Information Theoretic Limits on Non-cooperative Airborne Target Recognition by Means of Radar Sensors

by

Jacques Étienne Cilliers

A thesis submitted to the University College London for the degree of Doctor of Philosophy

Department of Electronic and Electrical Engineering
UCL

2018

Declaration

I, Jacques Étienne Cilliers confirm that the work presented in this thesis is my own. Where information has been derived from other sources, I confirm that this has been indicated in the thesis.

J.E. Cilliers

Abstract

The main objective of this research is to demonstrate that information theory, and specifically the concept of mutual information (MI) can be used to predict the maximum target recognition performance for a given radar concept in combination with a given set of targets of interest. This approach also allows for the direct comparison of disparate approaches to designing a radar concept which is capable of target recognition without resorting to choosing specific feature extraction and classification algorithms. The main application area of the study is the recognition of fighter type aircraft using surface based radar systems, although the results are also applicable to airborne radars.

Information theoretic concepts are developed mathematically for the analysis of the radar target recognition problem. The various forms of MI required for this application are derived in detail and are tested rigorously against results from digital communication theory. The results are also compared to Shannon's channel capacity bound, which is the fundamental limit on the amount of information which can be transmitted over a channel.

Several sets of simulation based experiments were conducted to demonstrate the insights achievable by applying MI concepts to quantitatively predict the maximum achievable performance of disparate approaches to the radar target recognition problem. Asymptotic computational electromagnetic code was applied to calculate the target's response to the radar signal for freely available geometrical models of fighter aircraft. The calculated target responses were then used to quantify the amount of information which is transmitted back to the radar about the target as a function of signal to noise ratio (SNR). The information content of the F-14, F-15 and F-16 were evaluated for a 480 MHz bandwidth waveform at 10 GHz as a baseline. Several ultra-wideband (UWB) waveforms, spanning 2-10 GHz, 10-18 GHz and 2-18 GHz, but which were highly range ambiguous, were evaluated and showed SNR gains of 0.5-2 dB relative to the baseline.

The effect of sensing the full polarimetric response of an F-18 and F-35 was evaluated and SNR gains of 5-7 dB over a single linear polarisation were measured. A Boeing 707 scale model (1:25) was measured in the University of Pretoria's compact range spanning 2-18 GHz and gains of 2 dB were observed between single and dual linear polarisations. This required numerical integration in 8004 dimensions, demonstrating the stability of the MI estimation algorithm in high dimensional signal spaces.

The information gained by including the difference channel signal of an X-band monopulse radar for the F-14 data set was approximately 3 dB at 50 km and increased to 4.5 dB at 2 km due to the increased target extent relative to the antenna pattern. This experiment necessitated the use of target profiles which were matched to the range of the target to achieve maximum information transfer.

Experiments were conducted to evaluate the loss in information due to envelope processing. For the baseline data set, SNR losses in the region of 7 dB were measured.

Linear pre-processing using the fast Fourier transform (FFT) and principal component analysis (PCA), before envelope processing, were compared and the PCA algorithm outperformed the FFT by approximately 1 dB at high MI values.

Finally, the expression for multi-target MI was applied in conjunction with Fano's inequality to predict the probability of incorrectly classifying a target. Probability of error is a critical parameter for a radar user. For the baseline data set, at P(error) = 0.001, maximum losses in the region of 0.6 to 0.9 dB were measured. This result shows that these targets are easily separable in the signal space.

This study was only the proverbial "tip of the iceberg" and future research could extend the results and applications of the techniques developed. The types of targets and configurations of the individual targets could be increased and analysed. The analysis should also be extended to describe effects internal to the radar such as phase noise, spurious signals and analogue to digital converters and external effects such as clutter and multipath. The techniques could also be applied to quantify the gains in target recognition performance achievable for multistatic radar, multiple input multiple output (MIMO) radar and more exotic concepts, such as the fusion of data from multiple monostatic microwave radars with multi-receiver multi-band passive bistatic radar (PBR) data.

Dedication

“I lift my eyes up to the mountains;
From whence shall my help come?
My help comes from the Lord,
Who made heaven and earth.”

Psalm 121:1-2

To my wife, Therese, for her continuous love and for persevering with me through the journey of research which became this thesis.

To my daughters Elise and Adeline for their love, understanding, inquisitiveness and companionship. Elise, thank you for writing the odd paragraph while I wasn’t looking.

To my parents, for their unselfish support over many years, and for putting up with the eccentricities of raising an engineer.

To my wife’s parents, Mr. and Mrs Westhof, for their support and encouragement.

In memory of Mr. Westhof who passed away on the 13th of March 2013.

Acknowledgements

I am indebted to many people for their advice and assistance towards the successful completion of this thesis.

In particular to my supervisors from UCL, Professor Chris Baker and Professor Karl Woodbridge, thank you for all the guidance and encouragement as well as all the proof reading and insightful comments.

Thank you to Chris Baker for mentoring me while relocating over three continents. Unfortunately I was never able to answer your original question: "What is the radar return telling us about the target?", but I hope that this study is at least the first step in understanding that question.

Thank you to Karl Woodbridge for keeping up the pressure and especially for all the help and guidance during the preparation for the PhD viva.

Thank you to Professor Hugh Griffiths for reviewing the thesis and for all the very insightful comments and suggestions.

Thank you to Professor Wimpie Odendaal for making the scaled Boeing model measurements possible and for reviewing the original research proposal.

A special thank you to Andre McDonald for many hours of philosophical debates about the derivation and interpretation of some of the equations in this thesis. Also, thank you for the loan of some excellent reference books, without which this study would not have been completed. Also, thanks for teaching me the finer points of railgun operation by example.

Thank you to Willie Nel and Francois Anderson for asking many questions in striving to understand the implications of this work. Your questions always forced me to think of new ways to explain the theory and to interpret the results.

Thank you to my good friends and colleagues Francois Maasdorp, Momil Jamil and Kevin Kloke for the mutual support and encouragement while we were all working on our PhDs and Masters degrees.

Finally, thank you to Professor Louis Linde for teaching me the fine art of signal processing and for introducing me to information theory.

TABLE OF CONTENTS

TABLE OF FIGURES

TABLE OF TABLES

TABLE OF SYMBOLS

Symbol	Description
$\mathbf{B}$	Vector Magnetic field in tesla (T).
c	Speed of light. (Defined as 299 792 458 m/s in a vacuum) (Approximately 2.997x10^8 m/s in air)
$\mathbf{E}$	Vector Electric field in volt per metre (V/m).
$E\{X\}$	Expected value of random variable X .
f	Frequency in hertz.
g	Conductivity in siemens per metre (S/m)
$H(X)$	Entropy of a discrete random variable X in bits.
$H(X\|Y)$	Condition entropy for discrete random variables X and Y.
i	Unit imaginary number, $i=\sqrt{-1}$
$I(X)$	Self information in bits.
$I(X,Y)$	Joint self information between two random variable in bits.
$I(X;Y)$	Mutual information between random variables X and Y in bits.
$\mathbf{J}_\mathrm{S}$	Current density in amperes/metre
λ	Wavelength in metres.
N_0	Noise power spectral density in W/Hz
P_e	Probability of error.
R	Range in metres.
σ	Standard deviation of Gaussian Noise OR radar cross section (RCS) in square metres. (depending on context)
$\sqrt{\sigma_{ij}}$	Square root of RCS. (complex valued)
ω	Radian frequency in rad/s.
W	Bandwidth in Hz

TABLE OF ABBREVIATIONS

3-D	Three-dimensional
ATR	Automatic Target Recognition
AWGN	Additive White Gaussian Noise
CAD	Computer Aided Drawing
CEM	Computational Electromagnetic
EDERI	Electronic Defence Evaluation Research Institute
EM	Electromagnetic
ESM	Electronic Surveillance Measures
FFT	Fast Fourier Transform
FSK	Frequency Shift Keying
GO	Geometrical Optics
JEM	Jet Engine Modulation
JTC	Joint Tracking and Classification
HBM	Helicopter Blade Modulation
HF	High Frequency
HRRP	High Range Resolution Profile
ISAR	Inverse Synthetic Aperture Radar
mD	micro Doppler
MFSK	Multiple Frequency Shift Keying
MI	Mutual Information
MIMO	Multiple Input Multiple Output
MMW	Millimetre Wave
MoM	Method of Moments
MTI	Moving Target Indicator
NCTI	Non-Cooperative Target Identification
NCTR	Non-Cooperative Target Recognition
OTHR	Over The Horizon Radar

PCA	Principle Component Analysis
PEC	Perfect Electrical Conductor
pdf	Probability density function
PO	Physical Optics
PRF	Pulse Repetition Frequency
PSK	Phase Shift Keying
QAM	Quadrature Amplitude Modulation
Q^2PSK	Quadrature-Quadrature Phase Shift Keying
RAM	Radar Absorbing Material
RCS	Radar Cross Section
RF	Radio Frequency
RV	Random Variable
SAR	Synthetic Aperture Radar
SBR	Shooting and Bouncing Rays
SNR	Signal to Noise Ratio
TEWA	Threat Evaluation and Weapons Assignment
UHF	Ultra High Frequency
UWB	Ultra Wideband
VHF	Very High Frequency

1 Introduction

Radar systems are the only sensors which can detect targets at long ranges and in any weather conditions. This property makes radar systems invaluable in several military and commercial applications, such as the detection and tracking of aircraft as well as satellites. In most applications of radar the ability to classify or recognize the targets of interest will enhance the situational awareness and decision making capability of the users of the radar system.

Due to the fact that clutter and multipath effects increase the complexity of the analysis for surface based targets, this study will focus on airborne targets, in scenarios where the clutter and multipath can be minimized by the radar’s antenna pattern. This scenario is often referred to as “blue sky” by radar designers. Once techniques have been developed for the “blue sky” scenario, the insight and techniques which have been developed can be applied to extend the analysis to the cases where clutter and multipath effects are present.

The following sections define the non-cooperative target recognition (NCTR) problem and its context as well as describing the recognition process from a radar point of view and a theoretical point of view.

1.1 Problem Definition

Modern ground-based and airborne radars are all-weather sensors that are capable of detecting airborne targets at ranges in excess of 200 km. The military, and to a lesser extent civilian, utility of such radars will be greatly increased if the radar can automatically determine the class or identity of the detected target. Non-cooperative target recognition (NCTR), is an active field of research that attempts to address this problem. No reliable solution to the problem, which allows for the recognition of all airborne targets, in all scenarios, has as yet been reported in the open literature. Some techniques, such as inverse synthetic aperture radar (ISAR), jet engine modulation (JEM) and helicopter blade modulation (HBM) have however been demonstrated in realistic scenarios. These algorithms often do not identify the exact aircraft, but group the targets into classes which have similar characteristic signatures. Some of the algorithms (e.g. JEM) are also only valid for a restricted set of target aspect angles. Due to the expense of obtaining measured data sets containing large numbers of targets, the performance of the algorithms is often simulated, or reported for a relatively small number of targets. Most of the literature on radar target recognition is based on the application of a feature extraction function followed by a classifier, which usually has to be trained. Many of these studies are based on measured data which limits the ability to predict the performance of an algorithm to the noise and distortion levels of the recorded target data.

The open literature does not contain any reports of radar systems which have been designed with NCTR as the radar’s primary function. There is also a distinct lack of theory regarding the prediction of NCTR algorithms' performance, which implies that a radar system cannot be designed or optimized for a specified level of target recognition performance. The current mindset to improving a radar's recognition performance involves an increase in its native resolution, in one or more of the dimensions in which it is able to sense a target. During the design phase of a radar system a designer's insight would be improved if theoretical and analytical techniques were developed which could address the following problems:

1. Whether NCTR is fundamentally possible in a given radar system.
2. What level of NCTR is possible (e.g. classification or identification).
3. How to predict the limit of NCTR performance if the system parameters and targets of interest are known.

The designer of the radar system requires techniques for predicting the recognition performance of a radar sensor during the design stage of the system. Insight into the interaction between the sensor and the recognition algorithm is a prerequisite for optimizing the overall system with regards to its recognition function. Before the detailed design of the radar is commenced upon, it is necessary for the design team to be able to compare various preliminary designs against one another. Analysis tools that enable designers to make tradeoffs between system concepts based on recognition performance could lead to radical departures from accepted norms in radar design principles.

Several new radar architectures and related processing techniques are currently emerging from radar research laboratories around the world under the broad grouping of Multiple Input Multiple Output (MIMO) radar. This grouping is a generalisation of older architectures such as bistatic and multistatic radars. If the correct analysis tools can be applied to such concepts to show that they exhibit the potential for significant improvement in recognition performance over monostatic systems, the development cost of such systems could be justified.

A set of theory that allows for the analysis and fair comparison of disparate radar systems in the abovementioned scenarios would thus allow designers to compare conceptual radar designs. If absolute performance bounds can be found which are impossible to exceed then the radar designer can use these to ascertain how far away the performance of proposed NCTR algorithms are from the maximum achievable level of performance.

1.2 Generalized Radar NCTR System

Most researchers interested in improving the recognition performance of a radar-based NCTR system focus on the recognition algorithm. The most important question is actually whether or not the returns from a target are classifiable in the first place. The next question would pertain to how the returns are altered by mechanisms internal to the radar, and its design parameters. The next step would be to consider the internal and external subsystems to the radar which support the recognition function. It is thus important to realize that the recognition performance of a radar-based NCTR system could be mediocre on a system level as a side effect of having neglected the subsystems and functions which are necessary to support the recognition function, despite having developed a recognition algorithm with extremely high levels of performance during initial testing phases. Some of these functions might not reside in the physical radar system, but exist in its external environment. Support systems can be disparate and examples range from equipment and software for the characterisation of potential targets to the training of the radar operators and the development of military doctrine. Figure 1.1 is a diagrammatic representation of a generalized radar based recognition system which indicates the elements which would influence the recognition performance as well as elements which would be responsible for the definition of metrics for adequate recognition performance for higher level systems such as a Threat Evaluation and Weapons Assignment (TEWA) system or the command and control system.

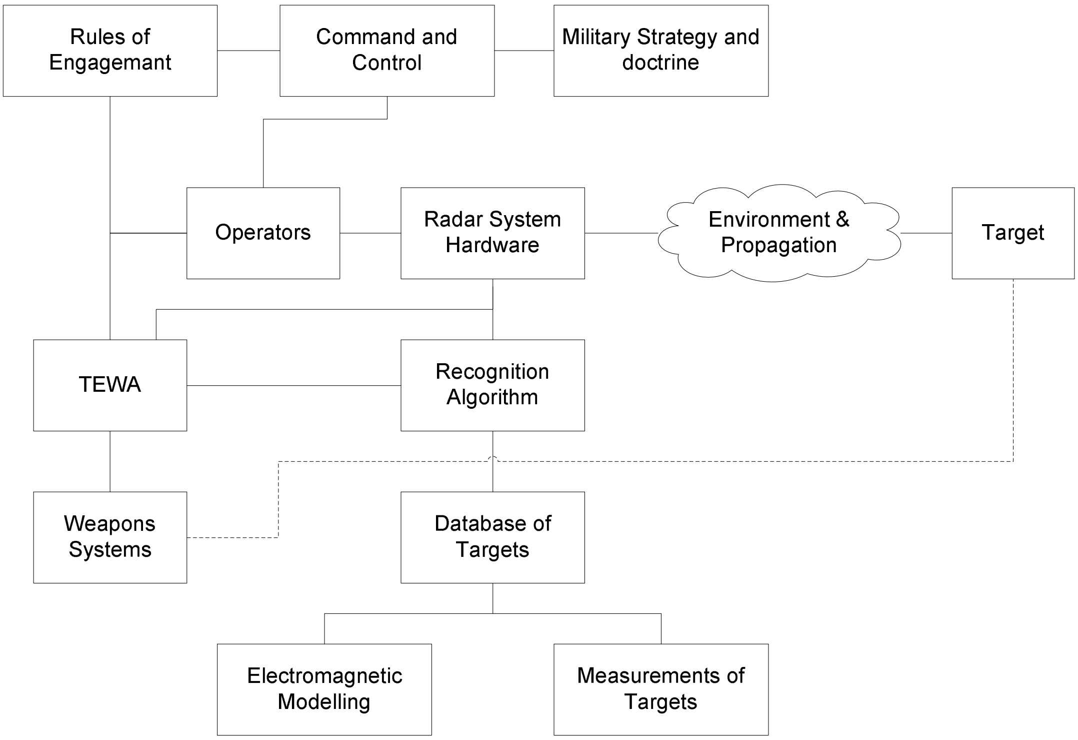


Figure 1.1: Simplified system context diagram showing the external environment in which the radar based recognition algorithm operates for a military application.

It should be noted that the purpose of this figure is to provide an overview of the system - it is not a rigorous representation of the system and the environment within which it exists and interacts with other entities. It should also be noted that many of these components are interdependent and that most of the links between the components are bidirectional. For example, if NCTR becomes available in a radar system, the command and control strategy might be updated, which in turn could lead to a different set of requirements being placed on the NCTR function.

In the figure the target database is shown as a separate entity to the recognition algorithm. In some cases these two components could be merged, for example when template matching is used to compare the radar's received signal to the signals stored in the database. In other cases, such as trained classification algorithms, the data base is implicitly included in the classification algorithm in the training step of the classifier. It is also not required that the whole database be available in a radar at a specific time as the threat targets could vary depending on the scenario.

This figure also highlights the fact that a successful NCTR system requires technical support from a support organisation that is competent in radar and radar-related fields such as the electromagnetic simulation of target signatures as well as the measurement of target signatures for the creation of the recognition database.

The radar system, the electromagnetic environment and the physical properties or characteristics of the target all contribute to the generation of a set of signals, received by the radar, which can be exploited for recognition purposes. The main focus of this study will be the analysis of the amount of information which a set of received signals conveys to the radar system about the class of target or the identity of the target being illuminated by the radar.

It should be noted that this study will not address techniques for the fusion of data from multiple radar systems or other sensor types such as optical imaging sensors.

1.3 Recognition Process

It is necessary to define the nomenclature relating to the various processing stages required in a radar target recognition system from a radar point of view as well as a theoretical point of view.

1.3.1 Radar Recognition Process

Two acronyms for the recognition of radar targets using automated techniques have developed over time. The first term is: “automatic target recognition” (ATR) which is used for air-to-ground scenarios such as synthetic aperture radar (SAR). The second term is “non-cooperative target recognition” which is used for ground-to-air and air-to-air scenarios. The remainder of this document will use the term NCTR as the research deals with the recognition of airborne targets.

According to Cohen, in Chapter 12 of [Long1992], the NCTR process can be seen as consisting of five functional blocks within the radar. These functions, which will be discussed in this section, are shown in the flow chart in Figure 1.2 below.

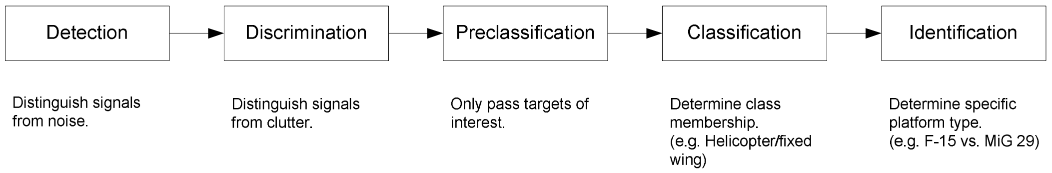


Figure 1.2: Functions which comprise the recognition process, after [Long1992].

Detection is the most fundamental level of the recognition process. It deals with the problem of distinguishing a target from thermal noise. Classical optimum detection theory is well developed, and can be applied in most design problems, except in cases where it is necessary to reduce the false alarm rate to effectively reduce the processing load on the recognition system. Discrimination is defined as the process of distinguishing potential targets from the surrounding clutter. This process is responsible for reducing the number of potential targets which are presented to higher levels of the recognition system, to reduce throughput requirements and ensure good recognition performance. An example of a discriminatory process is the classical moving target indication (MTI) technique. In some cases it is necessary to perform discrimination before detection, and in some cases the two functions are merged.

The function of pre-classification in a recognition system is to exclude from further consideration, those targets which pass the detection and discrimination stages, but which were not of sufficient interest to be passed on to the classification and identification algorithms. This process is also referred to as “alien separation”. For example, in a scenario in which the classification/identification algorithms have been trained to identify specific fixed wing military jet aircraft, slow targets such as moving ground vehicles and small propeller aircraft will be discarded by the pre-classifier.

Classification describes a process whereby the remaining targets are characterized as belonging to a certain class of vehicles, such as helicopter, fixed wing etc. In this case the physical attributes of the target play an important role in that they are usually the means by which targets can be grouped. The radar has to be able to measure these attributes in terms of commonalities and differences accurately to ensure the correct grouping of a target. Classification is thus the process of determining the nature of a potential threat, and possibly whether it is friendly or not.

Identification is the final and most difficult stage of the recognition process. The goal is to identify the specific platform as, for example, either an F-15 or a MIG-29. The algorithms used for this stage are similar to those used in the classification section, as both types of algorithms are trained on a representative subset of the targets of interest. In most cases some set of features is used to represent the targets and use is made of clustering, pattern recognition, or artificial intelligence techniques to determine the proper declaration. The ultimate goal of identification in the air defence scenario would be to identify the exact aircraft, down to its serial number. If this can be achieved, then the intention of the pilot can often be inferred. Once this decision has been made the target is referred to as “friend”, “foe” or “neutral”. In some literature, the identification of a specific aircraft is referred to as recognition, thus adding an extra function after classification.

More recently, extra terms have been added to discriminate between variants of a single target and variations of a single target’s configuration. The terms “characterisation” and “finger printing” are used to describe this type of discrimination [BlacknellGriffiths2003].

From the above description, it can be seen that the recognition process forms a hierarchical pyramid with an increasing level of certainty about the identity of target.

1.3.2 Theoretical Fields of Study Required for the Radar Based NCTR Process

This section introduces some of the theoretical fields of study which underpin NCTR. Figure 1.3 below shows a conceptual diagram of the recognition process in a radar system.

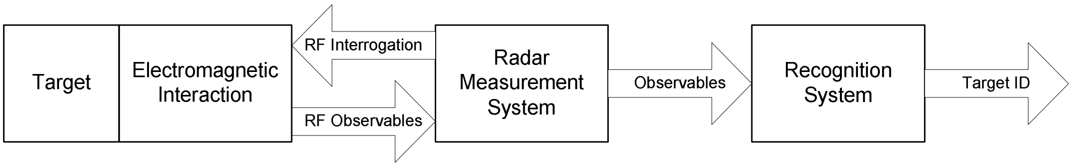


Figure 1.3: Conceptual target recognition process.

The radar system transmits a set of waveforms to interrogate the target in the electromagnetic domain. These waveforms interact with the target geometry, are then reradiated and consequently observed by the radar system. The radar system applies signal processing techniques to convert these measurements into observables which are then passed to the recognition system.

Most published NCTR techniques encompass some or all of the above processes, but the focus is often on the identification sub-system, which is described mathematically by the field of pattern recognition. This is due to the fact that the purpose of an NCTR algorithm is to match some set of measurements of a target to some set of stored parameters for each possible target. If the measurements match the stored parameters to within a certain measure of closeness then the target can be identified. The main processes contained in pattern recognition, and their purposes are shown diagrammatically in Figure 1.4 below.

The “Measurement of observables” process entails the measurement of some form of raw data. In a radar application it would entail measuring a set of parameters pertaining to a single target which can be sensed by the radar. “Feature extraction” is the process of projecting the multidimensional measured data set onto a subspace of a smaller dimension than the original data set. “Classification” is the process of creating a set of boundaries in the feature space. The “Recognition” process is responsible for assigning a class type to each of the sets defined in the classification step.

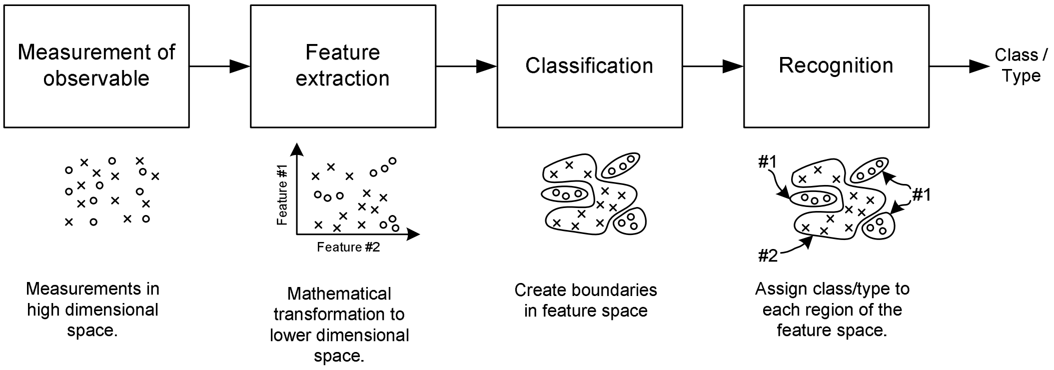


Figure 1.4: Generic pattern recognition block diagram.

A large body of knowledge exists in the open literature pertaining to feature extraction techniques, but often the designer will have to develop a new feature extraction technique for the problem at hand. Pattern recognition is an established field rooted in statistical decision theory [Duda2001]. The aim of the pattern recognition system is usually to minimize the probability of error, or to minimize the average cost of an error.

1.4 Information Theory

Information theory is the field of study which mathematically formalizes the relatively vague concepts of a “message” and the amount of “information” the message contains. The field came into being when Claude Shannon published his paper titled “A mathematical theory of communication” in 1948 [Shannon1948]. Shannon realized that the content of a message being communicated was uncoupled from the representation of that message. This realization ushered in the “digital age” underpinned by the idea that all information can be represented as zeros and ones. Shannon studied the amount of information contained in random variables and defined two limits: the data compression limit which predicts the minimum data representation limit for compression algorithms and the data transmission limit which predicts the maximum rate at which error-free transmission of data can be achieved. The data compression limit is defined by the entropy of a random variable which is a measure of the uncertainty of its outcomes. The transmission limit is called the “channel capacity” and is based on the definition of “mutual information”, which is the reduction in the uncertainty (entropy) of a random variable due to the observation of another random variable. Mutual information (MI) is a more general measure of the dependence between two random variables than most commonly used measures such as the correlation coefficient and other low order joint moments. The quantitative measure of the information, in bits (binary digits), provided about the event $x = a_k$ by the occurrence of the event $y = b_j$ is given by the logarithm of the ratio of the a posteriori probability to the a priori probability of a_k as follows [Gallager1968]

$$I_{X;Y}\left(a_k;b_j\right)=\log_2\frac{P_{X|Y}\left(a_k\,|\,b_j\right)}{P_X\left(a_k\right)},\tag{1.1}$$

where $X=\{a_1,\ldots,a_K\}$ and $Y=\{b_1,\ldots b_J\}$ are the sample spaces associated with a_k and b_j respectively. $P(a)$ is the standard function mapping an event, a, to its associated probability. Note that mutual information can be denoted using a semicolon, a colon or a comma, but use will be made of the semicolon notation in this thesis. The mean value of this expression is called the “average mutual information” and is given by:

$$I\left(X;Y\right)=\sum_{k=1}^{K}\sum_{j=1}^{J}P_{XY}\left(a_k,b_j\right)\log_2\frac{P_{X|Y}\left(a_k\,|\,b_j\right)}{P_X\left(a_k\right)}.\tag{1.2}$$

It will be shown in Chapter 4 that $I\left(X;Y\right)\geq 0$ and that it attains the value of zero only if X and Y are statistically independent. In this case the two variables do not depend on each other in any way, so there is no information to be gained about the one by observing the other.

If X is the input to a channel, and Y is its output, then the capacity, C, of this channel is defined as the maximum mutual information over all possible choices of the input probability distribution:

$$C=\max_{P_X(x)} I\left(X;Y\right),\tag{1.3}$$

and is measured in bits per channel usage. This is the maximum rate at which error free communication is possible over the channel. Channel capacity gave communication engineers a limit to benchmark their data transmission systems against and has in turn led to the development of technologies for high speed digital data transmission over wired and wireless media as well as technologies for digital data storage.

1.5 Aims and Objectives

The main objective of this research is to demonstrate that information theory, and specifically the concept of mutual information, can be used to compare the recognition performance of disparate approaches to designing a radar concept which is capable of target recognition without resorting to choosing specific feature extraction and classification algorithms. To this end it is necessary to develop a good understanding of the mathematical principles underpinning information theory, as well as the theory describing the interaction of the transmitted radar pulse with the target and the ways in which the radar can sense and exploit this interaction to classify or recognize the target. Once the mutual information has been calculated, it can be compared to the MI for various other recognition options, or use can be made of Fano’s inequality [Fano1961, Cover2006] to estimate the probability of error, which, in this case is the probability of incorrectly assigning a received signal to a specific target orientation and/or type of target.

1.6 Contributions

The main novel contribution of this thesis is the development of a technique for characterising the maximum achievable amount of information, specifically the mutual information, which can be extracted from high range resolution profiles of an airborne target for a specific radar waveform in the presence of additive white Gaussian noise. The most important quality of this information is that it directly determines the maximum achievable probability of correct classification (or lowest probability of error), through Fano’s inequality, for the given set of signals and for a given signal to noise ratio. It is impossible to design a classifier that achieves better performance than this upper bound.

Determination of the information content of different sets of signals also allows for the comparison of the amount of information which can be extracted from a target for various waveforms, or the comparison of the amount of information which can be extracted from various types of targets. In the latter case this can be converted to a probability of making a misclassification error between different targets. This error represents the lowest possible error probability which can be achieved for the specific radar waveform and thus serves as a benchmark for future researchers working on the development of target recognition systems and/or algorithms.

The calculation and comparison of the information content was demonstrated for three fighter targets in the paper titled “Evaluation of the information content of wideband and ultra-wideband radar returns from an F-14, F-15 and F-16 using asymptotic electromagnetic techniques”, in which it was shown that for the specific aircraft models, the F-15 required an increase in signal to noise ratio of approximately 3 dB to reach the same level of recognition performance. The amount of information which could be extracted from these three targets was also compared for disparate designs of high range resolution waveforms, and it was shown that surprisingly little information was lost if use is made of frequency stepped waveforms which are ambiguous in range. The paper titled “On the gain in recognition performance due to the addition of polarisation in an X-band High Range Resolution radar evaluated for F-18 and F-35 targets using asymptotic EM techniques” extended the technique to radars which can sense the full polarisation matrix and showed that signal to noise ratio gains in the region of 5 dB are achievable for full polarisation over a single linear polarisation. In this thesis the polarisation based performance improvement was also characterised for measurements of a 1:25 scale model of a Boeing 707 measured in a compact range, thus demonstrating the applicability of the information content estimation algorithm to measured data sets.

The above examples made use of the complex valued returns from the target and thus defined the absolute maximum classification performance which could be achieved. The second contribution was the modification of the abovementioned technique, for the estimation of the information content, which allowed for the effect of envelope processing on the information content to be characterised. It was shown that this very standard processing in a radar's signal processor can lead to a loss in signal to noise ratio required for the same level of recognition performance of approximately 7 to 9 dB.

All the examples above were focussed on the comparison of the amount of information which could be extracted from a single target. The third contribution was the extension and demonstration of the information estimation technique for the estimation of the probability of misclassification between various targets. It was shown that losses on the order of 0.5 to 1 dB can be expected if the classification algorithm is required to classify between three targets.

In all cases the rigorous derivations of expressions for mutual information have been included for the benefit of future researchers in this area.

1.6.1 Publications

The following publications were generated during this research:

- J.E. Cilliers, J.C. Smit, A. McDonald, C. Baker & K. Woodbridge, Evaluation of the information content of wideband and ultra-wideband radar returns from an F14, F15 and F16 using asymptotic electromagnetic techniques, IET International Radar Conference, 2012
- J.E. Cilliers, J.C. Smit, C. Baker & K. Woodbridge, On the gain in recognition performance due to the addition of polarisation in an X-band High Range Resolution radar evaluated for F-18 and F-35 targets using asymptotic EM techniques, 2015 IEEE Radar Conference (RadarCon), 2015, pp. 1296-1299
- J.E. Cilliers, J.C. Smit, C. Baker & K. Woodbridge, On the information gain obtainable by exploitation of the monopulse difference channel for an X-band high range resolution radar evaluated using asymptotic EM techniques, IEEE Radar Conference (Johannesburg), 2015, pp. 533-538
- J. Cilliers, J. Steyn, J. Smit, C. Pienaar & M. Pienaar, Considering CAD Model Accuracy for Radar Cross Section and Signature Calculations of Electrically Large Complex Targets, International Radar Conference, 2014
- J.E. Cilliers & J.C. Smit, On the trade-off between mainlobe width and peak sidelobe level of mismatched pulse compression filters for linear chirp waveforms, EuRAD Radar Conference, 2009, pp. 9-12

- J.C. Smit, J.E. Cilliers & E. Burger, Comparison of MLFMM, PO and SBR for RCS investigations in radar applications, IET International Conference on Radar Systems (Radar 2012), pp. 1-5
- F. Maasdorp, J. Cilliers, M. Inggs & C. Tong, Simulation and measurement of propeller modulation using FM broadcast band commensal radar, IET Electronics Letters, 2013, vol. 49, no. 23, pp. 1481-1482
- F. Maasdorp, J. Cilliers, M. Inggs & C. Tong, FM band commensal radar technology used for the detection of small aircraft and the measurement of propeller modulation, 2015 IEEE Radar Conference (RadarCon), pp. 664-668

1.7 Thesis Layout

The rest of this document is structured as follows: Chapter 2 is a high level overview of the electromagnetic theory which describes the interaction between the signal transmitted by the radar and the target. The aim of this chapter is to start with the underlying electromagnetic theory and from this to develop a list of target attributes which can be sensed by the radar. Chapter 3 gives an overview of various techniques, based on the discussion in Chapter 2, which other researchers have applied to the problem of radar target recognition. This summarises the context for the research in this study as well as giving some examples from the literature related to the problem of radar target recognition and approaches to solving it. Chapter 4 is a theoretical development of the information theoretic concepts required for the analysis of the radar target recognition problem and concludes the background material required for this study. This chapter is intended as an introduction to the field of information theory and focuses on the development of the theory required to define mutual information. Two non-standard examples are developed to show the interplay between entropy and mutual information. These examples were also chosen to show that mutual information is a more general measurement the correlation between random variables than standard techniques such as correlation coefficients. Based on the theory in Chapter 4, Chapter 5 gives detailed derivations of various forms of mutual information which will be required for the analysis of radar recognition problems. The aim is to derive the expressions for mutual information in enough detail that future researchers will be able to verify the final expressions. Where possible, the mutual information results are verified against standard results in the literature. These results are usually from digital communications theory as this is the area in which information theory was originally developed and has driven the achievable performance since its inception.

The next seven chapters present the results for various analyses which were conducted. Chapter 6 presents an overview of the various experiments which were conducted in the next six chapters and gives the types of targets used for each experiment. Chapter 7 focuses on the analysis of point scatterer type targets and the results are compared to the Shannon bound and optimal waveforms as an extra verification. In Chapter 8 the results of

MI analyses for three fighter targets are presented. Use is made of freely available computer aided design (CAD) models for the F-14, F-15 and F-16 fighter aircraft as examples of more realistic targets. The RCS of these targets is calculated using a computational electromagnetic computation (CEM) package which makes use of asymptotic EM techniques. The targets are analysed over a 480 MHz bandwidth at X-band. Chapter 9 extends this analysis for the same targets by making use of ultra-wideband waveforms in an attempt to extract more information about the target. These waveforms span the 2-18 GHz band. Chapter 10 extends the analysis to multi-channel radar receivers. The gain achievable from polarisation for the F-18 and F-35 CAD models is analysed. Measurements made of a 1:25 scale Boeing 707 model, in the University of Pretoria’s compact range, are also analysed over a 2-18 GHz frequency span. This analyses also showed that the MI calculation techniques are valid for very high dimensional problems as the highest dimension required for the 707 data was 8004. Lastly the extra information which can be gleaned from the difference channel of a monopulse radar is analysed for the F-14 model. All the preceding results were calculated using the complex valued target profiles. Chapter 11 presents the effect of envelope processing on the amount of data which can be extracted from the target return. The effect of linear pre-processing before calculating the envelope is also analysed. Chapter 12 makes use of Fano’s inequality to calculate the probability of error from the mutual information results for the F-14, F-15 and F-16. This concludes the results section and shows how the mutual information can be converted into a radar performance parameter which is important to the radar user. The results chapters show the applicability and insights achievable by applying the mutual information concepts to disparate approaches to the radar recognition problem. Chapter 13 presents the conclusions as well as discussing concepts for further research in this field.

2 Interaction of the Electromagnetic Field with a Target

To recognize an unknown target, the only physics which a radar designer can exploit is the interaction between the electromagnetic (EM) field produced by the radar and the target that is being interrogated. In most cases of interest, this interaction is of sufficient complexity to render it mathematically intractable even for a single frequency and single aspect angle. The normalised echo returned from the target to the radar is scaled by a factor known as its radar cross section (RCS). The RCS of a target is a function of the radar's frequency, bandwidth and polarization as well as the shape of the target, its orientation relative to the radar and the materials from which it is constructed. At microwave frequencies the RCS is highly dependent on the geometry of the target, and can thus also be influenced by changes in the target structure due to its interaction with the atmospheric turbulence, manoeuvre induced flexing, moving control surfaces and external stores. The illuminating and received signal can also be a function of the radar-target geometry, especially in cases where multipath can be encountered.

This section summarises the various scattering mechanisms and applicable electromagnetic theorems and is concluded by a discussion of electromagnetic phenomena which a radar could sense to enable recognition of an aerial target. The discussion of the EM phenomena is at a conceptual level and the reader is referred to standard texts such as [Knott2004] and [Jenn2005] for an in-depth discussion of RCS and techniques for the calculation thereof.

2.1 Electromagnetic Mechanism for Scattering

The radar is designed to transmit EM energy, usually in the form of a time limited pulse via its antenna. This energy propagates through the earth's atmosphere and is reflected from targets of interest as well as other objects in the environment which are not of interest to the radar, such as ground clutter. The reflected energy is converted by the radar's antenna to a signal which can be analysed by the radar's signal processor. The characteristics of the received signal, such as its amplitude and Doppler shift are then exploited to detect targets.

The interaction between the EM field which illuminates the target in conjunction with the target's geometry is responsible for creating the scattered field, which is then sensed by the radar's antenna. For all target recognition research, this is the fundamental physical principle which is of interest as the scattered field is a function of the target's geometry and the target's motion. The target's geometry and state of motion is thus encoded onto transmitted signal and reflected back to the radar. This signal is received by the radar and can be utilised for the detection of targets as well as the classification and recognition of targets, if enough information about the target has been transmitted back to the radar. The physics which governs EM fields and their interaction with the environment is dictated by Maxwell's equations, which are discussed in the following section. This will lead to the definition of the Radar Cross Section (RCS), which is a measure of the EM reflectivity of an object.

