3.33,2.79	2,1.41	1.43,0.78	1.11,0.35
2/2	110.0,108.6	14.44,13.11	6,4.69	3.47,2.18	2.35,0.83
2/3	62.63,60.87	9.87,8.31	4.67,3.16	3.00,1.46	2.23,0.55
3/3	1110.0,1107.8	51.48,49.28	14,11.92	6.38,4.32	3.72,1.46
2/4	46.90,44.84	8.41,6.66	4.29,2.63	2.90,1.24	2.22,0.51
3/4	425.0,422.1	25.76,23.24	8.77,6.42	4.84,2.48	3.38,0.79
4/4	1105.3,1104.2	174.94,171.81	30,27.09	10.55,7.75	5.24,2.27
2/5	39.08,36.75	7.72,5.81	4.13,2.36	2.87,1.16	2.22,0.50
3/5	243.3,239.9	18.39,15.59	7.27,4.66	4.46,1.85	3.34,0.64
4/5	3227.5,3223.7	68.98,65.51	15.78,12.62	7.17,4.05	4.56,1.13
5/5	110619,110860	586.5,582.4	62,58.22	16.50,12.94	6.94,3.24
2/6	34.42,31.84	7.34,5.28	4.06,2.21	2.86,1.13	2.22,0.50
3/6	166.4, 162.6	15.11,12.05	6.64,3.79	4.35,1.57	3.33,0.61
4/6	1502.4,1497.8	41.98,38.23	11.80,38.23	6.26,2.82	4.46,0.80
5/6	26010.8,26017.3	189.9,185.5	27.95,23.98	10.15,6.32	5.79,1.55
6/6	$> (10)^6, > (10)^6$	1958.2,1953.1	126,121.3	25.00,20.66	8.82,4.41
2/7	31.34,28.54	7.11,4.93	4.03,2.12	2.86,1.11	2.22,0.50
3/7	125.9,121.8	13.34, 10.03	6.33,3.28	4.31,1.44	3.33,0.61
4/7	871.5,866.5	30.75,26.76,	10.08,6.39	5.93,2.20	4.45,0.72
5/7	10202,10199	100.0,95.31	18.93,14.76	8.39,4.21	5.60,1.01
6/7	216521,217321	536.2,530.8	49.39,44.56	14.03,9.48	7.09,2.06
7/7	$> 7(10)^6, > 7(10)^6$	6530.7,6524.6	254,248.4	37.14,31.99	10.91,5.79

as Castella also pointed out, the standard deviations are very large compared to the means, especially when p is small. Therefore the distributions of the delay are quite broad, so that the mean value must be considered as a general trend and not a precise design value. The results again stress the dangers of basing design and analysis on first-order statistics alone.

6.5   A Network of Asynchronous Actuators

Consider again figure 6.1. As before, assume that the detector examines a radar field containing C cells with an a priori target concentration of π_1[6]. Those cells crossing the threshold are passed to a network of N_A actuators, each of which may or may not finish with the cell it is currently servicing. We assume that the entire network is updated once per service cycle, and that each available actuator is loaded only at the end of discrete service cycles. The network is assumed to have been operating long enough so that the steady-state probabilities for the processes describing the individual actuators and the entire network are valid. We also assume that the service times of the servers are accurately described by binomial distributions.

We calculate the probability that a new target, coming into the equilibrium environment, will be detected and find an available actuator. Note that we calculate the single look detection probability (that is, the target gets only one chance to make it into the actuator network to be confirmed. We can make the cumulative detection probability arbitrarily close to 1 by increasing the number of looks allowed.)

Let $p_0 = P(\text{hit}|H_0) = \pi_0 pf$ be the probability of a given cell being empty and causing a hit, and $p_1 = P(\text{hit}|H_1) = \pi_1 pd$ be the probability of a cell containing a target and causing a hit. Now let pe_0 and pe_1 be the probability of a server busy with, respectively, a false alarm or a target finishing with the cell in question.

To analyze the network, we require two state variables, the number of actuators busy with false alarms and the number of actuators busy with targets. Let

$P(s = i, j)_k = P(i$ actuators busy with false alarms and k actuators busy with targets at time $k)$

Then the equations of the Markov process describing the network are given by

$$P(s = i, j)_{k+1} = A_{i,j,l,m} P(s = l, m)_k \tag{6.31}$$

---

[6]For the examples in this section, we consider larger values of π_1 and ϱ than in section 6.2, since it is expected that the stages of this section will be found towards the end of a multi-stage detection chain.

