10-31-1992

An information theoretic approach to DSN evaluation

Vivek Mehrotra
New Jersey Institute of Technology

An Information Theoretic Approach to DSN Evaluation

by
Vivek Mehrotra

A Thesis
Submitted to the Faculty
of New Jersey Institute of Technology
in Partial Fulfillment of the Requirements for the Degree of Master of Science
Department of Electrical and Computer Engineering
October 1992

APPROVAL SHEET

An Information Theoretic Approach to DSN Evaluation

by

Vivek Mehrotra

Dr. Nirwan Ansari, Thesis Advisor
Assistant Professor of Electrical and Computer Engineering,
New Jersey Institute of Technology

Dr. Joseph Frank, Committee Member
Associate Professor of Electrical and Computer Engineering,
New Jersey Institute of Technology

Dr. Edwin Hou, Committee Member
Assistant Professor of Electrical and Computer Engineering,
New Jersey Institute of Technology

BIOGRAPHICAL SKETCH

Author: Vivek Mehrotra

Degree: Master of Science in Electrical Engineering

Date: October, 1992

Undergraduate and Graduate Education:

- Master of Science in Electrical Engineering, New Jersey Institute of Technology, Newark, NJ, 1992
- Bachelor of Engineering in Electrical and Electronics Engineering, Birla Institute of Technology and Science, Pilani, India, 1990

Major: Electrical Engineering

ABSTRACT

An Information Theoretic Approach to DSN Evaluation

by

Vivek Mehrotra

Evaluation of Distributed Sensor Networks (DSN's) for optimal detection using Bayesian cost formulation methods has been the objective of several studies. There have been a few studies on DSN evaluation using an information theoretic approach, wherein an assymetric channel models a detector. We look at the multi-sensor system at the receiver as a black-box and model it as an assymetric channel whose cross over probabilities depend on the probabilties of detection and of false alarm. These probabilities in turn depend on the thresholds of the local detectors and on the fusion rule used. We consider the case wherein the receiver has control over the value of the probability of the signal being present, by influencing transmitter coding. The probability to be used at the receiver is the one that solves the MED (Minimum Equivocation Detetion) problem. Minimizing the equivocation between the input and output is the same as maximizing the mutual information. The input probability, P_0, at the receiver that maximizes the mutual information is then achieved by having an encoder between the transmitter and the multi-sensor system at the receiver. The only factor in encoding is the proportion of 0's to 1's. Variable length codes with different ordering of 0's and 1's are possible to get the same performance. Results show that as we move towards an optimum ROC (Receiver Operating Characteristic) curve, the mutual information attains a higher value. The value of P_0 that helps attain this value is then derived by encoding the transmitter output.

ACKNOWLEDGEMENT

I would like to express my sincere gratitude to my advisor Prof. Nirwan Ansari for his guidance and friendship and for always motivating me to reach higher standards. I would like to thank him also for leading me to areas I had never known before.

Special thanks are due to Prof. Edwin Hou for providing me partial financial support.

I appreciate the friendly advice and suggestions from Prof. Zoran Siveski in doing my research.

I thank Prof. Joseph Frank and Prof. Edwin Hou for serving as members of the committee.

TABLE OF CONTENTS

1 INTRODUCTION 1

1.1 Literature Survey 1

1.2 Organization of the Report 4

2 BAYESIAN METHODS FOR DSN EVALUATION 5

2.1 Non-Optimal Solutions 7

2.2 Globally Optimal Solution 9

3 INFORMATION THEORETIC VIEWPOINT OF DETECTION THEORY 11

3.1 Common Grounds in Detection Theory and Information Theory 11

3.2 Controlling P_0 and P_1 at the Receiver to Maximize $I(H;u)$. 17

3.2.1 Why Control P_0 and P_1 at the Receiver 17

3.2.2 Encoding 18

4 RESULTS AND DISCUSSIONS 25

5 CONCLUSIONS 33

LIST OF FIGURES

2.1 Distributed Sensor System with Data Fusion. 6

3.1 (a) A Classical Binary Detection System. 12

3.1 (b) A Classical Information Transmission System. 12

3.2 Channel Modelling Multi-sensor System at the Receiver. 13

3.3 Complete System to Maximize Mutual Information 19

4.1 A Plot of $I(H;u)_{max}$ Against P_d and P_f. 29

4.2 Mutual Information vs P_0 30

4.3 Mutual Information vs P_0 . . . 31

4.4 Mutual Information vs P_0 32

CHAPTER 1
INTRODUCTION

A *Distributed Sensor Network* (DSN) can be defined as a set of spatially scattered intelligent sensors designed to obtain measurements from the environment, abstract relevant information from the data gathered, and to derive appropriate inferences from the data gathered. Distributed sensor networks depend on multiple processors to simultaneously gather and process information from many sources. Interest in these systems stems from a realization of the limitations imposed by relying on a single source of information to make decisions.

In recent years, there has been an increasing interest in distributed sensor systems. This interest has been sparked by the requirements of military surveillance systems. Signal detection with multiple sensors can be performed in two manners In the traditional method, the local sensors communicate all observations directly to a central detector where decision processing is performed. This method often requires a large bandwidth for the communication channel in order to obtain real-time results. In the second method, each sensor has an associated detector which decides locally whether a signal is detected or not. The local decisions are sent to a fusion center where they are combined for global decision making. This method does not require the large bandwidth of the first method. However, performance is degraded because the central processor does not receive all the information The advantages in cost, reliability, and communications bandwidth, however, may outweigh the loss of performance.

1.1 Literature Survey

Tenney and Sandell[1] were among the first to study the problem of detection with distributed sensors. Their analysis follows classical Bayesian theory. Using the minimum cost criterion, they determine that for a fixed fusion rule, the optimal structure for the local detectors is the likelihood ratio test when local observations are independent. Optimal thresholds of the local detectors are shown to be related by a set of coupled nonlinear algebraic equations. They did not however explicitly tackle the problem of fusion of these decisions in terms of development of optimal data fusion algorithms.

Chair and Varshney[2], using the minimum cost criterion, optimize the fusion center assuming fixed thresholds of the local detectors and independent local decisions. Individual decisions are weighted according to their reliability, i.e., the weights are a function of the probability of miss and the probability of false alarm of the individual detection, and comparing the weighted sum of these decision probabilities against a likelihood threshold to derive the global decision. A different type of generalization of the analysis was presented independently by Sadjadi[3]. He offered an optimum solution for the general case of m-ary hypotheses testing for n sensors but without explicit optimal fusion. A nonparametric centralized object recognition scheme is proposed by Demirbas[4]. This scheme uses object features collected by several sensors. Recognition is performed by a binary decision tree generated from a training set. This proposed scheme does not assume the availabilty of any probability density functions. The study by Reibman and Nolte[5] extended the previous studies[2][1] by simultaneous optimization of the local detectors and the global fusion processor. They show that globally optimal performance for the distributed detector system is obtained if neither the local thresholds nor the the fusion rule are set *a priori* but are chosen according to the criterion of optimality. This approach requires finding the minimum cost solution to a set of coupled nonlinear equations. The resulting fusion processor is a k-out-of-n logical function, where the value of k is not fixed a priori, but is chosen according to the optimality criterion. Krzysztofowicz and Long[6] formulated a Bayesian detection model for a distributed system of sensors, wherein each sensor provides the central processor with a detection probability rather than an observation vector or a detection decision.

In statistical decision theory, a variety of criteria are used for the optimization of detectors. The above studies attempt to optimize the system performance by Bayesian cost formulation. In applications where such costs are available and meaningful, they provide an excellent choice for system optization. In this, a fixed cost is assigned to each possible course of action, and the average cost is then minimized. However, in some applications, our interest is to maximize the amount of information transfer from the input side to the output side.

Our study investigates the problem of sensor fusion from an information theoretic point of view. Middleton[7] and Gabriele[8] used such a criterion for the design of an optimum decision system where they minimized the equivocation (or information loss) between the input and output. Gabriele offered a method of imposing information criteria for threshold setting in simple binary hypothesis tests. More recently, Hoballah and Varshney[9] proved that the detector that maximizes information transfer is a threshold detector. Further, they used the minimum equivocation criterion to derive optimum thresholds and fusion rules for the classical single detector and for two distributed detection topologies.

In our thesis, we consider the case where the respective input probabilities of the hypotheses at the receiver can be controlled by influencing transmitter coding. In such a case, our goal is to maximize the information transfer across a sensor system at the receiver.