2.1.1 Maxwell's Equations for Electromagnetic Fields in Free Space

The generation and propagation of EM fields as well as the interaction of EM fields with objects are predicted mathematically by Maxwell's equations, which he published in 1862, and are referred to as "classical field theory". These equations are not exact, but are an approximation of a subset of a more general theory known as "quantum electrodynamics". Luckily, this is not required for the calculation of EM fields required for the prediction or measurement of RCS. Maxwell's equations relate the EM fields to the sources which are responsible for their generation. These sources are either charge and current or the EM fields themselves if these fields are time variant. In free space, where there is no charge or current present, Maxwell's equations are given by the following set of four coupled first order differential equations:

$$\nabla \cdot \mathbf{E} = 0 , \tag{2.1}$$

$$\nabla \cdot \mathbf{B} = 0 , \tag{2.2}$$

$$\nabla \times \mathbf{E} = -\frac{\partial \mathbf{B}}{\partial t} , \tag{2.3}$$

$$\nabla \times \mathbf{B} = \mu_0 \varepsilon_0 \frac{\partial \mathbf{E}}{\partial t} , \tag{2.4}$$

where $\mathbf{E}$ is the electric field vector, $\mathbf{B}$ is the magnetic field vector. The constant ε_0 is the permittivity of free space and is defined as

$$\varepsilon_0 = 8.8541878176\ldots\times 10^{-12} \quad \text{F/m} . \tag{2.5}$$

The constant μ_0 is the permeability of free space and is defined as

$$\begin{aligned} \mu_0 &= 4\pi \times 10^{-7} \quad \text{H/m} \\ &\approx 1.2566370614\ldots\times 10^{-6} \quad \text{N/A}^2 . \end{aligned} \tag{2.6}$$

The set of equations can be decoupled by taking the curl of both sides of (2.1) and (2.4) and applying the following vector derivative identity:

$$\nabla \times (\nabla \times \mathbf{A}) = \nabla (\nabla \cdot \mathbf{A}) - \nabla^2 \mathbf{A} \, . \tag{2.7}$$

For the $\mathbf{E}$ field this process is simplified as follows:

$$\begin{aligned} \nabla \times (\nabla \times \mathbf{E}) &= \nabla \times \left(-\frac{\partial \mathbf{B}}{\partial t} \right) \\ \nabla (\nabla \cdot \mathbf{E}) - \nabla^2 \mathbf{E} &= -\frac{\partial}{\partial t} \nabla \times \mathbf{B} \end{aligned} \tag{2.8}$$

where time and space derivatives of $\mathbf{B}$ have been interchanged, if $\mathbf{B}$ is well behaved, which is usually the case. Now substituting (2.1) and (2.4) into (2.8) gives

$$\begin{aligned} \nabla (\nabla \cdot \mathbf{E}) - \nabla^2 \mathbf{E} &= -\frac{\partial}{\partial t} \nabla \times \mathbf{B} \\ -\nabla^2 \mathbf{E} &= -\mu_0 \varepsilon_0 \frac{\partial^2 \mathbf{E}}{\partial t^2} \\ \nabla^2 \mathbf{E} &= \mu_0 \varepsilon_0 \frac{\partial^2 \mathbf{E}}{\partial t^2} \end{aligned} \tag{2.9}$$

This is a single equation for $\mathbf{E}$, although it is a second order differential equation. Similarly, for the magnetic field, $\mathbf{B}$, the same approach can be followed to arrive at the following second order differential equation:

$$\begin{aligned} \nabla \times (\nabla \times \mathbf{B}) &= \nabla \times \left(\mu_0 \varepsilon_0 \frac{\partial \mathbf{E}}{\partial t} \right) \\ \nabla (\nabla \cdot \mathbf{B}) - \nabla^2 \mathbf{B} &= \mu_0 \varepsilon_0 \frac{\partial}{\partial t} \nabla \times \mathbf{E} \\ -\nabla^2 \mathbf{B} &= \mu_0 \varepsilon_0 \frac{\partial}{\partial t} \left(-\frac{\partial \mathbf{B}}{\partial t} \right) \\ \nabla^2 \mathbf{B} &= \mu_0 \varepsilon_0 \frac{\partial^2 \mathbf{B}}{\partial t^2} \end{aligned} \tag{2.10}$$

Equation (2.9) and (2.10) are three-dimensional versions of the wave equation which describe the propagation of EM waves in a homogeneous, linear medium. These equations are a necessary consequence of Maxwell's equations, but any solution to these equations also has to be confirmed to be a solution to Maxwell's equations as well. The fact that the wave equation is a consequence of Maxwell's equations does not imply that the converse is true. In a vacuum, each Cartesian component of $\mathbf{E}$ and $\mathbf{B}$ have to satisfy the three-dimensional wave equation, which for the z component is given by:

$$\nabla^2 f = \frac{1}{v^2}\frac{\partial^2 f}{\partial t^2}, \tag{2.11}$$

where v is the speed of the propagating wave. For the case of EM propagation in a vacuum, this is given by:

$$v = \frac{1}{\sqrt{\mu_0 \varepsilon_0}} = c = 299\ 792\ 458 \quad \text{m/s}, \tag{2.12}$$

which is the speed of light.

One of the most important consequences of Maxwell's equations is the wave equation, which governs the propagation of EM fields in media which are homogeneous and linear. One of the solutions to the set of equations is the complex exponential plane wave which is discussed in the next section.

2.1.2 Plane Wave Solution to Maxwell's Equations

If the EM fields are uniform over a plane which is orthogonal to the direction of propagation of the field, the field is called a "plane wave". This is a special case, but is usually applicable to radar scattering problems.

The targets are illuminated by an EM field which is a plane wave, and the radar senses the portion of this EM field which is scattered from the target. The plane wave assumption is valid as long as the target is in the far field of the radar's antenna, which is given by

$$R_{ff} = \frac{2D^2}{\lambda}, \tag{2.13}$$

where λ is the wavelength of the incident field, and D is the largest dimension of the antenna. The assumption is also made that propagation of the EM waves occurs in a vacuum, as the properties of the earth's atmosphere are close enough to those of a vacuum for the purpose of this study.

The single frequency or monochromatic solution to Maxwell's equation in free space for a plane wave is harmonic in both time and space, and can be represented mathematically by the following general form:

$$\mathbf{A}(R,t) = \mathbf{A}_0(R)\exp\left(-i\left(\omega t - kR\right)\right), \tag{2.14}$$

where $i = \sqrt{-1}$, $\omega = 2\pi f$ is the radian frequency in rad/s of the EM field of frequency in f (in hertz), R is the distance travelled by the field in metres and k is the wave number, given by $2\pi/\lambda$. The wavelength, λ, of the EM field is given by $\lambda = c/f$, where c is the speed of light in the medium of propagation. The function, $\mathbf{A}_0(R)$, contains a complex entry for each of the three orthogonal directions in a three-dimensional (3-D) vector space, although the physical, and hence measureable, fields are given by the real part of (2.14). The range dependency of $\mathbf{A}_0(R)$ allows for the decay in magnitude of fields as the range from the source of the field increases. The ωt term in the exponential is responsible for generating the time dependency of the EM field, and the kR term is responsible for generating the space dependency of the EM field.

Once the plane wave responsible for the propagation of the EM field has been defined the RCS of an object can be defined in terms of the incident and scattered far fields.

2.1.3 Definition of Radar Cross Section (RCS)

The RCS of a target is defined as the ratio of the power scattered from the target relative to the power incident on the target for the case where both are plane waves. More formally the IEEE defines it as: "*RCS is defined as* 4π *times the ratio of the power per unit solid angle scattered in a specified direction to the power per unit area in a plane wave incident on the scatterer from a specified direction. More precisely, it is the limit of that ratio as the distance from the scatterer to the point where the scattered power is measured approaches infinity.*" Mathematically, this reduces to

$$\sigma = \lim_{R\to\infty} 4\pi R^2 \left|\frac{\mathbf{E}^{scat}}{\mathbf{E}^{inc}}\right|^2, \tag{2.15}$$

where R is the range to the scattering object, "scat" denotes the scattered field and "inc" denotes the incident field. The scattered EM field either has to be calculated or measured.

The scattered field for an object of limited spatial dimension decays at a rate proportional to the inverse of the range, so as R is increased the ratio of the scattered to incident power will converge to a constant value. Practically, RCS measurements are made where the value of R is between tens of metres and thousands of meters.

The scattered field is the quantity that is sensed by the radar, so once it has been determined, it can be used to characterise the EM reflectivity of the target. The only remaining detail is how the incident fields interacts with the object to produce the scattered field, which is discussed in the next section.

2.1.4 Electromagnetic Scattering from Perfectly Conducting Bodies

In this study the assumption is made that the targets of interest are constructed from perfectly electrically conducting (PEC) materials. This assumption is valid for most fighter aircraft type targets, which are usually constructed from metallic alloys. It is however not valid for fighters or external components constructed from modern aeronautical materials such as carbon fibre or fighters having coatings of exotic materials which absorb RF energy such as radar absorbent material (RAM).

The target's geometry represents a boundary between free space, in which the incident EM field is propagating, and the target itself which is constructed from a PEC material, which has the characteristic that its conductivity, g, is infinite. Due to the high conductivity the material cannot support a filed, and all field quantities inside the material are zero. In the non-conducting medium, the following boundary conditions can be derived from Maxwell's equations:

$$\hat{\mathbf{n}} \times \mathbf{E} = 0 , \tag{2.16}$$

$$\hat{\mathbf{n}} \times \mathbf{H} = \mathbf{J}_S , \tag{2.17}$$

where $\hat{\mathbf{n}}$is a unit vector in the direction of the surface normal to the PEC surface and $\mathbf{J}_S$ is the current density in amperes/metre. The electric field constraint equation forces the component of the electric field which is parallel (tangential) to the PEC surface to zero. The fields in these two equations are the total fields which consist of the vector sum of the incident and reflected fields:

$$\hat{\mathbf{n}} \times \left(\mathbf{E}^{inc} + \mathbf{E}^{scat} \right) = 0 , \tag{2.18}$$

$$\hat{\mathbf{n}} \times \left(\mathbf{H}^{inc} + \mathbf{H}^{scat} \right) = \mathbf{J}_S . \tag{2.19}$$

Using Maxwell's equations, Stratton and Chu derived an integral equation for the scattered electrical and magnetic fields. Applying the constraints for a PEC body, the Stratton-Chu equations are given by

$$\mathbf{E}^{scat} = \int_S -j\omega\mu \mathbf{J}_S \psi + \left(\hat{\mathbf{n}} \cdot \mathbf{E} \right) \nabla \psi \, dS , \tag{2.20}$$

$$\mathbf{H}^{scat} = \int_S \mathbf{J}_S \times \nabla \psi \, dS . \tag{2.21}$$

where

$$\psi = \frac{e^{-jkr}}{4\pi R} . \tag{2.22}$$

is Green's function, and R is the distance from the point on the PEC body to the point at which the field is observed. These two equations are known as the electric field integral equation (EFIE) and the magnetic field integral equation (MFIE) respectively. The scattered electric field is thus a function of surface current and charges, whereas the scattered magnetic field is only a function of electric currents. Making use of the principle of conservation of charge,

$$\left(\hat{\mathbf{n}}\times\mathbf{E}\right)=\frac{\rho}{\varepsilon}=-\frac{j}{\omega\varepsilon}\left(\nabla\cdot\mathbf{J}_S\right), \tag{2.23}$$

and applying the boundary condition at the PEC's surface, the Stratton-Chu equations can be simplified to:

$$\mathbf{E}^{scat}=\int_S j\omega\mu\mathbf{J}_S\psi+\left(j/\omega\varepsilon\right)\nabla\cdot\mathbf{J}_S\,\nabla\psi\;dS, \tag{2.24}$$

$$\hat{\mathbf{n}}\times\mathbf{H}^{inc}=\tfrac{1}{2}\mathbf{J}_\mathrm{S}+\int_S\mathbf{J}_\mathrm{S}\times\nabla\,\psi\;dS. \tag{2.25}$$

These equations are now in terms of a single unknown, namely he surface current, and are used as a starting point for numerically solving for the surface current. The result is thus a numerical approximation of an exact formulation. Either of these equations can be solved for far field scattered fields, as the electric and magnetic fields are related to each other by

$$\mathbf{H}=\frac{1}{Z_0}\hat{\mathbf{k}}\times\mathbf{E}, \tag{2.26}$$

in the far field, where Z_0 is the characteristic impedance of the medium and $\hat{\mathbf{k}}$ is the unit vector in the direction of propagation. The standard numerical technique for the solution of one of the integral field equations is the method of moments (MoM), which requires the discretization of the PEC surface (usually into flat triangles). The problem is then formulated as a set of linear equations and then solved by means of matrix inversion. Once the surface currents have been determined the scattered fields can be calculated from either of the integral field equations. For most RCS analysis cases only the far field scattered fields are of interest, so the EFIE and MFIE can be written as:

$$\mathbf{E}^{scat}=\frac{-j\omega\mu}{4\pi R}e^{-jkR}\int_S\left(\mathbf{J}_S-\mathbf{J}_S\cdot\hat{\mathbf{R}}\right)e^{j\mathbf{k}\cdot\mathbf{r}}\;dS, \tag{2.27}$$

$$\mathbf{H}^{scat} = \frac{j\omega\varepsilon}{4\pi R} e^{-jkR} \int_S \sqrt{\tfrac{\mu}{\varepsilon}} \left(\mathbf{J}_S \times \mathbf{R} \right) e^{j\mathbf{k}\cdot\mathbf{r}} \, dS \, . \tag{2.28}$$

where R is the distance from the origin to the observation point and $\mathbf{r}$ is a coordinate on the surface of the PEC object. Once the surface currents have been solved, all the bistatic fields can be calculated for a single illumination direction. For the monostatic case, the currents have to be recalculated for each new direction.

The last two expressions ((2.27) and (2.28)) are sometimes referred to as “radiation integrals” due to the fact that they predict the radiation of a PEC body once the surface currents are known. The integration operation in these expressions is not uniquely invertible, so these equations are responsible for destroying information about the target. This will become more apparent once the data processing inequality has been introduced in Section 4.4.

2.1.5 Frequency Dependent Scattering Regimes

The electromagnetic interaction between an incident field and an object is classified into three regions which depend on the size of the target (L) relative to the wavelength of the incident field (λ). These are defined as follows:

- Rayleigh region ($\lambda \gg L$).
- Resonant region ($\lambda \approx L$.)
- Optical region ($\lambda \ll L$).

The classical comparison of these regions is depicted in Figure 2.1 below which is based on Figure 3.14 in [Knott2004, p. 96], for a sphere. A sphere is one of the few objects for which a closed form expression for its RCS can be derived. Maxwell’s equations are solved for a sphere by using the Mie scattering formalism, which results in an infinite sum of a function comprising of spherical Bessel and spherical Hankel functions. This graph was calculated by limiting the number of terms in the summation. The graph shows the RCS of a sphere, of radius r, relative to the projected area of the sphere which is given by πr^2.

Most manmade targets have external components and shapes of various sizes, which implies that the scattering from a target is often a collection of scattering from the three scattering regions.

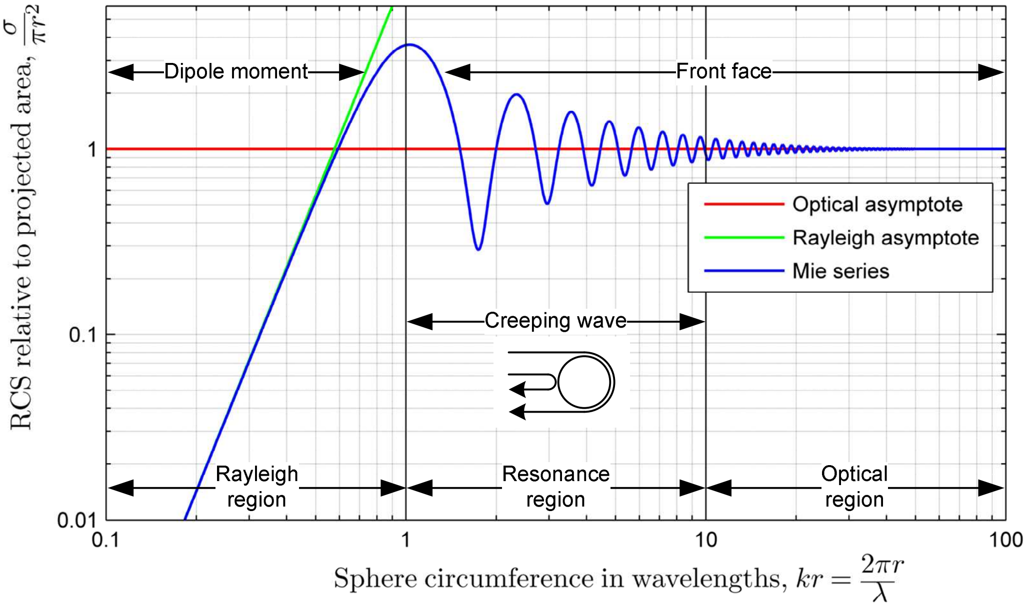


Figure 2.1: RCS of a perfectly conduction sphere showing the three scattering regimes, after [Knott2004].

2.1.6 Electromagnetic Theorems

In this section a select group of theorems pertaining to Maxwell's equations, which have bearing on RCS calculations as well as NCTR principles are discussed briefly.

2.1.6.1 Uniqueness Theorem

This theorem states that a solution to Maxwell's equations which satisfies the boundary conditions of the problem is a unique solution. This theorem guarantees that there is only one solution to a problem, and that the method of solution is actually inconsequential.

2.1.6.2 Superposition Theorem

If the medium of propagation is linear, the field intensity due two sources radiating simultaneously is equal to the sum of the field intensities of each source radiating in isolation. This theorem is a direct consequence of the linearity of Maxwell's equations.

2.1.6.3 Theorem of Similitude

This theorem provides scaling relationships between the various variables and constants in Maxwell's equations. This is very important to the field of NCTR and RCS prediction as scaled models of targets often have to be characterised due to the full scale targets not being available or due to the size constraints of measurements chambers.

In the equations that follow, unprimed variables represent full scale variables, and primed variables represent scaled variables. The scaling factors for time and length are defined as

q and p respectively. The scaling factors for the electric field and the magnetic field are given by α and β respectively. The effects of these scaling factors are tabulated below. The third column is for the special case, which occurs when $\alpha = \beta$ and $p = q$, which is referred to as the "geometrical model".

Table 1: Scaling factors for electromagnetic quantities, after [Jenn2005, p. 45]

Quantity	**General case**	**Geometrical model**
Time	$t' = t/q$	$t' = t/p$
Length	$L' = L/p$	$L' = L/p$
Wavelength	$\lambda' = \lambda/q$	$\lambda' = \lambda/p$
Frequency	$f' = qf$	$f' = pf$
Permittivity	$\varepsilon' = p\alpha\varepsilon/\beta q$	$\varepsilon' = \varepsilon$
Permeability	$\mu' = p\beta\mu/\alpha q$	$\mu' = \mu$
Conductivity	$\sigma' = p\alpha\sigma/\beta$	$\sigma' = p\sigma$
Current density	$J = pJ/\beta$	-
Power density	$W' = W/\alpha\beta$	-
Phase velocity	$\mu_p' = q\mu_p/p$	$\mu_p' = \mu_p$
Antenna gain	$G' = G$	-
Propagation constant	$\gamma' = p\gamma$	-
Impedance	$z = \beta\eta/\alpha$	-
Radar Cross Section	$\sigma' = \sigma/p^2$	$\sigma' = \sigma/p^2$

For the geometrical model, ε and μ are the same in both systems. Most of the other quantities are scaled by p and the RCS is scaled by p^2. This means that the same materials can be used for the model as were used in the full size object. The extremely important exception to this is the conductivity, which must be increased by the factor with which the length is decreased. This implies that the scaled model must be constructed from a material which is a better conductor than the full scale object. The conductivity of a model is often increased by polishing the surfaces.

2.1.6.4 Linearity

Maxwell's equations are linear as long as the medium is linear. This is the case for the majority of cases which are of interest from an NCTR point of view. The property of linearity implies that the effect of multiple sources can be calculated by means of the superposition of the effect of each source. The linearity also means that an object's interaction with an EM field can be described in terms of an impulse response in the time domain, or a frequency response in the frequency domain. The signal received by the radar is thus a convolution of the radar's transmit signal with the impulse response of the target. An example of the impulse response of a perfectly conducting sphere, of radius a is shown in the figure below. Due to the linearity of Maxwell's equations the frequency response of the sphere in Figure 2.1 and the impulse response of the sphere in Figure 2.2 below are equivalent descriptions of the interaction of the EM field with the sphere.

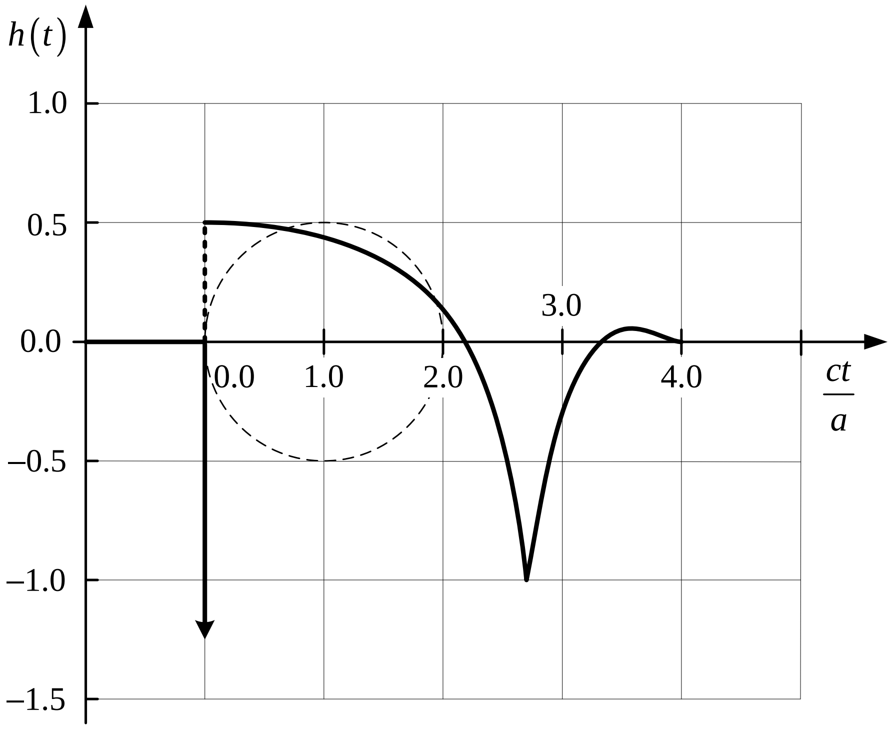


Figure 2.2: The "lesser spotted" impulse response of a conducting sphere of radius a, after [Astanin1997, p. 100].

2.2 Polarization

An EM field is fully described by its component electric (E) and magnetic (H) fields. The polarization of the field describes the locus of the E field vector at a point in space, as a function of time. If the vector oscillates along a straight line, the polarization is referred to as "linear", if the locus is an ellipse, the polarization is referred to as "elliptical" and if the locus is a circle the polarization is referred to as circular. For ground based radars it is usually convenient to refer to linear polarization as Vertical (V) or Horizontal (H), which indicates the polarization ellipse relative to the earth's surface. If the radar is circularly polarized, the polarization is classified as left-hand or right-hand, depending on the direction of spin of the E-field, seen in the direction of propagation. Any polarization can be described using two basis functions, so using V and H as basis functions any other form of polarization can be written as a linear combination of these two polarizations. This leads to the concept of a polarization matrix which is required to describe the arbitrary polarization which may be received from a target. The scattering matrix can be used to completely specify the relationship between the incident and scattered fields. Note that the scattering matrix is valid for any two orthogonal polarization types.

The incident-scattered field relationship is given by the following equation,

$$\begin{bmatrix} \mathbf{E}_1^{scat} \\ \mathbf{E}_2^{scat} \end{bmatrix} = \begin{bmatrix} S_{11} & S_{12} \\ S_{21} & S_{22} \end{bmatrix} \begin{bmatrix} \mathbf{E}_1^{inc} \\ \mathbf{E}_2^{inc} \end{bmatrix}, \tag{2.29}$$

where E_1 and E_2 are any two orthogonal polarizations, and "inc" refers to the incident field, and "scat" to the scattered field. The components of the scattering matrix, $\mathbf{S}$, are denoted by S_{ij} and are related to the RCS by the following expression:

$$S_{ij} = \frac{\sqrt{\sigma_{ij}}}{4\pi r^2}, \tag{2.30}$$

where $\sqrt{\sigma_{ij}}$ is a complex number and σ_{ij} is the RCS. The scattering matrix can thus be written as:

$$\begin{bmatrix} \mathbf{E}_1^{scat} \\ \mathbf{E}_2^{scat} \end{bmatrix} = \frac{1}{4\pi r^2} \begin{bmatrix} \sqrt{\sigma_{11}} & \sqrt{\sigma_{12}} \\ \sqrt{\sigma_{21}} & \sqrt{\sigma_{22}} \end{bmatrix} \begin{bmatrix} \mathbf{E}_1^{inc} \\ \mathbf{E}_2^{inc} \end{bmatrix}, \tag{2.31}$$

Which for the vertical and horizontal case gives:

$$\begin{bmatrix} \mathbf{E}_H^{scat} \\ \mathbf{E}_V^{scat} \end{bmatrix} = \frac{1}{4\pi r^2} \begin{bmatrix} \sqrt{\sigma_{HH}} & \sqrt{\sigma_{HV}} \\ \sqrt{\sigma_{VH}} & \sqrt{\sigma_{VV}} \end{bmatrix} \begin{bmatrix} \mathbf{E}_H^{inc} \\ \mathbf{E}_V^{inc} \end{bmatrix}, \tag{2.32}$$

Once the scattering matrix has been measured for a given set of polarization basis functions, it can be converted to any other polarization set without any loss of information. It is also interesting to note that for a single pulse and a fixed bandwidth, measurement of the polarization scattering matrix for each range bin on the target represents the maximum information which can be extracted by the radar [Cohen1991a].

2.3 Electromagnetic Theory and Approximations

Maxwell’s theory of electromagnetic waves addresses the prediction of a scattered field from a target. This theory provides very accurate predictions, but the mathematical equations that are associated with it can only be solved analytically for simple objects such as spheres and cylinders.

The Stratton-Chu equations were derived from Maxwell’s equations. The Stratton-Chu equations are integral equations, as the field variable to be solved occurs inside the argument of the integral. The Method of Moments (MoM) numerical technique can be applied to solve a discretized form of the problem at hand. The applicability of this technique is constrained by the available memory of the computer that is being used, as well as the electrical size of the object that is being analysed. The MoM technique becomes unstable as the size of the object increases beyond several hundreds of wavelengths of the incident field.

Due to the limitations of the MoM technique, approximations to the Stratton-Chu equations that are suited to the analysis of electrically large objects have been developed. These techniques are often termed high frequency approximations. The two most popular approximations are the Geometric Optics (GO) and Physical Optics (PO) approximations. Several extensions to each of these approximations have been developed in order to improve the accuracy with which the scattered fields may be predicted, but these extensions only apply to certain cases.

In some cases, useful results can be obtained from so-called “point scatterer” models. These models stem from the concept of a perfect radar target, which is a target that reradiates the energy incident on it isotropically. Insight into radar scattering problems can thus be obtained by approximating targets as a set of point scatterers on the skin of the target, or in positions which are not on the physical target but are positioned so as to simulate a specific effect (see [Rihaczek1996] for an in depth discussion of such models). Use should be made of two or more scatterers per wavelength. These models generally do not include shadowing or multiple reflections and are thus only used to obtain first order results. The target surface can also be approximated by a set of polarizable points. Each of the points can have a dipole moment in response to the local electric field, and the points can thus interact with each other by means of this field. The resulting set of equations usually has to be solved by numerical methods. These techniques are referred to as “coupled dipole” or “discrete dipole” approximations, and are more generally applicable than the point scatterer models.

2.4 Scattering Mechanisms

During the development of the theory for electromagnetic scattering and its various approximations a set of nomenclature has arisen. Some of this nomenclature is grounded in the actual theory of scattering and some in the approximations thereof. This situation has led to terms that describe mechanisms which were added to the approximations to correct for effects that were omitted from the initial formulation of the approximation. Figure 2.3 and Figure 2.4 below are diagrammatic representations of the scattering effects that contribute to the RCS of an object. These figures illustrate the diverse nature of the nomenclature related to scattering. The scattering mechanisms are briefly described in the following paragraphs, with reference to [Jenn2005] and [Knott2004].

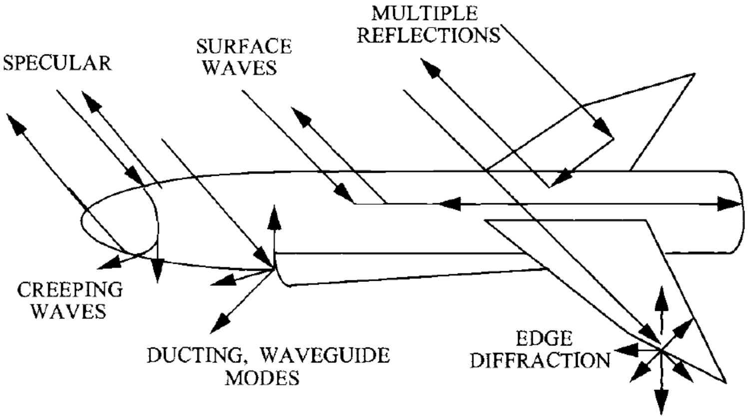


Figure 2.3: Scattering mechanisms from [Jenn2005], Fig. 1.25, p. 27. (© AIAA 2005)

Single and multiple reflections are usually the largest contributors to the RCS. Single or specular reflections occur when the local surface of the object is orthogonal to the direction of propagation of the impinging EM wave. Multiple reflections occur when the EM wave reflects off more than one surface in such a way that a large proportion of the incident energy is reflected back towards the radar (e.g. dihedrals and trihedrals).

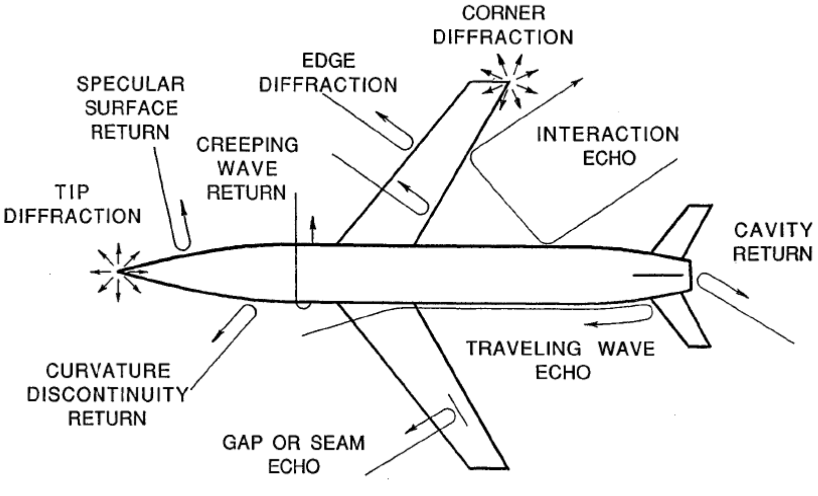


Figure 2.4: Scattering mechanisms from [Knott2004], Fig. 6.1, p. 227. (Reproduced with permission of the Institute of Engineering Technology.)

Diffraction refers to fields which are induced due to discontinuities in the target geometry, such as edges. The magnitude of this effect is usually smaller than that of specular reflections, but it has the property that the fields are produced over large angular regions.

The term travelling waves refers to currents which are induced on the surface of the target, but which travel along its surface. Currents that are induced in this manner can propagate around curved sections of the target and can therefore lead to fields that are produced from regions of the object which are not directly illuminated. Surface-induced currents lose energy as they radiate energy. As surface-induced currents reach a discontinuity in the surface, they are reflected and radiate energy.

Ducting occurs when incident EM fields are trapped in partially closed structures such as air inlets and tail pipes. The field is reflected multiple times within the cavity, and can therefore emerge over a wide angular region. Structures of this type are termed re-entrant structures.

For most targets of interest, these scattering effects may interact with one another in ways that are counter intuitive. Most of the scattering mechanisms also cause a change in the polarization of the EM field. Scattering effects and their interactions can thus be extremely complex, as is evident from this introductory discussion. As the designer of a target recognition algorithm would like to exploit these scattering effects, the techniques for analysing and measuring these effects need to be mature before the development of a target recognition algorithm can commence.

2.5 Radio Frequency (RF) Observables

From the preceding summary of some of the theoretical aspects of electromagnetic theory, this section contains a summary of which characteristics of a target can be observed using a radar. The target being interrogated by the radar has a physical state, and this state interacts with the impinging EM wave. This interaction is causes a reflected EM wave which is in turn sensed by the radar and thus influences which aspects of the target the radar can sense. The physical state of the target can be described by the following attributes:

1. Position.
2. Velocity.
3. Acceleration.
4. Orientation.
5. Visibility of rotating components.
6. Shape.
7. Material composition.

On a single pulse basis and for a single target orientation, the radar can inherently only measure the range to the target, and the polarisation response of the target as a function of range. If the radar forms a track on the target on a pulse-to-pulse basis, the dynamics of the target, namely velocity and acceleration, can be estimated. From these an educated guess at the orientation of the target can be made. For a target such as an airliner, this can be quite a good guess given the limited dynamics of the platform, but for a fighter, it could be climbing at a 45° angle, but might be at any roll angle relative to the radar. The single pulse measurement can be augmented by the addition of a measurement of the polarization matrix as a function of range. To gain further insight into the target type, the radar has to resort to processing multiple pulses received from the target. If use is made of a coherent radar system, the Doppler signature of the target can be sensed. This usually allows for the recognition of targets with rotating components such as helicopter rotors, propellers and jet engines. If the target trajectory exhibits a benign form of aspect change over a set of pulses, a two dimensional image of the target can be formed, in range and a dimension perpendicular to range, due to the different relative Doppler velocities of various components of the target. This technique is known as Inverse Synthetic Aperture Radar (ISAR) and can sometimes be effective in recognizing the outline of a target. The RF observables from the radar's point of view are thus summarized according to the following domains:

1. Target dynamics domain (Position, velocity and acceleration).
2. Magnitude domain.
3. Frequency domain.
4. Range / range resolution domain.
5. Polarization domain.
6. Doppler domain.
7. Image domain.

From this list it can be seen that some of the observables are more suited to deciding on the class of the target (e.g. target dynamics), whereas others, often in combination, are necessary for recognition (e.g. Polarization based imaging).

2.6 Conclusion

This section introduced some of the theoretical and practical aspects which are required to understand the interaction between the radar's transmitted pulse, in the form of an electromagnetic wave, and the target. A complete list of how aspects of this interaction can be exploited to classify or recognize targets using a radar system is given. To further understand the interaction between the radar's transmitted pulse and the target a brief discussion of the electromagnetic interaction of a plane wave with a target has shown this to be a very complex problem for which analytical solutions exist for only a few isolated cases.

3 Research Context

This section gives an overview of some of the more promising literature pertaining to radar based target recognition techniques. The focus of the literature search was the recognition of airborne targets, recognition performance prediction and fundamental limits on target recognition using a radar system. The structure of this section is based on the domains which the radar can sense as discussed in Section 2.5. Some images have been included from measurements, simulation and the literature to highlight the non-intuitive nature of the target return.

3.1 Literature Summaries and Books

There are four books which cover the theory and some practical aspects of target recognition using radar. P. Tait, [Tait2005], gives a high level introduction which covers most of the techniques which have been developed over the years. The two books by V.G. Nebabin [Nebabin1995] and Y.D. Sirman [Shirman2002], respectively, give interesting insights into the Russian approach to target recognition. Although both books are slightly dated they do contain a good summary of the approaches to NCTR. Shirman's book is more theoretical, especially as it is focussed on the EM simulation of targets. The book by Blacknell and Griffiths [BlacknellGriffiths2003] is more up to date and covers both the surface and air target cases.

Good overview papers have been published by Baum [Baum1994] and Cohen [Cohen1991a]. The paper by Baum addresses the system level in a radar system and is an excellent introductory paper to the whole topic of NCTR.

3.2 Target Dynamics Domain

The algorithms presented in this section focus on the concept of sharing information between the radar's tracking sub-system and the target identification sub-system. The tracking information can be used by the identification sub-system to reduce the database lookup required in aspect when matching target profiles. The target track can also be used to identify the target based on its flight envelope. If the target has been identified, or at least classified, the tracking sub-system can be optimized using the dynamics for this type of target.

The flight envelopes of a set of targets can also be used to gather information about the class of target. In some cases, e.g. if the target enters a portion of its flight envelope where it is the only target that can achieve this performance, recognition might also be possible. An example of a velocity-altitude plot for the envelopes of various aerial targets is given in the figure below. Dimensions such as turning ability could be added to this diagram.

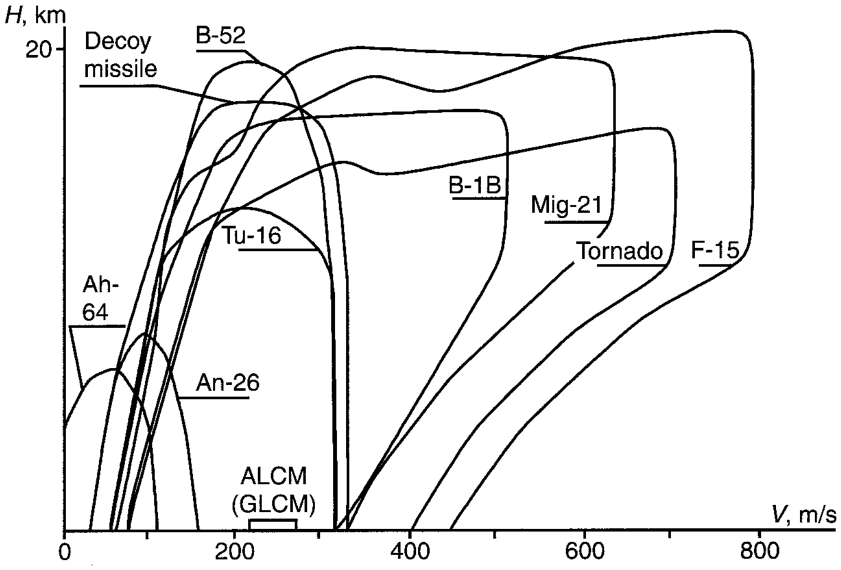


Figure 3.1: Example of a velocity-altitude envelope for various targets from [Shirman2002], Fig. 4.2, p. 133. (© Artech House)

The paper by S.P. Jacobs and J.A. O'Sullivan (who is a widely published researcher in the target identification field) addresses joint tracking and recognition for an S-band radar [Jacobs1997, Jacobs2000]. The paper also contains good measurements and simulations of the HRRP's of targets. The paper by Challa and Pullford [Challa2001] is interesting in that information gleaned from Electronic Surveillance Measures (ESM) sensors is combined with radar based measurements to increase the target recognition performance. This paper also compares two tracking schemes, the first being fusion of the tracking and recognition outputs, while the second scheme, called Joint Tracking and Classification (JTC), allows the tracking filter to aid the recognition process from a target dynamics point of view, and the recognition process can set the target dynamics model in the tracking filter. Use is made of a Bayesian approach to optimally combine the information sources. The numerical complexity of the system is high, and the authors suggest the use of particle filtering.

3.3 Long Term Statistics

The focus of this section is techniques which make use of longer term statistics of a radar target. Examples of these are RCS fluctuation and glint.

This concept is also prevalent in the field of passive radar where the aspect dependant variation of target magnitude is reduced due to the use of frequencies in the VHF band. This allows for recognition based on the magnitude history of the tracked target [Herman2002].

Borden [Borden1986] presents a method which makes use of the statistics of the target return over a narrow bandwidth and a restricted set of aspect angles for target identification. He also emphasises the use of the phase of the return signal by making use of the phase gradient. Some of the measured parameters relate to target size.

Another portion of the target dynamics includes the interaction of the target with the turbulence in the atmosphere. Shirman [Shirman2002, p. 40] presents a model for the simulation of these effects as well as some notes on the smoothing effects of pilots and auto-pilots. If the radar's tracking filters are accurate enough, it might be possible to derive information about the class of aircraft and possibly its loading, from its interaction with turbulence.

It would probably only be viable to estimate this kind of information from the longer time scale (on the order of 10's of seconds) of tracking data. This in turn would imply an extra time delay before a target's class is reported. A delay in the order of 10's of seconds will probably not be allowable in many military applications of radar.

3.4 Frequency Domain

A multi frequency technique is presented in [Ksienski1975]. In this technique, the radar transmits in the high Rayleigh and resonance regions of the target at a number of frequencies which are effectively phase locked. The authors show relatively good recognition performance for a set of four frequencies, when use is made of phase as well as two linear polarizations. The test target set consisted of a F104, F4, Mig19 and Mig21, which were tested for full azimuth rotation, and 7 discrete roll angles. It is shown that recognition based on magnitude only, is improved by the addition of phase information, and that this is further improved by the addition of polarization information. This experiment was repeated in [Lin1981] for two frequencies and eight targets. The performance achieved was 95% correct recognition for a noise level which was 20% of the signal level.