When the number of busy actuators is less than N_A,

$$A_{i,j,l,m} = \sum_{c_0=\max(0,l-i)}^{l} \mathcal{B}_{c_0}^{l}(pe_0)\mathcal{B}_{b_0}^{C-l-m}(p_0) \sum_{c_1=\max(0,m-j)}^{m} \mathcal{B}_{c_1}^{m}(pe_1)\mathcal{B}_{b_1}^{C-l-m-b_0}(p_1) \quad (6.32)$$

where

$$b_0 = i - l + c_0$$

and

$$b_1 = j - m + c_1$$

The interesting things, as far as analysis is concerned, happen when $i + j = N_A$, that is, when there are at least as many, and generally more, threshold crossings as actuators. Then

$$A_{i,j,l,m} = \sum_{c_0=\max(0,l-i)}^{l} \sum_{c_1=\max(0,m-j)}^{m} \sum_{x_0=\max(0,b_0)}^{C-l-m} \sum_{x_1=\max(0,b_1)}^{C-l-m-x_0} T_1 T_2 T_3 T_4 T_5 \quad (6.33)$$

where

$$T_1 = \mathcal{B}_{c_0}^{l}(pe_0)$$

is the probability of c_0 actuators busy with false alarms emptying,

$$T_2 = \mathcal{B}_{c_1}^{m}(pe_1)$$

is the probability of c_1 actuators busy with targets emptying,

$$T_3 = \mathcal{B}_{x_0}^{C-l-m}(p_0)$$

is the probability of x_0 new false alarms passing the threshold,

$$T_4 = \mathcal{B}_{x_1}^{C-l-m-x_0}(p_1)$$

is the probability of x_1 new targets passing the threshold, and

$$T_5 = \begin{cases} \dfrac{\dbinom{x_0}{b_0}\dbinom{x_1}{b_1}}{\dbinom{x_0+x_1}{b_0+b_1}} & \text{if } x_0 + x_1 \geq b_0 + b_1 \\ 1 & \text{else} \end{cases}$$

is the probability of a given cell finding an available actuator given that there are too many new potential cells.

Thus we can calculate the transition matrix and therefore the steady-state distributions of the Markoff process describing the network operation. Given that i actuators are busy with false alarms and j are busy with targets, the average carried load is given by

$$\lambda_S = \sum_{i,j} P_{ij} \left(\sum_{h_0=1}^{i} h_0 \mathcal{B}_{h_0}^i(pe_0) + \sum_{h_1=1}^{j} h_1 \mathcal{B}_{h_1}^j(pe_1) \right) \tag{6.34}$$

$$= \sum_{i,j} P_{ij}(i\,pe_0 + j\,pe_1) \tag{6.35}$$

Then as in (6.5), the probability of a cell not finding a server is given by

$$P_b = 1 - \frac{\lambda_S}{\xi_1 C} \tag{6.36}$$

The overall detection probability, given by

$$pd_{ov} = pd(1 - P_b) \tag{6.37}$$

is plotted in figure 6.10 as a function of $pe_0 = pe_1$ for various values of normalized output rate, ρ. (We will consider the effect of different pe_0 and pe_1 shortly). Since the signal-to-noise ratio, ϱ has been increased to 20 dB, the limiting factor in (6.37) is $(1 - P_b)$. In fact, we have included only curves for pd_{ov}, since the curves for $1 - P_b$ appear to be identical. The decrease in pd resulting from decreased ρ is overcome by the decrease in P_b. This is an important point: for large signal-to-noise ratios, the constraint on the resource becomes the crucial factor in the overall detection performance. As expected, pd_{ov} increases significantly with pe_0 and pe_1, since the actuators process the hits (and become available) more quickly.

We have assumed that the mean time to process targets and false alarms is identical in figure 6.10. To understand the effects of $pe_0 \neq pe_1$, consider figure 6.11. We have arbitrarily set $pe_0 = 1 - pe_1$ to illustrate the point. At $\rho = 0.2$, $p_1 \approx p_0$, and the decrease in pd_{ov} is symmetric in pe_0 and pe_1. As ρ increases, $p_0 > p_1$, so that overall detection probability is affected more by decreases in pe_0 than in pe_1. It is probably typical that $p_0 > p_1$. Therefore, it is important to use processors that do not waste too much time with false alarms.

In the next section we apply the above analysis to a network of trackers, each using a M/N rule for track initiation.

Figure 6.10: Overall Detection Probability as function of $pe_0 = pe_1$, for $\pi_1 = .005, C = 100, \varrho = 20$ dB, $N_A = 10$. The numbers on the curves are the values for ρ, the normalized output rate.