1.2 Organization of the Report

In Chapter 2, we talk about the classical Bayesian methods used in characterizing DSN's. In Chapter 3, an information theoretic approach to the distributed detection problem is discussed. In Chapter 4, maximization of the mutual information over *a priori* probabilities at the receiver is done. Results are presented too. The report ends with conclusions in Chapter 5.

CHAPTER 2
BAYESIAN METHODS FOR DSN EVALUATION

The binary hypothesis testing problem for n sensors can be formulated[5] as follows: The two hypothesis, H_0 and H_1, have *a priori* probabilities P_0 and P_1 respectively, where

H_0: $\mathbf{x}_i = \mathbf{n}_i$

H_1: $\mathbf{x}_i = \mathbf{s} + \mathbf{n}_i$

for $i = 1, \ldots, n$.

The observation vectors $\mathbf{x}_i$ are statistically independent from sensor to sensor. The observation vector at local sensor i, $\mathbf{x}_i$, has joint probability density function $P(\mathbf{x}_i|H_j)$ under hypothesis j, for $j = 0, 1$. At each sensor i, the decisions are given by:

$$u = \gamma(\mathbf{x}_i) = \begin{cases} 0. & \text{if detector i decides } H_0 \\ 1. & \text{if detector i decides } H_1 \end{cases}$$

The probability of false alarm P_{f_i} and the probability of detection P_{d_i} are used as measures of performance for each sensor i. We say that a false alarm has occurred, when the sensor decides that a signal is present when it actually is not present, i.e $P(u_i = 1|H_0) = P_{f_i}$. Detection occurs when the sensor decides that a signal is present, when it actually is present, i.e $P(u_i = 1|H_1) = P_{d_i}$.

Processing occurs at the local sensors, and then the fusion center (see Fig 2.1) receives the decisions of the sensors as its observations. Because individual sensor observations are independent and no communication occurs between local sensors, the local decisions are also statistically independent; hence, the joint probability density function of the "observation" at the fusion center under hypothesis j may

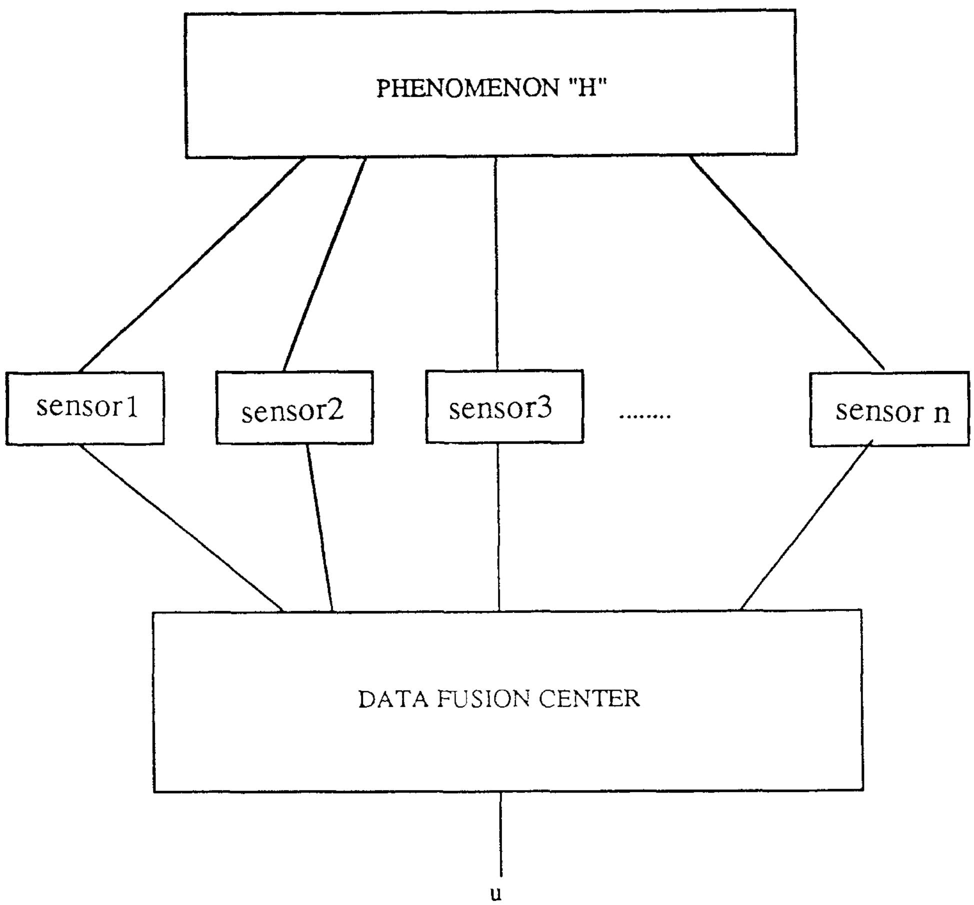


Fig.2.1 Distributed Sensor System with Data Fusion.

be written as $P(u_1, \ldots, u_n|H_j) = \prod_{i=1}^{n} P(u_i|H_j)$. The decision at the fusion center,

$$u = \gamma(u_1, u_2, \ldots, u_n) = \begin{cases} 0, & \text{if global decision } H_0 \\ 1, & \text{if global decision } H_1 \end{cases}$$

The action of the fusion rule can be summarized by $F(u_1, \ldots, u_n)$, the probability of choosing H_1 given local decisions $(u_1, \ldots, u_n)$. The criterion chosen for optimality is to minimize the overall cost of the global decision. If $J(u, H_j)$ is the cost of the fusion center choosing u when H_j is true, optimality demands $E(J(u, H_j))$, the total cost, be minimized. The functions $P(u_i|x_i)$ and $P(u|u_1, \ldots, u_n)$, which characterize the decision rule of the local detectors and the fusion detector, respectively, are chosen such that $E(J(u, H_j))$ reaches a global minimum.

2.1 Non-Optimal Solutions

Chair and Varshney[2] considered the optimization of the data fusion algorithm once the local detectors have already been designed. They show that the optimal fusion rule is given by the following likelihood ratio test, with β as the threshold:

$$\frac{P(u_1, \ldots, u_n|H_1)}{P(u_1, \ldots, u_n|H_0)} > \beta, \; H_1 \text{ is true} \tag{2.1}$$

and

$$\frac{P(u_1, \ldots, u_n|H_1)}{P(u_1, \ldots, u_n|H_0)} < \beta, \; H_0 \text{ is true} \tag{2.2}$$

where

$$\beta = \frac{P(H_0)[J(1, H_0) - J(0, H_0)]}{P(H_1)[J(0, H_1) - J(1, H_1)]}. \tag{2.3}$$

Because the decisions are statistically independent, the likelihood ratio is

$$\begin{aligned} \frac{P(u_1, \ldots, u_n|H_1)}{P(u_1, \ldots, u_n|H_0)} &= \prod_{i=1}^{n} \frac{P(u_i|H_1)}{P(u_i|H_0)} \\ &= \prod_{S_1} \frac{P(u_i = 1|H_1)}{P(u_i = 1|H_0)} \prod_{S_0} \frac{P(u_i = 0|H_1)}{P(u_i = 0|H_0)} \\ &= \prod_{S_1} \frac{P_{d_i}}{P_{f_i}} \prod_{S_0} \frac{1 - P_{d_i}}{1 - P_{f_i}} \end{aligned} \tag{2.4}$$

where S_1 is the set of all i such that $u_i = 1$, and S_0 is the set of all i such that $u_i = 0$. The decision rule is summarized by

$$\begin{aligned} F(u_1, \ldots, u_n) &= P(u = 1|u_1, \ldots, u_n) \\ &= \left[\begin{array}{l} 1, if \frac{P(u_1,\ldots,u_n|H_1)}{P(u_1,\ldots,u_n|H_0} \geq \beta \\ 0, if \frac{P(u_1,\ldots,u_n|H_1)}{P(u_1,\ldots,u_n|H_0} < \beta \end{array} \right. \end{aligned} \quad (2.5)$$

which depends on local perfomance characteristics, P_{f_i} and P_{d_i}, through (2.4) above. The above optimal fusion structure was designed assuming that the local detectors are specified, i.e., P_{f_i} and P_{d_i} are assumed fixed.