Chen [Chen1990] made use of a recognition technique based on the ratio of target magnitudes sensed at different frequencies and a Nearest Neighbour classifier. Only simulation results are presented, but the concept looks promising if extended to other RF observables such as polarization, especially if phase is included. A technique of this nature was investigated in [Stausberger1992] for HF Over the Horizon Radar (OTHR), which also included the effect of multipath propagation.

3.5 Range Domain

In 1965, Kennaugh and Moffat [Kennaugh1965] showed that the transient electric field reflected from an object carries information relating to the overall dimension, and approximate shape of the object. They proposed characterising an object's RCS as an impulse response, which probably led to the use of the range domain for the recognition of targets.

3.5.1 High Range Resolution Profiles (HRRP)

High Range Resolution (HRR) techniques make use of high bandwidth pulses to increase the range resolution to the point where a target is spanned by multiple range bins. This is possible by using just a single pulse and thus has the benefit of being insensitive to target motion. Another advantage is that the relative positions of the scattering centres are stable, and are thus less sensitive to carrier frequency and the specific waveform utilized. The heading of the target can be used to estimate the aspect at which the target is being observed. This reduces the search space during the recognition phase. A good overview of HRRP based recognition is given by Li and Yang, [Li1993], in which they show that increased resolution leads to better recognition performance. Zyweck and Brogner [Zyweck1996] present recognition results for two commercial airliners. The results are based on HRR data recorded by a radar in close proximity to an airport in Australia. A correct classification percentage of 93% is achieved using a relatively simple classifier. The authors also give several interesting images and HRR profiles and discuss several practical problems.

3.5.2 Transient Response

The transient response techniques are based on the assumption that the target is interrogated by the radar in a frequency band which corresponds to the resonant region of the target. The transient response of a target can then be divided into an early and late section. The early section describes the initial interaction of the pulse with the target up until steady state. The late section describes the target resonances once the driving (transmitted) pulse has passed the target. The so-called “Extinction pulse” or “E-pulse” techniques are based on the idea of generating a filter that cancels either the early or the late response of the target. Cancellation then indicates that the target has been recognized. The advantage of this technique is that it is claimed to be aspect independent, although no rigorous proof of this statement seems to exist (The introduction of [Ksienski1975] gives some insight though). Some promising laboratory results have however been published. The disadvantage of these techniques is that very wide bandwidths are required for the measurements in conjunction with the use of low frequency (HF to VHF) bands. Li et.al, [Li1998], discuss an interesting approach which uses the aspect independent late-time signal for initial classification and then the early-time response to search within the space generated by the initial estimate. Recently, some initial results have been presented by Aldhubaib and Shuley [Aldhubaib2009, Aldhubaib2010], on the use of natural resonance in conjunction with polarization for the recognition of simple objects. K-M Chen and D.P. Nyquist are prolific publishers in this area along with several collaborators and students. A good review paper is presented by Morgan [Morgan1988].

3.6 Polarization Domain

Polarization is an extra signal dimension which can be sensed by a radar and it has the potential to increase the detection and recognition performance of the radar. More theoretical and practical research with dual polarization radars is necessary before the benefits will be fully realised. Analysis of the polarization of a radar return can aid the identification of the various scattering mechanisms occurring on a target. Many of the full polarization radars are still in their developmental stages, except in the case of SAR systems.

For target recognition applications high resolution is necessary implying that polarization cannot be used as an upgrade path on low resolution radars to enable target recognition. The paper by Giuli, [Giuli1986], is an extremely useful overview of the applications of polarization in the radar field. Several papers describe the theoretical development of polarization dependant target RCS and some research on the use of polarization for target recognition. The paper by Steedly and Moses, [Steedly1991], contains some novel representations of data, although their measurements scale to the HF/low-VHF band.

Che and Yung describe a variable polarization system with very fast switching times. The system makes use of a ferrite phase shifter that can be operated at power levels of 30 kW [Che2000]. A high level description of the development and testing of a phased array MMW seeker radar is given in [Killen1989]. This radar is capable of producing HRR profiles with a resolution of 0.29 m as well as measuring the full polarization matrix. The design and calibration of the radar as well as several test procedures are discussed.

These papers show the continued interest in the addition of polarization into radar systems, specifically to improve the target recognition performance of a radar system.

3.7 Doppler Domain

A fixed target structure moving at a fixed speed relative to a radar will cause a frequency shift in the pulses received by the radar. Any structures on the target which are in motion relative to the main bulk of the target will show a Doppler shift relative to the main return. For airborne targets structures such as rotors, propellers and jet engines will thus add extra Doppler components to the received signal, which can be used to classify or identify the targets. This effect is also referred to as “micro-Doppler” (mD or µD) by some authors. Standard texts such as [Shirman2002] and [Tait2005] contain good summaries of techniques applicable to this domain. Some specific examples are the recognition of helicopters by means of the main rotor return discussed in [Cilliers2008, Cilliers2010] and measurements of micro-Doppler from propeller aircraft using a passive bistatic radar based on commercial FM transmissions [Maasdorp2013].

3.8 Imaging Domain

The techniques described in this section are aimed at generating images of the targets. In some publications these images are claimed to be of optical quality, but in general this is very difficult to achieve, given the complexities of the EM-target interaction. Most of these techniques require some form of mild rotation of the target to form an image, but have the advantage that if the target rotation is within the required bounds, the image resolution is independent of range. The first section, however, describes the imaging of targets using “real aperture” techniques which do not require target motion to form an image, but instead require large (hundreds of metres) antenna apertures.

3.8.1 Real Aperture Imaging

As opposed to synthetic aperture imaging, where target rotation is required to collect data from various target aspects, real aperture imaging makes use of large, usually sparse, arrays to form extremely narrow antenna beams and thus image a target. This has the advantage of being instantaneous in most cases, and not requiring target motion to form an image. The practical difficulties with these techniques include the physical setup of such a large array, clock and oscillator distribution, and calibration of the array.

Steinberg demonstrated this technique using an array consisting of 32 elements, with an antenna spacing of 1 m at X-band [Steinberg1988]. This setup gave a beam width of approximately 0.5 milliradian, which translates to a cross-range resolution of 1.5 m at a range of 3 km. The range resolution was approximately 1 m. Examples of two images created by the addition of five underlying images are shown in the figure below.

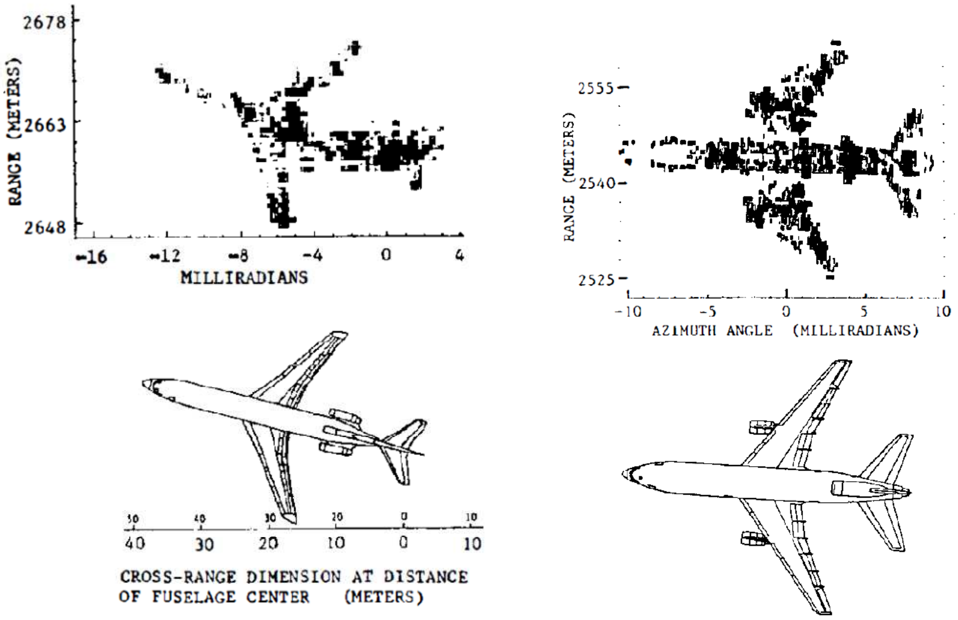


Figure 3.2: Example of real aperture images from [Steinberg1998]. Left Boeing 727, Right Lockheed L-1011. (© IEEE 1998)

The concept of real beam imaging has been extended by means of simulation studies to MIMO type radar configurations by Wang et.al. [Wang2010, Wang2010a]. This has the advantage of greatly reducing the amount of antennas and associated hardware required. It does however increase the processing power requirement. More recently, Zhu et. al have shown mathematically and in simulation, that 3D imaging is possible using MIMO-ISAR, and that this technique can greatly reduce the coherent processing interval required to form the ISAR image, which in turn reduces the effect of target motion on the image [Zhu2011]. They also claim that if the number of antennas is large enough, 3D imaging of an airborne target will be possible with a single pulse.

3.8.2 Inverse Synthetic Aperture Radar (ISAR)

The generation of an ISAR image relies on the relative rotational motion between the radar and the target. Due to the fact that this motion cannot be guaranteed by the radar, the quality of the ISAR image cannot be guaranteed. ISAR processing has the advantage that if the target motion is within the specification, the resulting image resolution is independent of the range of the target. ISAR images are generated by processing multiple HRR profiles of a target, which implies that the quality of the underlying HRR profiles must be high. The generation of an ISAR image also requires far more attention to the correction for target motion. ISAR images do however seem to offer the best data for NCTR as the geometry of the target is often apparent. A good discussion of target recognition based on ISAR images is given in [Kim2005].

ISAR techniques are also applied to turntable experiments and due to the controlled nature of these experiments, the results are usually very good. One of the disadvantages is that the geometry of the aircraft will not be exactly the same as its in-flight geometry due to the lack of aerodynamic loading.

The paper by Zyweck and Bogner contains some excellent imagery of a Mirage 3 generated using a turntable [Zyweck1994]. The dynamic range of the images is limited to 30 dB. In the figure below the left hand image shows the ISAR image for an aspect approximately 30° from head on. The right hand image is for the same aspect, but the engine intakes have been blocked with Radar Absorbing Material (RAM). The artefacts in the image labelled “#1” are caused by the engine inlet, and reflections of the engine compressor blades.

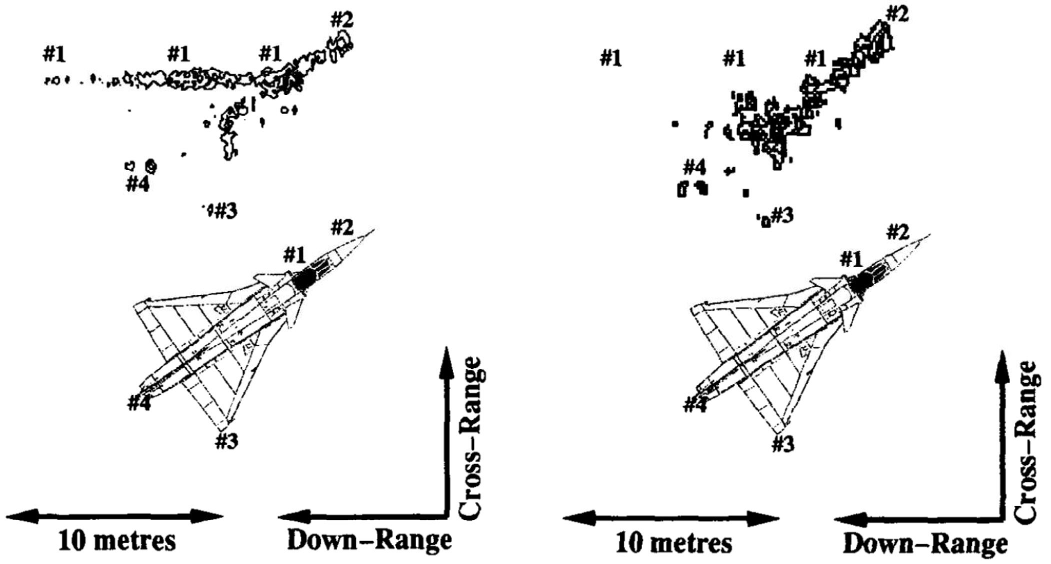


Figure 3.3: ISAR images of a Mirage III from [Zyweck1994]. In the right hand image the engine intakes have been plugged with RAM. (© IEEE 1994)

The authors also show an image of the plane with its engine being rotated electrically which gives a good idea of the type of artefacts which will exist for airborne targets.

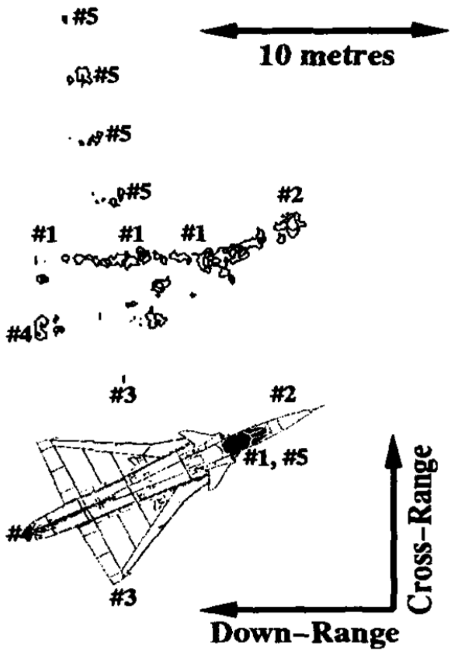


Figure 3.4: ISAR Image of a Mirage III with the engine being electrically rotated at normal operating speed from [Zyweck1994]. (© IEEE 1994)

Due to the relatively low PRF with which this data was measured, the engine modulation has been aliased in the Doppler frequency domain. The artefacts in the image labelled “#5” are caused by the rotation of the engine. Due to the fact that ISAR processing makes use of Doppler shifts to measure cross-range, these reflections appear as pixels which have been shifted in the cross range dimension, and do not fall within the physical bounds of the target.

Polarimetric ISAR incorporates the use of polarization in the ISAR process. Recently, Martorella et. Al. [Martorella2011] have published a model based Pol-ISAR recognition algorithm which achieves 100% recognition at a SNR of 5 dB, albeit for relatively simple targets in an anechoic chamber.

3.8.3 Pseudo 3D Techniques

3.8.3.1 Monopulse Imaging

The title of the solitary paper on this topic is “Enhanced range profiles for radar-based target classification using monopulse tracking statistics” by B.H. Borden [Borden 1995a]. The proposed technique makes use of the distribution of the monopulse angular measurements within a range “slice” of the target to form a 3D pseudo-image of the target. Only simulation results based on point scatterer target models are presented, and an example “image” is given below.

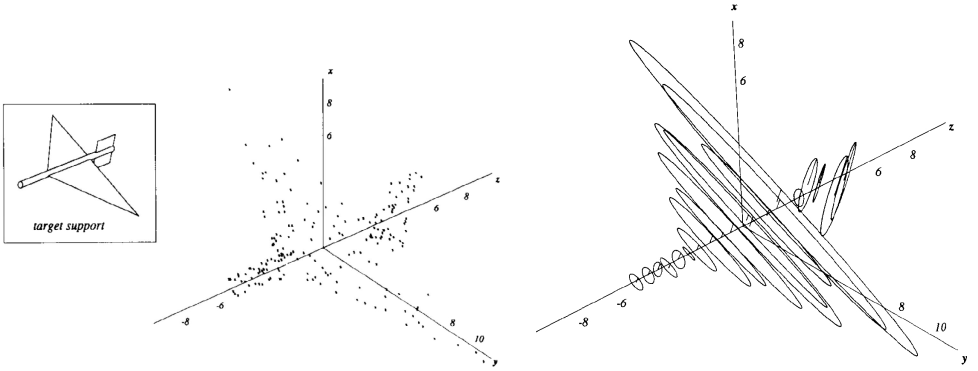


Figure 3.5: Physical target model, point scatterer model and reconstructed “image” for the Monopulse statistics technique from [Borden1995a]. (© IEEE 1995)

As this technique exploits the scintillation of the target, the target has to rotate approximately 5° to form the image and the data has to be collected over several hundred pulses (so this technique also falls into the “long term statistics” grouping). Unfortunately, the target has symmetry along its centre line which might increase the correlation between the physical target and the derived image.

3.8.3.2 Interferometric ISAR (IF-ISAR or InISAR)

This is a techniques which makes use of two receive antennas which are separated in elevation. This allows a single height and RCS to be estimated per cell in an ISAR image. A good overview of this technique is given by Xu and Narayanan [Xu2001], and an example of the image of a T-72 tank is given in the figure below.

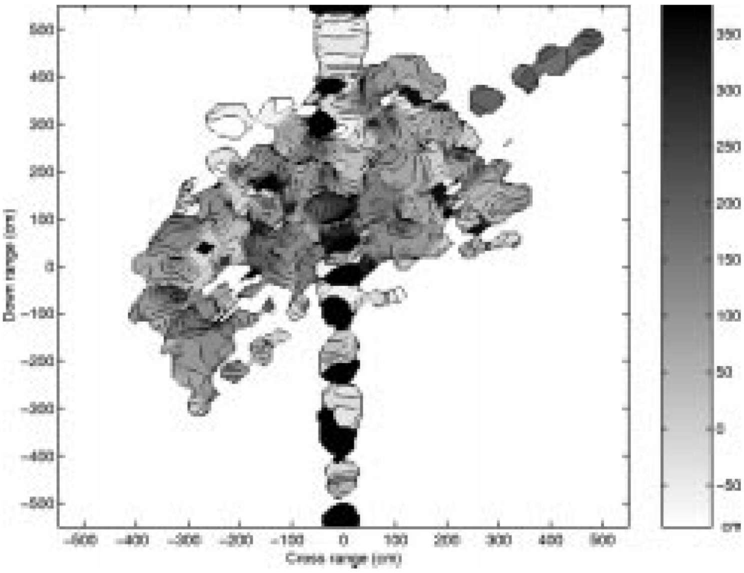


Figure 3.6: Interferometric ISAR image of a T-72 tank [from Xu2001]. The axis are in cm. The barrel of the tank (top right) is at 200 cm, which agrees with the physical height of the barrel. (© IEEE 2001)

Two problems can be experienced with this technique. The first is that phase unwrapping has to be implemented between two conventional ISAR images and the antenna spacing should thus be designed so that this is possible. The second is that if two scatterers exist in the same cell, a form of glint is present, which can invalidate the height calculation.

3.8.4 3D Inverse Synthetic Aperture Radar (3D-ISAR)

The techniques presented in this section are aimed at producing three-dimensional ISAR images. For this to be possible the target movement/geometry relative to the radar should exhibit elevation as well as azimuth movement.

The processing technique developed for ISAR was extended to 3-D ISAR by Seybold and Bishop [Seybold1996]. This entailed making measurements using a turntable with the radar antenna at various heights. Use had to be made of a calibration target included in the scene to align the phase of constituent 2-D ISAR images. The technique was demonstrated on a set of four corner reflectors set at different heights. An efficient implementation of a near field 3-D ISAR processor was published 2 years later by Fortuny, [Fortuny1998], who gave a very thorough discussion on the topic. A slightly different approach using UWB radar is given in [Kidera2010] where the surface of the object is reconstructed using the intersection of spheres.

A technique by Stuff [Stuff2002,Stuff2003] has also been published which is capable of extracting three dimensional scattering centres of a target. This technique is very promising, but hinges on being able to identify four stable scatterers in the received radar data and assuming that these are caused by a rigid body. If four stable scatterers can be identified, the algorithm only requires six high resolution radar pulses to solve for the relative positions of the scatterers in 3-D space. This result can then be used to estimate the trajectory of the rigid body relative to the radar and thus form a 3-D ISAR image. The technique does not uniquely determine the object orientation, and has the limitation that the radar's relative motion should not form an elliptical cone with its vertex on target. This means that it will fail for standard SAR or turntable ISAR type trajectories.

It should be noted that several conference papers appeared in 2009 and 2010 on various topics regarding 3-D radar imaging, thus confirming that this is currently a very active area of research.

3.9 Inversion of EM Problems

The preceding sections dealt with the analysis of EM scattering from an object under the assumption that its geometry and electrical properties are completely known. The problem of inferring the geometry of an object, and preferably its electrical properties, from scattered EM waves is referred to as the inverse problem. This type of problem is not unique to electromagnetism - several journals exist whose sole purpose is to report research in the field of so-called "inverse problems." In general, inverse problems result in a set of non-linear equations which are ill-posed. The reliable estimation of complex shapes in three dimensions still lies in the realm of impossibility.

A good review of inverse problems as applied to radar imaging is contained in the article by Borden [Borden2002]. The link between inverse problems and radar image formation by means of Inverse Synthetic Aperture Radar is discussed in detail and some restrictions to this technique, from a mathematical point of view are given. As examples of the difficulty involved with this type of problem, the following two figures are given. The first figure shows a single range profile return of a Boeing 727 airliner at an instantaneous bandwidth of 500 MHz at a carrier frequency of 9.5 GHz. The second figure shows an ISAR image of the same target (Note that the illumination in this case is from the bottom of the figure). It can be seen that this image shows scattering at points which are physically removed from the airframe structure. This is due to the re-entrant nature of the engine intakes. This highlights the need for better understanding of the EM effects as well as the implementation of such insight in the imaging algorithms. This also raises the question of whether it is necessary that a radar image reflects the physical shape of the object accurately, especially if use is made of automatic classification algorithms.

Some progress has also been made in the tomography based EM reconstruction of two dimensional objects [Angel2003, Fan2002, Huang2007]. This technique requires the object to be illuminated from one direction, and scattered fields to be recorded in fine angular steps around the target. The results given are also for relatively simple objects, which have a smooth exterior surface. As these techniques are still in their infancy, they are not yet applicable to radar type problems, but show promise if research is continued in this direction. It is however interesting to note that in 1969, R.M. Lewis published a paper with the title “Physical optics inverse diffraction” [Lewis1969] in which he showed that it was possible to reconstruct the illuminated portion of a target under the physical optics scattering approximation for a relatively limited set of observation angles and frequencies. Unfortunately, no evidence could be found that this technique has been tested on measured radar data.

Imaging of radar targets is by no means a new field, but the preceding discussion has shown that the models upon which such algorithms are based are often inadequate for target recognition purposes unless the recognition algorithm can take some of the imaging artefacts into account.

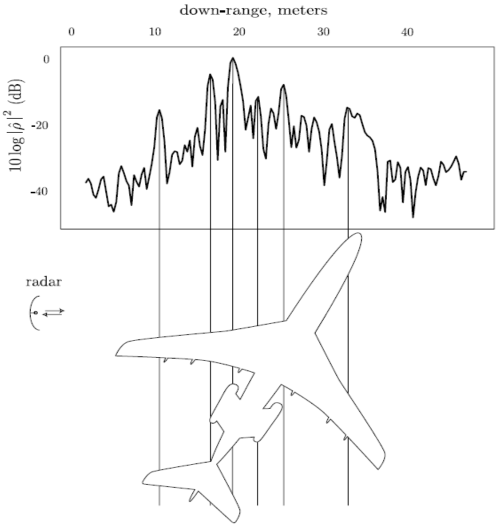


Figure 3.7: A single range profile of a Boeing 727 and an initial matching to structures on the airframe from [Borden2002]. The carrier frequency was 9.25 GHz and the instantaneous bandwidth was 500 MHz.

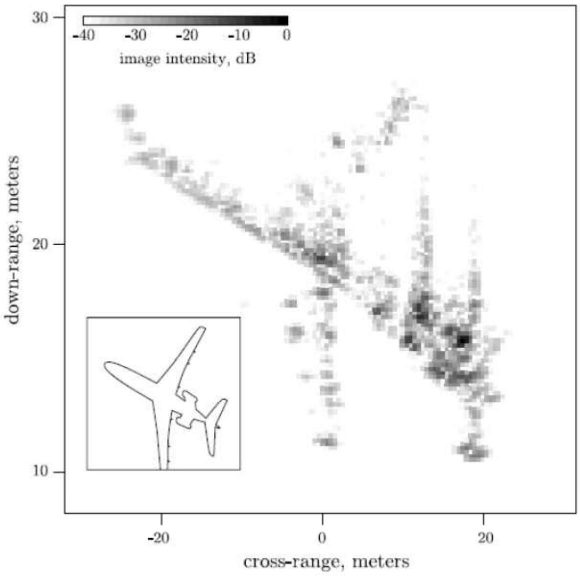


Figure 3.8 : ISAR image of a Boeing 727 from [Borden2002].

3.10 System Level Studies

This section contains descriptions of studies of the NCTR sub-system at a higher system level. These types of papers often hint at the reasons for specific design decisions.

Zwart et.al. published a paper on techniques for compensating for translational and rotational range migration of scatterers in HRR profiles [Zwart2003]. The authors make use of the RCS-prediction code called RAPPORT (developed by the TNO) to simulate HRR profiles for five large commercial aircraft. The paper addresses several practical problems and the authors show a final recognition error of 1.08% for the five targets.

An overview paper which focuses on FLIR based target recognition [Bhanu1986] contains several concepts which would be applicable to a radar based recognition system. A more philosophical overview paper, addressing ATR in the wider sense was published by Augustyn [Augustyn1992].

3.10.1 ATR for Ground Targets

Williams et.al. and Gross [Williams2000,Gross1999] describe a HRR based system for identifying ground targets in near real time. The [Williams2000] paper gives some very insightful results on the increase in classification performance as more target “looks” are obtained and target orientation is estimated. The algorithms are tested using the public domain Moving and Stationary Target Acquisition and Recognition (MSTAR) data set.

3.10.2 HRR vs. ISAR

The paper by Leslie Novak [Novak1991] shows that the recognition performance based on ISAR is better than that of HRR. Unfortunately the targets are ground vehicles, but the result is still a useful indication of what one could expect. It is interesting to note that use was made of polarimetric data to distinguish between odd-bounce and even-bounce returns.

3.10.3 Database Generation

For an NCTR algorithm to operate in a radar it is necessary to store information for all targets to be recognized from all aspect angles and then to match these to the observations made by the radar. It might also be necessary to have a central database which can generate recognition templates for different radars and algorithms from data acquired using measurement radar systems as well as data generated by means of simulation.

This section contains a discussion of literature which presents techniques which would be useful for creating signature databases for an NCTR system. These include simulation as well as measurement approaches.

Li et.al [Li1996a] show that as the range resolution increases, the angular matching region increases thus allowing for smaller databases. This study only made use of point scatterer models with some measurements of scaled models. The result however is counter-intuitive and warrants further investigation.

Two very good papers have been published by Bhalla, which deal with the generation of signature databases using SBR to directly generate a 3-D ISAR on which scattering centre extraction is performed using the CLEAN algorithm [Bhalla1996, Bhalla1997]. This allows for a 3-D representation of the target from which 1-D range profiles, or 2-D images can be constructed for recognition purposes. Unfortunately they do not compare actual recognition rates, only the accuracy of orientation estimation.

A paper by Zembower [Zembower1998] describes the establishment of a database in the USA to support target recognition and sensor fusion.

3.11 Conclusion

In this section some promising NCTR techniques from the open literature have been discussed. Some examples, of both simulated and measured data, are given to highlight the complexity of the interaction between the radar pulse and the target being observed. A discussion on the invertibility of Electromagnetic problems has shown that this leads to a set of non-linear, ill-conditioned equations which can only be solved under very limiting assumptions. It is thus not possible to directly invert the EM problem to achieve reliable target recognition. One remaining approach is to analyse the information content of radar signals which have been reflected from targets by applying concepts from Shannon’s information theory.

4 Information Theory

Information Theory was first published by Shannon in 1948 in his paper “A Mathematical Theory of Communication” [Shannon1948], which was later re-titled to “*The* Mathematical Theory of Communication”. The theory addresses two fundamental problems in communication theory, these being the maximum compression that can be applied to a source, and the maximum rate of data transmission over a given channel. The theory is general enough that it has been applied in fields as broad ranging as psychology, economics and micro-biology. It is thus conceivable that this theory can be applied to the radar target recognition problem to calculate the feasibility of target recognition and to compare the recognition performance of several vastly different radar designs.

4.1 The Meaning of “Information”

In natural language the meaning of the word “information” is related to the transfer of knowledge. This generally implies that it has meaning and value to the person receiving the information. Quantifying the amount or value of such information is very subjective as well as dependant on the receiver’s background information and current requirement for the information.

In information theory the definition of information is focussed on the underlying data used to represent the message rather than the meaning of the message. It thus ignores the actual message (possibly a subset of the knowledge being transferred) that is being transferred and rather characterises the representation of this data which has to be transmitted probabilistically. For example, in a coin flipping experiment, if a coin has a probability of heads appearing on a flip of 0.95, the observation of heads does not transfer nearly as much information to the observer as observing a tail. If a finite sequence of these flips had to be stored, one could store the flips directly as binary data, or one could store the positions at which tails occurred. It can be shown that the second option requires less data, thus showing that there is less information in the experiment than initially thought due to the trivial representation of the data. The maximum information for this experiment occurs when the probability of heads and tails are equal to ½. If this is the case then 1 bit of information is generated per flip, but if the probabilities are not equal, then, on average, less than one bit of information is generated per flip. This line of thought thus naturally leads to the concept of average information, and the fact that this quantity can take on a value of less than 1 bit.

Information theory makes use of the concept of entropy to describe the randomness of the output of such an experiment. For equally probable outcomes, the entropy is given by $-\log_2(p)$, where p is the probability of the outcomes. The entropy is thus a measure of the uncertainty of observing a specific event. In the above example, observing a tail is more unexpected, and thus has a higher self information (4.32 bits) than observing a head (0.07

bits). The entropy of the process is then the average value of the self information values, weighted by their respective probabilities, which gives an entropy of 0.29 bits, which is much lower than the highest possible entropy, which is 1 bit, for the equal probability case.

If the output of this experiment was to be observed in a noisy environment (i.e. there is noise added to the result in some undefined way, which has the effect of making a heads look like a tails and vice-versa in a probabilistic manner) the amount of information transferred can be expressed as the initial uncertainty about the outcome of the event minus the uncertainty after observing the event. This information gain is referred to as the mutual information between the input and output random variables which describe the random process, and the noise process. The rest of this chapter presents a formal development of these concepts, and some examples of the application thereof to radar problems.

4.2 Overview of the Application of Information Theory to Radar NCTR

The first application of information theory to radar seems to have been by Woodward [Woodward1953] in 1953. His analyses showed that the information increases over time as the radar observes a stationary target in stochastically stationary Gaussian noise. This information gain was shown to have a similar form to Shannon’s expression for channel capacity. Two researchers have more recently investigated the use of information theory for the analysis of radar NCTR performance. These are S.D. Briles [Briles1993] and M. Bell [Bell1988]. Both the authors show that it is possible to use the concepts of entropy and mutual information defined by Shannon to analyse radar recognition problems. The theoretical development by Bell was identified as being more understandable, and was used as the base for the theoretical introduction in the following sections. Bell also includes an initial discussion on techniques for the analysis of the information content of Synthetic Aperture Radar (SAR) images. Briles, on the other hand analyses the effect of increased bandwidth on the theoretical limit of recognition performance. This is however only conducted on very sparse measurements and for a target with a relatively simple geometry, but the trend is definitely that increasing bandwidth leads to increased recognition performance. Malas, Cortese and Ryan [Malas2015] combined Fano’s inequality with the data processing inequality to develop a technique for trading off the required fidelity amongst the various processing steps in a HRRP based radar target recognition system. Their analysis of a two target problem, using point scatterers, and generating six range bins shows that a four bit analogue to digital converter will suffice to maximize the information flow. Information theory has also been applied to optical imaging sensors by J.A. O’Sullivan [O'Sullivan1998] and some of this theory is applicable to the radar case. The following sections will give a theoretical description of information theory and how it is applied to radar systems.

4.3 Theoretical Introduction to Information Theory Applied to Radar

Information theory was originally conceptualized by Shannon for a generalized communication system consisting of a transmitter, a channel and a receiver where the main purpose of the system is to send data over the channel with as few bit errors as possible. To apply this concept to a radar system the retransmission of the radar's transmitted signal by the target is the equivalent of the transmitter in the communications system (i.e. the transmitted EM wave is modified by the target's physical characteristics and geometry and the scattered EM wave is then sensed by the radar). The propagation path back to the radar and the radar's front end form the channel, and the radar receiver forms the receiver. Importantly, the designer of the communications system can design the transmit waveforms, whereas the designer of the radar system has to try and exploit the target's reflected waveforms maximally. Once this mapping from the communication system to the radar system has been made, information theory can be applied to analyse the performance of the radar system, and in some cases to design parts of the radar system. An example of insight obtained by using this approach is given in [Bell1988] where it is shown that a radar system designed for optimal target detection differs significantly from one designed to recognize targets.

4.3.1 Expected Value of a Random Variable

The expected value of a random variable is used regularly in the theoretical sections of this document and is thus defined here. The expected value of the continuous random variable X is given by the following expression

$$E\{X\} = \int_{-\infty}^{\infty} x\, p(x)\, dx, \tag{4.1}$$

where $p(x)$ is the probability density function of the random variable. This is also the mean of the random variable. The expectation of the function $g(X)$ of a random variable is given by:

$$E\{g(X)\} = \int_{-\infty}^{\infty} g(x)\, p(x)\, dx. \tag{4.2}$$

For the case of a discrete random variable the integrals are replaced by summations.

4.3.2 Information Theory and Discrete RV's

This section describes the theoretical development of mutual information, which will be seen to be the critical concept in the information theoretical analysis of a radar system.

Let X be a discrete random variable (finite/countable) which takes on values from the set $R_X = \{x_1, x_2, \ldots\}$ and let $p(x) = P\{X = x\}$ be the probability density function (pdf) of X. To characterize the amount of information conveyed by an outcome of this random variable (RV) one would like to choose a measure which is inversely proportional to the probability of the occurrence of a specific event. This can also intuitively be seen as a measure of an observer's "surprise" at observing a specific event. If the event is very unlikely, the amount of "surprise" will be higher and the amount of information conveyed will thus be higher. Shannon defined the self information of an event as follows:

$$I(x) = -\log\left(p(x)\right) = \log\left(\frac{1}{p(x)}\right), \qquad p(x) \in [0,1] \tag{4.3}$$

where the base of the logarithm is 2, the units are bits and if the base is e, the units are nats [Proakis1989]. Bits will be used from this point onwards as the values are more intuitive. A comparison of the two measures is given in Figure 4.1, below.

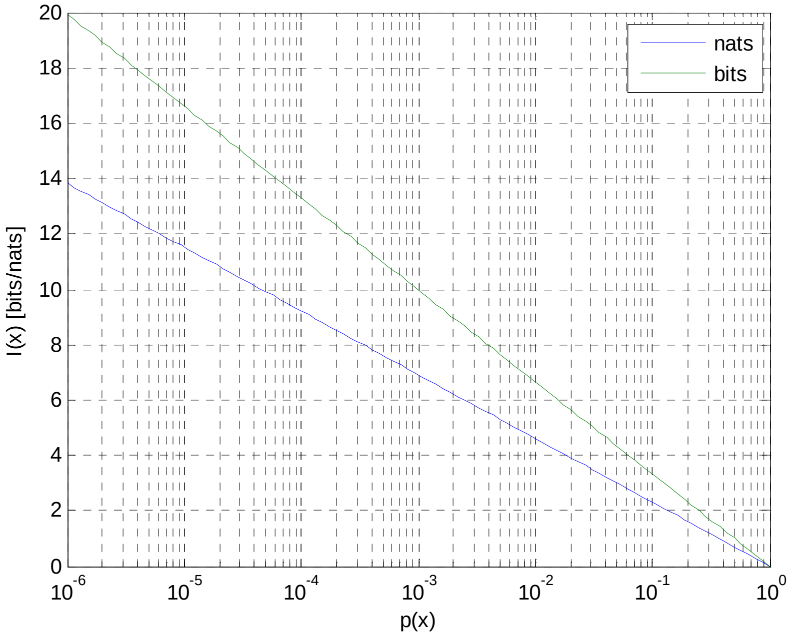


Figure 4.1: Self information as a function of probability of an event. (Note: asymptote for p(x) = 0)

It should be noted that this is not the only possible definition (e.g. Renyi information measures [Renyi1961]), but it is the one with the most useful analytical properties. The self information, $I(X)$, is a new random variable, as each original outcome, x, now has an associated $I(x)$. The entropy of X is defined as the expectation of $I(x)$:

$$\begin{aligned} H(X) &= E\{I(x)\} \\ &= -\sum_{x \in R_X} p(x)\log\left(p(x)\right), \end{aligned} \tag{4.4}$$

where $p\log(p)=0$ if $p=0$, which is derived from the following limit

$$\lim_{p \to 0^+} p\log(p) = 0. \tag{4.5}$$

The entropy is thus a measure of the average uncertainty of a random variable. The entropy of a RV is also the average code length, in bits, required to specify the outcome of a random variable.

This concept can be extended to the case for two (or more) random variables to characterize the entropy contained in two variables. If X and Y are jointly distributed discrete RV's from $R_X=\{x_1,x_2,\ldots\}$ and $R_Y=\{y_1,y_2,\ldots\}$ having a joint pdf $p(x,y)=P(X=x,Y=y)$, then the self information of joint occurrence of $X=x$ and $Y=y$ is given by

$$I(x,y) = -\log\left(p(x,y)\right) \tag{4.6}$$

The joint entropy of X and Y is then defined as the expected value the self information:

$$\begin{aligned} H(X,Y) &= E\{I(X,Y)\} \\ &= -\sum_{x \in R_X}\sum_{y \in R_Y} p(x,y)\log\left(p(x,y)\right). \end{aligned} \tag{4.7}$$

Some of the important properties of the entropy function are listed below:

1. Let $\mathbf{p}=\left[p(x_1),p(x_2),\ldots\right]$ be the pdf of X, then $H(X)$ is continuous in $\mathbf{p}$.
2. $H(X)\geq 0$. (Equality if all but one $p(x_j)$ equal zero.)
3. For a finite RV X with $R_X=\{x_1,x_2,\ldots,x_r\}$, $H(X)\leq\log(r)$. Equality if all $p(x_j)=1/r$. (Upper bound on $H(X)$)

4. If X and Y are jointly distributed RV's: $H(X,Y) \leq H(X) + H(Y)$, with equality only if X and Y are statistically independent.
5. $H(X)$ is a concave function of $\mathbf{p}$.

Property two implies that the entropy can only be zero if the RV consists only of one event, which thus has a probability of unity. Property three gives the maximum value which the entropy can obtain, and this is for the case when all outcomes of the experiment are equally likely. Property four implies that the joint entropy of two jointly distributed RV's is always less than the sum of the entropies of the constituent RV's, except in the case when the RV's are statistically independent, then the joint entropy is equal to the sum of the individual entropies. From this discussion it can be seen that the entropy leads to some very intuitive insights about the understanding of the "uncertainty measure" that was defined in (4.4).