6.6   A Network of M/N Rules: A Track While Scan System

We now consider a simplified example consisting of a network of track initiators, each implementing a M/N rule. A tracker requires processing beyond that of incrementing the counter for the M/N processing - Kalman filtering, track association and other functions are required (these will not be discussed here). For our purpose of illustrating points relevant to the thesis, we consider only the dynamics of the M/N detectors. Equivalently we assume stationary targets or perfect position prediction from the Kalman filtering.

The first problem we must address is the definition of pe_0 and pe_1 for a tracker. We could assume that a target remains in the tracker for a given length of time (for example, the transit time across the air-space under surveillance) and then is deleted. However, for our purposes, we assume that the tracker is concerned only with the problem of track initiation. Once a track is determined, the tracker (on the next scan) hands off the target to another
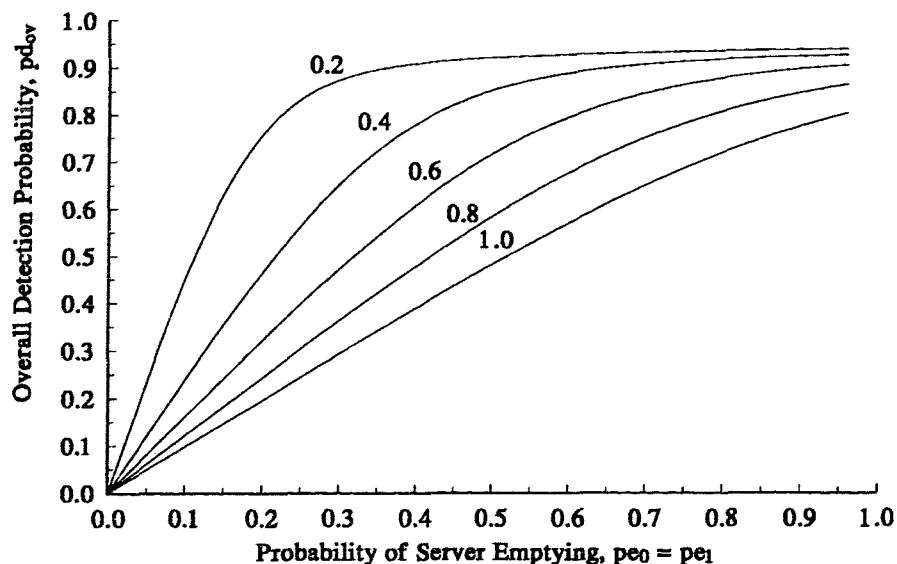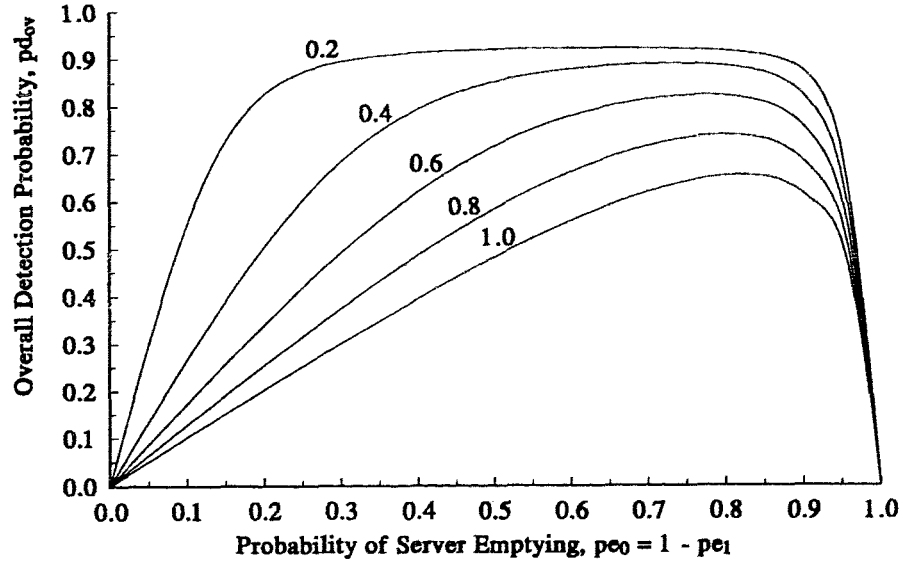
Figure 6.11: Overall Detection Probability as function of $pe_0 = 1 - pe_1$, for $\pi_1 = .005, C = 100, \varrho = 20$ dB, $N_A = 10$. The numbers on the curves are the values for ρ, the normalized output rate.

actuator (which we do not consider in this analysis), and becomes available again to process another potential track.