However, the optimal choice of P_{f_i} and P_{d_i} can be determined by optimizing the local detectors in terms of the fusion rule. This problem is considered by Tenney and Sandell[1]. Given a fixed fusion rule and independent local observations, they show that the optimum detector that satisfies the minimum cost criterion is the threshold detector with its likelihood ratio test given by:

$$\frac{P(\mathrm{x}_i|H_1)}{P(\mathrm{x}_i|H_0)} > \alpha_i, \; H_1 \text{ is true} \quad (2.6)$$

and

$$\frac{P(\mathrm{x}_i|H_1)}{P(\mathrm{x}_i|H_0)} < \alpha_i, \; H_0 \text{ is true} \quad (2.7)$$

where α_i is the local detector threshold.

The local detector threshold, α_i, depends on the fusion rule chosen and on the other local detector thresholds. Reibman and Nolte[5] show that for a fixed fusion rule, for the case of a 3-sensor system ($n = 3$), the optimal local detector thresholds are given by:

$$\alpha_i = f_i(\alpha_j, \alpha_k, F(u_i, u_j, u_k)) \quad (2.8)$$

for $i \neq j \neq k$. And i, j, $k \in \{1, n\}$. $F(p, q, r) = P(u = 1|u_1 = p, u_2 = q, u_3 = r)$. The equations are extended in an obvious manner for $n > 3$. Note that (2.8) gives only necessary conditions for optimal detection with a fixed fusion rule[1].

If multiple solutions exist for a particular fusion rule, the minimum cost solution must be determined.

2.2 Globally Optimal Solution

For optimal detection, neither the local detectors nor the fusion rule are *a priori* fixed, and equations given by (2.8) must be solved for each optimal fusion rule to determine the overall minimum cost solution. $F(u_1, \ldots, u_n)$ has 2^n variables and therfore can have 2^{2^n} assignments. Reibman and Nolte[5] also show that the assumptions made to derive equations given by (2.8) rule out the possibility of a number of assignments. The solution of the above equations can become computationally difficult; hence by making certain assumptions, solving for an overall minimum cost solution is considered as is described in [5].

The statistics of the observations entering each local detector, $P(x_i|H_j)$, are equivalent for all i given j Also, the local detector thresholds α_i are set equal to α for all i. With this constraint, the fusion likelihood ratio becomes a function of k, where k is the number of local detectors which decide H_1. It is given by

$$\frac{P(u_1, \ldots, u_n|H_1)}{P(u_1, \ldots, u_n|H_0)} = \frac{P_{d_i}(\alpha)}{P_{f_i}(\alpha)}^k \frac{1 - P_{d_i}(\alpha)}{1 - P_{f_i}(\alpha)}^{n-k} \tag{2.9}$$

The fusion rule is a "k-out-of-n" logical function, where k depends upon the overall threshold β. The dependence of the local threshold α can be expressed using β and k. i.e,

$$\alpha = \beta \frac{P_{f_i}(\alpha)}{P_{d_i}(\alpha)}^{k-1} \frac{1 - P_{f_i}(\alpha)}{1 - P_{d_i}(\alpha)}^{n-k} \tag{2.10}$$

The optimal values of α and k are found by solving equation (2.10) for each value of k and locating the (α, k) pair which produce the minimum cost solution for the given value of β.

Results for this simplified solution are shown in [5]. The slope of the Receiver

Operating Characteristic at any point is equal to β, the fusion processor threshold at that point.

CHAPTER 3
INFORMATION THEORETIC VIEWPOINT OF DETECTION THEORY

Gabrielle[8] first offered a method of analyzing the effect of imposing information criteria for threshold setting in simple binary tests. It considered such a test as a general binary assymetric channel whose transition probabilities are related to the probabilities of errors of the first and second kind in the test.

The test threshold value, which controls the tradeoff between these probabilities, is selected on the basis of the performance of the test as a communication channel between its input and output. Gabrielle showed that the maximization of the mutual information associated with a binary hypothesis test subject to the operating characteristic of the test determines the threshold.

Recently, Hoballah and Varshney[9] re-examined the problem of optimum detection for entropy-based cost functions. They considered the design of DSN's based on this cost function, i.e , the one which minimizes the information loss. This cost function can be used in applications where we are interested in the amount of information that we are able to transfer, rather than the nature of the information itself. In such situations, cost may be a variable (unlike fixed costs assigned to actions in Bayesian formulation methods), and entropy based cost functions may be more meaningful. The *goal* therefore becomes to maximize the amount of information transfer.

We study the DSN with data fusion. Studies on DSN with local inferences have been done too, and the more interested reader may refer to [9].

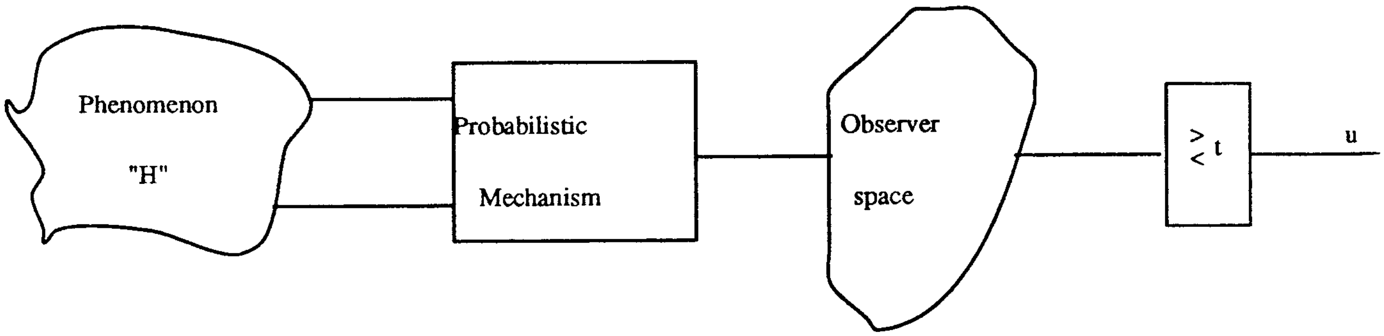


Fig.3.1(a) A Classical Binary Detection System.

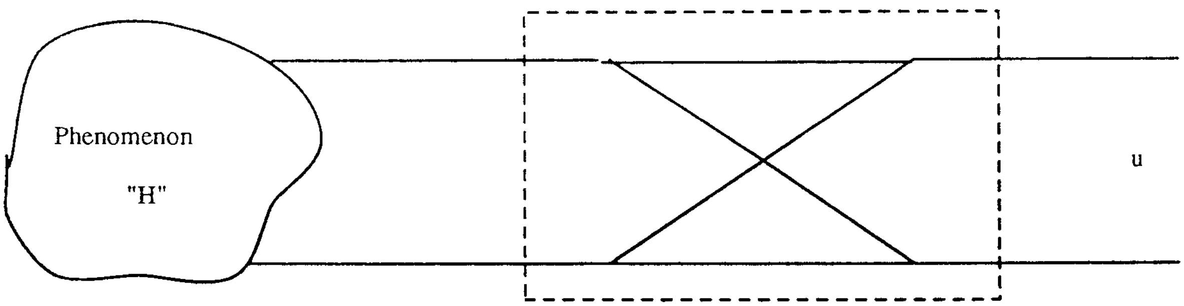


Fig.3.1(b) A Classical Information Transmission System.

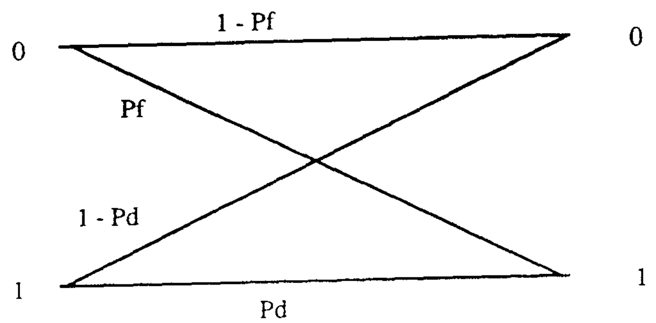


Fig.3.2 Channel Modelling Multi-sensor System at the Receiver.

3.1 Common Grounds in Detection Theory and Information Theory

As shown in Fig.3.1(a), the source in the detection problem is the phenomenon *H*. This is a random variable that can assume one of the two values: 0 or 1. *H* = i corresponds to the hypothesis H_i being present, i=0,1. This is followed by a data processor, that generates a likelihood ratio, which is then compared against a threshold, to give a final decision u. This output is a decision random variable which may again assume the value of 0 or 1. The source in the detection problem can be viewed as the information source in the information transmission problem. For the information transmission problem, (See Fig.3.1(b)), the source alphabet, H, is considered to be binary $\{0, 1\}$ We assume that the output of such a source is transmitted via a Binary Asymetric Channel, whose cross-over probabilities are determined by the probability of detection, P_d, and the probability of false alarm, P_f. The decisions may be looked at as the output of the communication channel in Fig 3.1(b) The channel modelling the multi-sensor system at the receiver is shown in Fig 3.2.