The definition of entropy can also be extended to conditional probabilities. The pdf of a RV given that the outcome of another RV has been observed is defined as:

$$p(y \mid x) = \frac{p(x,y)}{p(x)}, \tag{4.8}$$

where the RV, Y, has been condition on a specific value of the RV X being observed, that is Y is conditioned on $X = x$. The entropy of Y conditioned on $X = x$ can now be defined as

$$H(Y \mid X = x) = -\sum_{y \in R_y} p(y \mid x) \log\left(p(y \mid x)\right). \tag{4.9}$$

The conditional entropy: $H(Y \mid X)$ is defined by averaging $H(Y \mid X = x)$ over $x \in R_X$ as follows:

$$\begin{aligned} H(Y \mid X) &= E\left\{H(Y \mid X = x)\right\} \\ &= \sum_{x \in R_X} p(x) H(Y \mid X = x) \\ &= \sum_{x \in R_X} p(x) \left[-\sum_{y \in R_y} p(y \mid x) \log\left(p(y \mid x)\right)\right] \\ &= -\sum_{x \in R_X} p(x) \sum_{y \in R_y} p(y \mid x) \log\left(p(y \mid x)\right) \\ &= -\sum_{x \in R_X} \sum_{y \in R_y} p(x,y) \log\left(p(y \mid x)\right) \end{aligned} \tag{4.10}$$

Similarly, the conditional entropy of X given that $Y = y$ is defined as:

$$H\left(X \mid Y\right) = -\sum_{x \in R_X} \sum_{y \in R_y} p\left(x, y\right) \log\left(p\left(x \mid y\right)\right) \tag{4.11}$$

Now that entropy, joint entropy and conditional entropy have been defined, the concept of Mutual Information (MI) can be defined as:

$$I\left(X;Y\right) = H\left(X\right) - H\left(X \mid Y\right). \tag{4.12}$$

Mutual information is thus a measure of the difference between the a priori uncertainty in X and the uncertainty in X after observing Y. Note that this can also be written as [Cover2006]

$$I\left(X;Y\right) = H\left(X\right) + H\left(Y\right) - H\left(X,Y\right), \tag{4.13}$$

and is thus a measure of the amount of information that Y provides about X. Mutual information is also symmetrical, which can be shown by the following manipulation:

$$\begin{aligned} I\left(X;Y\right) &= H\left(X\right) - H\left(X \mid Y\right) \\ &= H\left(Y\right) - H(Y \mid X) \\ &= I\left(Y;X\right). \end{aligned} \tag{4.14}$$

This symmetry property lead to the name “mutual” information. MI can be defined in terms of the joint and marginal distributions of two RV’s as follows:

$$\begin{aligned} I\left(X;Y\right) &= I\left(Y;X\right) \\ &= \sum_{x \in R_X} \sum_{y \in R_y} p\left(x, y\right) \log\left(\frac{p\left(x, y\right)}{p\left(x\right) p\left(y\right)}\right) \end{aligned} \tag{4.15}$$

From this definition, it can be seen that MI quantifies the distance between the joint distribution random variables, $p\left(x, y\right)$, and their joint distribution if they were independent, $p\left(x, y\right) = p\left(x\right) p\left(y\right)$. MI thus measures the amount of information that X and Y have in common, that is, how much knowledge of the one reduces uncertainty about the other.

Mutual information has the following properties:

1. $I\left(X;Y\right) \geq 0$. $I\left(X;Y\right) = 0$ if and only if X and Y are statistically independent.

2. $I(X;Y) = I(Y;X)$.

3. $I(X;Y)$ is a concave function of the input probabilities $p(x)$.

4. $I(X;Y)$ is a convex function of the conditional probabilities $p(y|x)$.

The first property can be interpreted as follows: If X and Y are statistically independent then neither variable carries any information regarding the other, and the MI is thus zero. If Ycarries any information concerning X, then the MI is larger than zero. It is impossible to achieve a negative MI for two discrete RV's. The maximum value of MI is thus achieved when the one RV carries all the information about the other. This means that $X = Y$, i.e. observing the one outcome completely specifies the other. In this case $p(x,y)$ only exists on the diagonal $y = x$. For each point on this diagonal, $p(x) = p(y)$ and $p(x,y) = p(x)$. This means that:

$$\begin{aligned} I(X;Y) &= \sum_{x\in R_X} \sum_{y\in R_y} p(x,y)\log\left(\frac{p(x,y)}{p(x)\,p(y)}\right) \\ &= \sum_{x\in R_x} p(x)\log\left(\frac{p(x)}{p(x)\,p(x)}\right) \\ &= \sum_{x\in R_x} p(x)\log\left(\frac{1}{p(x)}\right) \\ &= H(X) \end{aligned} \tag{4.16}$$

and the maximum value of MI is the entropy of either of the RVs, when they are equal to each other.

The expression given in (4.15) for the MI is the average value of the MI. Each term in the summation is a RV in its own right and also has the following properties: variance, moments of all orders and a moment generating function [Gallager1968, p. 18].

Having defined mutual information, the concept of channel capacity, C, can now be defined as the maximum value of $I(X;Y)$ over all probability distributions $p(x)$.

$$C = \max_{p(x)} \{ I(X;Y) \} \tag{4.17}$$

This gives a numerical value for the largest rate at which information can be transferred over the channel. With proper encoding, information can be sent across the channel at any rate less than C with an arbitrarily small probability of error. This embodies the fundamental limit which Shannon calculated in 1948, which has led to vastly improved digital communication rates. The expression in (4.17) is an expression for the channel capacity of a discrete channel, implying that there are a fixed number of input and output states, and a set of transition probabilities between these states. This simplification is useful in digital communication settings such as forward error correction, but for the radar case where the receiver is limited by additive white Gaussian noise the concept of channel capacity first has to be extended to continuous probability density functions. In the next section the analogous concepts to entropy and mutual information for continuous RVs will be developed, leading to the definition of channel capacity for this case.

The underlying concepts supporting the derivation of channel capacity are shown diagrammatically in Figure 4.2.

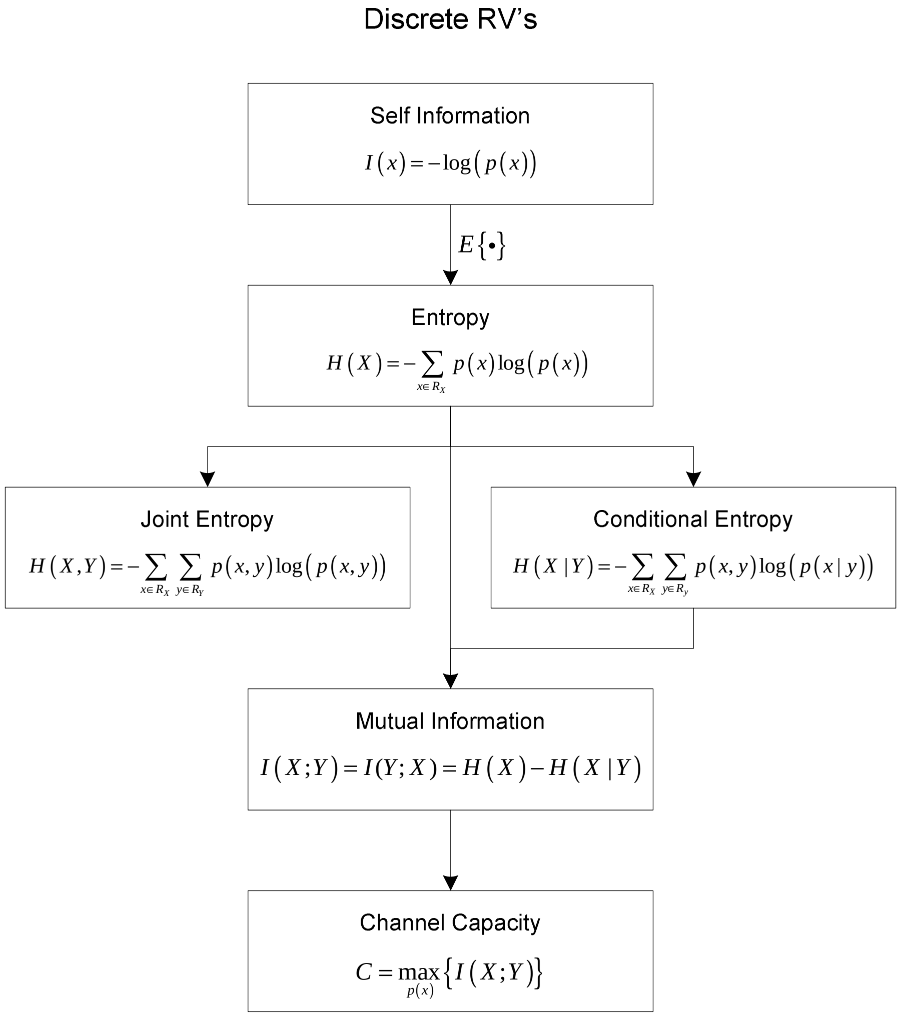


Figure 4.2: Diagrammatic representation of the concepts required to define channel capacity.

In the next section two examples of the application of information theory will be presented which highlight the interaction between the various information theoretic quantities which have been discussed up to this point.

4.3.3 Examples Demonstrating Correlation vs Information Theory

The following examples were developed to illustrate the application and interpretation of information theoretic principles discussed in this chapter. The examples were also chosen to highlight the superiority of using mutual information as an estimate of correlation over the traditional correlation coefficient. The correlation coefficient of two random variables is defined as:

$$\rho = \frac{\operatorname{cov}(X,Y)}{\sigma_X \sigma_Y}, \tag{4.18}$$

Where σ_X and σ_Y are the standard deviations of the random variables X and Y respectively and the covariance of the two random variables is defined as:

$$\operatorname{cov}(X,Y) = E\left\{ (X - \mu_X)(Y - \mu_Y) \right\}, \tag{4.19}$$

where μ_X and μ_Y are the means of the random variables. For both these examples the correlation coefficient is exactly zero, but the MI shows that there is correlation between the variables. The problem chosen is quite general, but in radar terms it can be seen as an analysis of the correlation between the in-phase (I) and quadrature phase (Q) channels.

4.3.3.1 Example 1

For the first example, the joint pdf of two random variables was constrained to lie on the unit circle and consist of 64 equiprobable impulse functions. The impulse functions were spaced equally in angle. The joint pdf as well as the two marginal pdfs are shown in Figure 4.3 below. Note that the marginal pdfs have been translated to make the figure more readable. The I channel has been mapped to the x-axis, and the Q channel to the y-axis.

The joint pdf can be written as

$$\begin{aligned} p(x,y) &= \sum_{k=1}^{N} P(X = x_k, Y = y_k)\,\delta(x - x_k, y - y_k) \\ &= \frac{1}{64}\sum_{k=1}^{64} \delta\big(x - \sin(\theta_k), y - \cos(\theta_k)\big) \end{aligned}, \tag{4.20}$$

where $\theta_k = 2\pi\left(k - \tfrac{1}{2}\right)/k,\ \ k \in \{1, 2, \ldots, 64\}$ and the marginal pdfs can be written as

$$\begin{aligned} p(x) &= \sum_{k=1}^{N} P(X = x_k)\,\delta(x - x_k) \\ &= \frac{1}{32}\sum_{k=1}^{32} \delta\left(x - \sin(\theta_k)\right) \end{aligned}, \tag{4.21}$$

and

$$\begin{aligned} p(y) &= \sum_{k=1}^{N} P(Y = y_k)\,\delta(y - y_k) \\ &= \frac{1}{32}\sum_{k=1}^{32} \delta\left(y - \cos(\theta_k)\right) \end{aligned}, \tag{4.22}$$

where the 1/32 is due to there always being two impulse functions which align during the calculation of the marginal pdf.

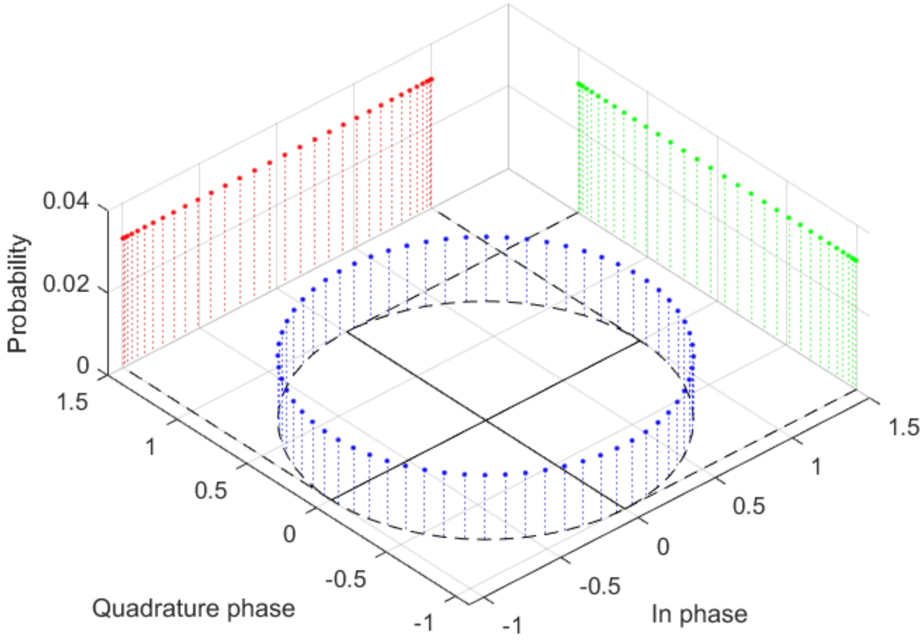


Figure 4.3: Joint and marginal densities for example 1. Blue indicates the joint pdf and red and green indicate the two marginal pdfs.

The covariance for this case can be derived as follows

$$\begin{aligned}\operatorname{cov}(X,Y) &= E\{(X-\mu_X)(Y-\mu_Y)\}\\ &= E\{XY\}\\ &= E\left\{\frac{1}{32^2}\sum_{k=1}^{32}\delta(x-\sin(\theta_k))\delta(y-\cos(\theta_k))\right\}\\ &= \frac{1}{32^2}E\{\sin(\theta_k)\cos(\theta_k)\}\\ &= \frac{1}{32^2}E\left\{\frac{1}{2}\sin(2\theta_k)\right\}\\ &= 0\end{aligned}, \tag{4.23}$$

For the information theory calculations, the MI between the two random variables will be calculated as well as several of the entropies, which will then be used to confirm the value calculated for the MI. First the MI is calculated as follows

$$\begin{aligned} I(X;Y) &= \sum_{x\in R_X}\sum_{y\in R_y} p(x,y)\log\left(\frac{p(x,y)}{p(x)\,p(y)}\right)\\ &= \sum_{j=1}^{32}\sum_{k=1}^{32}\begin{cases}\frac{1}{64}\log_2\left(\frac{1/64}{1/32\times 1/32}\right) & p(x_j,y_k)\neq 0\\ 0 & p(x_j,y_k)=0\end{cases}\\ &= 64\left(\tfrac{1}{64}\log_2(16)\right)\\ &= 4 \quad \text{bits}\\ &= \mathrm{I}(Y;X)\end{aligned} \tag{4.24}$$

where use is made of the fact that $p(x,y)$ is non-zero at 64 of the points contained in the $32\times 32=1024$ possible locations in the x-y plane. Next the entropies of I and Q are calculated:

$$\begin{aligned} H(X) &= -\sum_{x\in R_X} p(x)\log_2(p(x))\\ &= -\sum_{j=1}^{32}\frac{1}{32}\log_2\left(\frac{1}{32}\right)\\ &= 5 \quad \text{bits}\\ &= \mathrm{H}(Y)\end{aligned} \tag{4.25}$$

This result shows that the entropy of observing only the I or only the Q channel is 5 bits, or alternatively 5 bits are required to unambiguously identify the outcome of a single channel. This stems from the fact that there are 32 equiprobable coordinates to describe either the I or the Q channel. The MI between the I and Q channel is 4 bits, which means that observing the output of the I channel will also give 4 bits of information about the Q channel (and vice-versa). For this example the remaining uncertainty of 1 bit in the Q channel represents the uncertainty as to whether the Q channel signal lies on the upper or lower semi-circle. Once the I coordinate is fixed, there are only two options for the Q coordinate, and hence only a single bit is required to unambiguously identify a single Q coordinate. This is the definition of the conditional entropy of Q given that I has occurred, and is calculated from (4.12) as follows:

$$\begin{aligned} I(X;Y) &= H(Y) - H(Y \mid X) \\ H(Y \mid X) &= H(Y) - I(X;Y) \\ &= 5 - 4 \\ &= 1 \quad \text{bit} \end{aligned} . \qquad (4.26)$$

Note the conditional entropy of I given that Q has occurred has the same numerical value of 1 bit. To confirm the MI calculation, use is made of (4.13), which is repeated below.

$$I(X;Y) = H(X) + H(Y) - H(X,Y) , \qquad (4.27)$$

The first two entropy terms have been calculated already and all that remains is to calculate the joint entropy:

$$\begin{aligned} H(X,Y) &= -\sum_{x \in R_X} \sum_{y \in R_Y} p(x,y) \log\left(p(x,y)\right) \\ &= -\sum_{j=1}^{32} \sum_{k=1}^{32} \begin{cases} \frac{1}{64} \log_2\left(\frac{1}{64}\right) & p(x_j, y_k) \neq 0 \\ 0 & p(x_j, y_k) = 0 \end{cases} . \\ &= -64\left(\tfrac{1}{64} \log_2\left(\tfrac{1}{64}\right)\right) \\ &= 6 \quad \text{bits} \end{aligned} \qquad (4.28)$$

which means that 6 bits of information are required to unambiguously identify a single (I.Q) coordinate. This is due to the fact that there are 64 unique coordinates which are all equiprobable. Now, substituting the entropy results into the expression for the MI:

$$\begin{aligned} I(X;Y) &= H(X) + H(Y) - H(X,Y) \\ &= 5 + 5 - 6 \\ &= 4 \quad \text{bits} \end{aligned} , \qquad (4.29)$$

which agrees with the original calculation which was based on the definition of MI.

In summary, this example has shown that even though the correlation coefficient is exactly zero, there is still I large amount of information contained in the one random variable about the other random variable. Once the value of one of the random variable is known, the coordinate of the other random variable is known except for its sign. The unknown sign results in 1 bit of uncertainty.

4.3.3.2 Example 2

The second example is an extension of the first to a slightly more complex case. The joint pdf was chosen to lie on a Lissajous curve defined by:

$$\begin{aligned} &x = \sin\left(2\theta_k\right) \\ &y = \cos\left(\theta_k\right) \\ &\theta_k = 2\pi\left(k - \tfrac{1}{2}\right)/k, \quad k \in \{1, 2, \ldots, 64\} \end{aligned}, \tag{4.30}$$

This was chosen so that there are always exactly 2 positions on the joint pdf with the same y-coordinate and 4 positions with the same x-coordinate. This choice changes the marginal pdf for the x-coordinate to become:

$$p(x) = \frac{1}{16}\sum_{k=1}^{16} \delta\left(x - \sin\left(2\theta_k\right)\right). \tag{4.31}$$

The y-coordinate remains unchanged from the first example and is given by:

$$p(y) = \frac{1}{32}\sum_{k=1}^{32} \delta\left(y - \cos\left(\theta_k\right)\right). \tag{4.32}$$

The marginal pdfs and the joint pdf are shown in Figure 4.4 below.

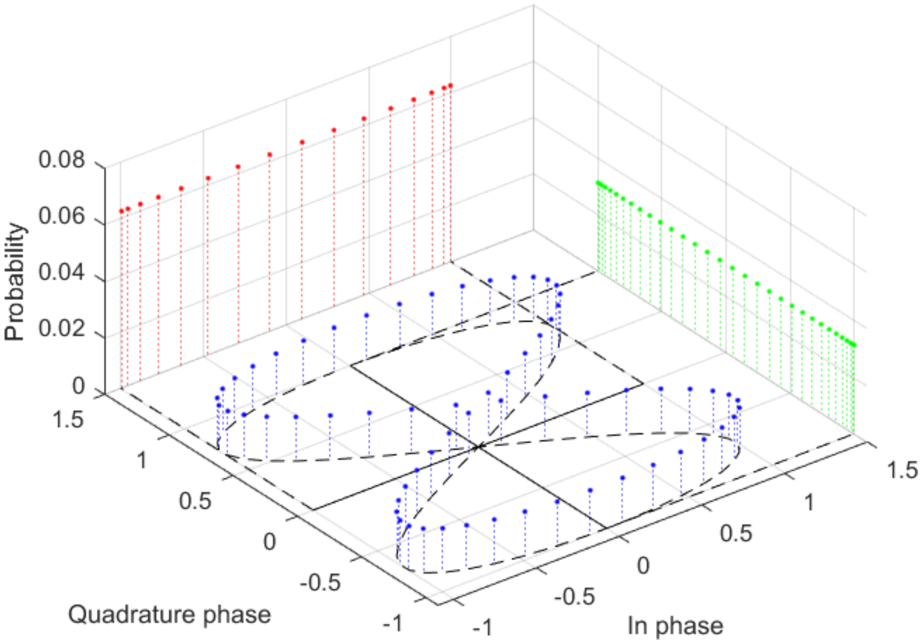


Figure 4.4: Joint and marginal distributions for Example 2. Blue indicates the joint pdf and red and green indicate the x and y marginal pdf respectively.

Once again, due to the symmetry of the pdfs, the covariance is exactly zero.

$$\begin{aligned}\operatorname{cov}(X,Y) &= E\{(X-\mu_X)(Y-\mu_Y)\} \\ &= E\{X\,Y\} \\ &= E\left\{\frac{1}{16\times 32}\sum_{k=1}^{32}\delta\left(x-\sin\left(2\theta_k\right)\right)\delta\left(y-\cos\left(\theta_k\right)\right)\right\} \\ &= \frac{1}{512}E\{\sin(2\theta_k)\cos(\theta_k)\} \\ &= \frac{1}{512}E\{2\sin(\theta_k)\cos^2(\theta_k)\} \\ &= 0\end{aligned}, \qquad (4.33)$$

The MI on the other hand can be calculated as follows:

$$\begin{aligned} I(X;Y) &= \sum_{x\in R_X}\sum_{y\in R_y} p(x,y)\log\left(\frac{p(x,y)}{p(x)\,p(y)}\right) \\ &= \sum_{j=1}^{16}\sum_{k=1}^{32}\begin{cases} \frac{1}{64}\log_2\left(\frac{1/64}{1/16\times 1/32}\right) & p(x_j,y_k)\neq 0 \\ 0 & p(x_j,y_k)=0 \end{cases} \\ &= 64\left(\tfrac{1}{64}\log_2(8)\right) \\ &= 3 \quad \text{bits} \\ &= \mathrm{I}(Y;X) \end{aligned} \tag{4.34}$$

The fact that the MI has fallen from 4 bits to 3 bits means that there is still correlation between the two random variables, but it has been reduced due to the extra uncertainty in the one dimension. The entropy of the I channel is derived as follows:

$$\begin{aligned} H(X) &= -\sum_{x\in R_X} p(x)\log_2(p(x)) \\ &= -\sum_{j=1}^{16}\frac{1}{16}\log_2\left(\frac{1}{16}\right) \\ &= 4 \quad \text{bits} \end{aligned} \tag{4.35}$$

which means that less bits are required to represent all 16 possibilities than in the first example. The entropy of the Q channel is once again given by

$$\begin{aligned} H(Y) &= -\sum_{x\in R_X} p(y)\log_2(p(y)) \\ &= -\sum_{j=1}^{32}\frac{1}{32}\log_2\left(\frac{1}{32}\right) \quad , \\ &= 5 \quad \text{bits} \end{aligned} \tag{4.36}$$

which is the same as the first example. Next the conditional entropies are calculated, but whereas these were equal in the first example they differ in this example. The conditional entropy of I given that Q has occurred is given by:

$$\begin{aligned} H(Y\,|\,X) &= H(Y) - I(X;Y) \\ &= 5-3 \quad . \\ &= 2 \quad \text{bits} \end{aligned} \tag{4.37}$$

This is due to there being 4 possible Q values for each possible I value, thus resulting in 2 bits of uncertainty remaining after I has been observed. The conditional entropy of Q given that I is known is calculated as follows:

$$\begin{aligned} H(X\,|\,Y) &= H(X) - I(X;Y) \\ &= 4-3 \\ &= 1 \quad \text{bit} \end{aligned} \tag{4.38}$$

This is due to the fact that once the Q channel has been observed, there are only two possible outcomes for the I channel, thus resulting in a single bit of uncertainty.

Once again, the MI calculation is confirmed by making use of the entropies and the joint entropy, which is calculated as follows:

$$\begin{aligned} H(X,Y) &= E\{I(X,Y)\} \\ &= -\sum_{x\in R_X}\sum_{y\in R_Y} p(x,y)\log\left(p(x,y)\right) \\ &= -\sum_{j=1}^{16}\sum_{k=1}^{32}\begin{cases} \frac{1}{64}\log_2\left(\frac{1}{64}\right) & p(x_j,y_k)\neq 0 \\ 0 & p(x_j,y_k)=0 \end{cases}. \\ &= -64\left(\tfrac{1}{64}\log_2\left(\tfrac{1}{64}\right)\right) \\ &= 6 \quad \text{bits} \end{aligned} \tag{4.39}$$

which means that 6 bits are still required to unambiguously represent 64 equiprobable coordinates in the I-Q plane. The MI is now calculated in terms of the entropies and joints entropies as follows:

$$\begin{aligned} I(X;Y) &= H(X) + H(Y) - H(X,Y) \\ &= 4+5-6 \qquad , \\ &= 3 \quad \text{bits} \end{aligned} \tag{4.40}$$

which agrees with the initial calculation of the MI for this example.

4.3.4 Information Theory and Continuous Random Variables

For a continuous RV X , on $\mathbb{R}$, with pdf $f(x)$, the differential entropy is defined as

$$\begin{aligned} h(X) &= -\int_{-\infty}^{\infty} f(x)\log\left(f(x)\right)dx \\ &= \lim_{\Delta x\to 0}\left\{-\sum_k f(x_k)\log\left(f(x_k)\right)\right\} \end{aligned} \tag{4.41}$$

Note that this is not the limit of the Rieman sum as that would give two terms, the second of which would be ∞. This term is ignored and only the first term is retained, hence the use of the term “differential”. For discrete RV’s, only the probability of each outcome determines the entropy. This is however not the case the entropy of a continuous RV. If a continuous RV is multiplied by a constant, a, then the differential entropy is given by:

$$h(aX) = h(X) + \log(|a|) \tag{4.42}$$

Differential entropy is thus a measure of the entropy of a continuous RV relative to the entropy of a uniformly distributed RV. From (4.42) it can be seen that if the distribution is more concentrated than a uniform distribution, a negative entropy can result.

In the same way that entropy was developed for discrete RV’s in the previous section, it can be extended using the definition above to continuous RV’s. The expressions for each case are summarized below.

Joint differential entropy is defined as

$$h(X,Y) = -\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(x,y)\log(f(x,y)) \tag{4.43}$$
,

conditional differential entropy is defined as,

$$h(X|Y) = -\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(x,y)\log(f(x|y)) \tag{4.44}$$

and mutual information is defined as

$$I(X;Y) = h(X) - h(X|Y) \tag{4.45}$$

It is important to note that the two ∞ terms which were discussed during the definition of differential entropy have cancelled each other out in the definition of MI for continuous variables. The mutual information is thus not called “differential mutual information”, but just “mutual information”.

It should be noted that all the concepts defined in this and the preceding section can be generalized to vectors of random variables. The vector version of mutual information can then be used to derive the channel capacity [Shannon1948]: Firstly, the maximum mutual information between the input and output of a Gaussian noise channel, of bandwidth W and an average power constraint of P_{av} is given by

$$\max_{p(x)} I(\mathbf{X}_N;\mathbf{Y}_N) = WT\log_2\left(1+\frac{P_{av}}{WN_0}\right) \tag{4.46}$$
,

where N_0 is the power spectral density of the noise. The channel capacity per unit time is obtained by dividing by T:

$$C = W \log_2\left(1+\frac{P_{av}}{WN_0}\right), \tag{4.47}$$

where the units of C are bits/s. The channel capacity is the fundamental limit on the rate of error free transmission for a channel system is power and bandwidth constrained and operates in the presence of AWGN. It should be observed that as the SNR tends to infinity, so does the channel capacity, but if the bandwidth tends to infinity, then the capacity tends to an asymptotic value of [Proakis1989]

$$C_\infty = \frac{P_{av}}{N_0}\log_2(e) = \frac{P_{av}}{N_0 \ln(2)}. \tag{4.48}$$

The expression for the channel capacity has a linear dependence on bandwidth, but a logarithmic dependence on signal to noise ratio. It is therefore easier to increase the channel's capacity by increasing the bandwidth than by increasing the transmitter's power. Shannon's coding theorem then states that as long as the transmission rate is lower than the channel capacity, communication is possible at an arbitrarily small probability of error.

Expression (4.46) gives the channel capacity for a continuous waveform channel, and is thus applicable to the radar case. In terms of signal propagation the radar channel and communication channel are similar in that both are limited by additive noise (mostly from the receiver components), and increasing the signal to noise ratio increases the channel capacity. The distinguishing difference is that the communications engineer has control over the design of the probability distribution, $p(x)$ as well as the waveforms, whereas in the radar case this is a function of the relative angles between the radar and the target, and is thus not a parameter which can be controlled by the radar engineer. The extra difference is that the propagation environment for a communication system can be free space, a wired connection or even a data storage device, whereas for the radar the propagation environment is nearly always free space. The possible exception to this is ground penetrating radar. The effect of these different propagation environments is the loss in signal strength, which is accounted for by the term P_{av} being a measure of the received power.

4.3.4.1 Example: Continuous Communication Channel

In this section a basic example is discussed to show the use of some of the concepts defined in this section. It is known that a Gaussian RV has the highest possible entropy for a RV, under constrained power. If such a RV is to be communicated over a noisy channel, what would the performance of such a system be? The problem is depicted in the figure below.

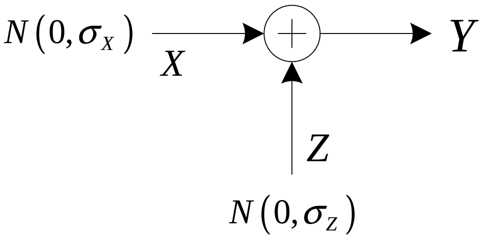


Figure 4.5: Example of transmission of a Gaussian RV over a channel corrupted by Gaussian noise.

For a Gaussian, zero mean RV, S, with variance σ^2, the entropy is given by:

$$h(S)=\frac{1}{2}\log\left(2\pi e\sigma^2\right). \tag{4.49}$$

The entropy of the output of the channel is given by:

$$h(Y)=\frac{1}{2}\log\left(2\pi e\left(\sigma_X^2+\sigma_Z^2\right)\right), \tag{4.50}$$

and it can be shown analytically that $h(Y\,|\,X)=h(Z)$. This is interpreted as the entropy of the output, given full knowledge of the input, is the entropy of the noise, which is given by:

$$h(Y\,|\,X)=\frac{1}{2}\log\left(2\pi e\sigma_Z^2\right). \tag{4.51}$$

The mutual information between input and output is given by:

$$\begin{aligned} I(X;Y)&=h(Y)-h(Y\,|\,X) \\ &=\frac{1}{2}\log\left(1+\frac{\sigma_X^2}{\sigma_Z^2}\right), \end{aligned} \tag{4.52}$$

which is also the channel capacity for this case, and the variance ratio is equivalent to the signal to noise ratio. The channel capacity is attained by any Gaussian RV, regardless of its mean, with variance σ_X^2. This example has been simulated for input distributions of both Gaussian as well as uniform RV's, and the results are shown in the figure below.

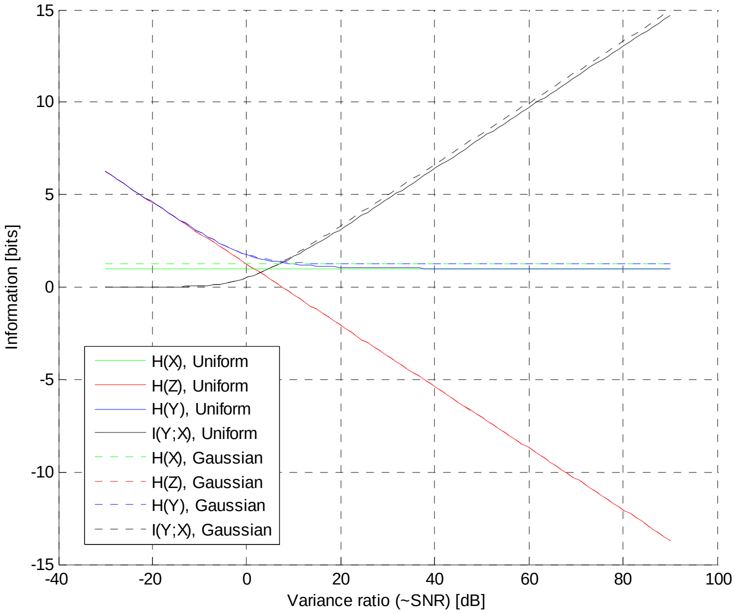


Figure 4.6 : Entropy and mutual information comparison for uniform and Gaussian source.

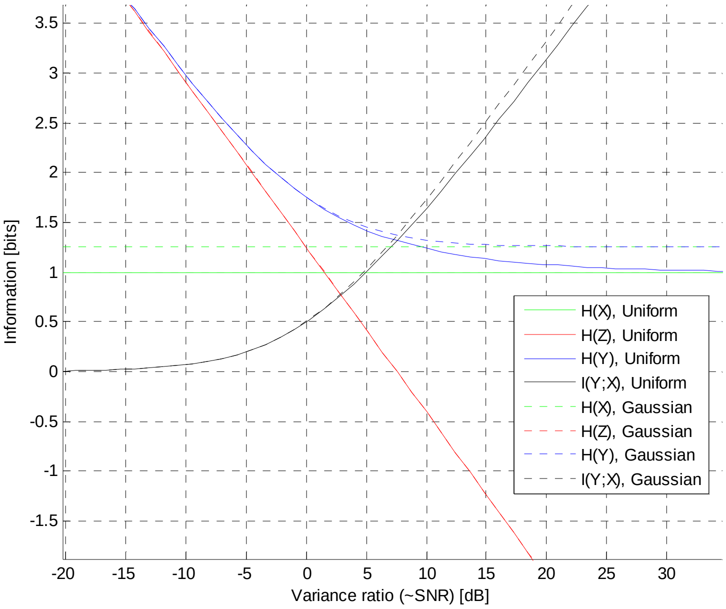


Figure 4.7: Entropy and mutual information comparison for uniform and Gaussian source: zoom of transition regions.

In these examples the Gaussian input RV case always has a higher entropy as well as mutual information than the uniform RV case. This is due to the fact that the Gaussian distribution has the highest entropy for a given RV when the variance of the RV is fixed.

4.3.5 Mutual Information for the Discrete Input Continuous Output Channel

In many communication examples, the input to the channel over which data is to be transmitted is chosen from a set of discrete symbols. The channel adds noise to this input where the pdf of the noise is a continuous RV. Let X be the discrete input random variable, which has N possible outcomes denoted by x_i, and Y is described by its marginal pdf $p(y)$ If X and Y are statistically dependent, then $p(y)$ can be expressed as

$$p(y) = \sum_{i=1}^{N} p(y|x_i)\, p(x_i) \tag{4.53}$$

The information provided about the input event $X = x_i$ by observing the output event $Y = y$ is given by

$$\begin{aligned} I(x_i, y) &= \log_2\left(\frac{p(y|x_i)\, p(x_i)}{p(y)\, p(x_i)}\right) \\ &= \log_2\left(\frac{p(y|x_i)}{p(y)}\right) \end{aligned} \tag{4.54}$$

which is sometimes referred to as “point information”. The average mutual information between X and $Y$is given by [Proakis1989, p.75] and [Gallager1968, p.33]

$$I(X,Y) = \sum_{i=1}^{N} \int_{-\infty}^{\infty} P(x_i)\, p(y|x_i) \log_2\left(\frac{p(y|x_i)}{p(y)}\right) dy \tag{4.55}$$

4.3.6 Mutual Information and Radar Measurement Performance

This section describes the link between mutual information, and the measurement performance of a radar system. In this case “measurement” refers to the measurement of a specific parameter, such as the range of the target or the polarization matrix of the target. The conceptual system is shown in the figure below.

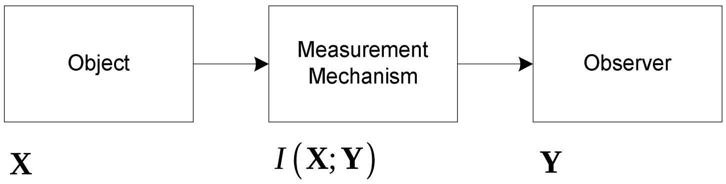


Figure 4.8: Conceptual radar measurement process.

In this figure, $\mathbf{X}$ and $\mathbf{Y}$ are random vectors, and have a joint distribution. The measurement process maps $\mathbf{X}$ into $\mathbf{Y}$ stochastically. The greater the mutual information between these two variables, the more information is obtained about the object.

Maximum number of partitions, N , of R_X is given by in [Bell1988, p. 47, Eq. 2.36]

$$N = \left\lfloor 2^{I(\mathbf{X};\mathbf{Y})} \right\rfloor, \tag{4.56}$$

for an arbitrarily small probability of error. This is based on Shannon's theorem for the noisy channel.

Rate distortion theory can now be used to relate measurement error to mutual information. The distortion (or error) is denoted as $d(\mathbf{x},\mathbf{y})$ where $\mathbf{x}$ is the parameter vector and $\mathbf{y}$ is the measurement vector. This is a non-negative function defined for all pairs of $\mathbf{x} \in R_X$ and $\mathbf{y} \in R_Y$. The average distortion, or equivalently the fidelity criterion is given by

$$D = E\left\{d(\mathbf{x},\mathbf{y})\right\} \tag{4.57}$$

The rate distortion function is defined as

$$R(\delta) = \min\left\{I(\mathbf{X},\mathbf{Y}) : D \le \delta\right\}. \tag{4.58}$$

This represents a minimization over all mechanisms which satisfy $D \le \delta$. The rate distortion function gives the minimum possible rate at which information must be transferred by a measurement mechanism to have an average distortion D less than or equal to δ. A smaller δ thus requires a larger value of $R(\delta)$. $R(\delta)$ is a non-increasing function of δ.

4.3.6.1 Example: Rate distortion function for a Gaussian noise channel

Using the same setup as the previous example (see Figure 4.5), a designer would like to measure the scalar parameter X. Again, assume that X and Z are statistically independent. Use can be made of the mean square error as the distortion measure, which is given as:

$$d(x,y) = (x-y)^2 \quad (4.59)$$

then

$$\begin{aligned} D &= E\left\{(x-y)^2\right\} \\ &= E\left\{(x-(x+z))^2\right\} \\ &= E\left\{(-z)^2\right\} \\ &= E\left\{Z^2\right\} \\ &= \sigma_Z^2 \end{aligned} \quad (4.60)$$

The mutual information between X and Y is given by:

$$I(X;Y) = \frac{1}{2}\log\left(1+\frac{\sigma_X^2}{\sigma_Z^2}\right) \quad (4.61)$$

The resulting rate distortion function is given by

$$R(\delta) = \frac{1}{2}\log\left(\frac{\sigma_X^2}{\delta}\right) \quad (4.62)$$

Therefore, if more precision is required in the measurement, a larger minimum rate of information transfer by the measurement mechanism is required. An increase in the mutual information thus gives an increase in the number of separable classes, as well as an increase in the precision of the measurements.

4.4 Data Processing Inequality

The data processing inequality is a fundamental limit on what can be achieved by processing raw data. In summary, it states that no algorithm can improve the inferences that can be made from a set of data, and this in turn implies that once information has been destroyed it cannot be recovered.

Use is made of a Markov chain to prove the theorem. A Markov chain is a mathematical model for a process in which knowledge of the current state of a system is all that is needed to be able to predict the behaviour of the system into the future. This means that the future of the system is independent of the past.

The random variables X, Y and Z form a Markov chain, denoted by $X \to Y \to Z$, if Z depends only on Y and Z is independent of X. The joint pdf of the variables can then be written as:

$$p(x, y, z) = p(x)\, p(y|x)\, p(z|y) \quad (4.63)$$

The data processing inequality theorem then states that:

If $X \to Y \to Z$, then $I(X;Y) \geq I(X;Z)$ (4.64)

This can be interpreted as seeing X as being the input to a channel, Y as the output of the channel and Z as the result of some processing that has been done on the data Y (i.e. $Z = f(Y)$) in an attempt to estimate X. The theorem then implies that there is no function, $f(\bullet)$, or algorithm that can increase the amount of information that Y conveys about X. In fact, the chances are good that information will be destroyed due to the processing and in the best case the information will remain constant.

4.5 Relationship Between Probability of Error and Mutual Information

Mutual information can be used to gain insight into classification problems by analysing the probability of error and the number of partitions into which the input random variable's space can be divided.

The maximum number of partitions N, of R_X which can be identified by observation of the output of the channel is given by

$$N = \left\lfloor 2^{I(X,Y)} \right\rfloor. \quad (4.65)$$

This expression has been applied to the example of the transmission of a Gaussian RV over a channel corrupted by additive white Gaussian noise (AWGN) (See section 4.3.4.1). The number of partitions into which the output space can be classified has been calculated using (4.65) as a function of the SNR, and the result is plotted in the figure below.