For example, consider the 2/3 decision rule of figure 6.7. In order to use it for our purposes here, the line with probability 1 from state 2 to itself must be replaced with a line from state 2 to state 0. The equations governing the trackers are then

$$
\begin{bmatrix} P(0) \\ P(1) \\ P'(2) \\ P(3) \end{bmatrix}_k = \begin{pmatrix} q & 0 & q & 1 \\ p & 0 & 0 & 0 \\ 0 & q & 0 & 0 \\ 0 & p & p & 0 \end{pmatrix} \begin{bmatrix} P(0) \\ P(1) \\ P(2) \\ P(3) \end{bmatrix}_{k-1} \tag{6.38}
$$

The equilibrium distribution of the tracker states may be found from the steady-state solution to (6.38), (i.e. as $k \rightarrow \infty$). We separately calculate the steady-state probabilities for a tracker under the two hypotheses.

The probability of a busy tracker becoming available (i.e. returning to state 0) can then be calculated given the steady-state probabilities from

$$pe_0 = \sum_{s \neq 0} P(\text{in state } s | H_0) P(\text{transition to state } 0 | \text{ in state } s, s \neq 0, H_0) \qquad (6.39)$$

where $P(\text{in state } s | H_0)$ is given from the steady-state probabilities of (6.38) when no target is present, and $P(\text{transition to state } 0 | \text{ in state } s, s \neq 0, H_0)$ is given by the normalizing the $(0, s)^{th}$ entries of the transition matrix in (6.38) over $s \neq 0$. Note that pe_1 is defined in a similar manner, but with a target assumed present.

Having found pe_0 and pe_1 for an individual tracker, we are now ready to consider the dynamics of a network of trackers.

The probability that a given cell will cross the threshold and will be confirmed in the following $N - 1$ cycles is given by[7]

$$P_{c1} = p \sum_{l=M-1}^{N-1} \binom{N-1}{l} p^l (1-p)^{N-1-l} \qquad (6.40)$$

The probability that a given cell will pass the first test and subsequently be rejected is given by $p - P_{c1}$. Finally, we are in a position to calculate the single look detection probability of a target; it is the product of the single look confirmation probability as given by (6.40) and the single look tracker availability given by (6.36):

$$pd_{ov} = P_{c1}(1 - P_b) \qquad (6.41)$$

The overall detection probability is plotted in figure 6.12 for a 20 dB Rayleigh target in Gaussian noise under various conditions. There are a number of things to note:

1. The less selective rule (2/5) is superior at lower ρ; however the values over which it is superior are for values of pd_{ov} which are too low to be useful. At higher ρ values, more false alarms are let in, and the 2/5 rule takes longer to process them than the 3/5.

2. Decreasing π_1 while maintaining the expected number of targets, $\pi_1 C$ constant provides better performance over all values of ρ. This is because decreasing π_1 increases pd, as discussed in section 3.2.1.

---

[7]Note that this is not quite the same as letting $k = N$ in equation (6.25), which doesn't assume a threshold crossing in the first look.
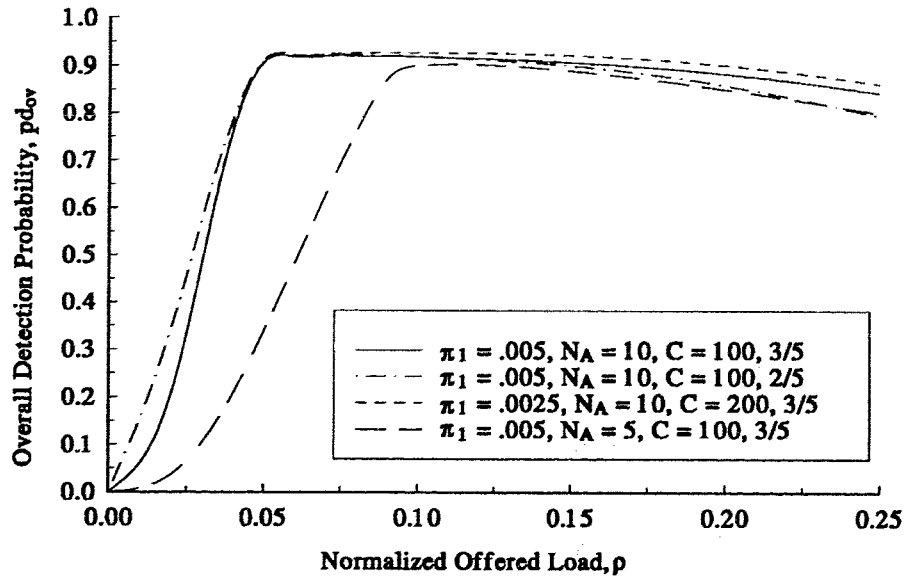
Figure 6.12: Overall Detection Probability for 20 dB Rayleigh targets in Gaussian Noise

3. Decreasing the number of trackers, N_A causes the peak in pd_{ov} to be shifted to a larger value of ρ. This is because pd is the factor causing the crash in pd_{ov}. The value for ξ_1 for the crash in pd_{ov} is the same, but the normalized rate ρ value, when normalized by N_A, is shifted outwards.