We consider detection problems where we minimize the equivocation between the input and the output, i.e., the Minimum Equivocation Detection (MED) problem. In this case, we are interested in minimizing an average cost, which is not a constant, but a function of the *a posteriori* probability of H given u. The average Bayesian risk is defined as $E\{J(u, H)\}$ as seen in the previous chapter, where *J(u,H)* is a constant for a given pair u and H. For the MED problem, the average cost is the conditional entropy of H given u, i.e., $h(H|u) = E\log[\frac{1}{P(H|u)}]$.

If the *a priori* probabilities P_0 and P_1 are known, Hoballah and Varshney[9] have obtained decision rules to minimize the average cost $h(H|u)$ Shannon[12] had demonstrated how information can be reliably transmitted over a noisy communication channel by first considering a measure of the amount of information about the transmitted *message* contained in the observed output of the channel. To do so, he defined the notion of mutual information between input events and output events, H and u, in our case. It is given by $I(H;u) = h(H)$ - $h(H|u)$

For the channel model shown in Fig.3.2, the mutual information $I(H;u)$ is given by:

$$I(H;u) = \sum_{H}\sum_{u} P(H,u) \log \frac{P(H|u)}{P(H)}$$

and the *a posteriori* probabilities are

$$P(u=0) = P_0(1-P_f) + (1-P_0)(1-P_d) = \xi_0 \tag{3.11}$$

and

$$P(u=1) = P_0 P_f + (1-P_0)P_d = \xi_1 \tag{3.12}$$

As $h(H)$, the entropy of H, is constant when P_0 and P_1 are known, minimizing $h(H|u)$ is equivalent to maximizing $I(H;u)$. Thus the minimization of equivocation is equivalent to the maximization of mutual information.

Different Receiver Operating Characteristics (ROC) curves are associated with different decision rules for a given detection system. The ROC corresponding to the optimum Bayesian detection rule lies above the $P_d = P_f$ line. As shown in Chapter 2, the optimum performance of the detection system is obtained when neither the local decision rules nor the fusion rule is fixed *a priori*. ROC's corresponding to nonoptimum decision rules will lie partly or entirely below the ROC corresponding to the optimum decision rule. Therefore, for any given probability of false alarm $P_{f'} \in \{0,1\}$, the corresponding probability of detection $P_{d'}$ on the optimum ROC and the one on the non-optimum ROC, $P_{d''}$ satisfy the following relation:
For all $P_{f'}$, such that $P_{f'} \in \{0,1\}$, $P_{d_{opt}'} \geq P_{d_{nonopt}''}$.

It has been proved that[10] given *a priori* probabilities P_0 and P_1, for each value of P_f (or P_d), the mutual information $I_{min}(H;u)$ is a concave upward function of the transition probabilities $P(u|H)$. Using this fact, it has been shown[9] that the minimum mutual information is achieved at the point where $P_d = P_f$.

From the above findings, since $I(H;u)$ is a concave upward function of P_d for a given value $P_{f'}$ of P_f, and the minimum value of $I(H|u)$ is achieved by choosing $P_d = p_{f'}$, then for values of P_d such that $P_d \geq P_{f'}$, $I(H;u)$ is an increasing function of P_d.

Therefore, if $P_{d'} \geq P_{f'}$, then $I(P_{f'}, P_{d'}) \geq I(P_{f'}, P_{d''})$. i.e., for a fixed value $P_{f'}$, the value of P_d obtained by maximization of P_d also maximizes $I(H;u)$. Therefore, the pair (P_d, P_f) which maximizes $I(H;u)$ lies somewhere on the ROC of the optimum Bayesian Detection System.

Tenney and Sandell[1], in their pioneering work on DSN's had shown that in a binary hypothesis test, when sensors observe independently, each sensor uses an independent, local likelihood ratio test, but with thresholds determined via a coupled computation. Hoballah and Varshney[9] came up with a similar result using information theoretic methods. They showed that given P_0 and P_1, and the conditional densities $p(\mathrm{x}|H_j)$, $j = 0,1$; the optimum detector that maximizes the mutual information is the threshold detector. The optimum MED detector can be implemented as a threshold detector with the following likelihood ratio test:

$$\frac{P(\mathrm{x}|H_1)}{P(\mathrm{x}|H_0)} > t, \;\; H_1 \text{ is true} \tag{3.13}$$

and

$$\frac{P(\mathrm{x}|H_1)}{P(\mathrm{x}|H_0)} < t, \;\; H_0 \text{ is true} \tag{3.14}$$

with

$$t = \frac{-P_0[\log\frac{\xi_0}{\xi_1} - \log\frac{(1-P_f)}{P_f}]}{(1-P_0)[\log\frac{\xi_0}{\xi_1} - \log\frac{(1-P_d)}{P_d}]} \tag{3.15}$$

where ξ_0 and ξ_1 are given by equations (3.11) and (3.12) respectively.

The above was the classical detection problem utilizing a single sensor. And the result obtained matches the results obtained in[8]. For the case of two detectors, and then for the N detector case, optimum thresholds have been obtained using information theoretic methods in [9].

3.2 Controlling P_0 and P_1 at the Receiver to Maximize $I(H;u)$

The receiver of a multi-sensor system can be modelled as a channel as has been validated by studies by Gabriele[8] and Hoballah and Varshney[9]. The cross over probabilities of this information channel depend on P_f and P_d. From Fig 3.1(b), it can be seen that the binary information channel models a wide range of multisensor systems. i.e., the channel can be thought of as a black box which has $H = 0$ or $H = 1$ as its input, and $u = 0$ or $u = 1$ as its output. In other words, for a wide range of local detector thresholds and fusion rules, we can literally get a unique channel.

As can be seen from the ROC curves obtained in [5], ROC's of non-optimum decision rules will lie *partly* or entirely below the ROC corresponding to the optimum decision rule. Hence, a particular P_d-P_f combination does not uniquely say that a particular decision rule is being used. We hereafter consider the multi-sensor system at the receiver to be an *Asymetric Binary Channel* and consider a special case where we try to maximize $I(H;u)$ across the channel.

3.2.1 Why Control P_0 and P_1 at the Receiver

Gabriele[8] hinted about maximizing $I(H;u)$ by influencing transmitter coding.

A suitable measure for efficiency of transmission is by making a comparison between the actual rate and the upper bound of the rate of transmission of information for a given channel. In a discrete channel with prespecified noise characteristics, i.e., with a given transition probability matrix, the rate of information transmission depends on the source that drives the channel. In the network analogy, one could specify the load and determine the class of transducers that would match the given load to a specified class of sources. The maximum (or the upper bound) of the rate of information transmission corresponds to a proper matching of the source and the channel. This ideal characterization of the source in turn depends on the probability transition characteristics of the given channel.

3.2.2 Encoding

Encoding is frequently used in a wide variety of cases as a transformation procedure operating on the input signal prior to its entry into the communication channel. The main purpose of coding, is, in general, to improve the efficiency of the communication process in some sense.

In our system, (See Fig.3.3) we use encoding between the transmitter and the receiver so as to increase the amount of information across the multi-sensor system at the receiver that is modelled as an asymmetric channel. The channel of communication usually deals with symbols of some specified list. This list is generally referred to as the *alphabet* of the communication language. By an independent

source, we mean a device that selects messages at random from a discrete message ensemble $\{m_1, \ldots, m_N\}$ with prescribed probabilities $\{p(m_1), \ldots, p(m_n)\}$

In our system, a transmitter is an independent source and it either sends a 0 or a 1 depending upon voltage levels, as is usually done in data communications. In other words, we have a binary event represented by the two hypothesis, Hyp_0 and Hyp_1, which have *a priori* probabilities Pr_0 and Pr_1 respectively.