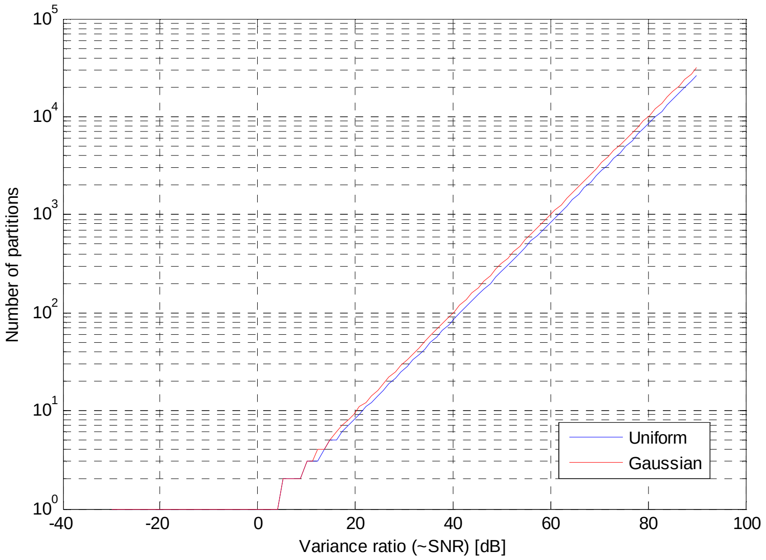


Figure 4.9: Number of partitions for arbitrarily small probability of error.

If the number of partitions calculated using (4.65) is less than the number of possible outcomes of the input discrete RV, then errors in the decoding of the received signal will be made. For this example the input space is continuous, so an infinite number of partitions are required to represent it, thus requiring the mutual information to be infinite if the source is to be perfectly reconstructed.

Fano's inequality relates the probability of incorrectly estimating the value of the input random variable, X, given an observation of the output random variable, Y, to the conditional entropy $H\left(X \mid Y\right)$. The probability of error is defined as

$$P_e = P\left(\hat{X}\left(y\right) \neq X\right), \tag{4.66}$$

where $\hat{X}$ is the estimate of the input RV made at the output of the channel. The strictest version of Fano's inequality is given by [Cover2006, p. 39]

$$H\left(P_e\right) + P_e \log_2\left(N-1\right) \geq H\left(X \mid Y\right), \tag{4.67}$$

where $H\left(P_e\right)$ is the entropy of the RV associated with the error event. The inequality is a lower bound on the probability of error for an N-class multiple hypothesis testing problem. The inequality is defined for an input space, R_X, which is a discrete random variable. This inequality was originally derived as a lower bound, but has been extended to a lower and upper bound by making use of Renyi's entropy [Erdogmus2004]. Fano's inequality can be related to mutual information by expanding the conditional entropy:

$$\begin{aligned} H(P_e)+P_e\log_2(N-1) &\geq H(X\,|\,Y) \\ H(P_e)+P_e\log_2(N-1) &\geq H(X)-I(X;Y). \end{aligned} \tag{4.68}$$

This equation cannot be solved algebraically for P_e, so use has to be made of numerical techniques at this point. The expression in (4.68) can in turn be solved for the mutual information

$$\begin{aligned} I(X;Y) &\leq H(X)-H(P_e)-P_e\log_2(N-1) \\ &= f(P_e), \end{aligned} \tag{4.69}$$

And the function $f(P_e,N)$ can be inverted to give a curve of the form

$$P_e \leq f^{-1}\left(I(X;Y)\right), \tag{4.70}$$

which can then be used to relate mutual information to the probability of error. The inverse function is plotted in Figure 4.10 for several values of N. The function is well behaved except at the point where the MI approaches 0 where the graph has an inflection point. For this study values of P_e which are close to zero are of interest, and this maps to the high MI region of graph, thus avoiding the inflection problem. This graph confirms the very direct relationship between mutual information and the probability of error, and shows that high values of MI can be used to imply low probabilities of error.

A weaker form of Fano’s inequality is given by

$$\begin{aligned} P_e &\geq \frac{H(X)-I(X;Y)-1}{\log_2(N)} \\ &= \frac{\log_2(N)-I(X;Y)-1}{\log_2(N)} \end{aligned} \tag{4.71}$$

where P_e has been solved for explicitly by making the assumption that $H(P_e)=1$ and that N is large enough that $\log_2(N+1)\cong\log_2(N)$. The line for this equation was added to Figure 4.10 for the $N=1024$ case. It can be seen that the approximation is accurate in the $P_e=0.5$ region, but looses accuracy at high and low values of MI. The stricter bound will thus be used in the remainder of this study. Mutual information can thus be used by a classification system designer to estimate the recognition performance for a specific set of targets in the presence of noise.

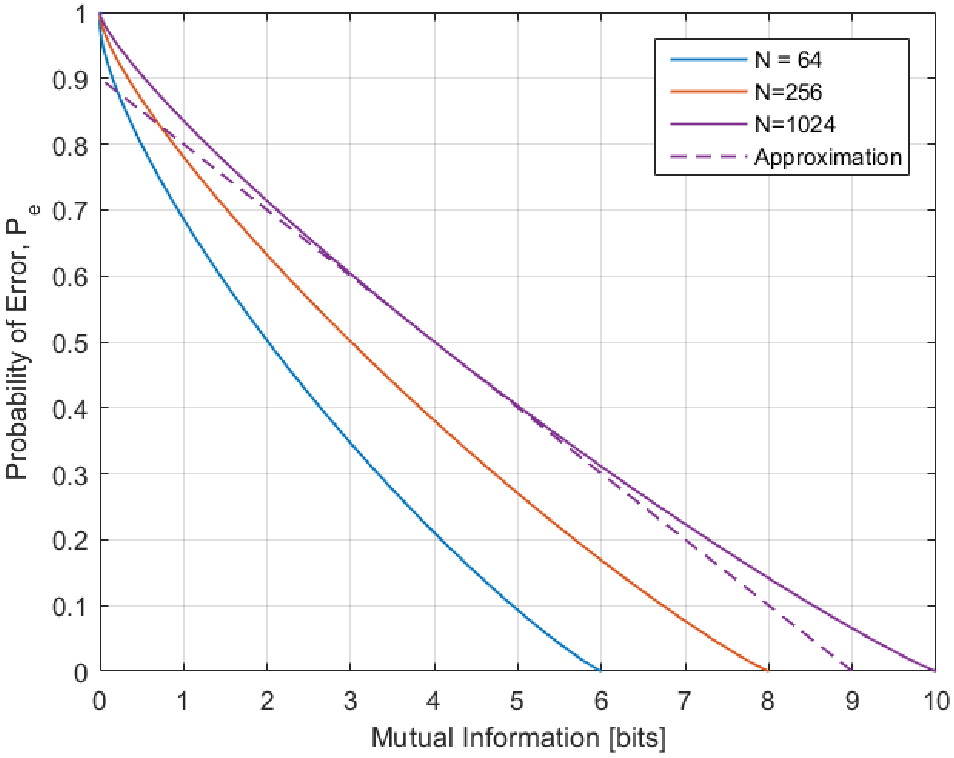


Figure 4.10: Examples of inverse mapping functions for the calculation of probability of error from mutual information.

4.6 Two Applications of Information Theory to the Target Recognition Problem

This section gives two examples from the literature on the use of information theory to analyse the recognition performance of a radar system.

4.6.1 Information Measure of Recognition Quality (IMRQ)

This example is based on a discussion in [Shirman2002], pp 58. Assume an "alphabet" of K targets each with probability of occurrence P_k, the entropy before recognition is given by

$$H_{BR} = -\sum_{k=1}^{K} P_k \log_2\left(P_k\right) \tag{4.72}$$

Let $P_{ik} = P_{i|k} P_k$ be the probability of having made a decision in favour of target i when target k is actually present. The entropy after a decision has been made is given by:

$$H_{AR} = -\sum_{i=1}^{K}\sum_{k=1}^{K} P_{ik} \log_2\left(P_{i|k}\right) \tag{4.73}$$

The change in entropy from before to after recognition is named the "information measure of recognition quality" (IMRQ) and is given by:

$$\begin{aligned} I &= H_{BR} - H_{AR} \\ &= \log_2(K) + \frac{1}{K}\sum_{i=1}^{K}\sum_{k=1}^{K} P_{i|k}\log_2\left(P_{i|k}\right) \end{aligned} \tag{4.74}$$

If the probabilities of correct and incorrect recognition are constant for all combinations:

$$P_{i|i} = P_{correct} \tag{4.75}$$

and

$$P_{i|k} = P_{incorrect} = \frac{1}{K-1}\left(1 - P_{correct}\right) \tag{4.76}$$

then the coarse IMRQ can be defined as:

$$I_{coarse} = \log_2(K) + P_{correct}\log_2\left(P_{correct}\right) + \left(1 - P_{correct}\right)\log_2\frac{1 - P_{correct}}{K-1} \tag{4.77}$$

This equation is graphed as a function of $P_{correct}$ in the figure below. The dashed lines in the figure show the entropy of the problem, H_{BR}, before recognition.

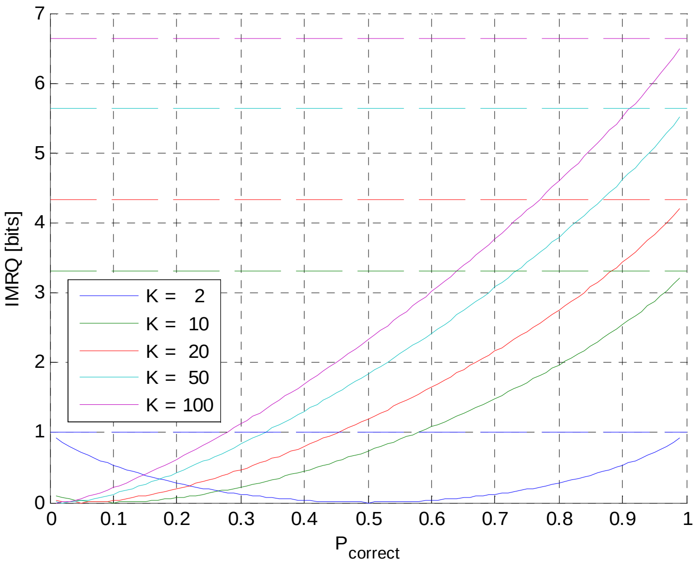


Figure 4.11: Information measure of recognition quality (IMRQ) as a function of the probability of correct classification.

This graph gives a rough estimate of the value required of the mutual information to recognize a given number of targets, at a given probability of correct recognition. The

IMRQ is plotted below as a function of K for various values of $P_{correct}$. Note that K can be interpreted as the number of target profiles which have to be distinguished from each other.

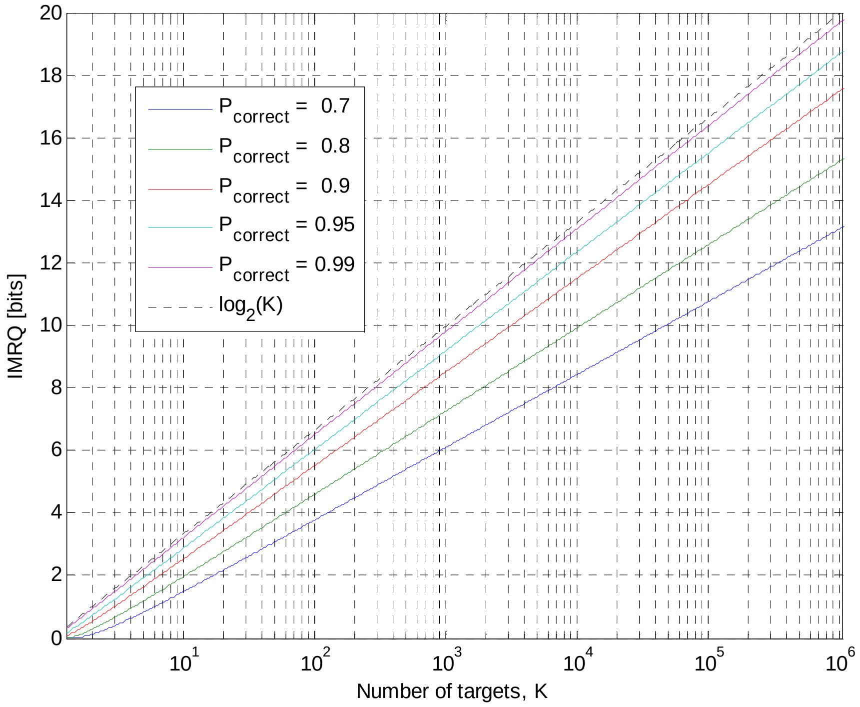


Figure 4.12: Information measure of recognition quality (IMRQ) as a function of the number of target profiles, K.

This gives a good idea of the loss in correct recognition, for a loss in mutual information. The dashed black trace indicates the entropy of the target ensemble and thus gives the upper bound on the IMRQ.

4.6.2 Amount of Information Required for Recognition

This example closely follows the discussion in [Nebabin1995, p. 24]. The original paper where the mathematics is derived is unfortunately written in Russian, and does not seem to be readily available.

The initial equation is of the same form as Shannon's channel capacity formula, but the technique for the derivation of the second expression is unknown. The amount of data required for the recognition is given by:

$$\begin{aligned} I^* &= \Delta FT \log_2\left(1+q_v^2\right) \\ &= \Delta FT \log_2\left(1+\frac{2\left(\ln m_s - \ln P_{error}\right)}{1-P_{correct}}\right) \end{aligned} \tag{4.78}$$

where, ΔF is the radar bandwidth, T is the data reception and processing time, P_{error} is the probability of erroneous recognition, $P_{correct}$ is the probability of correct recognition, m_s is the number of signals to be recognized (given by the product of the number of targets and the number of aspect angles for recognition), E_m is the average energy of the recognized signals, N_0 is the noise power spectral density, and

$$q_v = \sqrt{2E_m / N_0} \tag{4.79}$$

is the voltage signal to noise ratio. For a single pulse with a bandwidth of 40 MHz, a PRI of 0.1 ms (15 km ambiguous range), a single target described by 4096 azimuth waveforms in 180°, $P_{error} = 0.01$ and $P_{correct} = 0.99$, the required amount of data for recognition is 30.95×10^3 bits. This relatively large answer seems to stem from the long data processing time. If this time is reduced to the equivalent of 6 range bins (which spans a 20 m target at this bandwidth, the required information is 68.02 bits, which seems more realistic. The most plausible explanation of this equation is that the author calculated the signal dimensionality given by the product of the bandwidth and the data reception time. This was then used along with the probabilities of error and correct recognition to establish a multi-level signalling scheme on a multidimensional signal set with the spacing between the signal points calculated to meet the given probabilities. This spacing can then be used to define the signal to noise ratio, which could then be used to calculate the channel capacity. The exact definition of T will thus have to be found before this approach can be used successfully.

4.7 Conclusion

This section contains a survey of the theory required to use information theory to analyse a radar based NCTR system. Some bounds were given on the amount of information required for correct recognition, but an analysis similar to that in [Briles1993] will have to be conducted to characterise the amount of information that might be available in the backscattered return from a target. The data processing inequality was also introduced, which states that no amount of clever processing can increase the amount of information known about a random variable. This highlights the fact that information theory is probably the best way to analyse a NCTR system before any of the radar's signal processing and recognition algorithms have been designed. The most promising approach seems to be the characterisation of the mutual information between the target's electromagnetic return and the signal which is received by the radar, followed by the conversion of this numerical value to a probability of error by making use of Fano's inequality.

5 Development of the Mutual Information Based Approach for the Analysis of Radar Recognition Problems

This section contains the development of techniques for the evaluation of the information content of radar signals. This can then be used to characterize the maximum attainable recognition performance for various radar signal processing algorithms.

5.1 Abstraction of the Radar Recognition Problem

In this section the radar recognition problem is reduced to its simplest form to allow the problem to be addressed analytically. To this end, the radar target and radar system is mapped to a communication topology and thus analysed.

In a communication system the designer can design the waveforms which are to be sent over a communication channel. The channel then possibly deforms these signals which are then received by the receiver which is usually noise limited. The radar target can be likened to the transmitter as it is capable of generating an ensemble of waveforms when interrogated by a pulse from the radar's transmitter. If the target's orientation is unknown, its orientation can be seen as a random variable. The backscattered waveform is generated via a one to one mapping between the target orientation and the waveform. The target can thus be seen as a random process capable of generating a waveform based on its orientation relative to the radar. These waveforms are however not under the control of the radar designer, and the waveforms are thus not optimal for information transfer. The propagation environment between the target and the radar can be seen as the channel and the radar receiver then receives the waveform, but is also noise limited as is the case for the communication system.

The radar problem can thus be analysed as a non-optimal communication system. The factor which limits the performance of a communication system is the amount of mutual information between the transmitted signal set and the received signal set. To analyse the radar problem the amount of mutual information between the target waveforms and the waveforms received at the radar by its receiver will have to be estimated using the following expression:

$$I(X,Y)=\sum_{i=1}^{N}\int_{-\infty}^{\infty}P(x_i)\,p(y|x_i)\log_2\left(\frac{p(y|x_i)}{p(y)}\right)dy \qquad (5.1)$$

This expression was used by Ungerboeck [Ungerboeck1982] to characterise the achievable capacities of various standard communication schemes. He then used information gains between the schemes to show how modulation and coding could be combined to give a performance gain in bit error probability.

5.2 Derivation of Mutual Information for Real and Complex Variables

Ungerboeck's original paper [Ungerboeck1982] is relatively sketchy on the details of the derivation of mutual information. Due to this the derivation is expanded in more detail in this section.

Assuming perfect synchronization, and symbol period T, the output of the channel is given by

$$y_n = a_n + z_n \tag{5.2}$$

where a_n is the real- or complex-valued discrete input signal at time nT and z_n is a Gaussian noise sample with zero mean and variance σ^2 along each dimension. The SNR is defined as

$$\begin{aligned} SNR &= \frac{E\left\{\left|a_n^2\right|\right\}}{E\left\{\left|z_n^2\right|\right\}} \\ &= E\left\{\left|a_n^2\right|\right\} \big/ D\sigma^2 \end{aligned} \tag{5.3}$$

where D is the number of dimensions. Normalised signal power is assumed

$$E\left\{\left|a_n^2\right|\right\} = 1 \, . \tag{5.4}$$

The mutual information for a discrete memoryless channel with continuous outputs is given by:

$$I\left(a_k, y\right) = \sum_{k=0}^{N-1} P(k) \int_{-\infty}^{\infty} p\left(y \middle| a_k\right) \cdot \log_2 \left\{ \frac{p\left(y \middle| a_k\right)}{\sum_{i=0}^{N-1} P(i)\, p\left(y \middle| a_i\right)} \right\} dy \tag{5.5}$$

where N is the number of symbols and $P(\bullet)$ denotes the a-priori probability of the symbols. Note that the integration in (5.5) is written as a single integral, but the number of integrations is equal to the number of dimensions in the input waveforms. Single integrals are used throughout the document to reduce the amount of clutter in the equations.

As the interference is additive, the output pdf given an input symbol can be written as:

$$p(y \mid a_k) = \frac{1}{(2\pi\sigma^2)^{D/2}} \exp\left(\frac{-\left|y - a_k\right|^2}{2\sigma^2} \right) \tag{5.6}$$

where D is the number of dimensions. Note that the vector norm operator used in the exponential can be written as:

$$-\frac{1}{2}\sum_{j=1}^{D} \frac{\left(y_j - a_{kj}\right)^2}{\sigma^2} = \frac{-\left|y - a_k\right|^2}{2\sigma^2} \tag{5.7}$$

Note also that because $y = a_k + z$,

$$\begin{aligned} p(y \mid a_k) &= \frac{1}{(2\pi\sigma^2)^{D/2}} \exp\left(\frac{-\left|y - a_k\right|^2}{2\sigma^2} \right) \\ &= \frac{1}{(2\pi\sigma^2)^{D/2}} \exp\left(\frac{-\left|a_k + z - a_k\right|^2}{2\sigma^2} \right) \\ &= \frac{1}{(2\pi\sigma^2)^{D/2}} \exp\left(\frac{-\left|z\right|^2}{2\sigma^2} \right) \\ &= p(z) \end{aligned} \tag{5.8}$$

Thus $p(z)$ has a Gaussian distribution with zero mean and variance σ^2 along each dimension. Assume that all input events are equally likely, i.e.

$$P(a_k) = \frac{1}{N} \tag{5.9}$$

Substituting and simplifying, this gives:

$$\begin{aligned} I(a_k, y) &= \sum_{k=0}^{N-1} P(k) \int_{-\infty}^{\infty} p\left(y \middle| a_k\right) \cdot \log_2 \left\{ \frac{p\left(y \middle| a_k\right)}{\sum_{i=0}^{N-1} P(i)\, p\left(y \middle| a_i\right)} \right\} dy \\ &= \frac{1}{N} \sum_{k=0}^{N-1} \int_{-\infty}^{\infty} p\left(y \middle| a_k\right) \cdot \log_2 \left\{ \frac{1}{N} \sum_{i=0}^{N-1} \exp\left(\frac{\left|y - a_i\right|^2 - \left|y - a_k\right|^2}{2\sigma^2} \right) \right\} dy \\ &= \frac{1}{N} \sum_{k=0}^{N-1} \int_{-\infty}^{\infty} p\left(y \middle| a_k\right) \cdot \left[\log_2(N) - \log_2 \left\{ \sum_{i=0}^{N-1} \exp\left(\frac{-\left|a_k + z - a_i\right|^2 + \left|a_k + z - a_k\right|^2}{2\sigma^2} \right) \right\} \right] dy \\ &= \frac{1}{N} \sum_{k=0}^{N-1} \int_{-\infty}^{\infty} p\left(y \middle| a_k\right) \cdot \log_2(N)\, dy \quad - \quad \frac{1}{N} \sum_{k=0}^{N-1} \int_{-\infty}^{\infty} p\left(y \middle| a_k\right) \log_2 \left\{ \sum_{i=0}^{N-1} \exp\left(\frac{-\left|a_k + z - a_i\right|^2 + \left|z\right|^2}{2\sigma^2} \right) \right\} dy \end{aligned} \tag{5.10}$$

Now, substituting $p\left(y\middle|a_k\right)=p(z)$, and then making a change of variable using $y=a_k+z$ which gives $dy=dz$:

$$\begin{aligned}
I(a_k,y) &= \frac{1}{N}\sum_{k=0}^{N-1}\int_{-\infty}^{\infty} p\left(y\middle|a_k\right)\cdot\log_2(N)\,dy \quad - \quad \frac{1}{N}\sum_{k=0}^{N-1}\int_{-\infty}^{\infty} p\left(y\middle|a_k\right)\log_2\left\{\sum_{i=0}^{N-1}\exp\left(\frac{-\left|a_k+z-a_i\right|^2+\left|z\right|^2}{2\sigma^2}\right)\right\}dy \\
&= \frac{1}{N}\sum_{k=0}^{N-1}\int_{-\infty}^{\infty} p(z)\cdot\log_2(N)\,dy \quad - \quad \frac{1}{N}\sum_{k=0}^{N-1}\int_{-\infty}^{\infty} p(z)\log_2\left\{\sum_{i=0}^{N-1}\exp\left(\frac{-\left|a_k+z-a_i\right|^2+\left|z\right|^2}{2\sigma^2}\right)\right\}dy \\
& \frac{1}{N}\sum_{k=0}^{N-1}\int_{-\infty}^{\infty} p(z)\cdot\log_2(N)\,dz \quad - \quad \frac{1}{N}\sum_{k=0}^{N-1}\int_{-\infty}^{\infty} p(z)\log_2\left\{\sum_{i=0}^{N-1}\exp\left(\frac{-\left|a_k+z-a_i\right|^2+\left|z\right|^2}{2\sigma^2}\right)\right\}dz \\
&= \frac{\log_2(N)}{N}\sum_{k=0}^{N-1}1 \quad - \quad \frac{1}{N}\sum_{k=0}^{N-1}E\left\{\log_2\left[\sum_{i=0}^{N-1}\exp\left(\frac{-\left|a_k+z-a_i\right|^2+\left|z\right|^2}{2\sigma^2}\right)\right]\right\} \\
&= \frac{\log_2(N)}{N}N \quad - \quad \frac{1}{N}\sum_{k=0}^{N-1}E\left\{\log_2\left[\sum_{i=0}^{N-1}\exp\left(\frac{-\left|a_k+z-a_i\right|^2+\left|z\right|^2}{2\sigma^2}\right)\right]\right\} \\
&= \log_2(N) \quad - \quad \frac{1}{N}\sum_{k=0}^{N-1}E\left\{\log_2\left[\sum_{i=0}^{N-1}\exp\left(\frac{-\left|a_k+z-a_i\right|^2+\left|z\right|^2}{2\sigma^2}\right)\right]\right\}
\end{aligned} \tag{5.11}$$

This expression agrees with the expression arrived at by Ungerboeck. The expected value is over the Gaussian pdf, $p(z)$. Notice that this equation has a maximum value of $\log_2(N)$ and decreases as the amount of noise increases. Ungerboeck then made use of Monte Carlo integration to evaluate this expression.

Originally this expression was only applied to one and two dimensional signal sets. It is thus necessary to ensure that it can be applied to multi-dimensional problems.

5.3 Multi-Dimensional Mutual Information

In this section the result of a detailed derivation of the Mutual information for a discrete input continuous output channel is given. The full derivation has been given in Appendix A. The final result is:

$$I(\mathbf{a_k},\mathbf{y})=\log_2(N)-\frac{1}{N}\sum_{k=0}^{N-1}E\left\{\log_2\left[\sum_{i=0}^{N-1}\exp\left(-\frac{\left|\mathbf{a_k}+\mathbf{z}-\mathbf{a_i}\right|^2-\left|\mathbf{z}\right|^2}{2\sigma^2}\right)\right]\right\}, \tag{5.12}$$

where the variables have been replaced with vectors. This expression simplifies to that given by Ungerboeck for 2-D variables, but the derivation has been conducted rigorously to ensure that this expression can be applied to multi-dimensional problems.

5.4 Envelope Only Based Expressions for Mutual Information

In this section the expressions are derived for the mutual information in the cases where the radar signal processing generates the envelope or envelope squared of the received signal as its output signal.

5.4.1 Derivation of Mutual Information for Envelope Processing

In this section an expression is derived for the MI for a complex valued radar signal which has been processed by the radar's signal processing to obtain the envelope of the signal. This processing, which is applied to each range bin, is given by:

$$\begin{aligned} r_n &= \left|a_n + z_n\right| \\ &= \sqrt{\left(I_n + z_{In}\right)^2 + \left(Q_n + z_{Qn}\right)^2} \end{aligned} \qquad (5.13)$$

where n is the sample index and I_n and Q_n are the real and imaginary parts respectively for the n^th data sample, and $z_{In} + iz_{Qn}$ is the corresponding additive Gaussian noise sample.

This derivation starts from the general form for the MI of a discrete input continuous output channel:

$$\begin{aligned} I\left(a_k, r\right) &= \sum_{k=0}^{N-1} P\left(k\right) \int_{-\infty}^{\infty} p\left(r\left|a_k\right.\right) \cdot \log_2 \left\{ \frac{p\left(r\left|a_k\right.\right)}{\sum_{i=0}^{N-1} P\left(i\right) p\left(r\left|a_i\right.\right)} \right\} dr \\ &= \log_2\left(N\right) + \frac{1}{N} \sum_{k=0}^{N-1} \int_{-\infty}^{\infty} p\left(r\left|a_k\right.\right) \cdot \log_2 \left\{ \frac{p\left(r\left|a_k\right.\right)}{\sum_{i=0}^{N-1} p\left(r\left|a_i\right.\right)} \right\} dr \end{aligned} \qquad (5.14)$$

where the input is a vector of complex valued samples and the output is given by (5.13) above.

The pdf of the envelope of a received sample is given by a Ricean distribution. The Ricean pdf is obtained by a change of variable induced by the function $R = \sqrt{Y}$ performed on a non-central Chi-squared, 2 degrees of freedom RV, and is given by:

$$p\left(r\right) = \frac{r}{\sigma^2} \exp\left(-\frac{r^2 + s^2}{2\sigma^2}\right) I_0\left(\frac{rs}{\sigma^2}\right) \qquad (5.15)$$

where $I_0\left(\bullet\right)$ is the modified Bessel function of the first kind of order zero, and

$$s^2 = m_1^2 + m_2^2 \tag{5.16}$$

If each data sample is assumed to be independent, then:

$$\begin{aligned} p\left(r\,|a_k\right) &= \prod_{j=0}^{N-1} p\left(r_j\,|a_{kj}\right) \\ &= \prod_{j=0}^{N-1} \frac{r_{kj}}{2\sigma^2}\exp\left(-\frac{s_{kj}^2 + r_{kj}^2}{2\sigma^2}\right) I_0\left(\frac{r_{kj}s_{kj}}{\sigma^2}\right) \\ &= \prod_{j=0}^{N-1}\left[\frac{\sqrt{\left(a_{Ikj}+z_{Ij}\right)^2 + \left(a_{Qkj}+z_{Qj}\right)^2}}{2\sigma^2}\times\right. \\ &\quad \exp\left(-\frac{a_{Ikj}^2 + a_{Qkj}^2 + \left(a_{Ikj}+z_{Ij}\right)^2 + \left(a_{Qkj}+z_{Qj}\right)^2}{2\sigma^2}\right)\times \\ &\quad \left. I_0\left(\sqrt{\left(a_{Ikj}+z_{Ij}\right)^2 + \left(a_{Qkj}+z_{Qj}\right)^2}\,\frac{\sqrt{a_{Ikj}^2 + a_{Qkj}^2}}{\sigma^2}\right)\right] \end{aligned} \tag{5.17}$$

Note that there is no change of variable for this case, and that samples from $p\left(r\,|a_k\right)$ will be generated directly. The inner probability ratio can then be expanded as follows:

$$
\begin{aligned}
\frac{p\left(y\middle|a_k\right)}{\sum\limits_{i=0}^{N-1} p\left(y\middle|a_i\right)} &= \frac{\prod\limits_{j=0}^{N-1} \frac{r_{kj}}{2\sigma^2}\exp\left(-\frac{s_{kj}^2+r_{kj}^2}{2\sigma^2}\right) I_0\left(\frac{r_{kj}s_{kj}}{\sigma^2}\right)}{\sum\limits_{i=0}^{N-1}\prod\limits_{j=0}^{N-1} \frac{r_{kj}}{2\sigma^2}\exp\left(-\frac{s_{ij}^2+r_{kj}^2}{2\sigma^2}\right) I_0\left(\frac{r_{kj}s_{ij}}{\sigma^2}\right)} \\
&= \frac{\prod\limits_{j=0}^{N-1} \exp\left(-\frac{s_{kj}^2+r_{kj}^2}{2\sigma^2}\right) I_0\left(\frac{r_{kj}s_{kj}}{\sigma^2}\right)}{\sum\limits_{i=0}^{N-1}\prod\limits_{j=0}^{N-1} \exp\left(-\frac{s_{ij}^2+r_{kj}^2}{2\sigma^2}\right) I_0\left(\frac{r_{kj}s_{ij}}{\sigma^2}\right)} \\
&= \frac{\prod\limits_{j=0}^{N-1} I_0\left(\frac{r_{kj}s_{kj}}{\sigma^2}\right)}{\sum\limits_{i=0}^{N-1}\prod\limits_{j=0}^{N-1} \exp\left(\frac{s_{kj}^2+r_{kj}^2}{2\sigma^2}\right)\exp\left(-\frac{s_{ij}^2+r_{kj}^2}{2\sigma^2}\right) I_0\left(\frac{r_{kj}s_{ij}}{\sigma^2}\right)} \\
&= \frac{\prod\limits_{j=0}^{N-1} I_0\left(\frac{r_{kj}s_{kj}}{\sigma^2}\right)}{\sum\limits_{i=0}^{N-1}\prod\limits_{j=0}^{N-1} \exp\left(\frac{s_{kj}^2+r_{kj}^2}{2\sigma^2}-\frac{s_{ij}^2+r_{kj}^2}{2\sigma^2}\right) I_0\left(\frac{r_{kj}s_{ij}}{\sigma^2}\right)} \\
&= \frac{\prod\limits_{j=0}^{N-1} I_0\left(\frac{r_{kj}s_{kj}}{\sigma^2}\right)}{\sum\limits_{i=0}^{N-1}\prod\limits_{j=0}^{N-1} \exp\left(\frac{s_{kj}^2-s_{ij}^2}{2\sigma^2}\right) I_0\left(\frac{r_{kj}s_{ij}}{\sigma^2}\right)} \\
&= \frac{\prod\limits_{j=0}^{N-1} I_0\left(\frac{r_{kj}s_{kj}}{\sigma^2}\right)}{\sum\limits_{i=0}^{N-1}\prod\limits_{j=0}^{N-1} \exp\left(\frac{s_{kj}^2-s_{ij}^2}{2\sigma^2}\right) \prod\limits_{j=0}^{N-1} I_0\left(\frac{r_{kj}s_{ij}}{\sigma^2}\right)} \\
&= \frac{\prod\limits_{j=0}^{N-1} I_0\left(\frac{r_{kj}s_{kj}}{\sigma^2}\right)}{\sum\limits_{i=0}^{N-1}\left(\exp\left(\frac{1}{2\sigma^2}\sum\limits_{j=0}^{N-1}\left(s_{kj}^2-s_{ij}^2\right)\right)\prod\limits_{j=0}^{N-1} I_0\left(\frac{r_{kj}s_{ij}}{\sigma^2}\right)\right)}
\end{aligned}
\tag{5.18}
$$

Substituting this into the expression for MI:

$$
\begin{aligned}
I\left(a_k,r\right) &= \sum_{k=0}^{N-1} P\left(k\right)\int_{-\infty}^{\infty} p\left(r\middle|a_k\right)\cdot\log_2\left\{\frac{p\left(r\middle|a_k\right)}{\sum_{i=0}^{N-1}P\left(i\right)p\left(r\middle|a_i\right)}\right\}dy \\
&= \log_2\left(N\right)+\frac{1}{N}\sum_{k=0}^{N-1}\int_{-\infty}^{\infty} p\left(r\middle|a_k\right)\cdot\log_2\left\{\frac{\prod_{j=0}^{N-1} I_0\left(\frac{r_{kj}s_{kj}}{\sigma^2}\right)}{\sum_{i=0}^{N-1}\left(\exp\left(\frac{1}{2\sigma^2}\sum_{j=0}^{N-1}\left(s_{kj}^2-s_{ij}^2\right)\right)\prod_{j=0}^{N-1} I_0\left(\frac{r_{kj}s_{ij}}{\sigma^2}\right)\right)}\right\}dr \\
&= \log_2\left(N\right)+\frac{1}{N}\sum_{k=0}^{N-1} E\left\{\log_2\left[\frac{\prod_{j=0}^{N-1} I_0\left(\frac{r_{kj}s_{kj}}{\sigma^2}\right)}{\sum_{i=0}^{N-1}\left(\exp\left(\frac{1}{2\sigma^2}\sum_{j=0}^{N-1}\left(s_{kj}^2-s_{ij}^2\right)\right)\prod_{j=0}^{N-1} I_0\left(\frac{r_{kj}s_{ij}}{\sigma^2}\right)\right)}\right]\right\}
\end{aligned}
\qquad (5.19)
$$

where

$$
\begin{aligned}
s_{ij} &= \sqrt{a_{Iij}^2 + a_{Qij}^2} \\
s_{kj} &= \sqrt{a_{Ikj}^2 + a_{Qkj}^2} \\
r_{kj} &= \sqrt{\left(a_{Ikj}+z_{Ij}\right)^2 + \left(a_{Qkj}+z_{Qj}\right)^2}
\end{aligned}
\qquad (5.20)
$$

The following simplification of the fraction in (5.19) was derived to reduce the computational complexity:

$$
\begin{aligned}
\frac{\prod_{j=0}^{N-1} I_0\left(\frac{r_{kj}s_{kj}}{\sigma^2}\right)}{\sum_{i=0}^{N-1}\prod_{j=0}^{N-1}\exp\left(\frac{s_{kj}^2-s_{ij}^2}{2\sigma^2}\right)I_0\left(\frac{r_{kj}s_{ij}}{\sigma^2}\right)}
&= \frac{\prod_{j=0}^{N-1}\exp\left(-\frac{s_{kj}^2}{2\sigma^2}\right)\prod_{j=0}^{N-1} I_0\left(\frac{r_{kj}s_{kj}}{\sigma^2}\right)}{\sum_{i=0}^{N-1}\prod_{j=0}^{N-1}\exp\left(-\frac{s_{ij}^2}{2\sigma^2}\right)I_0\left(\frac{r_{kj}s_{ij}}{\sigma^2}\right)} \\
&= \frac{\prod_{j=0}^{N-1}\exp\left(\frac{-s_{kj}^2}{2\sigma^2}\right)\prod_{j=0}^{N-1} I_0\left(\frac{r_{kj}s_{kj}}{\sigma^2}\right)}{\sum_{i=0}^{N-1}\prod_{j=0}^{N-1}\exp\left(\frac{-s_{ij}^2}{2\sigma^2}\right)I_0\left(\frac{r_{kj}s_{ij}}{\sigma^2}\right)} \\
&= \frac{\exp\left(-\sum_{j=0}^{N-1}\frac{s_{kj}^2}{2\sigma^2}\right)\prod_{j=0}^{N-1} I_0\left(\frac{r_{kj}s_{kj}}{\sigma^2}\right)}{\sum_{i=0}^{N-1}\left(\exp\left(-\sum_{j=0}^{N-1}\frac{s_{ij}^2}{2\sigma^2}\right)\prod_{j=0}^{N-1} I_0\left(\frac{r_{kj}s_{ij}}{\sigma^2}\right)\right)}.
\end{aligned}
\qquad (5.21)
$$

Taking the logarithm of this expression, to the base 2, gives the following expression:

$$\begin{aligned}
&\log_2\left(\frac{\exp\left(-\sum_{j=0}^{N-1} s_{kj}^2\right)\prod_{j=0}^{N-1} I_0\left(\frac{r_{kj}s_{kj}}{\sigma^2}\right)}{\sum_{i=0}^{N-1}\left(\exp\left(-\sum_{j=0}^{N-1} s_{ij}^2\right)\prod_{j=0}^{N-1} I_0\left(\frac{r_{kj}s_{ij}}{\sigma^2}\right)\right)}\right) \\
&=\log_2\left(\exp\left(-\sum_{j=0}^{N-1} s_{kj}^2\right)\right)+\log_2\left(\frac{\prod_{j=0}^{N-1} I_0\left(\frac{r_{kj}s_{kj}}{\sigma^2}\right)}{\sum_{i=0}^{N-1}\left(\exp\left(-\sum_{j=0}^{N-1} s_{ij}^2\right)\prod_{j=0}^{N-1} I_0\left(\frac{r_{kj}s_{ij}}{\sigma^2}\right)\right)}\right) \\
&=-\frac{\sum_{j=0}^{N-1} s_{kj}^2}{\ln(2)}+\log_2\left(\frac{\prod_{j=0}^{N-1} I_0\left(\frac{r_{kj}s_{kj}}{\sigma^2}\right)}{\sum_{i=0}^{N-1}\left(\exp\left(-\sum_{j=0}^{N-1} s_{ij}^2\right)\prod_{j=0}^{N-1} I_0\left(\frac{r_{kj}s_{ij}}{\sigma^2}\right)\right)}\right)
\end{aligned} \tag{5.22}$$

which can then be substituted into the expression for MI to give:

$$\begin{aligned}
I(a_k, y) &= \log_2(N)+\frac{1}{N}\sum_{k=0}^{N-1} E\left\{-\frac{\sum_{j=0}^{N-1} s_{kj}^2}{\ln(2)}+\log_2\left(\frac{\prod_{j=0}^{N-1} I_0\left(\frac{r_{kj}s_{kj}}{\sigma^2}\right)}{\sum_{i=0}^{N-1}\left(\exp\left(-\sum_{j=0}^{N-1} s_{ij}^2\right)\prod_{j=0}^{N-1} I_0\left(\frac{r_{kj}s_{ij}}{\sigma^2}\right)\right)}\right)\right\} \\
&=\log_2(N)+\frac{1}{N}\sum_{k=0}^{N-1}\left\{-\frac{\sum_{j=0}^{N-1} s_{kj}^2}{\ln(2)}+E\left\{\log_2\left(\frac{\prod_{j=0}^{N-1} I_0\left(\frac{r_{kj}s_{kj}}{\sigma^2}\right)}{\sum_{i=0}^{N-1}\left(\exp\left(-\sum_{j=0}^{N-1} s_{ij}^2\right)\prod_{j=0}^{N-1} I_0\left(\frac{r_{kj}s_{ij}}{\sigma^2}\right)\right)}\right)\right\}\right\} \\
&=\log_2(N)-\frac{1}{N\ln(2)}\sum_{k=0}^{N-1}\sum_{j=0}^{N-1} s_{kj}^2+\frac{1}{N}\sum_{k=0}^{N-1} E\left\{\log_2\left(\frac{\prod_{j=0}^{N-1} I_0\left(\frac{r_{kj}s_{kj}}{\sigma^2}\right)}{\sum_{i=0}^{N-1}\left(\exp\left(-\sum_{j=0}^{N-1} s_{ij}^2\right)\prod_{j=0}^{N-1} I_0\left(\frac{r_{kj}s_{ij}}{\sigma^2}\right)\right)}\right)\right\}
\end{aligned} \tag{5.23}$$