In general, we expect to have multiple chances to detect a target. The detection probability, given that we have L looks (the maximum valid value of L will be a function of the target dynamics) is a complicated function of the time it takes the tracker to reach a decision (which is a function of the tracker initiation and deletion rules) and the tracker availability function.

The probability of a tracker emptying is a function of how long the tracking rule takes to confirm a target or to rule out a false alarm. The number of cycles required to make a decision will depend on the decision made. For a detection, the decision delay is between

M and N and is given by

$$\tau_1 = \sum_{l=M}^{N} l \frac{p^M(1-p)^{l-M}}{P_{c1}} \begin{pmatrix} l-2 \\ M-2 \end{pmatrix} \tag{6.42}$$

where as for the no detection case, the decision delay is between $N - M + 1$ and N and is given by

$$\tau_0 = \sum_{l=N-M+2}^{N} l \frac{(1-p)^{N-M+1}p^{l-(N-M-1)}}{p - P_{c1}} \begin{pmatrix} l-2 \\ N-M \end{pmatrix} \tag{6.43}$$

Both of these delays must be less than the time over which the statistics of the environment significantly change. Clearly τ_1 will be a decreasing function of pd and τ_0 will be an increasing function of pf. The exact values for the mean processing delays are, as we learned in section 6.4, only guidelines.

The effect of the track initiation rule is seen figure 6.12. For the 2/5 rule, between four and five returns are required for a tracker to rule out a false alarm, assuming the first return is a hit. However, the 3/5 rule requires only between three and five returns. Of course, there is a trade-off involved; a 3/5 rule will take longer to detect targets, and may miss small targets.

6.7 Conclusions

We have covered a lot of material in this chapter. Beginning with a simple queue, where targets and false alarms are assumed to be treated identically by the actuator, we showed that the value of ρ, the normalized output rate, corresponding to the maximum overall detection probability is a function of target concentration, the maximum queue length, the number of actuators and the number of cells scanned per actuator cycle. For low values of π_1, the overall detection probability finds it maximum somewhere $\rho \approx 0.8 \rightarrow 1$, and is a relatively flat function over that range. As the a priori probability of a target increases, the effect of reducing the output rate from the detector stage becomes more significant, and the resulting overall detection probability is not so flat around its maximum. Increasing the maximum queue length increases the overall detection probability, but may increase the mean queuing delay beyond limits set by the rate of change of the environment. Increasing the number of servers also increases pd_{ov}, but may be prohibitively expensive.

One way of providing a guaranteed rate out of a test is to eliminate the threshold altogether and rank the returns, picking the largest to be passed on to the actuator. Such techniques introduce losses when compared to threshold tests, but the losses are not large and decrease with increased signal-to-noise ratio.

If one wants to derive significance from a test that provides a constant rate, a M/N detector may be used. New results for the mean and variance of the time required to confirm a target were presented and underscore the danger of merely quoting means, since the variances are very large for strict tests when the single-look detection probability is low.

Often targets and false alarms require different amounts of processing by the processor. The Markoff model considered stresses the importance of designing systems that do not waste excessive time dealing with false alarms.

Chapter 7

Conclusions

Given the relative maturity of detection theory, it is perhaps surprising that the resource-constraint theory has not been promoted earlier. In contrast to existing detection theory, which either pre-supposes more a priori information than is typically available in the radar problem (as in cost function approaches) or else distances the theory from practical measures altogether (as in the Neyman-Pearson approach), resource-constrained detection is both practically motivated and practically implemented. It manages, in the presence of ignorance of the underlying statistics, to achieve adequate and quite often optimal detection performance.

The theory arises from looking anew at the problem of detection, and realizing that detectors almost always are, in fact, attention-directing devices for resource limited actuators. By returning to the most basic level of hypothesis testing, we have gained an understanding of how detection systems that cue resource-constrained actuators should be operated.