Hyp_0: 0 is transmitted

Hyp_1: 1 is transmitted

By encoding, we map our given set of *messages* $[m_0, m_1]$ into a new set of encoded messages $[c_0, c_1]$ so that the transformation is one-to-one. Although, gen

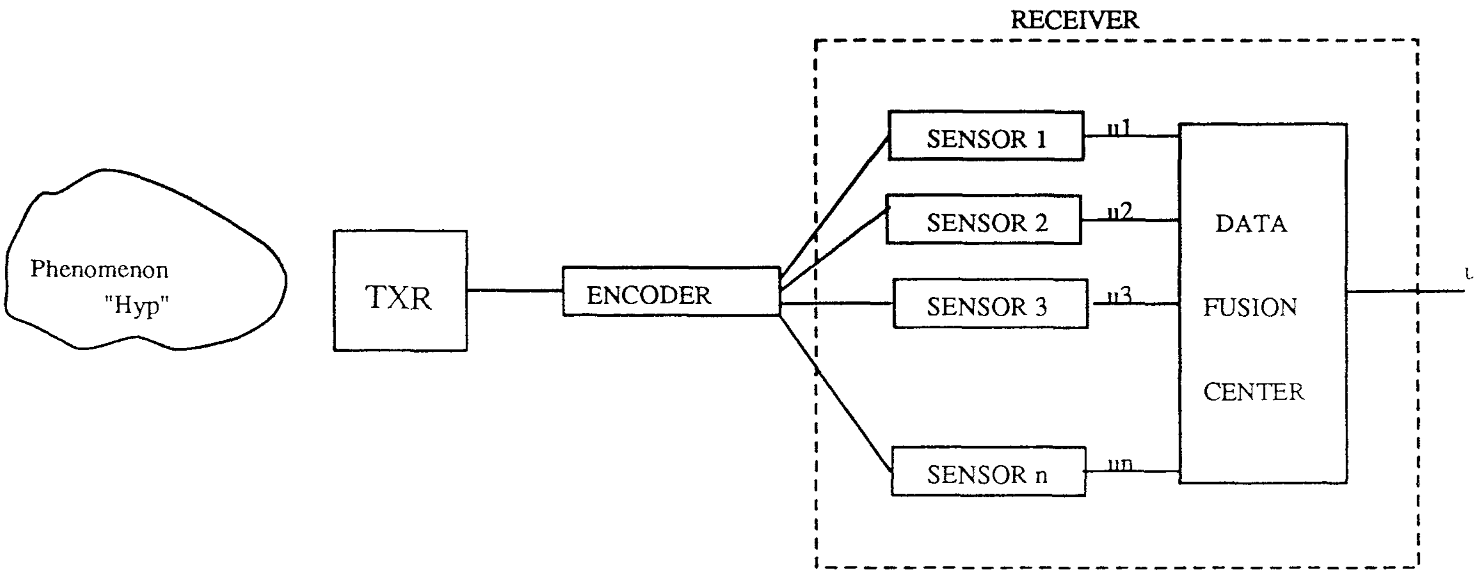


Fig.3.3 Complete System to Maximize Mutual Information

erally.By efficiency, we wish to improve the "efficiency" of the "transmission", it is possible to devise codes for any other desired purpose, as in our case, to improve the amount of information across the multi-sensor system at the receiver. And this is done without relevance to the transmission efficiency in our adopted sense. For the sake of clarity, we redefine the hypotheses and also the *a priori* probabilities at the receiver. However, we will assume that the signal does not get corrupted during the encoding process. i.e., if a 0 (or 1) is sent, a 0 (or 1) is received. We now have two hypothesis denoted by H_0 and H_1, which have *a priori* probabilities P_0 and P_1 respectively.

H_0: 0 is received

H_1: 1 is received

Efficient encoding, depending upon its application, improves certain cost functions. In our case, as has been emphasized before, the *goal* is to maximize the information transfer at the receiver. And for the MED problem that we are considering, the cost is given as a function of the *a posteriori* probability of H given u, where H = i corresponds to the hypothesis H_i being present (at the receiver), i=0,1. The output of the multi-sensor system at the receiver is the decision random variable u, which may again assume the value of 0 or 1.

The average cost of our MED problem is the conditional entropy of H given u, i.e., $h(H|u) = E\log[\frac{1}{P(H|u)}]$. The mutual information between input and output events, H and u, is given by

$I(H;u) = h(H) - h(H|u)$.

We wish to minimize the equivocation over all input probabilities P_0 (or P_1). Hence, $h(H)$ is not going to remain fixed. But minimizing the equivocation still implies maximizing the mutual information since $h(H)$ must be as low as possible when compared to $I(H;u)$ to get the minimum equivocation, i.e., to solve the MED problem. Hence, our goal of maximizing the mutual information in this case too remains unchanged.

In our study, as we are controlling P_0 by encoding at the transmitter, we are increasing the average cost per *message*, where the average cost per *message* is given by:

$$R_t = \sum_{i=1}^{N} p(m_i) \times t_i \tag{3.16}$$

where t_i is the duration, and $p(m_i)$ is the probability of *message* i. We confine ourselves to the simplest case when all symbols have identical cost. Thus, the average cost per *message* becomes proportional to the average of n_i , the number of symbols per *message* (or the average length of the *message*).

As we try to maximize the mutual information over the input probabilities at the receiver by encoding, we are obviously increasing the average cost per *message* as defined by equation (3.16).To achieve our *goal*, this is unavoidable. As we have said before, "efficiency" of "transmission" can be open to subjective interpretations.

Our transmitter has two *messages*. Let m_0 represent a 0 and m_1 represent a 1. Both *messages* are equiprobable. $[M] = [m_0, m_1]$

$[P(M)] = [Pr_0, Pr_1]$

$Pr_0 = Pr_1 = \frac{1}{2}$

We wish to encode the *messages* into *words* selected from a binary alphabet with a one-to-one correspondence.

Implementation

Our encoding is motivated by the need to control the P_0 at the receiver to maximize $I(H; u)$ across the multisensor system. (See Fig. 3.3) The encoded *word* is transmitted as a whole at any point of time. i.e., all the bits in the *word* are transmitted in parallel.

For different channel specifications. i.e., for different probabilities of detection and false alarm, we get the P_0 at which $I(H;u)$ reaches a maximum. *$I(H;u)$ is a convex $\bigwedge$ function in the space of the input probability distributions.* Once we get the input probability P_0 that maximizes the mutual information $I(H;u)$, we try to encode the message at the transmitter in an appropriate way so as to get that P_0 at the receiver. The coding is not unique as will be shown later by an example. The bits can be ordered differently in the same string to produce the same effect. Also, the codes can be of variable length to produce the same effect. What matters in the encoded word is *the proportion of the number of zeroes to the number of ones* to produce a particular effect.

For a paricular (P_d, P_f) pair, if we get the result that the mutual information is maximized in such a way that $P_0 = 0.75$, and if m_0 is transmitted, the following encoding can be performed.

$$m_0 \quad 0001$$

We encode m_0 alone, because as we have said before. we assume that if a 0 is transmitted, a 0 is received and that an encoder does not corrupt the transmission. It is easy to see that this encoding is not unique. Variable length codes with different orderings of 0's and 1's are possible for the same result.

To maximize $I(H;u)$ over all input probability distributions, we use the method that was suggested by S.Muroga[13]. Maximizing $I(H;u)$ is the same as maximizing the rate of transmission of information of the binary channels. See the channel model shown in Fig. 3.2.

The following is the method: First, one uses auxiliary variables Q_0 and Q_1, which satisfy the following equations:

$$\begin{aligned} (1-P_f)Q_0 + P_fQ_1 &= (1-P_f)\log(1-P_f) + P_f\log P_f \\ (1-P_d)Q_0 + P_dQ_1 &= (1-P_d)\log(1-P_d) + P_d\log P_d \end{aligned} \tag{3.17}$$

The rate of transmission of information $I(h;u)$ can be written as:

$$\begin{aligned} I(H;u) &= h(u) - h(u|H) \\ &= -[p_0\prime \log p_0\prime + p_1\prime \log p_1\prime] + \\ &\quad P_0[(1-P_f)\log(1-P_f) + P_f \log P_f] + \\ &\quad P_1[(1-P_d)\log(1-P_d) + P_d \log P_d] \end{aligned} \tag{3.18}$$

where $p_0\prime$ and $p_1\prime$ are the probabilities of receiving 1 and 0 at the output, respectively.