The numerical stability of this expression was found to restrict the SNR range over which the MI could be calculated due to overflow problems when using the standard modified Bessel function implementation in Matlab. It was found that improved numerical stability could be achieved if a scaled version of the modified Bessel function is used [Abromowitz1972, pp. 375], which is defined as follows:

$$I_0'(z) = I_0(z)\exp(-|z|) \tag{5.24}$$

The modified Bessel functions can thus be replaced using the following expression:

$$I_0(z) = \exp(|z|)\, I_0'(z) \tag{5.25}$$

The derivation is restarted from Eq. (5.21):

$$
\begin{aligned}
\frac{\prod_{j=0}^{N-1} I_0\left(\frac{r_{kj}s_{kj}}{\sigma^2}\right)}{\sum_{i=0}^{N-1}\prod_{j=0}^{N-1}\exp\left(\frac{s_{kj}^2-s_{ij}^2}{2\sigma^2}\right) I_0\left(\frac{r_{kj}s_{ij}}{\sigma^2}\right)} &= \frac{\prod_{j=0}^{N-1}\exp\left(-\frac{s_{kj}^2}{2\sigma^2}\right) I_0\left(\frac{r_{kj}s_{kj}}{\sigma^2}\right)}{\sum_{i=0}^{N-1}\prod_{j=0}^{N-1}\exp\left(-\frac{s_{ij}^2}{2\sigma^2}\right) I_0\left(\frac{r_{kj}s_{ij}}{\sigma^2}\right)} \\
&= \frac{\prod_{j=0}^{N-1}\exp\left(\frac{-s_{kj}^2}{2\sigma^2}\right)\exp\left(\frac{r_{kj}s_{kj}}{\sigma^2}\right) I_0'\left(\frac{r_{kj}s_{kj}}{\sigma^2}\right)}{\sum_{i=0}^{N-1}\prod_{j=0}^{N-1}\exp\left(\frac{-s_{ij}^2}{2\sigma^2}\right)\exp\left(\frac{r_{kj}s_{ij}}{\sigma^2}\right) I_0'\left(\frac{r_{kj}s_{ij}}{\sigma^2}\right)} \\
&= \frac{\prod_{j=0}^{N-1}\exp\left(\frac{2r_{kj}s_{kj}}{2\sigma^2}-\frac{s_{kj}^2}{2\sigma^2}\right) I_0'\left(\frac{r_{kj}s_{kj}}{\sigma^2}\right)}{\sum_{i=0}^{N-1}\prod_{j=0}^{N-1}\exp\left(\frac{2r_{kj}s_{ij}}{2\sigma^2}-\frac{s_{ij}^2}{2\sigma^2}\right) I_0'\left(\frac{r_{kj}s_{ij}}{\sigma^2}\right)} \\
&= \frac{\prod_{j=0}^{N-1}\exp\left(\frac{2r_{kj}s_{kj}-s_{kj}^2}{2\sigma^2}\right) I_0'\left(\frac{r_{kj}s_{kj}}{\sigma^2}\right)}{\sum_{i=0}^{N-1}\prod_{j=0}^{N-1}\exp\left(\frac{2r_{kj}s_{ij}-s_{ij}^2}{2\sigma^2}\right) I_0'\left(\frac{r_{kj}s_{ij}}{\sigma^2}\right)}
\end{aligned} \tag{5.26}
$$

The denominator can result in large arguments in the $\exp(\bullet)$ function, so the upper $\exp(\bullet)$ expression was moved into the denominator to further stabilize the expression:

$$
\begin{aligned}
\frac{\prod_{j=0}^{N-1}\exp\left(\frac{2r_{kj}s_{kj}-s_{kj}^2}{2\sigma^2}\right) I_0'\left(\frac{r_{kj}s_{kj}}{\sigma^2}\right)}{\sum_{i=0}^{N-1}\prod_{j=0}^{N-1}\exp\left(\frac{2r_{kj}s_{ij}-s_{ij}^2}{2\sigma^2}\right) I_0'\left(\frac{r_{kj}s_{ij}}{\sigma^2}\right)} &= \frac{\prod_{j=0}^{N-1} I_0'\left(\frac{r_{kj}s_{kj}}{\sigma^2}\right)}{\sum_{i=0}^{N-1}\prod_{j=0}^{N-1}\exp\left(\frac{s_{kj}^2-2r_{kj}s_{kj}+2r_{kj}s_{ij}-s_{ij}^2}{2\sigma^2}\right) I_0'\left(\frac{r_{kj}s_{ij}}{\sigma^2}\right)} \\
&= \frac{\prod_{j=0}^{N-1} I_0'\left(\frac{r_{kj}s_{kj}}{\sigma^2}\right)}{\sum_{i=0}^{N-1}\prod_{j=0}^{N-1}\exp\left(\frac{s_{kj}^2+2r_{kj}\left(s_{ij}-s_{kj}\right)-s_{ij}^2}{2\sigma^2}\right) I_0'\left(\frac{r_{kj}s_{ij}}{\sigma^2}\right)}
\end{aligned} \tag{5.27}
$$

This expression was found to be stable, except in some specific cases where the SNR exceeds 45 dB.

5.4.2 Mutual Information for Envelope Squared Processing

Due to the fact that this derivation follows the same approach as that used in the previous section, the detailed derivation is given in Appendix B and only the final expressions for the MI is given here.

The expression for the envelope squared case is given by:

$$\begin{aligned} I(a_k, y) &= \sum_{k=0}^{N-1} P(k) \int_{-\infty}^{\infty} p(y|a_k) \cdot \log_2 \left\{ \frac{p(y|a_k)}{\sum_{i=0}^{N-1} P(i)\, p(y|a_i)} \right\} dy \\ &= \log_2(N) + \frac{1}{N} \sum_{k=0}^{N-1} E\left\{ \log_2 \left[\frac{\prod_{j=0}^{N-1} I_0\left(\sqrt{y_{kj}} \frac{s_{kj}}{\sigma^2} \right)}{\sum_{i=0}^{N-1} \left(\exp\left(\frac{1}{2\sigma^2} \sum_{j=0}^{N-1} \left(s_{kj}^2 - s_{ij}^2 \right) \right) \prod_{j=0}^{N-1} I_0\left(\sqrt{y_{kj}} \frac{s_{ij}}{\sigma^2} \right) \right)} \right] \right\} \end{aligned} \tag{5.28}$$

where

$$\begin{aligned} s_{ij} &= \sqrt{a_{Iij}^2 + a_{Qij}^2} \\ s_{kj} &= \sqrt{a_{Ikj}^2 + a_{Qkj}^2} \\ y_{kj} &= \left(a_{Ikj} + z_{Ij} \right)^2 + \left(a_{Qkj} + z_{Qj} \right)^2 . \end{aligned} \tag{5.29}$$

If the following substitution is made:

$$r_{kj} = \sqrt{y_{kj}} \, . \tag{5.30}$$

which is the definition given in (5.20), then the expression for the MI based on envelope squared values is given by:

$$I(a_k, y) = \log_2(N) + \frac{1}{N} \sum_{k=0}^{N-1} E\left\{ \log_2 \left[\frac{\prod_{j=0}^{N-1} I_0\left(\frac{r_{kj} s_{kj}}{\sigma^2} \right)}{\sum_{i=0}^{N-1} \left(\exp\left(\frac{1}{2\sigma^2} \sum_{j=0}^{N-1} \left(s_{kj}^2 - s_{ij}^2 \right) \right) \prod_{j=0}^{N-1} I_0\left(\frac{r_{kj} s_{ij}}{\sigma^2} \right) \right)} \right] \right\} \tag{5.31}$$

This approach was analysed as a cross-check on the envelope only HRR derivation. One would expect the two to give the same MI as the processing only differs by the square root function, which is an invertible function.

The expressions derived above, in particular (5.27), can be used to compare the performance of envelope only processing to that of phase and envelope processing. This in turn allows for the evaluation of the amount of information lost due to processing which discards phase information by computing the envelope of the of the signal.

5.5 1-D Multi-Target Channel

All the preceding sections have focussed on the case of only a single target's information transfer. This section derives the mutual information for the case where two targets can potentially be present and transmit their signature waveforms over the channel. This corresponds to the scenario where multiple targets of different types could be present in the radar's surveillance volume and the radar's recognition sub-system has to test a received signal against multiple targets in its database. Two different targets at different relative aspect angles to the radar could have very similar HRRP's, thus leading to scenarios where misclassification is possible. This section investigates an approach to the analysis of this type of problem. The analysis starts with a simplified version of the problem and then extends this approach to the generalised problem.

If the following symbols are possible: {-1; 1; b}, with probabilities 0.25, 0.25 and 0.5. One could derive the MI for all three symbols, which should limit at a maximum value of 1.5 bits for high SNR. The MI value could then be broken into sub-expressions - one for target #1 (+1,-1) and one for target #2 (b). The approach for all three symbols is taken, by starting with the expression for mutual information below and the parts of the expression pertaining to the target of interest are isolated.

$$\begin{aligned} I(\{a_k;b\},y) &= \sum_{k=0}^{N-1} P(a_k)\int_{-\infty}^{\infty} p(y|a_k)\cdot\log_2\left\{\frac{p(y|a_k)}{p(y)}\right\}dy + P(b)\int_{-\infty}^{\infty} p(y|b)\cdot\log_2\left\{\frac{p(y|b)}{p(y)}\right\}dy \\ &= \frac{1}{4}\sum_{k=0}^{N-1}\int_{-\infty}^{\infty} p(y|a_k)\cdot\log_2\left\{\frac{p(y|a_k)}{p(y)}\right\}dy + \frac{1}{2}\int_{-\infty}^{\infty} p(y|b)\cdot\log_2\left\{\frac{p(y|b)}{p(y)}\right\}dy \end{aligned} \tag{5.32}$$

Expanding p(y):

$$\begin{aligned} p(y) &= \sum_{k=0}^{N-1} P(a_k)\,p(y|a_k) + P(b)\,p(y|b) \\ &= \frac{1}{4}\sum_{k=0}^{N-1} p(y|a_k) + \frac{1}{2}p(y|b) \end{aligned} \tag{5.33}$$

Expanding p(y|c):

$$p(y|c) = \frac{1}{\sqrt{2\pi}\sigma}\exp\left(\frac{-(y-c)^2}{2\sigma^2}\right) \tag{5.34}$$

Substituting into p(y):

$$\begin{aligned} p(y) &= \frac{1}{4}\sum_{k=0}^{N-1} p(y|a_k) + \frac{1}{2}p(y|b) \\ &= \frac{1}{4}\sum_{k=0}^{N-1}\frac{1}{\sqrt{2\pi}\sigma}\exp\left(\frac{-(y-a_k)^2}{2\sigma^2}\right) + \frac{1}{2}\frac{1}{\sqrt{2\pi}\sigma}\exp\left(\frac{-(y-b)^2}{2\sigma^2}\right) \end{aligned} \tag{5.35}$$

For the first term, using the transformation $y = a_k + z$:

$$\begin{aligned}\frac{1}{4}\sum_{k=0}^{N-1}\int_{-\infty}^{\infty} p(y|a_k)\cdot\log_2\left\{\frac{p(y|a_k)}{p(y)}\right\}dy &= \frac{1}{4}\sum_{k=0}^{N-1}\int_{-\infty}^{\infty} p(z)\cdot\log_2\left\{\frac{p(z)}{p(a_k+z)}\right\}dz \\ &= \frac{1}{4}\sum_{k=0}^{N-1} E\left\{\log_2\left[\frac{p(z)}{p(a_k+z)}\right]\right\}\end{aligned} \tag{5.36}$$

Expanding the fraction:

$$\begin{aligned}\frac{p(z)}{p(y)} &= \frac{p(z)}{\frac{1}{4}\sum_{i=0}^{N-1} p\big((a_k+z)|a_i\big)+\frac{1}{2}p\big((a_k+z)|b\big)} \\ &= \frac{\frac{1}{\sqrt{2\pi}\sigma}\exp\left(\frac{-(z)^2}{2\sigma^2}\right)}{\frac{1}{4}\sum_{i=0}^{N-1}\frac{1}{\sqrt{2\pi}\sigma}\exp\left(\frac{-(a_k+z-a_i)^2}{2\sigma^2}\right)+\frac{1}{2}\frac{1}{\sqrt{2\pi}\sigma}\exp\left(\frac{-(a_k+z-b)^2}{2\sigma^2}\right)} \\ &= \left[\left(\exp\left(\frac{(z)^2}{2\sigma^2}\right)\right)\left(\frac{1}{4}\sum_{i=0}^{N-1}\exp\left(\frac{-(a_k+z-a_i)^2}{2\sigma^2}\right)+\frac{1}{2}\exp\left(\frac{-(a_k+z-b)^2}{2\sigma^2}\right)\right)\right]^{-1} \\ &= \left[\frac{1}{4}\sum_{i=0}^{N-1}\exp\left(\frac{-(a_k+z-a_i)^2+(z)^2}{2\sigma^2}\right)+\frac{1}{2}\exp\left(\frac{-(a_k+z-b)^2+(z)^2}{2\sigma^2}\right)\right]^{-1}\end{aligned} \tag{5.37}$$

Substituting back gives:

$$\begin{aligned}&\frac{1}{4}\sum_{k=0}^{N-1}E\left\{\log_2\left[\left[\frac{1}{4}\sum_{i=0}^{N-1}\exp\left(\frac{-(a_k+z-a_i)^2+(z)^2}{2\sigma^2}\right)+\frac{1}{2}\exp\left(\frac{-(a_k+z-b)^2+(z)^2}{2\sigma^2}\right)\right]^{-1}\right]\right\} \\ &= -\frac{1}{4}\sum_{k=0}^{N-1}E\left\{\log_2\left[\frac{1}{4}\sum_{i=0}^{N-1}\exp\left(\frac{-(a_k+z-a_i)^2+(z)^2}{2\sigma^2}\right)+\frac{1}{2}\exp\left(\frac{-(a_k+z-b)^2+(z)^2}{2\sigma^2}\right)\right]\right\}\end{aligned} \tag{5.38}$$

This expression gives the amount of information transferred in the presence of the extra symbol b, when only the {+1;-1} inputs are allowed. If b was included, the MI would limit at 1.5 bits for high SNR. This implies that b adds information to the whole process. If b is seen as an extra target of a different class, then observing the channel output when b is transmitted is giving information about the second target, although the receiver is only set up to receive {+1,-1}. The total MI for this channel requires the addition of the 2nd term:

$$\frac{1}{2}\int_{-\infty}^{\infty} p(y|b)\cdot\log_2\left\{\frac{p(y|b)}{p(y)}\right\}dy \tag{5.39}$$

using the transformation $y=b+z$:

$$\begin{aligned}\frac{1}{2}\int_{-\infty}^{\infty} p(y|b)\cdot\log_2\left\{\frac{p(y|b)}{p(y)}\right\}dy &= \frac{1}{2}\int_{-\infty}^{\infty} p(z)\cdot\log_2\left\{\frac{p(z)}{p(z+b)}\right\}dz \\ &= \frac{1}{2}E\left\{\log_2\left\{\frac{p(z)}{p(z+b)}\right\}\right\}\end{aligned} \tag{5.40}$$

Expanding the fraction:

$$\begin{aligned}
\frac{p(z)}{p(y)} &= \frac{p(z)}{\frac{1}{4}\sum_{i=0}^{N-1} p\left((b+z)\middle|a_i\right) + \frac{1}{2} p\left((b+z)\middle|b\right)} \\
&= \frac{\frac{1}{\sqrt{2\pi}\sigma}\exp\left(\frac{-(z)^2}{2\sigma^2}\right)}{\frac{1}{4}\sum_{i=0}^{N-1}\frac{1}{\sqrt{2\pi}\sigma}\exp\left(\frac{-(b+z-a_i)^2}{2\sigma^2}\right)+\frac{1}{2}\frac{1}{\sqrt{2\pi}\sigma}\exp\left(\frac{-(b+z-b)^2}{2\sigma^2}\right)} \\
&= \left[\left(\exp\left(\frac{(z)^2}{2\sigma^2}\right)\right)\left(\frac{1}{4}\sum_{i=0}^{N-1}\exp\left(\frac{-(b+z-a_i)^2}{2\sigma^2}\right)+\frac{1}{2}\exp\left(\frac{-(z)^2}{2\sigma^2}\right)\right)\right]^{-1} \\
&= \left[\frac{1}{4}\sum_{i=0}^{N-1}\exp\left(\frac{-(b+z-a_i)^2+(z)^2}{2\sigma^2}\right)+\frac{1}{2}\exp\left(\frac{-(z)^2+(z)^2}{2\sigma^2}\right)\right]^{-1} \\
&= \left[\frac{1}{2}+\frac{1}{4}\sum_{i=0}^{N-1}\exp\left(\frac{-(b+z-a_i)^2+(z)^2}{2\sigma^2}\right)\right]^{-1}
\end{aligned} \tag{5.41}$$

Substituting back gives:

$$\begin{aligned}
\frac{1}{2}E\left\{\log_2\left\{\frac{p(z)}{p(z+b)}\right\}\right\} &= \frac{1}{2}E\left\{\log_2\left[\frac{1}{2}+\frac{1}{4}\sum_{i=0}^{N-1}\exp\left(\frac{-(b+z-a_i)^2+(z)^2}{2\sigma^2}\right)\right]^{-1}\right\} \\
&= -\frac{1}{2}E\left\{\log_2\left[\frac{1}{2}+\frac{1}{4}\sum_{i=0}^{N-1}\exp\left(\frac{-(b+z-a_i)^2+(z)^2}{2\sigma^2}\right)\right]\right\}
\end{aligned} \tag{5.42}$$

Now adding this to the first term gives:

$$\begin{aligned}
I(\{a_k;b\},y) &= \frac{1}{4}\sum_{k=0}^{N-1}\int_{-\infty}^{\infty} p\left(y\middle|a_k\right)\cdot\log_2\left\{\frac{p\left(y\middle|a_k\right)}{p(y)}\right\}dy+\frac{1}{2}\int_{-\infty}^{\infty} p\left(y\middle|b\right)\cdot\log_2\left\{\frac{p\left(y\middle|b\right)}{p(y)}\right\}dy \\
&= -\frac{1}{4}\sum_{k=0}^{N-1}E\left\{\log_2\left[\frac{1}{4}\sum_{i=0}^{N-1}\exp\left(\frac{-(a_k+z-a_i)^2+(z)^2}{2\sigma^2}\right)+\frac{1}{2}\exp\left(\frac{-(a_k+z-b)^2+(z)^2}{2\sigma^2}\right)\right]\right\} \\
&\quad -\frac{1}{2}E\left\{\log_2\left[\frac{1}{2}+\frac{1}{4}\sum_{i=0}^{N-1}\exp\left(\frac{-(b+z-a_i)^2+(z)^2}{2\sigma^2}\right)\right]\right\}
\end{aligned} \tag{5.43}$$

which is the expression for the total amount of information which can be transferred across this channel.

5.6 Multidimensional Multi-Target Case

This model assumes that other targets are interferers (i.e. not of interest), which might be “close” to the signal constellation coordinates of the target which has been designated mathematically as the primary target (or target of interest). This means that the presence of the other target signals destroys information about the primary target. This allows the similarity between the returns from two targets to be measured and compared. It is also assumed that all targets are equally likely and that there are M competing non-primary targets. A diagrammatic representation of the channel is shown in the figure below, the a_k represent signal vectors of the target of interest, and the b_j signal vectors for the interfering targets. The approach taken is first to derive the total mutual information for all the targets. The information terms relating to the primary target are then identified to give the mutual information for the primary target.

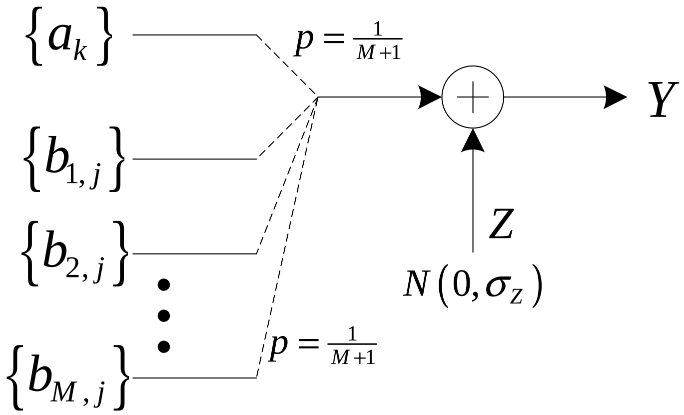


Figure 5.1: Generalized multi-target channel.

The initial expression for the MI is:

$$\begin{aligned} I\left(\{a_k;b_{m,j}\},y\right) &= \sum_{k=0}^{N-1} P(a_k)\int_{-\infty}^{\infty} p\left(y|a_k\right)\cdot\log_2\left\{\frac{p\left(y|a_k\right)}{p(y)}\right\}dy+\sum_{m=1}^{M}\sum_{j=0}^{N-1}\int_{-\infty}^{\infty} P\left(b_{m,j}\right)p\left(y|b_{m,j}\right)\cdot\log_2\left\{\frac{p\left(y|b_{m,j}\right)}{p(y)}\right\}dy \\ &= P(a)\sum_{k=0}^{N-1}\int_{-\infty}^{\infty} p\left(y|a_k\right)\cdot\log_2\left\{\frac{p\left(y|a_k\right)}{p(y)}\right\}dy+P(b)\sum_{m=1}^{M}\sum_{j=0}^{N-1}\int_{-\infty}^{\infty} p\left(y|b_{m,j}\right)\cdot\log_2\left\{\frac{p\left(y|b_{m,j}\right)}{p(y)}\right\}dy \end{aligned} \tag{5.44}$$

Expanding $p(y)$:

$$\begin{aligned} p(y) &= \sum_{i=0}^{N-1} P(a_i)\,p\left(y|a_i\right)+\sum_{m=1}^{M}\sum_{j=0}^{N-1} P\left(b_{m,j}\right)p\left(y|b_{m,j}\right) \\ &= P(a)\sum_{i=0}^{N-1} p\left(y|a_i\right)+P(b)\sum_{m=1}^{M}\sum_{j=0}^{N-1} p\left(y|b_{m,j}\right) \end{aligned} \tag{5.45}$$

Expanding $p(y|c)$:

$$p\left(y|c\right)=\frac{1}{\sqrt{2\pi}\sigma}\exp\left(\frac{-\left(y-c\right)^2}{2\sigma^2}\right) \tag{5.46}$$

Substituting into $p(y)$:

$$\begin{aligned} p(y) &= P(a)\sum_{i=0}^{N-1} p\left(y\middle|a_i\right) + P(b)\sum_{m=1}^{M}\sum_{j=0}^{N-1} p\left(y\middle|b_{m,j}\right) \\ &= P(a)\sum_{i=0}^{N-1}\frac{1}{\sqrt{2\pi\sigma}}\exp\left(\frac{-\left(y-a_i\right)^2}{2\sigma^2}\right) + P(b)\sum_{m=1}^{M}\sum_{j=0}^{N-1}\frac{1}{\sqrt{2\pi\sigma}}\exp\left(\frac{-\left(y-b_{m,j}\right)^2}{2\sigma^2}\right) \end{aligned} \tag{5.47}$$

For the first term of (5.44), using the transformation $y = a_k + z$:

$$\begin{aligned} P(a)\sum_{k=0}^{N-1}\int_{-\infty}^{\infty} p\left(y\middle|a_k\right)\cdot\log_2\left\{\frac{p\left(y\middle|a_k\right)}{p(y)}\right\}dy &= P(a)\sum_{k=0}^{N-1}\int_{-\infty}^{\infty} p(z)\cdot\log_2\left\{\frac{p(z)}{p\left(a_k+z\right)}\right\}dz \\ &= P(a)\sum_{k=0}^{N-1}E\left\{\log_2\left[\frac{p(z)}{p\left(a_k+z\right)}\right]\right\} \end{aligned} \tag{5.48}$$

Expanding the fraction:

$$\begin{aligned} \frac{p(z)}{p\left(a_k+z\right)} &= \frac{p(z)}{P(a)\sum_{i=0}^{N-1} p\left(a_k+z\middle|a_i\right) + P(b)\sum_{m=1}^{M}\sum_{j=0}^{N-1} p\left(a_k+z\middle|b_{m,j}\right)} \\ &= \frac{\frac{1}{\sqrt{2\pi\sigma}}\exp\left(\frac{-(z)^2}{2\sigma^2}\right)}{P(a)\sum_{i=0}^{N-1}\frac{1}{\sqrt{2\pi\sigma}}\exp\left(\frac{-\left(a_k+z-a_i\right)^2}{2\sigma^2}\right) + P(b)\sum_{m=1}^{M}\sum_{j=0}^{N-1}\frac{1}{\sqrt{2\pi\sigma}}\exp\left(\frac{-\left(a_k+z-b_{m,j}\right)^2}{2\sigma^2}\right)} \\ &= \left[\left(\exp\left(\frac{(z)^2}{2\sigma^2}\right)\right)\left(P(a)\sum_{i=0}^{N-1}\exp\left(\frac{-\left(a_k+z-a_i\right)^2}{2\sigma^2}\right) + P(b)\sum_{m=1}^{M}\sum_{j=0}^{N-1}\exp\left(\frac{-\left(a_k+z-b_{m,j}\right)^2}{2\sigma^2}\right)\right)\right]^{-1} \\ &= \left[P(a)\sum_{i=0}^{N-1}\exp\left(\frac{-\left(a_k+z-a_i\right)^2+(z)^2}{2\sigma^2}\right) + P(b)\sum_{m=1}^{M}\sum_{j=0}^{N-1}\exp\left(\frac{-\left(a_k+z-b_{m,j}\right)^2+(z)^2}{2\sigma^2}\right)\right]^{-1} \end{aligned} \tag{5.49}$$

Substituting back gives:

$$\begin{aligned} P(a)\sum_{k=0}^{N-1}E\left\{\log_2\left[\frac{p(z)}{p\left(a_k+z\right)}\right]\right\} &= P(a)\sum_{k=0}^{N-1}E\left\{\log_2\left[\left[P(a)\sum_{i=0}^{N-1}\exp\left(\frac{-\left(a_k+z-a_i\right)^2+(z)^2}{2\sigma^2}\right)\right.\right.\right. \\ &\qquad \left.\left.\left. +P(b)\sum_{m=1}^{M}\sum_{j=0}^{N-1}\exp\left(\frac{-\left(a_k+z-b_{m,j}\right)^2+(z)^2}{2\sigma^2}\right)\right]^{-1}\right]\right\} \\ &= -P(a)\sum_{k=0}^{N-1}E\left\{\log_2\left[P(a)\sum_{i=0}^{N-1}\exp\left(\frac{-\left(a_k+z-a_i\right)^2+(z)^2}{2\sigma^2}\right) + P(b)\sum_{m=1}^{M}\sum_{j=0}^{N-1}\exp\left(\frac{-\left(a_k+z-b_{m,j}\right)^2+(z)^2}{2\sigma^2}\right)\right]\right\} \end{aligned} \tag{5.50}$$

This is the desired expression for the first term, which is also the required MI associated with the target of interest. Some further simplifications might be possible, depending on the details of the problem being analysed. Now, for the second term of (5.44) , using the transformation $y = b_{m,j} + z$:

$$\begin{aligned}P(b)\sum_{m=1}^{M}\sum_{j=0}^{N-1}\int_{-\infty}^{\infty} p\left(y\middle|b_{m,j}\right)\cdot\log_2\left\{\frac{p\left(y\middle|b_{m,j}\right)}{p(y)}\right\}dy &= P(b)\sum_{m=1}^{M}\sum_{j=0}^{N-1}\int_{-\infty}^{\infty} p(z)\cdot\log_2\left\{\frac{p(z)}{p\left(b_{m,j}+z\right)}\right\}dz \\ &= P(b)\sum_{m=1}^{M}\sum_{j=0}^{N-1}E\left\{\log_2\left[\frac{p(z)}{p\left(b_{m,j}+z\right)}\right]\right\}\end{aligned}\tag{5.51}$$

Expanding the fraction gives:

$$\begin{aligned}\frac{p(z)}{p\left(b_{m,j}+z\right)} &= \frac{p(z)}{P(a)\sum_{i=0}^{N-1}p\left(b_{m,j}+z\middle|a_i\right)+P(b)\sum_{m=1}^{M}\sum_{j=0}^{N-1}p\left(b_{m,j}+z\middle|b_{m,j}\right)} \\ &= \frac{\frac{1}{\sqrt{2\pi}\sigma}\exp\left(\frac{-(z)^2}{2\sigma^2}\right)}{P(a)\sum_{i=0}^{N-1}\frac{1}{\sqrt{2\pi}\sigma}\exp\left(\frac{-\left(b_{m,j}+z-a_i\right)^2}{2\sigma^2}\right)+P(b)\sum_{m=1}^{M}\sum_{j=0}^{N-1}\frac{1}{\sqrt{2\pi}\sigma}\exp\left(\frac{-\left(b_{m,j}+z-b_{m,j}\right)^2}{2\sigma^2}\right)} \\ &= \left[\left(\exp\left(\frac{(z)^2}{2\sigma^2}\right)\right)\left(P(a)\sum_{i=0}^{N-1}\exp\left(\frac{-\left(b_{m,j}+z-a_i\right)^2}{2\sigma^2}\right)+P(b)\sum_{m=1}^{M}\sum_{j=0}^{N-1}\exp\left(\frac{-(z)^2}{2\sigma^2}\right)\right)\right]^{-1} \\ &= \left[P(a)\sum_{i=0}^{N-1}\exp\left(\frac{-\left(b_{m,j}+z-a_i\right)^2+(z)^2}{2\sigma^2}\right)+P(b)\sum_{m=1}^{M}\sum_{j=0}^{N-1}\exp\left(\frac{-(z)^2+(z)^2}{2\sigma^2}\right)\right]^{-1} \\ &= \left[P(a)\sum_{i=0}^{N-1}\exp\left(\frac{-\left(b_{m,j}+z-a_i\right)^2+(z)^2}{2\sigma^2}\right)+P(b)\sum_{m=1}^{M}\sum_{j=0}^{N-1}\exp(0)\right]^{-1} \\ &= \left[P(a)\sum_{i=0}^{N-1}\exp\left(\frac{-\left(b_{m,j}+z-a_i\right)^2+(z)^2}{2\sigma^2}\right)+M.N.P(b)\right]^{-1}\end{aligned}\tag{5.52}$$

Substituting back gives:

$$\begin{aligned}P(b)\sum_{m=1}^{M}\sum_{j=0}^{N-1}E\left\{\log_2\left[\frac{p(z)}{p\left(b_{m,j}+z\right)}\right]\right\} &= P(b)\sum_{m=1}^{M}\sum_{j=0}^{N-1}E\left\{\log_2\left[\left[P(a)\sum_{i=0}^{N-1}\exp\left(\frac{-\left(b_{m,j}+z-a_i\right)^2+(z)^2}{2\sigma^2}\right)+M.N.P(b)\right]^{-1}\right]\right\} \\ &= -P(b)\sum_{m=1}^{M}\sum_{j=0}^{N-1}E\left\{\log_2\left[P(a)\sum_{i=0}^{N-1}\exp\left(\frac{-\left(b_{m,j}+z-a_i\right)^2+(z)^2}{2\sigma^2}\right)+M.N.P(b)\right]\right\}\end{aligned}\tag{5.53}$$

which is the desired result for the second term. Substituting the two terms back into the original expression gives:

$$\begin{aligned}I\left(\{a_k;b_{m,j}\},y\right) &= P(a)\sum_{k=0}^{N-1}\int_{-\infty}^{\infty}p\left(y\middle|a_k\right)\cdot\log_2\left\{\frac{p\left(y\middle|a_k\right)}{p(y)}\right\}dy+P(b)\sum_{m=1}^{M}\sum_{j=0}^{N-1}\int_{-\infty}^{\infty}p\left(y\middle|b_{m,j}\right)\cdot\log_2\left\{\frac{p\left(y\middle|b_{m,j}\right)}{p(y)}\right\}dy \\ &= -P(a)\sum_{k=0}^{N-1}E\left\{\log_2\left[P(a)\sum_{i=0}^{N-1}\exp\left(\frac{-\left(a_k+z-a_i\right)^2+(z)^2}{2\sigma^2}\right)+P(b)\sum_{m=1}^{M}\sum_{j=0}^{N-1}\exp\left(\frac{-\left(a_k+z-b_{m,j}\right)^2+(z)^2}{2\sigma^2}\right)\right]\right\} \\ &\quad -P(b)\sum_{m=1}^{M}\sum_{j=0}^{N-1}E\left\{\log_2\left[P(a)\sum_{i=0}^{N-1}\exp\left(\frac{-\left(b_{m,j}+z-a_i\right)^2+(z)^2}{2\sigma^2}\right)+M.N.P(b)\right]\right\}\end{aligned}\tag{5.54}$$

This is the expression for the total mutual information over the channel if all targets have to be catered for by the radar's recognition sub-system.

5.7 Radar Related Derivations

This section contains some of the derivations which will be required to set up the MI experiments correctly from a radar point of view.

5.7.1 Calculation of Receiver Noise as a Function of Resolution with Pulse Compression

The analysis in this section is required to relate the actual noise power in a radar system to the noise power in the mutual information calculations.

The noise power in the radar receiver is given by

$$P_n = kTB \tag{5.55}$$

The signal to noise ratio gain due to pulse compression is given by

$$G_{PC} = \tau_p B \tag{5.56}$$

This means that the signal power can be kept constant, and the noise power can be reduced by this amount, giving the noise power after pulse compression as:

$$P_{n\_PC\_out} = \frac{P_n}{G_{PC}} = \frac{kTB}{\tau_p B} = \frac{kT}{\tau_p} \tag{5.57}$$

This gives the noise power for each sample in the HRRP.

Note: One needs to be very careful with the SNR calculation as a function of resolution and the number of samples in the HRRP. For example for a 6 bin, very low resolution setup, each bin has nearly exactly the same signal (i.e. low resolution pulse, which has been highly oversampled). This would give the 6 bin experiment an unfair advantage over the one bin experiment as the correlation receiver would give the 6 bin case a factor of 6 improvement in SNR due to coherent integration. On the other hand, the single sample case could have had a much lower analogue bandwidth, possibly even more than 6 times less, but at least 6 times less.

5.7.2 SNR Calculation for Disparate Waveforms

In some cases the MI will be calculated for two disparate waveforms so it is important to ensure that the two graphs of SNR versus MI can be compared. This calculation addresses the case where double the amount of samples (time or frequency) have been measured by a radar to create the HRRP.

Signal energy and power are respectively defined as follows:

$$\begin{aligned} E_x &= T\sum_{n=0}^{N-1} x^2\left[n\right] \\ P_x &= \frac{1}{N}\sum_{n=0}^{N-1} x^2\left[n\right] \end{aligned} \tag{5.58}$$

and the total noise power is given by:

$$P_n = \sum_{n=0}^{N-1} \sigma^2 = N\sigma^2 . \tag{5.59}$$

The variance of the sum of independent random variables is the sum of their individual variances.

Now for two cases, the first having M and the second 2M samples:

$$\begin{aligned} SNR_1 &= \frac{\frac{1}{M}\sum_{n=0}^{M} x_1^2\left[n\right]}{M\sigma^2} = \frac{P_{x_1}}{M\sigma^2} \\ SNR_2 &= \frac{\frac{1}{2M}\sum_{n=0}^{2M} x_2^2\left[n\right]}{2M\sigma^2} = \frac{P_{x_2}}{2M\sigma^2} \end{aligned} \tag{5.60}$$

If it is assumed that the power of the two received signals are equal and the same noise is added per dimension, then the second SNR is 3 dB lower than the first.

If the SNR’s are equal, as they would be when plotting the MI on the same system of axis, then:

$$\sigma_2^2 = \tfrac{1}{2}\sigma_1^2 , \tag{5.61}$$

which means that the noise variance per dimension for the second case is half of that for the first case. The respective noise powers for each case can now be calculated as:

$$\begin{aligned} P_{n1} &= \sum_{n=0}^{M-1} \sigma_1^2 = M\sigma_1^2 \\ P_{n2} &= \sum_{n=0}^{2M-1} \sigma_2^2 = 2M\sigma_2^2 = 2M\left(\tfrac{1}{2}\sigma_1^2\right) = M\sigma_1^2 \end{aligned} \tag{5.62}$$

showing that at equal SNRs for the two cases the total noise power has remained constant.

5.7.3 Stability of a Radar Return as a Function of Target Azimuth

If a target is resolved in high range resolution slight azimuth movement of the target will cause the HRRP to fluctuate rapidly. If there are two scatterers of equal magnitude in a HRRP bin, separated by a cross-range distance of d_{CR}, and the target rotates in azimuth by θ_{az}, then the two way path length difference is given by

$$\begin{aligned} R_{2-way} &= 2\,\Delta R \\ &= 2\left(d_{CR}\sin\left(\theta_{az}\right)\right) \\ &= 2d_{CR}\sin\left(\theta_{az}\right). \end{aligned} \tag{5.63}$$

This value has to be limited to a fraction of a wavelength to ensure that the magnitude value remains relatively constant. As a worst case, by allowing the distance to vary by a quarter of a wavelength gives the following restriction on θ_{az}

$$\begin{aligned} &\lambda/4 = 2d_{CR}\sin\left(\theta_{az}\right) \\ &\lambda/4 = 2d_{CR}\theta_{az} \qquad \text{for } \theta_{az} \ll 1 \\ &\theta_{az} < \frac{\lambda}{8d_{CR}}. \end{aligned} \tag{5.64}$$

For example, this implies that a 20 m cross-range target illuminated at 10 GHz should be sampled in angle at a minimum of every 0.0107° in a zimuth.