The form of the optimal rate-constrained test is a likelihood ratio test, with its threshold set to match the hit rate out of the detector to the processing rate of the subsequent actuator. The resulting test is practical, since the hit-rate is both controllable and observable, unlike the false alarm rate which is only controllable. No a priori information (other than the processing rates of the system stages, which are design parameters) is required to set the threshold, and there is no need to arbitrarily choose a design false alarm rate or cost function.

A new relational measure, the SLOC function, allows us to understand how various tests should be cascaded together for optimal performance. There is no need for the individual tests to be identically distributed, or synchronous. A dynamic programming argument

showed that, provided the SLOC numbers of the individual tests are properly ordered, the overall detection performance for a multi-stage system operating under any sensible criteria is maximized when each stage but the last is operated under the rate-constraint. When the tests are not properly ordered, the resulting detection probability will be less than could be achieved in theory, with perfect knowledge of the radar context (distributions, target concentration etc.); however, at least the test can still be implemented.

We have presented SLOC curves for a number of radar target models in Gaussian noise, and for Rayleigh fading and constant targets in K-Clutter. The shapes of the curves vary considerably for the different models, and for noncoherent and coherent integration. The SLOC function is a decreasing function of the signal-to-noise ratio (SNR); therefore faster, lower SNR tests which cue slower, higher SNR tests will always have at least some region of operation where the rate-constraint is optimal. A new processing architecture consisting of a noncoherent test cueing a coherent test has been shown to achieve significant reductions in processing complexity, with only moderate costs in detection performance.

Resource-constraint theory, along with the rest of decision theory, is based on first order statistics. We have presented techniques for dealing with the variance inherent in the output rate of a test. Ranking the returns and picking the largest to be processed is a viable option, introducing small detection losses relative to use of a threshold. Alternately, a queue may be introduced between the detection stage and following processor. There is a trade-off between the output rate of the test normalized to the maximum processing rate of the actuator following the test, and the probability of a hit finding a place in the queue. The length of the queue may be increased to decrease the probability of a full queue; however, the mean queuing delay may then be excessively long given the stationarity limits on the target and radar environment. A new analytical framework has been introduced for investigating the detection performance of a system consisting of a number of actuators which require different processing loads for false alarms than targets. The results underscore the need for algorithms that are able to quickly dispense with false alarms.

Because of its dependence on the processing rate of the subsystems, resource-constraint theory is a useful tool for both radar systems design and analysis. We have presented examples of the theory's application at three different system levels: at the command and control level in the obstacle avoidance example of section 3.4, at the data processing level in the track-while-scan example of section 6.6, and at the signal processing level in the

noncoherent/coherent integration system of section 5.3. The rate-constraint criterion has been used by radar designers previously, but without the analytical justification provided by our work.

7.1 General Applications of the Theory

In this thesis we have emphasized the application of resource-constraint theory to radar detection theory. However, the principles are applicable in a diversity of fields. Hierarchical decision techniques are often used in large systems such as corporations, in medical diagnoses and in the detection systems of natural organisms.

Armed with an understanding of what characteristics to look for in the decision sequence, we can design them into the process. (i.e. What tests should be applied, and in which order?)

For example, consider a company setting out to hire a new employee for a popular position in an age of rampant unemployment. The personnel department has finite resources to apply in its search. The first test should be able to screen many applicants, and at minimal cost. For example, if the position is for a recent graduate, the Grade Point Average may be used as a simple test statistic. The threshold is simply varied to permit the required number of applicants to make the first cut. The next stage could be an aptitude test, with the test results used as an indicator of suitability. The threshold applied to the test again depends on the number of applicants that can be screened by the next stage. The final stage might consist of an interview[1], at which point the exact "test statistic" to apply becomes subjective. Presumably the interviewer is the best judge of an applicants suitability.

Note how well the personnel department's procedure fits the resource constraint model. Each test provides more specific information regarding a candidate's potential, but requires more resources than the previous test.

Clearly resource-constrained detection provides more than just insight into how best to operate a series of radar sub-systems; it provides a new paradigm for multi-stage detection processes operating with limited processing resources. By looking afresh at the problem through the lens of practical application, and not being satisfied with statistical approaches that require more information than is available, we have developed a needed addition to detection theory.

---

[1] Indeed a series of interviews could be used.

Bibliography

[1] J. Neyman and E. Pearson, "On the problem of the most efficient tests of statistical hypotheses," *Phil. Trans. Roy. London Series A*, vol. 231, pp. 289–337, 1933.

[2] I. Good, *Probability and the Weighing of Evidence*. Griffin, 1950.

[3] L. Savage, *The Foundations of Statistics*. Wiley, 1954.

[4] P. Woodward, *Probability and Information Theory, with Applications to Radar*. Pergamon Press Ltd., 1953.