Next we introduce Q_0 and Q_1 into (3.3), through (3.2):

$$\begin{aligned} I(H;u) &= -(p_0\prime \log p_0\prime + p_1\prime \log p_1\prime) + \\ &\quad [P_0(1-P_f) + P_1(1-P_f)]Q_0 + [P_0P_f + P_1P_d]Q_1 \end{aligned} \tag{3.19}$$

Thus,

$$I(H;u) = -(p_0' \log p_0' + p_1' \log p_1') + p_0'Q_0 + p_1'Q_1$$

The maximization of $I(H;u)$ is done with respect to $p_0\prime$ and $p_1\prime$, the probabilities at the output. In order to do this, we may use the method of Lagrangian multipliers. This method suggests maximizing the function

$$I(H;u) = -(p_0' \log p_0' + p_1' \log p_1') + p_0'Q_0 + p_1'Q_1 + \mu(p_0' + p_1')$$

through a proper selection of the constant number μ. Taking partial derivatives of $I(H;u)$ w.r.t p_0' and p_1' respectively, we get the following equations:

$$\begin{aligned} -(\log e + \log p_0') + Q_0 + \mu = 0 \\ -(\log e + \log p_1') + Q_1 + \mu = 0 \end{aligned} \tag{3.20}$$

Solving, μ must satisfy the following condition.

$$\mu = -Q_0 + (\log e + \log p_0') = -Q_1 + (\log e + \log p_1')$$

where e is the base that is used for the log function. In our case e = 2. The maximum mutual information is found to be:

$$I(H;u)_{max} = Q_0 - \log p'_0 = Q_1 - \log p'_1.$$

The values of Q_0 and Q_1 may be obtained from the set of equations 3.17). But note that

$$p'_i = \exp(Q_i - I_{max}), i = 0, 1$$

Thus,

$$\begin{aligned} I(H;u)_{max} &= \log[\exp(Q_0) + \exp(Q_1)] \\ &= \log[2^{Q_0} + 2^{Q_1}] \end{aligned} \tag{3.21}$$

The maximum mutual information $I(H;u)_{max}$ of a binary channel is greater than zero except when $P_d = P_f$.

CHAPTER 4
RESULTS AND DISCUSSIONS

In this chapter, we get the probability P_0 at the receiver that maximizes the mutual information $I(H;u)$ for different local detector rules and fusion rules. Or more simply, for different channel specifications, i.e., for different P_d and P_f.

We can get the maximum mutual information from Muroga's method as we have seen in Chapter 3. The probability P_0 at which this maximum is attained can be obtained from

$$I(H;u) = \sum_H \sum_u P(H,u) \log \frac{P(H|u)}{P(H)}$$

Expanding this, we get the following expression:

$$\begin{aligned} I(H;u) \quad = \quad & P_0(1-P_f) \log \frac{1-P_f}{P_0(1-Pf)+(1-P_0)(1-P_d)} + \\ & P_0 P_f \log \frac{P_f}{P_0 P_f + (1-P_0)P_d} + \\ & (1-P_0)(1-P_d) \log \frac{(1-P_d)}{P_0(1-P_f)+(1-P_0)(1-P_d)} + \\ & (1-P_0)P_d \log \frac{P_d}{P_0 P_f + (1-P_0)P_d} \end{aligned}$$

Once we get the appropriate P_0, we derive this P_0 at the receiver end by encoding the transmitter output. Curves for certain specifications of the multi-sensor system are shown, and using the P_0 at which maximum information transfer occurs, we show how encoding at the transmitter can be performed so as to achieve our goal of solving for minimum equivocation detection, or, as we have shown in Chapter 3, for achieving maximum information transfer across the multi-sensor system at the receiver.

The way this works is as follows:

(DSN with fusion is considered. (See Fig 2.1))

- Using Bayesian methods, the multi-sensor system is evaluated for different decision rules.

- Using the Bayesian methods as discussed in Chapter 2, the multisensor system at the receiver is tuned for globally optimal performance, i.e., neither the local detector rules nor the fusion rules are *a priori* fixed.

- For the above Bayesian methods, we get P_d vs P_f values. As has been discussed in Chapter 3, ROC's corresponding to nonoptimum decision rules will lie partly or entirely below the ROC coresponding to the optimum decision rule.

- We now consider the multi-sensor system at the receiver as a black box, and model it as a channel, with cross over probabilities determined by P_d and P_f. (See Fig. 3.2)

- To achieve our goal of maximizing the mutual information $I(H;u)$ across the receiver, we get the value of P_0 at the receiver that maximizes $I(H;u)$.

- We then encode the 0 or 1 at the transmitter output in such a way that we get the desired P_0 at the receiver. (See Fig 3.3)

For a few P_d and P_f values, $I(H;u)_{max}$ is obtained (See Table 4.1). The P_0 at which it is achieved is then derived, as is shown by a few examples, and plots of $I(H;u)$ vs P_0 are shown. For the Binary Assymeric Channel that we use to model the sensor system at the receiver, we see that the P_0 at which the maximum $I(H;u)$ is reached, is around 0.5 (which is the ideal one as in a Binary Symmetric Channel).

For values of P_d and P_f that are likely to lie on a more optimum ROC curve, more information transfer occurs across the multi-sensor system at the receiver. A plot of $I(H;u)_{max}$ against P_d and P_f is shown in Fig.4.1.

Consider the channel with $P_d = 0.415922$, and $P_f = 0.463543$. $I(H;u)$ as expected has a maximum of close to 0. (See Table 4.1). Consider the channel with more optimum performance, with $P_d = 0.983481$ and $P_f = 0.535422$. $I(H;u)_{max}$ is equal to 0.242. (See Fig.4.2) This particular (P_d, P_f) combination is on a more optimum ROC curve than the previous one. A P_0 of 0.4 can help attain this. Hence if a 0 is transmitted, it can be coded as 00111 or 01011. Similarly, if a 1 is transmitted, it too can be coded in the same way. However, as we have said before, the cost given by (3.16) is increased.

Considering a $P_d = 0.947670$ and a $P_f = 0.171744$ (obviously a pretty good combination), we find that a maximum $I(H;u)$ of 0.512 is attained. (See Fig 4.3) A P_0 of about 0.47 does the job, which means that if a 0 or 1 is transmitted, it can be coded into 1010101010101010101010101, or another combination with the same proportion of 0's and 1's.

Considering yet another very good combination of $P_d = 0.979207$ and $P_f = 0.126442$, a maximum $I(H;u)$ of 0.647554 is achieved. (See Fig 4.4) The P_0 that helps achieve this is about 0.47, and therefore a similar coding at the transmitter output can be performed as in the previous case.

Q1	Q2	Pf	Pd	I(H;u)|max
-0.499870	-1.472198	0.175749	0.513893	0.094385
-0.748329	-1.212865	0.534558	0.308665	0.038023
-0.742645	-0.271592	0.171744	0.947670	0.512073
-0.721024	-0.945401	0.226441	0.702262	0.171162
-0.388766	-1.624136	0.124726	0.494796	0.121938
-0.265216	-2.060255	0.389664	0.083903	0.100110
-0.554084	-1.627750	0.368087	0.277238	0.006739
-2.042133	-0.089173	0.535422	0.983481	0.242311
-1.760266	-0.487231	0.646502	0.765716	0.012336
-2.144013	-0.369931	0.780272	0.767173	0.000100
-0.601412	-0.689082	0.151939	0.822988	0.355454
-0.842373	-1.021082	0.314699	0.625505	0.071045
-1.247390	-0.336470	0.917245	0.346917	0.278829
-0.879593	-1.109260	0.401172	0.519783	0.010140
-1.703124	-0.489870	0.785437	0.606785	0.027454
-3.346772	-0.140684	0.869960	0.931564	0.007744
-2.116875	-0.328166	0.674550	0.866563	0.038573
-1.583723	-0.547366	0.581919	0.758433	0.025582
-0.671292	-1.424320	0.355651	0.389265	0.000777
-0.740953	-0.648708	0.826964	0.200241	0.305937
-0.834115	-1.183912	0.463543	0.415922	0.001562
-0.607225	-0.136090	0.126442	0.979207	0.647554
-0.888398	-0.221499	0.958494	0.212646	0.483296
-1.157372	-0.714268	0.409075	0.737496	0.081133