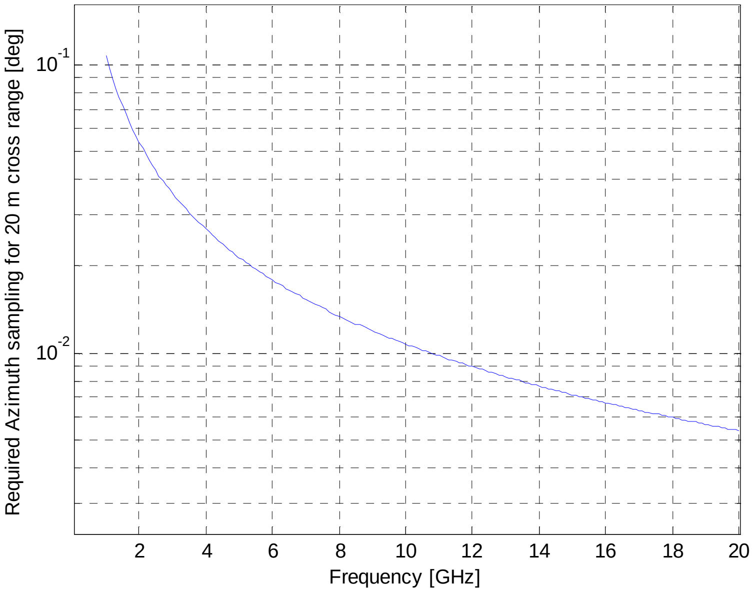


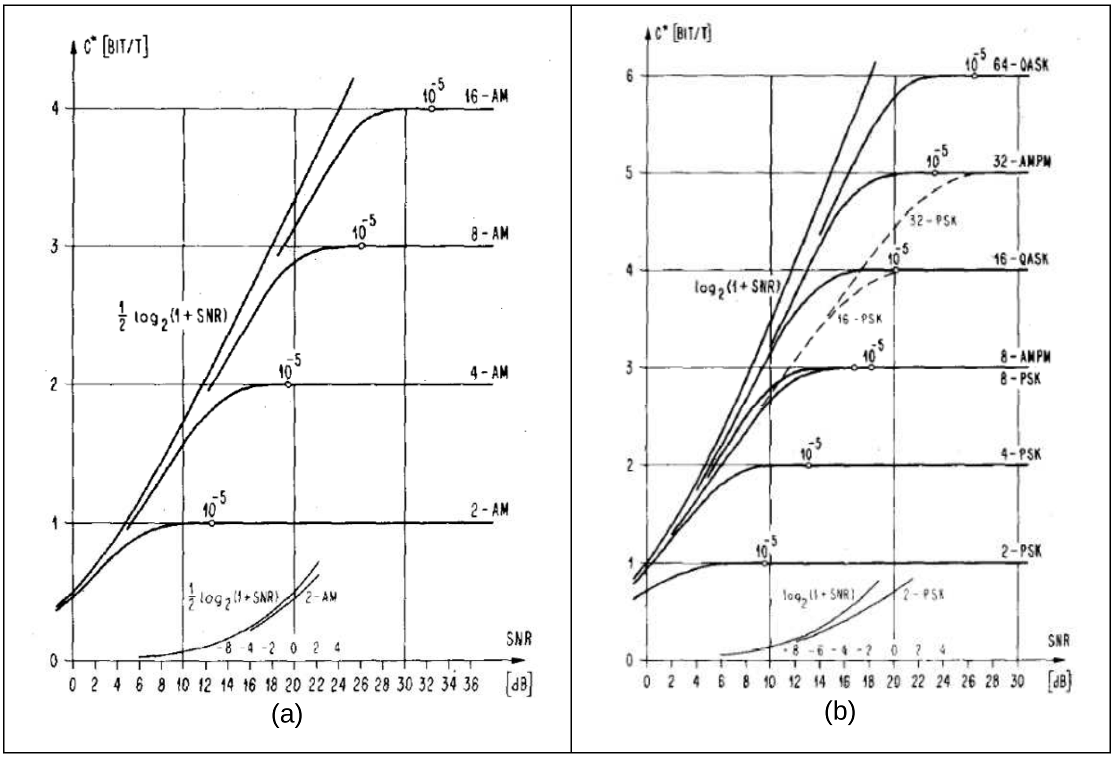
Figure 5.2: Required sampling interval as a function of frequency for a target having a cross range dimension of 20 m.

5.8 Validation

This section presents some validation results for the preceding sections. The aim is to compare the simulation results obtained with results from literature where possible.

5.8.1 Validation of Ungerboeck Derivation

The following figures show the results from Ungerboeck's original paper [Ungerboeck1982] (Figure 5.3) as well as simulation results for a subset of the same signalling schemes (Figure 5.4). It should be noted that Quadrature-Quadrature Phase Shift Keying (Q^2PSK) [Saha1989, Cilliers2002] has also been added although there are no published graphs for comparison. The Q^2PSK modulation format makes use of two orthogonal carriers as well as two orthogonal waveforms thus creating a 4-dimensional signalling space. The modulation format maps 4 bits onto a single symbol and makes use of the vertices of a hypercube as signalling points in the signal space. The technique has a theoretical spectral efficiency of 4 bits/s/Hz, which is double that of standard Quadrature Phase Shift Keying (QPSK).



(a) (b)

Figure 5.3: Channel Capacity for bandlimited AWGN channels for 1-D (a) and 2-D (b) modulation techniques, taken from [Ungerboeck1982]. (© IEEE 1982)

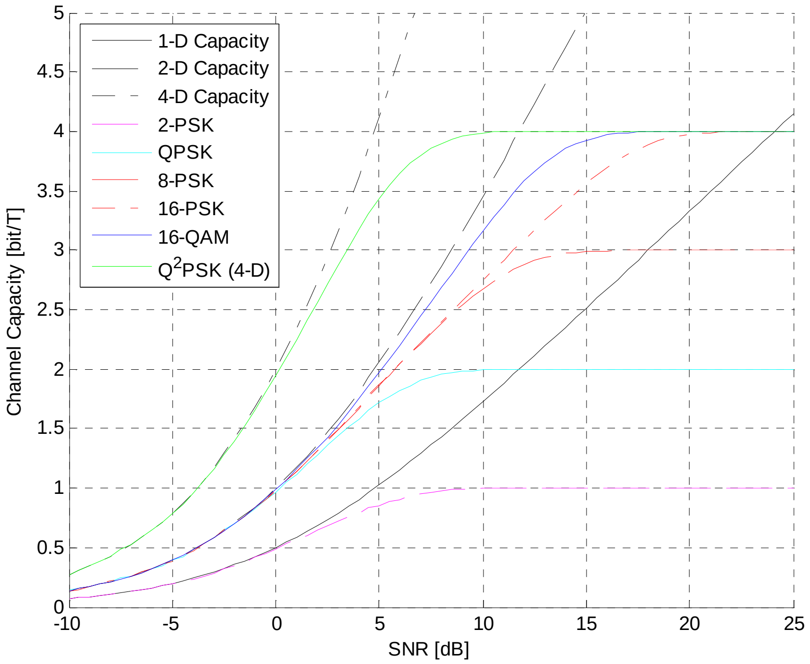


Figure 5.4: Numerical results using 10 000 samples for the E{} operation.

Another test was conducted to exercise the multidimensional MI derivation. For this test, 256 waveforms were generated by drawing samples from a Gaussian distribution. Each of the waveforms consisted of 12 samples. This was chosen because it conceptually agrees with the approach taken by Shannon to derive the expression for channel capacity. The curve should thus follow the 12-D capacity curve closely, but should flatten off to a constant value of $\log_2(256) = 8$ bits for high values of SNR.

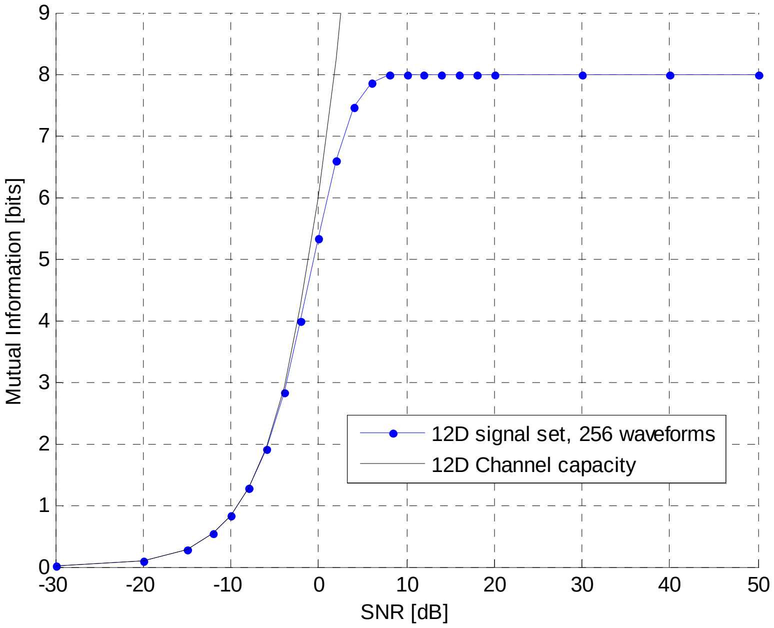


Figure 5.5: Mutual information for 256 waveforms, each 12D, with random samples drawn from a Gaussian distribution for each waveform.

The results in this section demonstrate the validity of the multi-dimensional mutual information calculation.

5.8.2 Validation of the Envelope Only Mutual Information Derivation

The envelope only version of MI is required to analyse non-coherent communication systems. This approach is used in some communication schemes such as frequency shift keying (FSK) when use is made of noncoherent detectors. The capacity of multiple frequency shift keyed (MFSK) was analysed by Butman et al. for noncoherent communication with NASAs planetary probes for planets with turbulent and dispersive atmospheres such as Venus, Jupiter and Saturn [Butman1973a, Butman1973a]. They present a graph of the capacity of various MFSK schemes relative to the coherent channel capacity [Butman1973a], which is repeated in Figure 5.6: . To compare the envelope only MI derivation with their results, manipulation of the SNR is required. Butman defines SNR as follows:

$$SNR = \frac{ST}{N_0} = \frac{\alpha^2}{2} \tag{5.65}$$

where S is the power in the received signal, T is the duration of a symbol and N_0 is the noise spectral density in W/Hz.

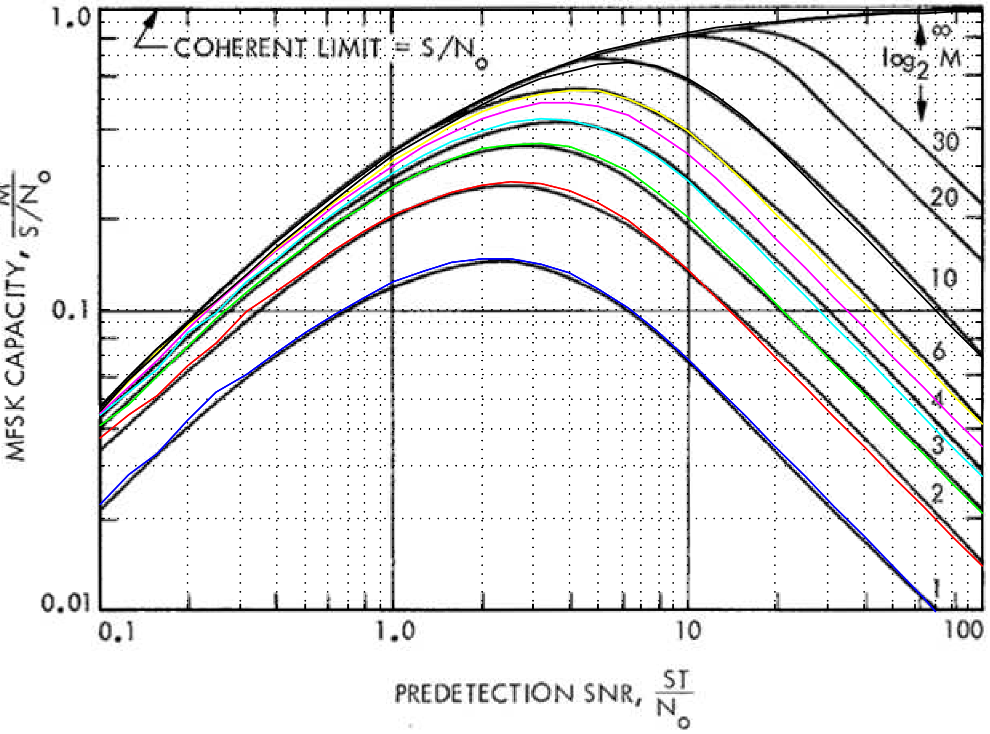


Figure 5.6: Capacity of various MFSK schemes overlaid with results for log2(M) = {1,2,3,4,5,6,10}. Background graph taken from [Butman1973a], p. 93, Fig. 3.

During the duration of a symbol and in a bandwidth of B Hz, a maximum number of

$$M = 2BT \tag{5.66}$$

orthogonal signals can be distinguished, thus allowing for the transmission of $\log_2(M)$ bits per symbol, thus $T = T_{bit}\log_2(M)$. Substituting into (5.65) gives

$$SNR = \frac{S\log_2(M)}{N_0} \tag{5.67}$$

If T_{bit} is set to unity. This equation can be solved for S/N_0 S/N0 which is required for the calculation of the coherent limit in

$$\frac{S}{N_0} = \frac{SNR}{\log_2(M)} \tag{5.68}$$

The wideband capacity of the coherent Gaussian channel is given by:

$$C_\infty = \frac{S}{N_0}\log_2(e) \qquad \text{bits/s} \tag{5.69}$$

where *e* is the base of the natural logarithm. The normalised wideband capacity for MFSK is given by Butman as:

$$\frac{C(\alpha)}{C_\infty} = \frac{2}{\alpha^2}\int_0^\infty x\exp\left(-\frac{x^2+\alpha^2}{2}\right)I_0(\alpha x)\ln\left(I_0(\alpha x)\right)dx - 1 \tag{5.70}$$

which has to be evaluated numerically. For the comparison in Figure 5.6: use was made of the following approximation given by Butman:

$$\frac{C(\alpha)}{C_\infty} = \frac{SNR}{2+SNR} \tag{5.71}$$

From Figure 5.6: it can be seen that the Monte-Carlo based curves match the graph from Butman very well. Some of the discrepancies are due to the Monte Carlo nature of the result, but the graph from the original article is also slightly warped. The deltas are however small enough to confirm that the derived expression and Monte-Carlo integration are producing the correct results.

5.9 Conclusion

This chapter started with the presentation of the expression for the mutual information for a channel with a discrete set of input signals but a continuous output space. Starting from this expression the expression for the MI for a channel with discrete vector inputs and continuous outputs was derived. This expression was validated against MI curves published by Ungerboeck [Ungerboeck1982] for various 2-dimensional signal sets.

Starting from the general MI expression again, the expressions for the MI when the receiver makes use of envelope detection and envelope squared detection were derived. These were validated against curves published by Butman [Butman1973].

Expressions were also derived for the case where the receiver is searching for a specific set of signals, but other interfering sets of signals can also be present. The principle was first illustrated using a 1-dimensional input space, and then extended to multidimensional input spaces. This allows the MI approach to be used in scenarios where multiple radar targets have to be recognised, and each can have a multidimensional response. It is thus referred to as the “Multidimensional multi-target case.”

The three main expressions are summarised as follows:

MI for single target, multi-dimensional input space:

$$I\left(\mathbf{a_k},\mathbf{y}\right)=\log_2\left(N\right)-\frac{1}{N}\sum_{k=0}^{N-1}E\left\{\log_2\left[\sum_{i=0}^{N-1}\exp\left(-\frac{\left|\mathbf{a_k}+\mathbf{z}-\mathbf{a_i}\right|^2-\left|\mathbf{z}\right|^2}{2\sigma^2}\right)\right]\right\}$$

MI for single target, multi-dimensional input space after envelope processing:

$$I\left(a_k,y\right)=\log_2\left(N\right)-\frac{1}{N\ln\left(2\right)}\sum_{k=0}^{N-1}\sum_{j=0}^{N-1}s_{kj}^2+\frac{1}{N}\sum_{k=0}^{N-1}E\left\{\log_2\left(\frac{\prod_{j=0}^{N-1}I_0'\left(\frac{r_{kj}s_{kj}}{\sigma^2}\right)}{\sum_{i=0}^{N-1}\prod_{j=0}^{N-1}\exp\left(\frac{s_{kj}^2+2r_{kj}\left(s_{ij}-s_{kj}\right)-s_{ij}^2}{2\sigma^2}\right)I_0'\left(\frac{r_{kj}s_{ij}}{\sigma^2}\right)}\right)\right\}$$

MI for Multi-target target, multi-dimensional input space:

$$\begin{aligned}I\left(\{\mathbf{a_k};\mathbf{b_{m,j}}\},\mathbf{y}\right)=&-P\left(a\right)\sum_{k=0}^{N-1}E\left\{\log_2\left[P\left(a\right)\sum_{i=0}^{N-1}\exp\left(\frac{-\left(\mathbf{a_k}+\mathbf{z}-\mathbf{a_i}\right)^2+\left(\mathbf{z}\right)^2}{2\sigma^2}\right)+P\left(b\right)\sum_{m=1}^{M}\sum_{j=0}^{N-1}\exp\left(\frac{-\left(\mathbf{a_k}+\mathbf{z}-\mathbf{b_{m,j}}\right)^2+\left(\mathbf{z}\right)^2}{2\sigma^2}\right)\right]\right\}\\&-P\left(b\right)\sum_{m=1}^{M}\sum_{j=0}^{N-1}E\left\{\log_2\left[P\left(a\right)\sum_{i=0}^{N-1}\exp\left(\frac{-\left(\mathbf{b_{m,j}}+\mathbf{z}-\mathbf{a_i}\right)^2+\left(\mathbf{z}\right)^2}{2\sigma^2}\right)+M.N.P\left(b\right)\right]\right\}\end{aligned}$$

In the next set of chapters these expressions will be used to analyse various radar target recognition problems.

6 Overview of Information Theoretic Analysis of Selected Radar Recognition Problems

This chapter gives an introduction to the results section of this thesis. The results will be presented in the groupings summarised in Table 2 below.

Table 2: Summary of experiments.

Exp #	Section	Short description	Targets
1	7.1.1	Effect of number of azimuth waveforms	Point targets
2	7.1.2	Effect of mainlobe shape	Point targets
3	7.1.3	Effect of range resolution	Point targets
4	7.1.4	Effect of target position relative to origin	Point targets
5	7.2	Effect of multiple targets	Point targets
6	7.3	HRR vs ISAR Processing	Point targets
7	8.1	Single target results	F-14, F-15, F-16
8	8.1.1	Effect of multiple targets	F-14, F-15, F-16
9	8.2	Effect of restricted azimuth sectors	F-14, F-15, F-16
10	8.3	High azimuth resolution analysis	F-15
11	9.1	Comparison of wideband and ultra wideband (8 GHz) waveforms	F-14, F-15, F-16
12	9.2	Comparison using 2-18 GHz waveforms	F-14, F-15, F-16
13	10.1	Gain in information content by using polarisation	F-18, F-35
14	10.2	Information content of measured Boeing 707 (1:25 scale)	Scaled B 707
15	10.3	Information content of monopulse sum and difference channels	F-14
16	11.1	Information content after envelope processing	F-14, F-15, F-16
17	11.2	Effect of pre-processing on information content after envelope processing	F-14, F-15, F-16
18	12.1	Multi-target information using single target and modified versions of the same target	F-14
19	12.2	Multi-target information and probability of error estimation.	F-14, F-15, F-16

These sections contain a selection of radar target recognition problems which have been analysed using the techniques developed in the previous section. All the problems in this section are aimed at the analysis of recognition performance based on High Range Resolution Profiles (HRRP) of the target. The problem has also been simplified by reducing the number of required aircraft orientations to the zero elevation plane in the target's coordinate system. This is equivalent to assuming that the aircraft is flying straight and level at relatively long ranges and low elevation angles. In all the experiments the assumption has been made that the target is in the far field of the radar's antenna, and is thus illuminated by a plane wave.

The initial set of experiments were conducted using point scatterer type models, with no occlusion of scatterers. The second set of experiments were conducted for data sets generated by an radar signature and RCS modelling tool, called SigmaHat [Smit2012a, Smit2012, Cilliers2014], that has been developed in the Radar group in the Defence, Peace, Safety and Security (DPSS) division of the CSIR. This software tool makes use of the high frequency approximations discussed in Section 2.5. and implements Physical Optics (PO) in combination with the shooting and bouncing ray (SBR) technique for multiple reflections. The latest version also includes edge scattering, but this functionality was not available when the simulations for this study were being executed. Use is also made of measured data of a 1:25 scale model of a Boeing 707, which was measured in the compact range at the University of Pretoria.

Most of the results which are presented are plots of mutual information (MI) versus SNR. The MI graphs are generated using Monte-Carlo integration and an internal accuracy monitoring algorithm has been implemented which guarantees the accuracy of the MI result per SNR value. Most were run at an accuracy of 0.01 bits or better, specifically as this value was small enough that error bars would not be necessary on the plots. Error bars would unnecessarily clutter already crowded plots. If the MI was run at a different accuracy to 0.01 bits, this is clearly stated in the experiment description.

7 Experiments using Point Scatterer Models

This chapter describes several experiments which made use of a point scatterer model of the target. This is a simplified model which assumes that a target can be represented as a collection of non-interacting scatterers which scatter an incident EM field omnidirectionally. This chapter also serves as a further validation of the MI estimation algorithms as the results are compared to the Shannon bound as this is a fundamental performance limit which is impossible to exceed.

7.1 Recognition of a Single Target - Effect of Range Resolution, Number of Azimuth Waveforms and Target Position

In this experiment a target was constructed of 6 point scatterers in the x-y plane. This setup does not make for a very realistic target, but it will allow some principles of the mutual information approach to be tested. The first 3 points were randomly distributed over a 15 m by 15 m square, but were reflected around the x-axis to create the second set of 3 points. This induces the left-right symmetry that most fixed wing aircraft exhibit. The amplitude of all the scattering points was unity and the phase zero. To generate the HRRPs, the scatterers were first projected onto the radar's line of sight and then the output waveform of the pulse compressor was created at the coordinate of each scatterer on the range line. Use was made of a triangular pulse compression waveform (i.e. the mainlobe was triangular) with zero sidelobes. The resulting HRRP's were calculated at a carrier frequency of 500 MHz (chosen to be low so that the highest bandwidth case is close to the limit of the narrowband approximation), and are plotted in a ring around the point scatterer target (scatterers denoted by red dots) in Figure 7.1 below. The dotted blue lines represent the boundaries between range bins. The range bins can be visualised by drawing tangents to the range bin circles perpendicular to the down range direction of the radar. The values on the colour bar to the right of Figure 7.1 represent voltage. In Figure 7.2 the resulting HRRP's are again plotted in a ring around the point scatterer target but with the amplitude scale in dB. Note the fast variation of the HRRP as a function of azimuth for angles in the region of 0 degrees and 180 degrees, where the cross range length of the target is at its maximum. The variation in the upper and lower region of the HRRP ring is much slower due the reduced cross range projection of the target. Figure 7.3 was included to demonstrate the effect of decreasing the range resolution by a factor of two, while maintaining the original sampling rate.

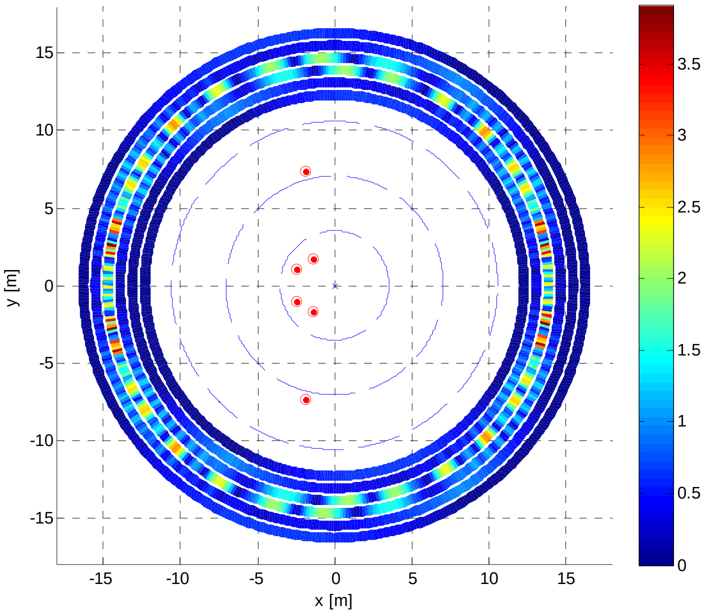


Figure 7.1: Scatterers and ring plot of HRRPs for target #1 for a range resolution of 3.536 m.

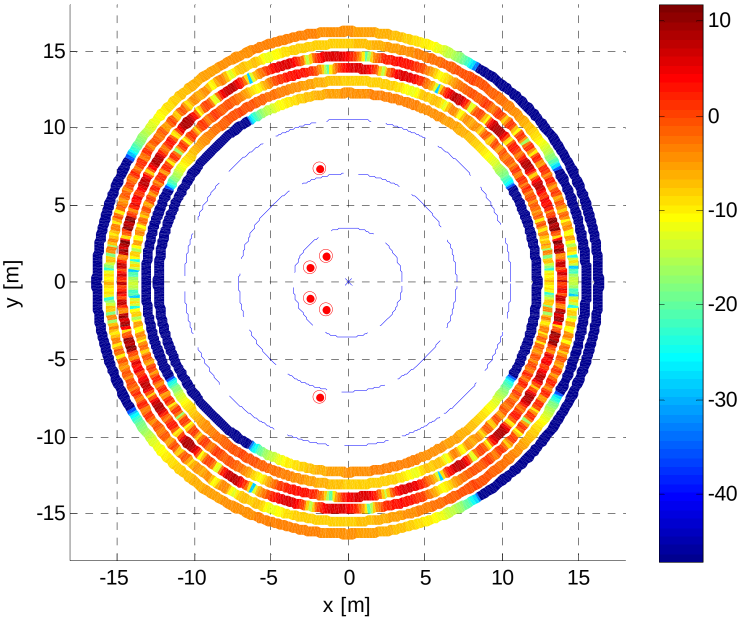


Figure 7.2: Scatterers and ring plot of HRRPs (in dB) for target #1 for a range resolution of 3.536 m.

To demonstrate the use of mutual information several experiments were conducted to evaluate the effect of the number of azimuth waveforms, the type of pulse compression mainlobe and range resolution on the amount of information transferred about the target to the radar. For the experiments in this section the MI estimation algorithm was setup to estimate the MI to within a standard deviation of below 0.01 bits.

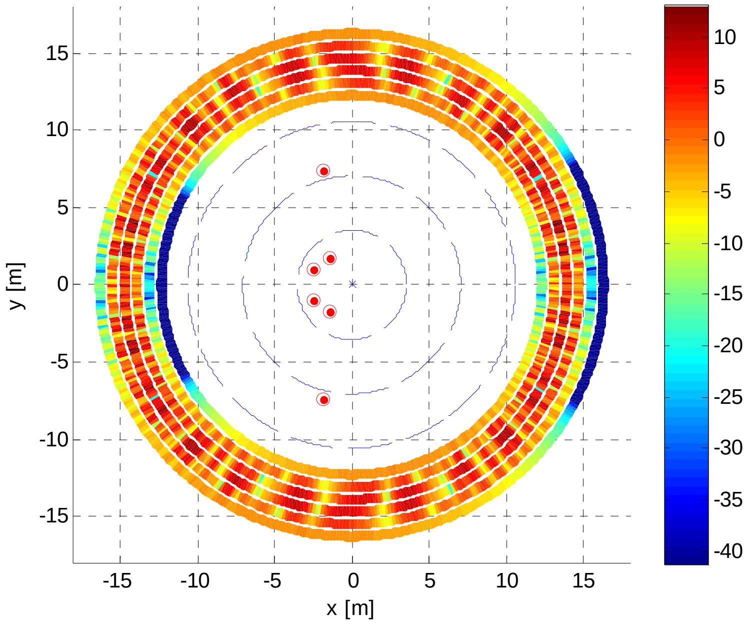


Figure 7.3: Scatterers and ring plot of HRRPs (in dB) for target #1 for a range resolution of 7.072 m.

It should be noted that the MI calculation assumes that the radar can accurately estimate the phase and magnitude of the target, and that the radar has a bank of filters matched to each of the target HRRP's. This observation leads to the insight that the absolute RCS of the target conveys information to the radar, so the HRRP profiles should not be normalized to unit energy as is often the case in the pre-processing steps to a NCTR algorithm. The MI graphs thus indicate the best achievable performance for such a radar.

7.1.1 Experiment #1: Effect of the Number of Waveforms in Azimuth

Using Target #1, sets of HRRP's were generated with increasing numbers of waveforms in the azimuth dimension. The numbers of waveforms were chosen to be powers of two so that the final value of the MI at high SNR corresponds to an integer number of bits. This also allows comparison of the MI results with multi-dimensions digital communication signalling strategies. The results are summarized in the figure below for a range resolution of 4 m, and 8 complex range bins spanning the target. The starting point was chosen as 512 waveforms, equally spaced over 360°, which was compared to 256 equally spaced waveforms over 180°. These two setups should give the same amount of mutual information, 8 bits, due to the exact symmetry of the target. This was true for the 256 waveform case, which gives a final MI value of 8 bits, to within the floating point precision of the computer. The 512 waveform case, however, gave a final MI value of 8.0039 bits. The discrepancy was shown to be a slight mismatch between the waveforms on either side of the target's axis of symmetry, which was caused machine precision effects during the generation of the HRRP's. From this simulation onwards, only one half of the HRRP's were used, corresponding to the set of waveforms contained in 180°. The figure also includes the Shannon channel capacity bounds for 8 and 16 dimensional signal sets, as well as two optimal digital communication signalling strategies. The most optimal waveform to transmit data in a power constrained scenario is by making use of Gaussian noise type waveforms. This is however not practical due to limitations of power amplifiers, so communication engineers make use of multi-dimensional signal constellations, the most effective of which is antipodal signalling (transmitting a -1 or +1 per signal dimension). The performance of both the abovementioned strategies is included in the graph for comparison.

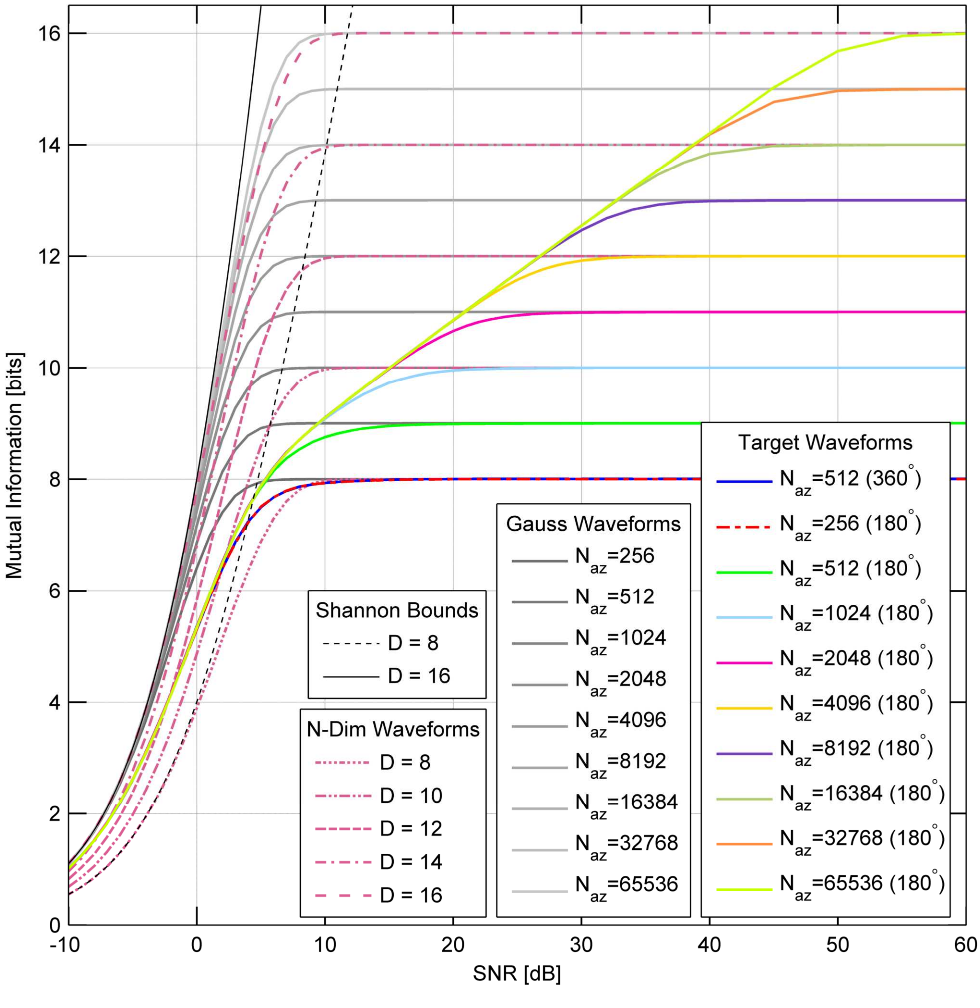


Figure 7.4: Mutual information versus SNR for various experiments for a 6 point scatterer model having symmetry about the y-axis. The Shannon bound as well as two optimal signalling strategies are included for comparison. The transmit frequency was 500 MHz, and the range resolution for the 8 bin case was 4.0 m.

All of the cases where the number of azimuth waveforms is more than 1024 waveforms, the MI graphs have a "knee" in the region between SNRs of 4 and 9 dB. This reduction in the amount of information gain is due to increasing correlation between adjacent waveforms as the angular sampling rate is increased. The Euclidean distance between the waveforms reduces, which leads to an increase in the SNR at which two adjacent waveforms become distinguishable from one another. For a specific number of azimuth waveforms, there is a gain of approximately 5 dB at the point where MI for the current set of waveforms flattens out, but the waveform sets with more azimuth samples continue to increase.

It is also interesting to note that all the graphs for 512 azimuth samples and above, pass through the 8 bit level at the same point. This point is at a SNR which is approximately 10 dB lower than the SNR at which the 256 azimuth sample waveforms reach the 8 bit level. This implies that there is a 10 dB performance gain which can be realised by having 512 or more matched filters in the radar to identify this target.

7.1.2 Experiment #2: Effect of Mainlobe Shape

To evaluate the effect of the mainlobe shape (in the range dimension) of the waveform on the information transferred from the target to the radar, simulations were conducted for 256 and 512 waveforms spanning 180°. Three types of mainlobe were compared for the two waveform sets: an impulse function, a triangular function and a sinc() function. From the results presented in the Figure 7.5 below, the performance of the three approaches is relatively similar.

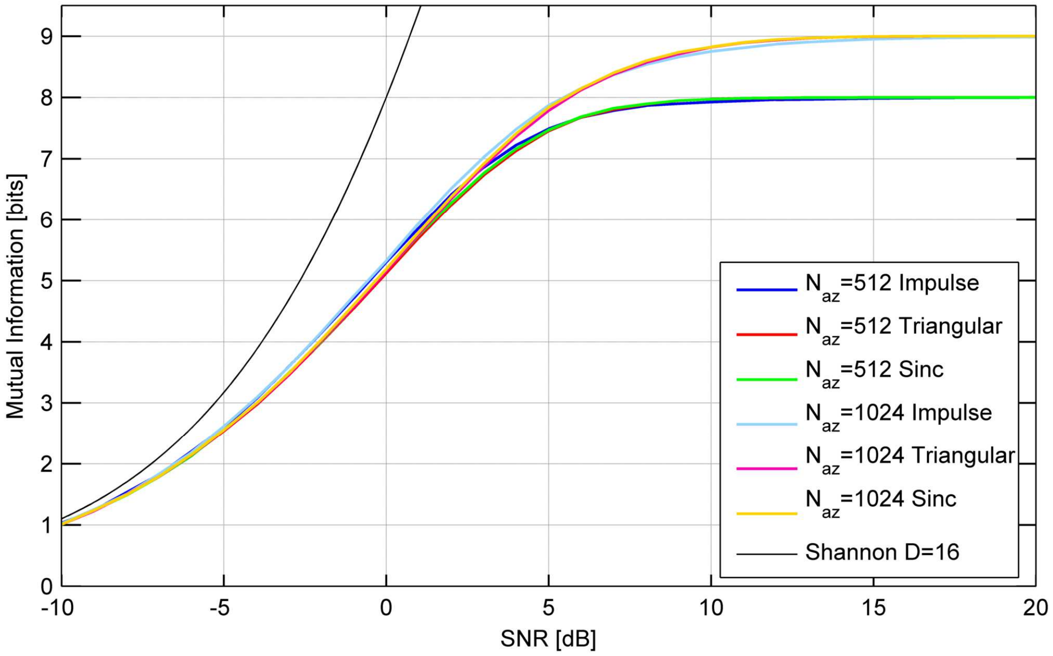


Figure 7.5: Mutual information versus SNR for various mainlobe functions in the range domain. The transmit frequency was 500 MHz, and the range resolution for the 8 bin case was 4.0 m.

The Impulse function outperforms the other two types by 0.2 dB in the region surrounding an SNR of 0 dB. When the MI starts reaching its final value, the situation reverses, and the other two types of mainlobe outperform the impulse type by approximately 2 dB. The impulse function is not realisable in a radar system, and was only used here as it should have the least detrimental effect on the information transfer process.

7.1.3 Experiment #3: Effect of Range Resolution

To investigate the effect of range resolution, experiments were setup with varying mainlobe widths in the time domain. An example of the result is shown in Figure 7.6 below for 1024 azimuth waveforms in 180°. The highest resolution was limited to 0.2 m as this requires a bandwidth of 750 MHz, which is probably technologically stretching for a centre frequency of 500 MHz. For this setup, the sampling rate was set to take a sample every 0.2 m, which resulted in oversampling of all the waveforms for range resolutions coarser than 0.2 m. The oversampling gain has translated the set of curves towards lower SNR values. The Curve for the Shannon bound was calculated for a 300 dimensional space due to the fact that there are 150 complex range bins for this setup, regardless of the range resolution.

From the result in Figure 7.6 it can be seen that finer range resolution increases the MI, and consequently the recognition performance of the radar. The MI curves for the 0.2 m and 0.25 m resolution cases are nearly indistinguishable. This means that 20% improvement in the range resolution has not added any extra information about the target.

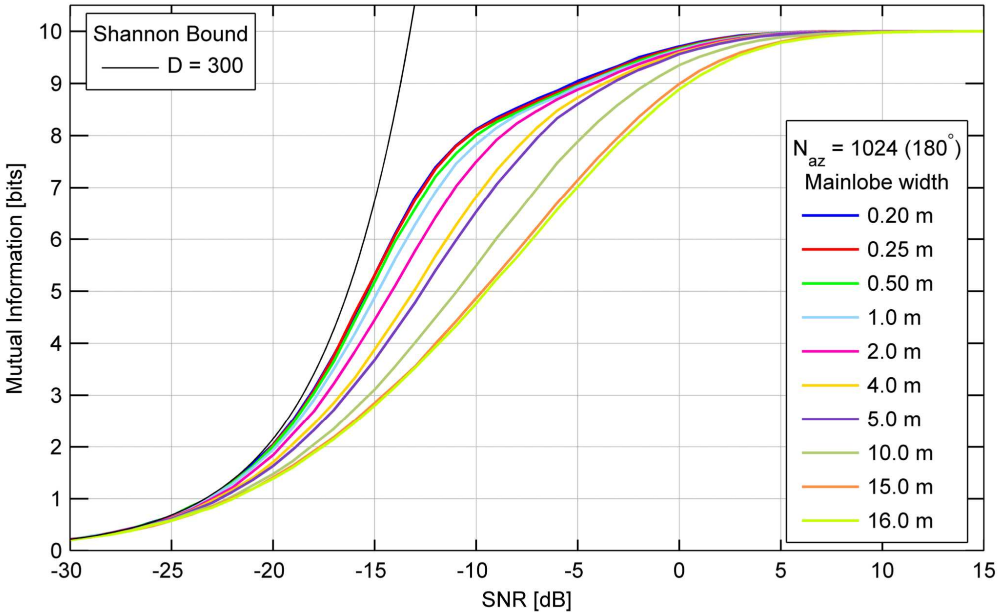


Figure 7.6: Comparison of MI for various values of mainlobe width in the time domain for 1024 waveforms spanning 180°.

It is also interesting to note that as the range resolution is halved, the amount of gain reduces as is shown in Figure 7.7. This figure shows that there are gains of up to 1.4 dB for increasing the resolution from 4 m to 2 m, but when a resolution of 0.5 m is reached, there is only a gain of 0.4 dB to be gained by increasing the resolution to 0.25 m. This shows that there is an upper limit to the amount of information which can be extracted by making use of finer range resolutions.