[5] D. Middleton, *An Introduction to Statistical Communications Theory*. McGraw-Hill, 1960.

[6] G. Trunk, *Radar Handbook, 2 ed. (edited by M. Skolnik)*, ch. 8. McGraw-Hill, 1990.

[7] P. Bogler, *Radar Principles with Applications to Tracking Systems*. Wiley, 1990.

[8] Z. Kurniawan and P. McLane, "Parameter optimization for an integrated radar detection and tracking system," *IEE Proceedings, Part F*, vol. 132, pp. 36–44, Feb. 1985.

[9] H. Resnikoff, *Illusions of Reality*. Springer-Verlag, 1989.

[10] R. Conant, "Laws of information which govern systems," *IEEE Trans. on Systems, Man, & Cybernetics*, vol. SMC-6, pp. 240–255, Apr. 1976.

[11] S. Hovanessian, *Introduction to Sensor Systems*. Artech House, 1988.

[12] C. Metz, "Basic principles of ROC analysis," *Seminars in Nuclear Medicine*, vol. VIII, pp. 283–298, Oct. 1978.

[13] P. Schweizer, W. Petlevich, P. Haley, and J. Oravec, "Image processing architecture for AUV mine-hunters," *Sea Technology*, pp. 55–61, Apr. 1992.

[14] D. Norman and D. Bobrow, "On data-limited and resource-limited processes," *Cognitive Psychology*, vol. 7, pp. 44–64, 1975.

[15] C. Lee and J. Chao, "Optimal local decision space partitioning for distributed detection," *IEEE Trans. on Aerospace & Electronic Systems*, vol. 25, pp. 536–543, July 1989.

[16] J. Bird, "Calculating the performance of linear and square-law detectors," *Submitted to IEEE Trans. on Aerospace & Electronic Systems*, 1992.

[17] G. Dillard and C. Antoniak, "A practical distribution-free detection procedure for multiple-range-bin radars," *IEEE Trans. on Aerospace & Electronic Systems*, vol. AES-6, pp. 629–635, Sept. 1970.

[18] F. Nathanson, *Radar Design Principles, 2 ed.* McGraw-Hill, 1991.

[19] M. Cherikh and P. Kantor, "Counterexamples in distributed detection," *IEEE Trans. on Information Theory*, vol. 38, pp. 162–165, January 1992.

[20] H. Poor, *An Introduction to Signal Detection and Estimation.* Springer-Verlag, 1989.

[21] J. Di Franco and W. Rubin, *Radar Detection.* Prentice-Hall, 1968.

[22] M. Denn, *Optimization by Variational Methods.* McGraw-Hill, 1969.

[23] J. Egan, *Signal Detection Theory and ROC Analysis.* Academic Press, 1975.

[24] L. Scharf, *Statistical Signal Processing: Detection, Estimation and Time Series Analysis.* Addison-Wesley, 1991.

[25] J. Proakis, *Digital Communications.* McGraw-Hill, 1983.

[26] I. Hoballah and P. Varshney, "An information theoretic approach to the distributed detection problem," *IEEE Trans. Information Theory*, vol. IT-35, pp. 988–994, Sept. 1989.

[27] R. Schuchard and R. Massof, "Decision making with clinical diagnostic tests in terms of statistical decision theory and information theory," *Applied Optics*, vol. 29, pp. 1436–1444, Apr. 1990.

[28] F. Kanaya and K. Nakagawa, "On the practical implication of mutual information for statistical decision making," *IEEE Trans. on Information Theory*, vol. 37, pp. 1151–1155, July 1991.

[29] R. Blahut, *Principles and Practice of Information Theory*. Addison Wesley, 1987.

[30] M. Skolnik, *Introduction to Radar Systems*. McGraww-Hill, 1962.

[31] D. Schleher, *MTI and Pulsed Doppler Radar*. Artech House, 1991.

[32] A. Wald, *Sequential Analysis*. Dover, 1973.

[33] J. Marcum, "A statistical theory of target detection by pulsed radar," *IRE Trans. Information Theory*, vol. IT-6, pp. 145–267, Apr. 1960.

[34] R. Lacoss, "Distributed mixed sensor aircraft tracking," in *Proc. of 1987 American Control Conf.*, pp. 1827–1830, 1987.

[35] R. Viswanathan, S. Thomopoulos, and R. Tumuluri, "Optimal serial distributed decision fusion," *IEEE Trans. on Aerospace & Electronic Systems*, vol. AES-24, pp. 366–375, July 1988.