Table 4.1 Maximum Mutual Information Tabulated using Muroga's Method

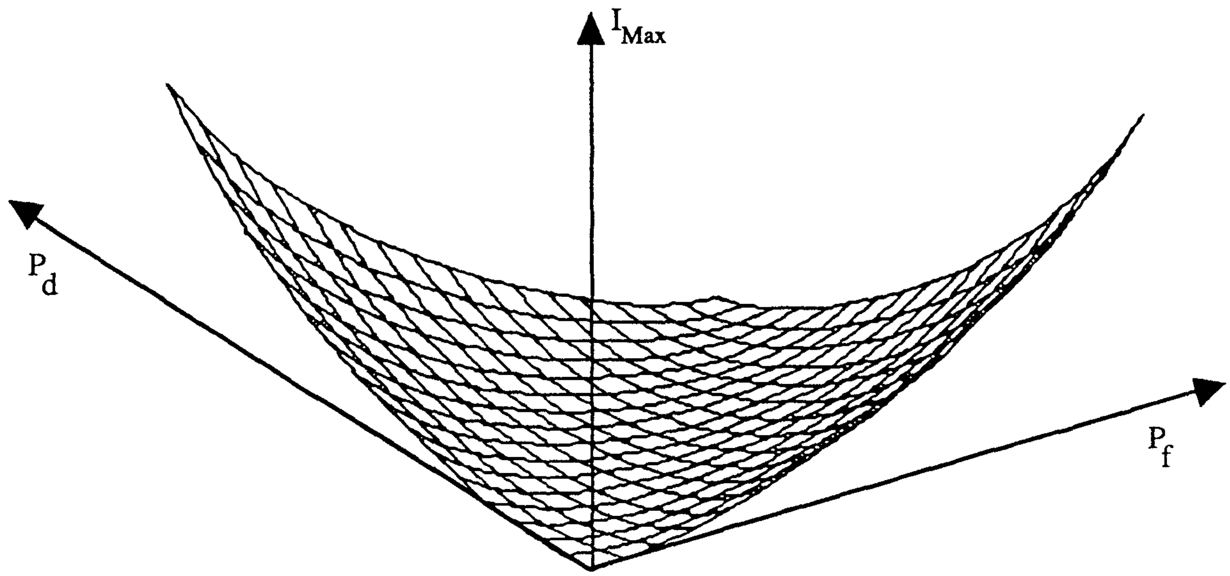
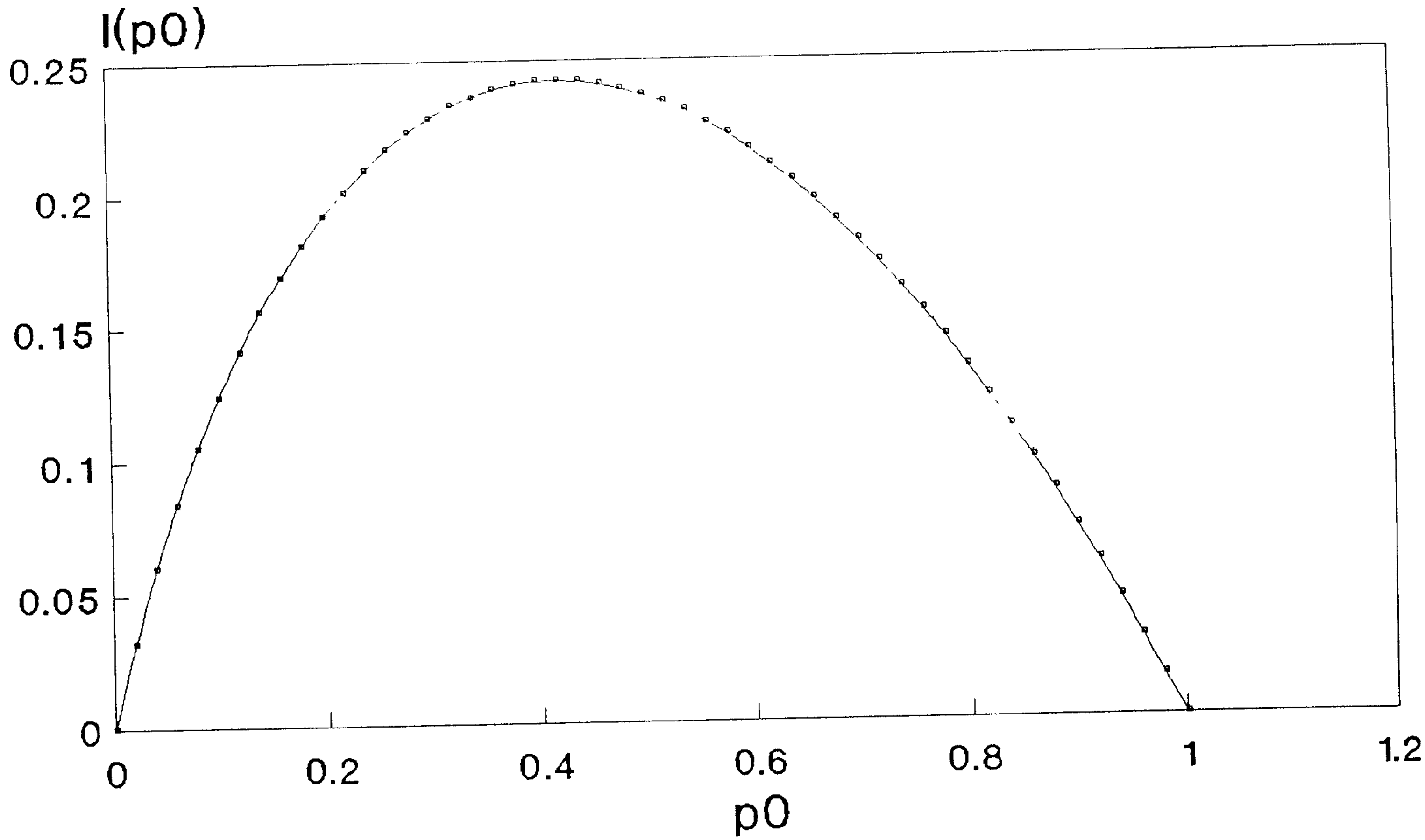


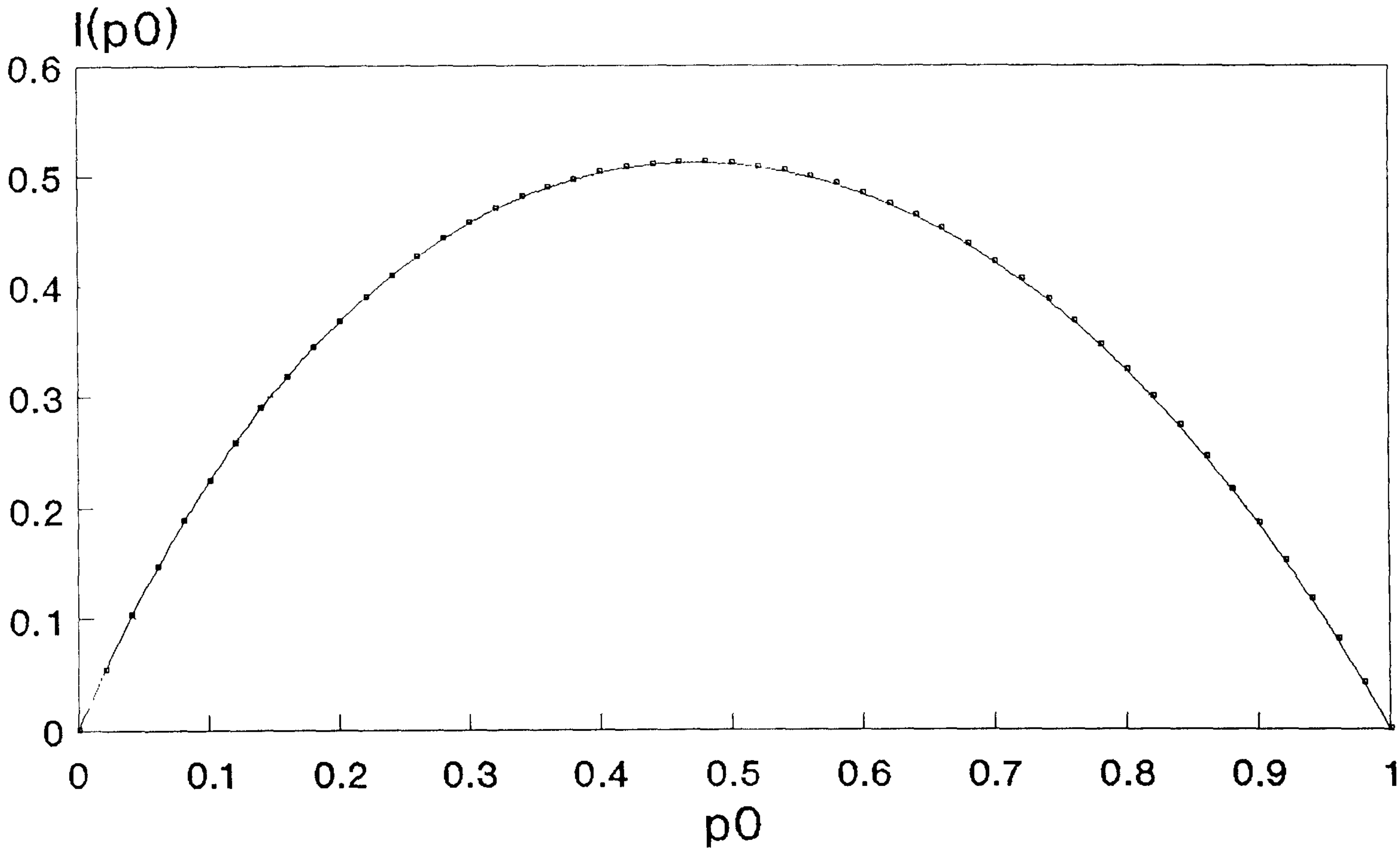
Fig.4.1 A Plot of $I(H;u)_{max}$ Against P_d and P_f.