[36] S. Thomopoulos, R. Viswanathan, and D. Bougoulias, "Optimal distributed decision fusion," *IEEE Trans. on Aerospace & Electronic Systems*, vol. AES-25, pp. 761–765, Sept. 1989.

[37] P. Rothman and R. Denton, "Fusion or confusion: knowledge or nonsense?," *SPIE Data Structures and Target Classification*, vol. 1470, pp. 2–12, 1991.

[38] G. Nemhauser, *Introduction to Dynamic Programming*. Wiley and Sons, 1966.

[39] L. Mitten, "Composition principles for synthesis of optimal multistage processes," *Operations Research*, pp. 610–619, 1964.

[40] P. Swerling, "Probability of detection for fluctuating targets," *IRE Trans. Information Theory*, vol. IT-6, pp. 296–308, Apr. 1960.

[41] J. Jao and M. Elbaum, "First-order statistics of a non-Rayleigh fading signal and its detection," *Proc. IEEE*, vol. 66, pp. 781–789, July 1978.

[42] P. Scholefield, "Statistical aspects of ideal radar targets," *Proc. IEEE*, vol. 55, pp. 587–590, Apr. 1967.

[43] N. Levanon, *Radar Principles*. Wiley, 1988.

[44] C. J. Barrett, *Principles of Modern Radar (edited by J.L. Eaves and E.K. Reedy)*, ch. 11. Van Nostrand Reinhold Co. Inc., 1987.

[45] D. Meyer and H. Mayer, *Radar Target Detection - Handbook of Theory and Practice*. Academic Press, 1973.

[46] K. Crump, "Numerical inversion of Laplace transforms using a Fourier series approximation," *Journal of the Association of Computing Machinery*, vol. 23, Jan. 1976.

[47] J. Shapiro, R. Reinhold, and D. Park, "Performance of peak detecting laser radars," *Proc. SPIE*, vol. 663, pp. 38–56, 1986.

[48] W. Press, B. Flannery, S. Teukolsky, and W. Vetterling, *Numerical Recipes in C: The Art of Scientific Computing*. Cambridge University Press, 1988.

[49] J. Jao, "Amplitude distribution of composite terrain radar clutter and the K-Distribution," *IEEE Trans. on Antennas & Propagation*, vol. AP-32, pp. 1049–1062, Oct. 1984.

[50] E. Jakeman and P. Pusey, "A model for non-Rayleigh sea echo," *IEEE Trans. on Antennas & Propagation*, vol. AP-24, pp. 806–814, Nov. 1976.

[51] R. Fante, "Detection of multiscatter targets in K-Distributed clutter," *IEEE Trans. on Antennas & Propagation*, vol. AP-32, pp. 1358–1363, Dec. 1984.

[52] A. Farina and F. Studer, "A review of CFAR detection techniques in radar systems," *Microwave Journal*, pp. 115–128, Sep. 1986.

[53] H. Finn and R. Johnson, "Adaptive detection mode with threshold control as a function of spatially sampled clutter level samples," *RCA Review*, vol. 29, pp. 414–464, Sept. 1968.

[54] R. Nitzberg, "Clutter map CFAR analysis," *IEEE Trans. on Aerospace & Electronic Systems*, vol. AES-22, pp. 419–421, July 1986.

[55] J. Echard, "Estimation of radar detection and false alarm probabilities," *IEEE Trans. on Aerospace & Electronic Systems*, vol. 27, pp. 255–260, Mar. 1991.

[56] W. Chu, "Buffer behaviour for Poisson arrivals and multiple synchronous constant outputs," *IEEE Trans. on Computers*, vol. C-19, pp. 530–534, June 1970.

[57] J. Little, "A proof of the queuing formula L = λW," *Operations Research*, vol. 9, pp. 383–387, 1961.

[58] A. Papoulis, *Probability, Random Variables, and Stochastic Processes*. McGraw-Hill, 1984.

[59] J. Saniie, K. Donohue, and N. Bilgutay, "Order statistic filters as postdetection processors," *IEEE Trans. Acoustics, Speech and Signal Processing*, vol. ASSP-38, pp. 1722–1732, October 1990.

[60] K. Gerlach and G. Andrews, "Cascaded detector for multiple high-prf pulse doppler radars," *IEEE Trans. Aerospace & Electronic Systems*, vol. AES-26, pp. 768–773, September 1990.

[61] S. Blackman, *Multiple-Target Tracking with Radar Applications*. Artech House, 1986.

[62] F. Castella, "Sliding window detection probabilities," *IEEE Trans. on Aerospace & Electronic Systems*, pp. 815–819, Nov. 1976.