Fig.4.2 Mutual Information vs P0



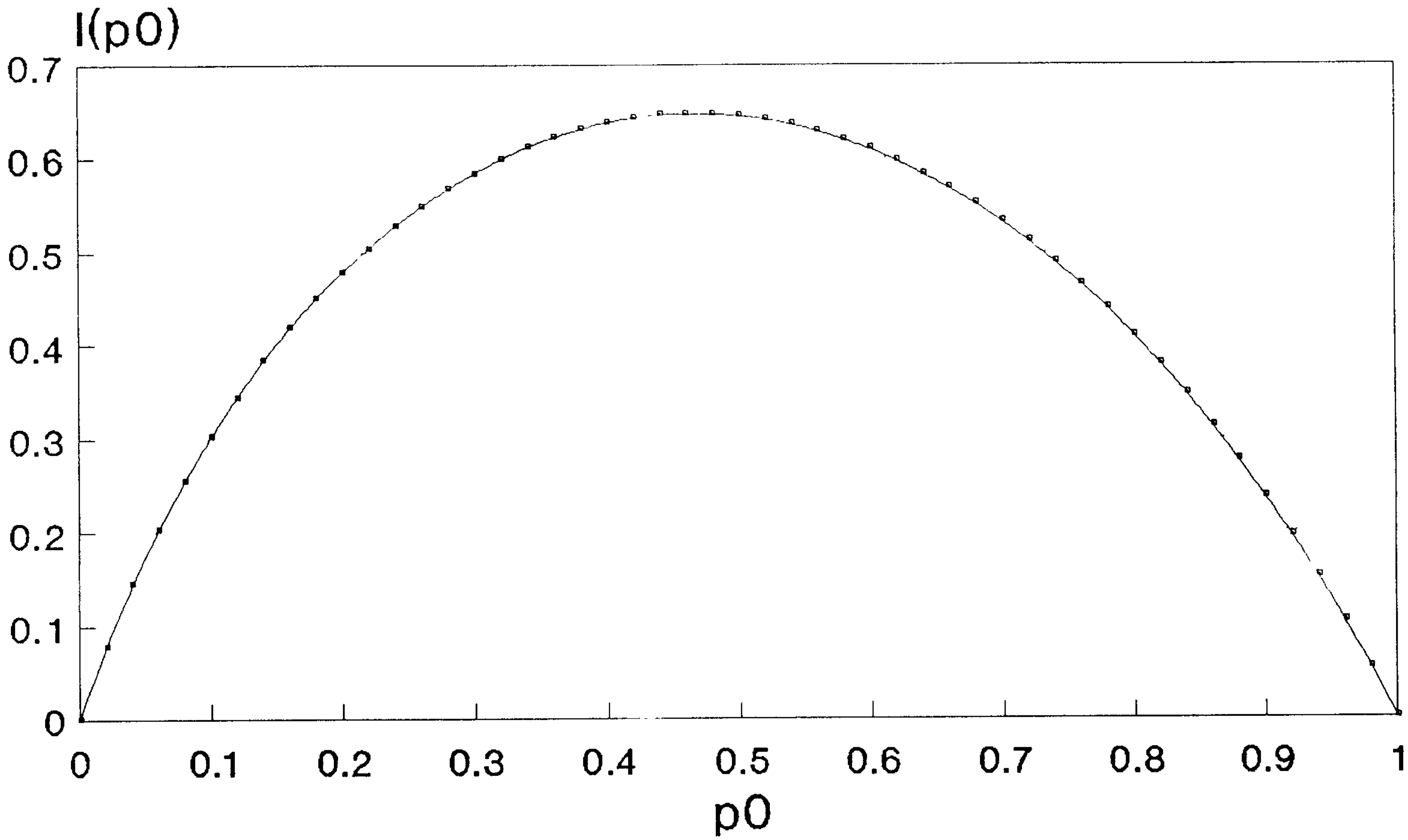
Pd = 0.983481
Pf = 0.535422

Fig.4.3 Mutual Information vs P0



Pd = 0.947670
Pf = 0.171744

Fig.4.4 Mutual Information vs P0



Pd = 0.979207
Pf = 0.126442

CHAPTER 5
CONCLUSIONS

A Distributed Sensor Network at the receiver is modelled as an assymetric channel, and the local detector rules and fusion rules decide the cross over probabilities of the channel. Our goal, being to maximize the amount of information transfer across the multi-sensor system at the receiver. By simulation, we get a sample of co-ordinates on the ROC curve and calculate the maximum information possible. The *a priori* probability, P_0, that helps attain this goal is then derived, assuming that the receiver has control over the value of the probability of the signal being present. This apriori probability is then obtained by encoding at the transmitter output. The only factor in encoding is the proportion of 0's to 1's. Variable length codes and variable ordering of 0's and 1's is possible. Results show that as we move towards an optimum ROC curve, the mutual information attains a higher value. The value of P_0 that helps attain this maximum is then derived by encoding the transmitter output.

BIBLIOGRAPHY

[1] Tenney,R.R., and N.R.Sandell."Detection with Distributed Sensors." *IEEE Trans. Aerospace and Electronic Systems.* vol. AES-17, pp. 501-510, 1981.

[2] Chair,Z., and P.K.Varshney. "Optimal Data Fusion in Multiple Sensor Detection Systems." *IEEE Trans. Aerospace and Electronic Systems.* vol. AES-22, pp. 98-101, 1986.

[3] Sadjadi,F.A. "Hypotheses testing in a Distributed Environment." *IEEE Trans. Aerospace and Electronic Systems.* vol. AES-22, pp. 134-137, Mar. 1986.

[4] Demirbas,K. "Distributed Sensor Data Fusion with Binary Decision Trees." *IEEE Trans. Aerospace and Electronic Systems.* vol.25, no.5, pp. 643-649, September 1989.

[5] Reibman,A.R and L.W.Nolte. "Optimal Detection and Performance of Distributed Sensor Systems." *IEEE Trans. Aerospace and Electronic Systems.* vol. AES-23, No.1, pp. 24-30, Jan 1987.

[6] Krzysztofowicz,R and D.Long. "Fusion of Detection Probabilities and Comparison of Multisensor Systems" *IEEE Trans. Sys., Man and Cybernetics.* vol. SMC-20, pp 665-677, May/June 1990.

[7] Middleton.D. *An Introduction to Statistical Communication Theory.* New York: McGraw-Hill, 1960.

[8] Gabrielle,T.L. "Information Criterion for Threshold Determination." *IEEE Trans. Information Theory* vol. IT-6, pp. 484-486, Oct. 1966

[9] Hoballah,I.Y., and P.K.Varshney. "An Information Theoretic Approach to the Distributed Detection Problem." *IEEE Trans. on Information Theory.* vol. IT-35, pp. 998-994, Sept 1989.

[10] McEliece,R.J. *The theory of information and coding.* Encyclopedia of Mathematics, vol.3, Reading, MA:Addison-Wesley,1977.

[11] Reza,F.M. *An Introduction to Information Theory.* New York: McGraw-Hill, 1961.

[12] Shannon,C.E. "Communication in the Presence of Noise." *Proc. IRE.* vol.37, pp. 10-21, 1949.

[13] Muroga,S. "On the Capacity of a Discrete Channel." *J. Phys. Soc. Japan.* 8L484, 1953.

[14] Viterbi,A.J and J.K.Omura. *Principles of Digital Communication and Coding.* McGraw-Hill Book Company, New York, 1979.