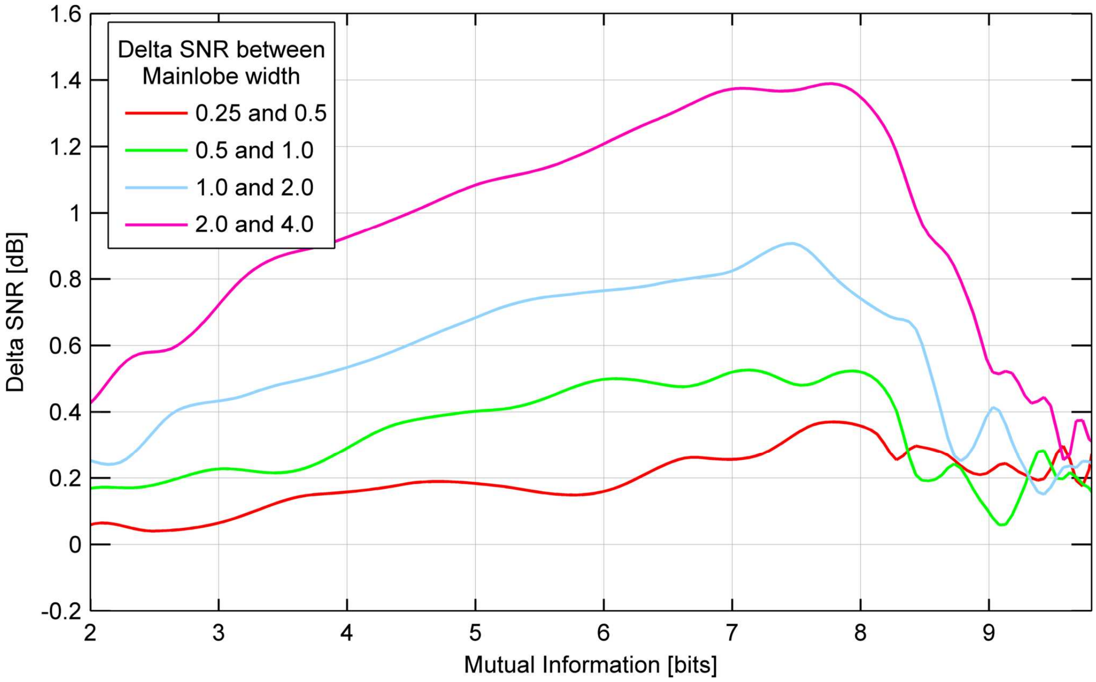


Figure 7.7: Comparison of gain in SNR versus mutual information for halving of range resolution.

To further investigate the effect of range resolution and higher numbers of waveforms in azimuth some more traces were added to the result above. The numbers of waveforms were extended to 8192 and 32768 in 360°, and lower resolutions of 50 m, 100 m, and 150 m were simulated. The result is plotted in Figure 7.8 below, where the simulation were only run over 180° in azimuth due to the target's symmetry around the y-axis. The 100 m result was found to be only slightly better than the 150 m result, so the 100 m result was excluded from the plot for clarity. Some of the lower resolutions were also excluded to make the plot less cluttered. The y-axis was also limited at an MI value of 6 bits for clarity. The traces enter the lower part of the plot in 5 discrete groups, each corresponding to one of the 5 resolutions. The traces limit at 3 distinct values, which correspond to the number of azimuth waveforms. The 150 m resolution is low enough that it starts to approximate the case where the HRRP bins are all the same value.

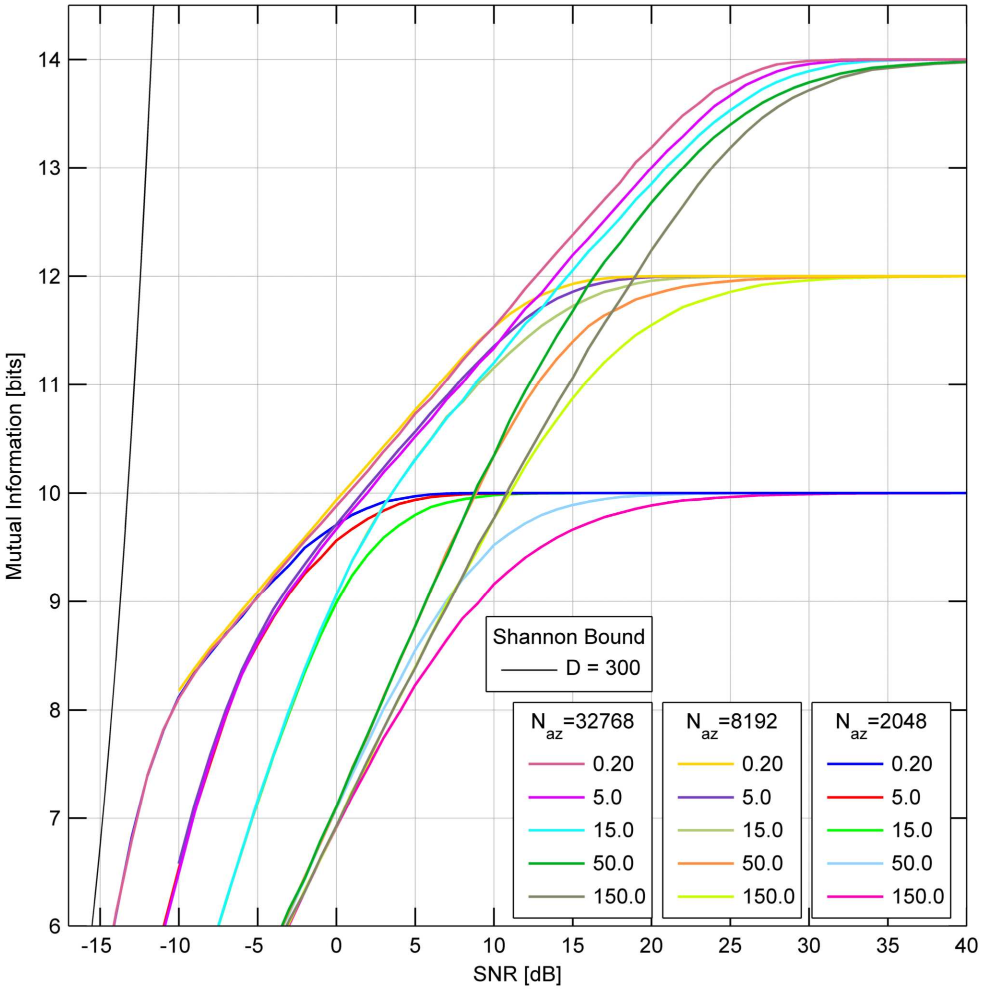


Figure 7.8: Comparison of the mutual information versus SNR for various combinations of range resolution and the number of waveforms in azimuth for the 6 point scatterer model.

An example of one of the higher resolution HRRP's is shown as a ring plot, in the figure below.

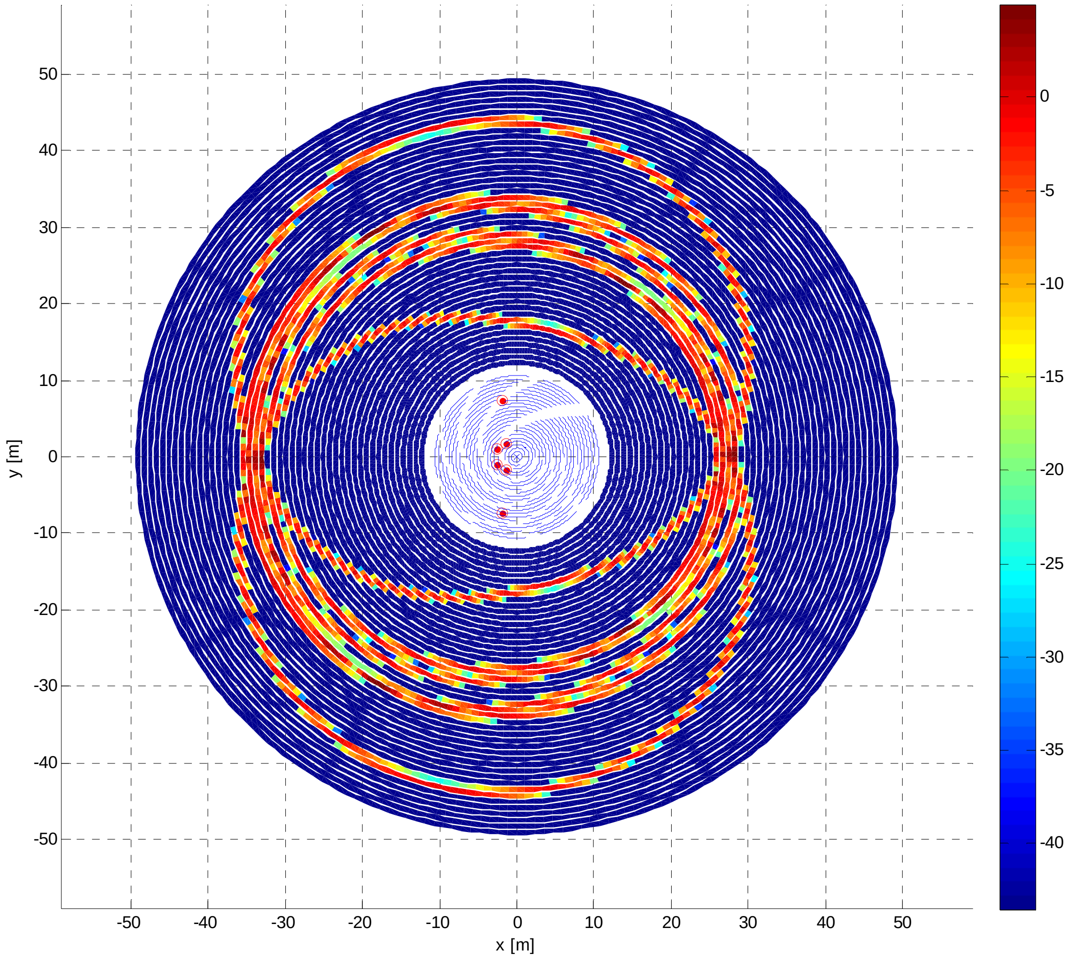


Figure 7.9: Scatters and and ring plot of HRRPs (in dB) for target #1 at a range resolution of 0.451 m (332.3 MHz).

7.1.4 Experiment #4: Effect of Target Position Relative to Origin

Given the assumptions inherent in the MI calculation, it was deemed necessary to run an analysis to ascertain the sensitivity of the MI calculation to the position of the target's centre of rotation. As an example, the target from the preceding analysis was used, but the scatterers were moved by subtraction of the mean of the positions of the scatterers. The resulting HRRP plot is given in the figure below. Use was made of the 1024 waveform data set.

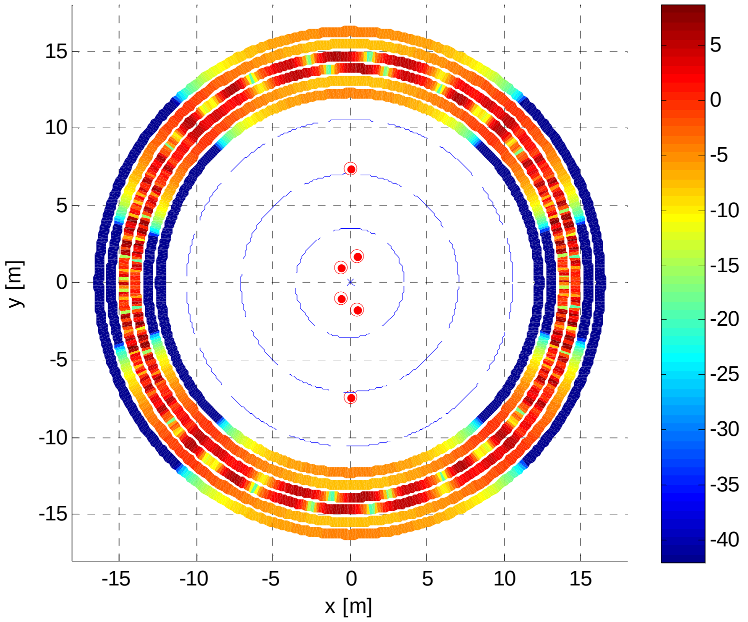


Figure 7.10: Scatters and ring plot of HRRPs (in dB) for target #1, which has been centred, at a range resolution of 3.536 m.

It is evident that the structure of the HRRP's is more symmetrical than the original version. The comparative plot of the MI before and after the translation is given below.

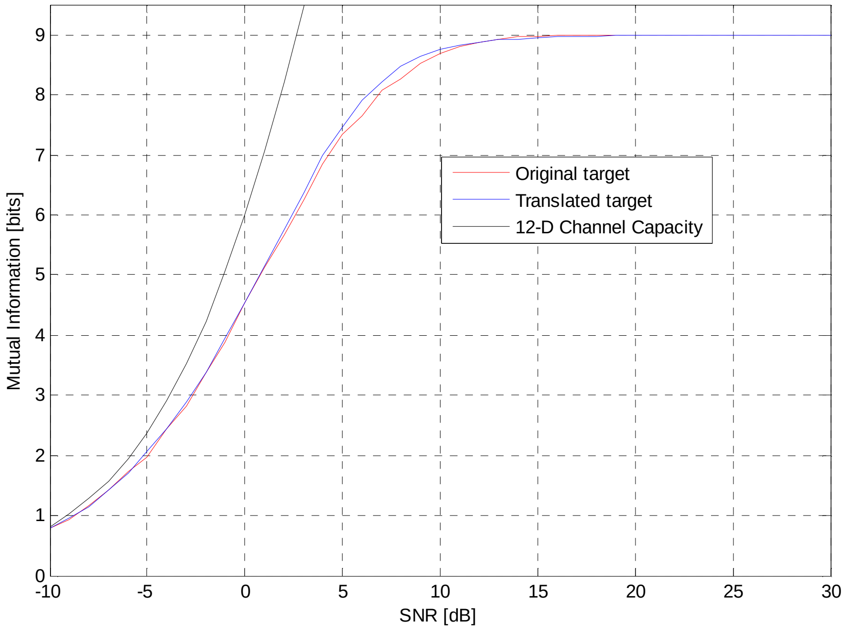


Figure 7.11: Comparative MI plot versus SNR for the translated target.

The difference between the two traces on this plot is within the tolerance of the MI algorithm, which was set to approximately 0.2 bits for this test.

7.2 Experiment #5: Multiple Targets

In this experiment 16 random targets were generated each consisting of 6 scatterers of unit magnitude and zero phase. The scatterers were uniformly distributed over a 15 m by 15 m area and each target was generated with 512 waveforms over 360 degrees in azimuth. The comparative graph of the mutual information is given below.

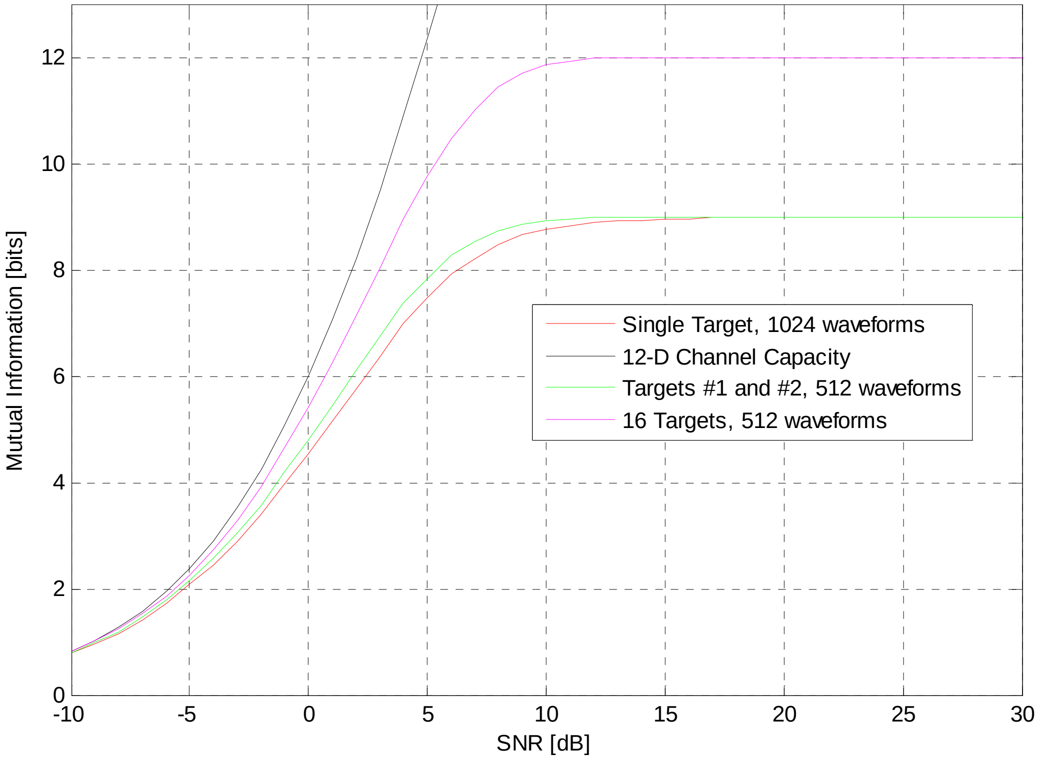


Figure 7.12: MI graph versus SNR for 16 targets, each represented by 512 waveforms.

Due to the symmetry of the targets, each target only has 256 unique waveforms, which equates to a maximum of 8 bits of information. When the waveforms from 16 of these targets are analysed together there should be a 4 bit growth in the amount of information. This is evident from the trace for the 16 target case which reaches a limiting value of 12 bits. It is interesting to note that all the targets can be separated when the SNR reaches approximately 15 dB. From the two curves for a single target and two targets, it seems that two targets contain slightly more information than a single target with double the number of waveforms. This effect is probably due to the higher level of correlation that would be expected in a single target as a function of angle as compared to two targets which have independent generation processes. This section eventually led to the multi-target mutual information approach which was derived in the theoretical section.

7.3 Experiment #6: HRR Versus ISAR Processing for Recognition

In this section a comparison between the information content of HRRP and ISAR is made. The data processing inequality predicts that the two techniques should have the same performance as long as the transformation used to create the ISAR image is invertible. To this end use was only made of an FFT to create the ISAR image and no windowing was applied. A set of 512 HRRP's covering 180° was used, with 16 HRRP's per ISAR image and 32 ISAR images over the 180° sector. The setup was chosen in such a way that the cross-range resolution of the ISAR image was equal to or better than the Range resolution of the HRRP's. For this experiment, the cross range resolution was calculated as:

$$R_{res\_c} = \frac{\lambda}{2\Delta\theta} = \frac{0.6}{2\times\left(\frac{16}{512}\pi\right)} = 3.06 \text{ m}, \tag{7.1}$$

which is approximately 13% better than the 3.5 m range resolution of the HRRP's. The resultant graph for the MI is shown in the figure below.

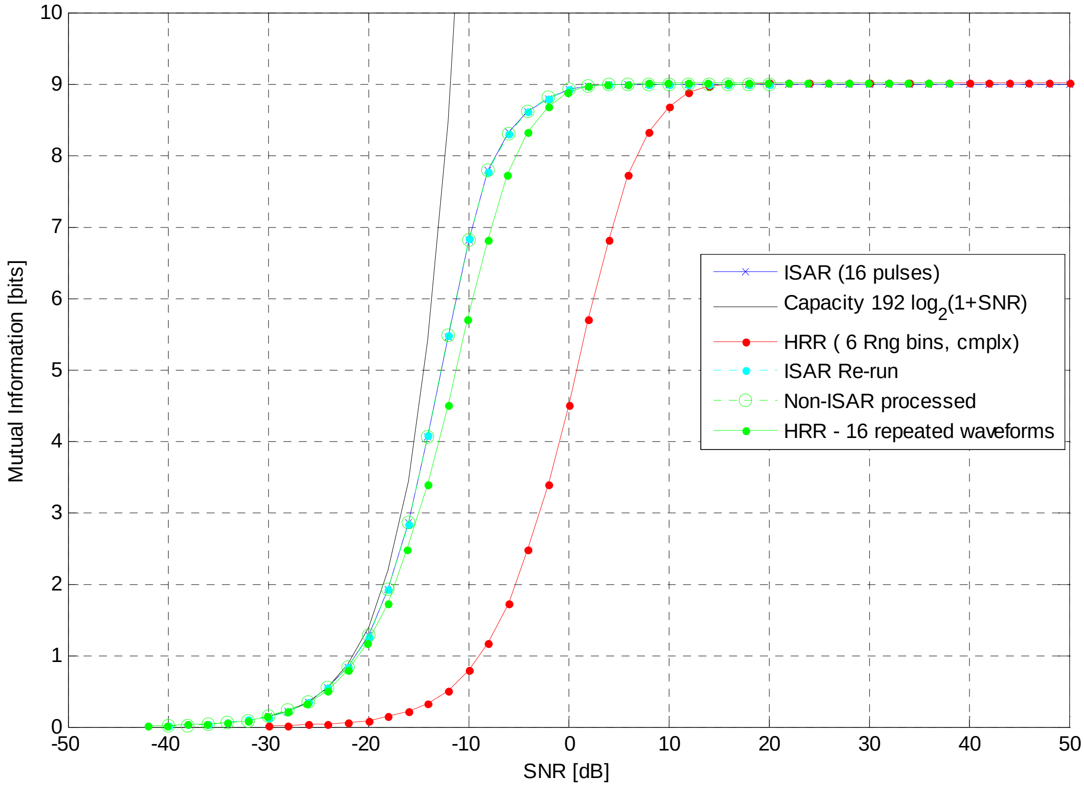


Figure 7.13: Comparison of MI versus SNR for HRR and ISAR processing.

The red trace labelled "HRR (6 Rng bins, cmplx)" is the MI result for a single HRRP. This is used as a benchmark to show the improvement obtained by processing more pulses. In this result a single HRRP is repeated 16 times to make its integration gain comparable with that of the 16 pulse ISAR processing. The ISAR image contains slightly more information than the repeated HRRP due to the fact that it makes use of 16 adjacent HRRP's in the data set to simulate the effect of platform rotation. If these same 16 pulses are used for HRR processing, effectively by creating a 16 times longer matched filter for the target, the same performance as that of the ISAR will be achieved. The difference is that the HRRP recognition can be made independent of platform rotation, whereas the ISAR requires constant angular velocity to form an image. Two independent simulations were run for the ISAR case, the second is labelled “ISAR Re-run”.

The MI curve for the ISAR and 16 consecutive HRRP case has moved much closer to the channel capacity curve. The increase of approximately 12 dB corresponds to the coherent integration gain over 16 pulses.

7.4 Conclusion

This section has demonstrated the insight which is achievable by applying MI to relatively simple point scatterer models. The section also served as a further verification of the MI calculation algorithm as most results are compared to their respective Shannon bounds.

Firstly, all the point target models used in this section exhibited symmetry around y-axis so only half the HRRP's were unique. The maximum value of the MI was thus one bit lower than the number of azimuth waveforms. The effect of the position of target relative to the origin was also shown to have a very small effect, if any.

Figure 7.4 showed that if each HRRP was chosen to consist of range bins with values chosen from a Gaussian distribution this resulted in the highest information transfer from the target to the radar. The Gaussian waveforms were within 5-6 dB of Shannon bound and N-dimensional waveforms, which make use of the vertices of a hypercube to transmit a -1 or +1 per dimension, were about 1 dB worse than the Gaussian waveforms for the highest dimensional problem (16 Dimensions). The HRRP waveforms showed much lower performance as well as a distinctive “knee” where the rate at which the MI was increasing slowed due oversampling in the azimuth domain. This results in a reduction in the Euclidean distance between waveforms as oversampling increases. A loss of 5 dB relative to Gaussian for 256 waveforms, and 45 dB for 65536 azimuth waveforms was observed.

Three mainlobe shapes (impulse, triangular and sinc) were tested and showed very similar results. It was expected that the impulse would have the best performance, but this is only the case at low to intermediate MI values. At high MI the more realistic mainlobe functions have slightly better performance.

A detailed analysis of the effect of range resolution showed that there is definitely a trend of diminishing returns by going to higher resolutions than 0.5 m. On the low side there is not a significant fall off in performance when reducing the resolution below 15 m. For increasing range resolution, the increase in the intermediate MI range is quite pronounced, and approximately an 8 dB SNR gain between 0.2 m and 16 m was observed, but this gain narrows down to approximately 5 dB at high MI values. For the examples analysed, finer range resolution always lead to improved performance, but increasing the number azimuth waveforms for the low resolution case decreases the resolution induced performance loss at high MI values. Interestingly, it was shown that the full MI can be extracted, even for resolutions as low as 150m, although the extra SNR required for this is as high as 25 dB. This is due to the fact that the input signal vectors for the MI calculation are complex valued, and the range to the centre point of the target is constant, which implies that the signal space is a highly oversampled set of samples which are each just an in-phase (I) and quadrature-phase (Q) sample. The MI graph thus represents the case where the range to the target is known by the radar to within a very small fraction of a wavelength. The radar can thus exploit the amplitude and phase information of a single sample to recognize the target.

There is definitely a loss in SNR (i.e. more SNR required) as the number of waveforms in azimuth is increased. This is due to the extra SNR required to discriminate between adjacent waveforms which become more and more similar as the number of azimuth samples is increased.

The performance of ISAR processing was shown to result in exactly the same amount of MI as just processing the same HRRPs that were used to form the ISAR image. This is expected as the data processing inequality states that information can only remain the same or be reduced due to processing. Because the ISAR process is FFT based, it is a reversible process, so the MI for both cases should be equal.

8 Information Content of an F-14, F-15 and F-16 using High Range Resolution Profiles

In this section the information content of more realistic targets are compared. The HRRP's of the three aircraft, the F-14, F-15 and F-16 are shown in Figure 8.1, Figure 8.2 and Figure 8.3 respectively.

Models of these aircraft, which were developed for flight simulators, were downloaded from the internet. These models were then analysed using a high frequency radar cross section code called SigmaHat to obtain their frequency responses at a set of discrete frequencies. This data could then be converted to HRRP's if required.

The outer shell of the aircraft was modelled as a perfectly electrically conducting (PEC) material. This means that effects caused by the canopy, radomes and radar absorbing material (RAM) are modelled incorrectly due to a lack of information pertaining to their electrical properties in these models. The version of SigmaHat used implemented Physical Optics (PO) and shooting and bouncing rays (SBR), but not edge diffraction.

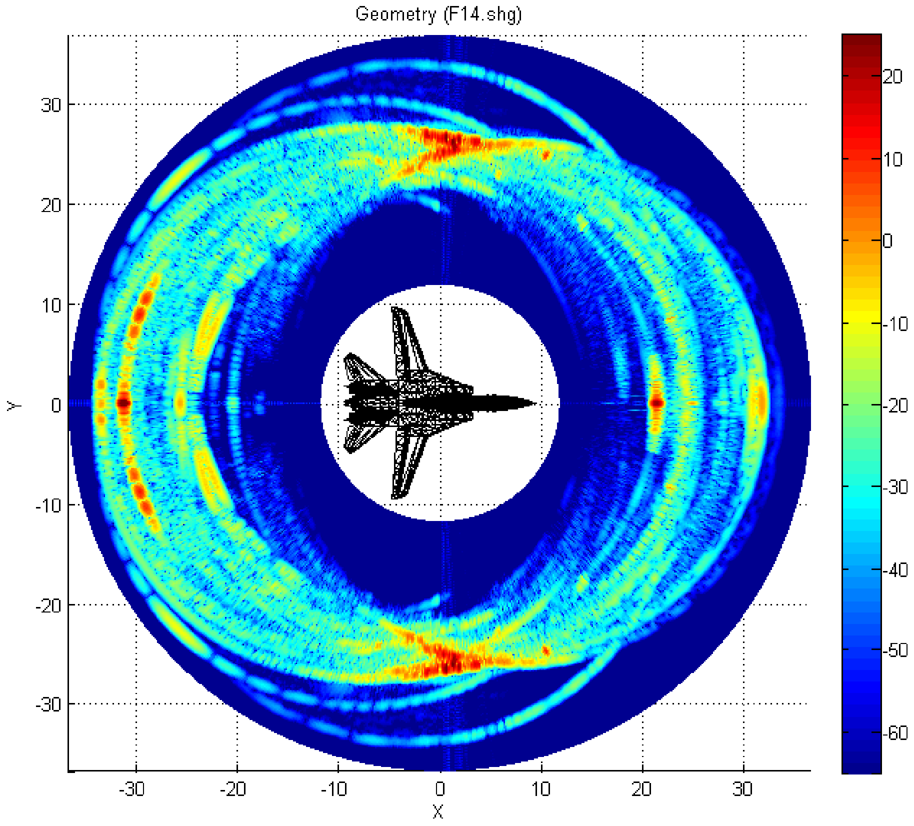


Figure 8.1: F-14 HRRP ring plot.

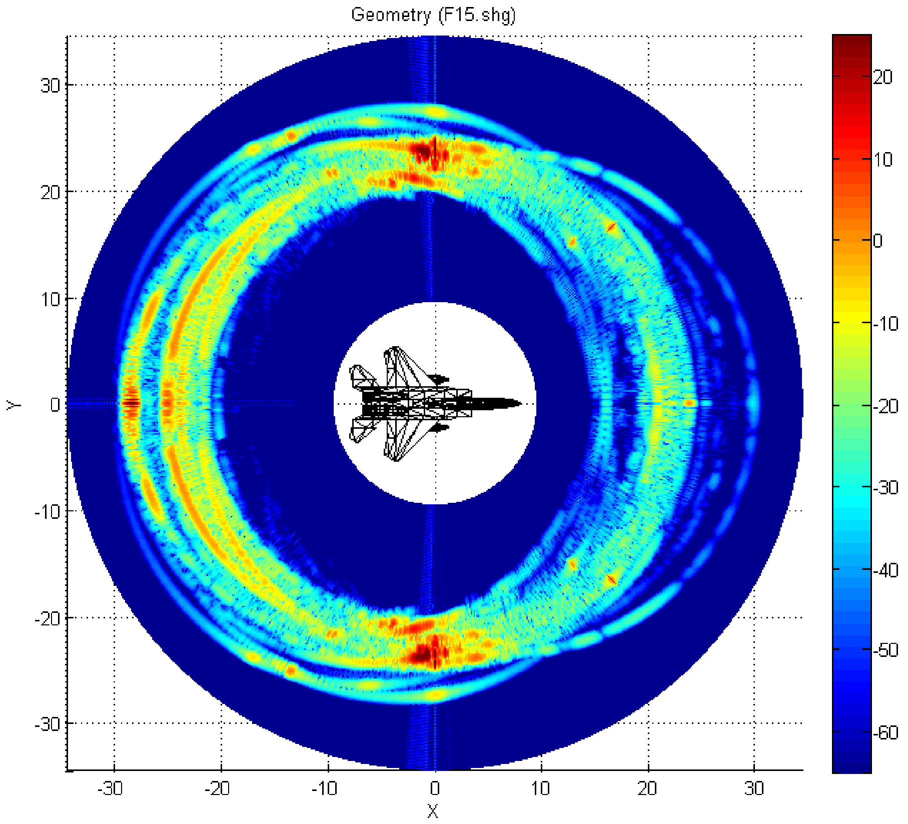


Figure 8.2: F-15 HRRP ring plot.

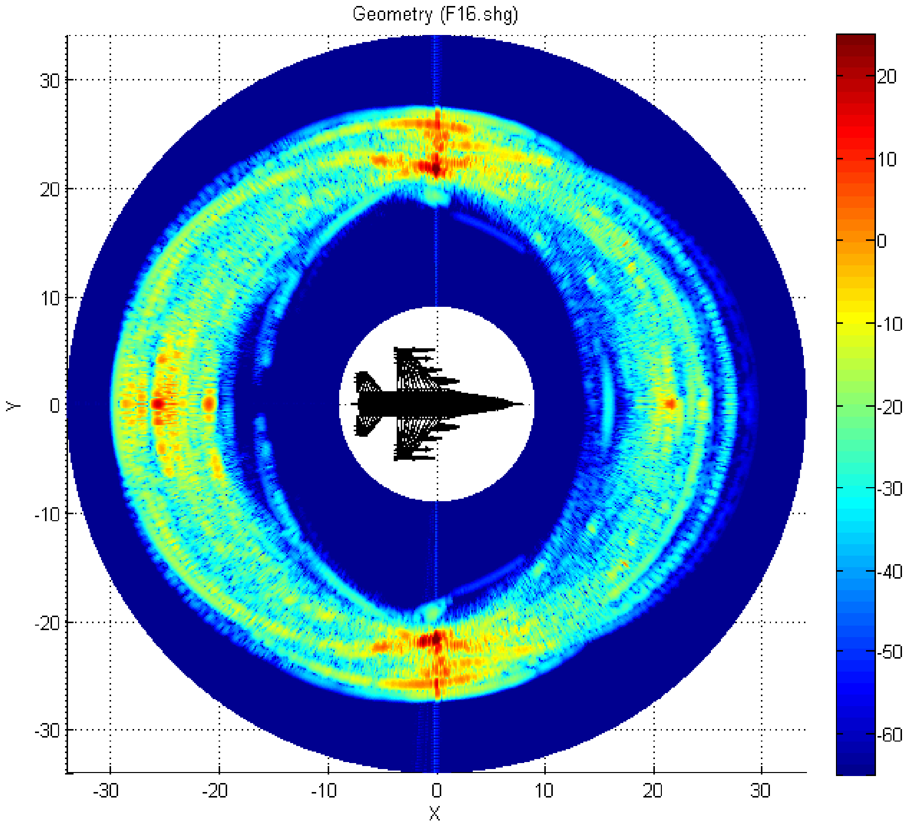


Figure 8.3: F-16 HRRP ring plot.

The targets' signatures were calculated for vertical polarization on transmit and receive and each target was illuminated at a set of 81 frequencies centred at 10 GHz in steps of 6 MHz (i.e. 9.76 GHz to 10.24). This gives a total bandwidth of 480 MHz, which in turn gives a range resolution of 0.3125 m. Each target was simulated at 1024 azimuth angles at an elevation angle of 0°. The EM software made use of physical optics (PO) and shooting and bouncing rays (SBR) with up to five reflections.

8.1 Experiment #7: Single Target Results

The initial mutual information graph for the three targets is shown in Figure 8.4 below. In this graph a trace has been added for the MI of a set of waveforms of the same dimension as the simulated target data, but with each waveform drawn from a Gaussian distribution.

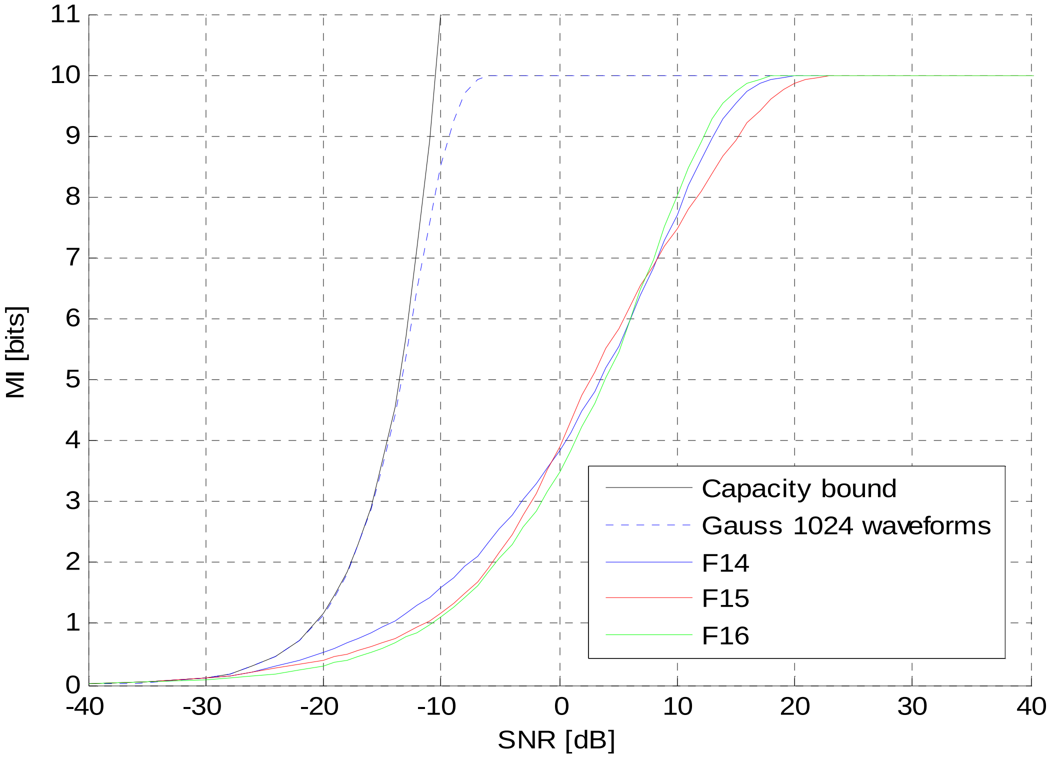


Figure 8.4: Comparison of the MI versus SNR for the F-14, F-15 and F-16 models.

The next figure (Figure 8.5) is a zoomed view of the same graph as the MI reaches its maximum value. The SNR's at which each of the traces reaches 99.95% of its maximum value are indicated by means of labels. This value of MI can be calculated, using Fano's inequality, to correspond to a probability of error of $P_e = 2.1157 \times 10^{-4}$ or approximately one in 4700. The values for all five traces are tabulated below, where the SNR has been rounded to one decimal place, and the traces have been ordered by increasing SNR. Each of these simulations was run relative to the average power in a single set of waveforms for each specific target.

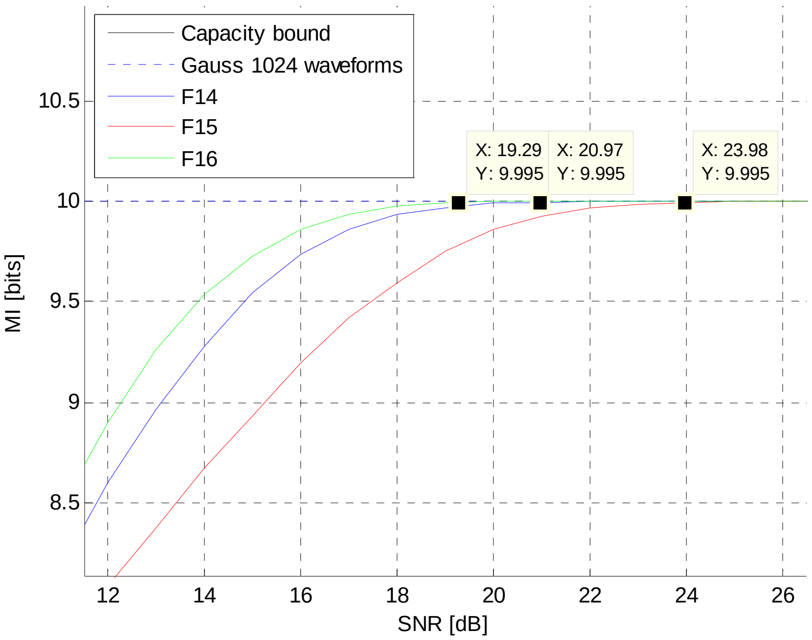


Figure 8.5: Zoomed view showing detail of the MI graph as it reaches its maximum value.

From the table below it can be seen that the Gaussian set of waveforms only perform approximately 5 dB worse than what the Shannon bound predicts. The three fighter aircraft however are between 25 dB and 30 dB worse than the Shannon bound. This gives a good indication of how suboptimal the waveforms which are reflected from the targets are as compared to communication signals.

Table 3: Comparison of SNR as MI reaches 99.95% of its maximum value.

	Capacity Bound	Gauss 1024 waveforms	F-16	F-14	F-15
MI [bits]	9.995	9.995	9.995	9.995	9.995
SNR [dB]	-10.5	-5.2	19.3	21.0	24.0

The F-16 reaches the maximum MI value at the lowest SNR of the three. This is probably due to the fact that the model has numerous under-wing stores loaded, which introduce extra scattering compared to the F-15 which only has drop-tanks, and the F-14 which has no external stores. The CAD model for the F-15 is not as detailed as the other two models, which might explain why it is the most difficult to classify.

8.1.1 Experiment #8: Effect of Multiple Targets

To further investigate the information transfer for these targets, all three sets of HRRP's were concatenated and the MI recalculated. To concatenate the target responses the three original matrices containing 1024 signatures of 81 samples each (81x1024 matrix) were concatenated to form a single matrix containing 3072 signatures of 81 samples each (81x3072 matrix). The MI result is shown in Figure 8.6 below. This would imply a situation where all three targets are equally likely, and the radar system is setup up to try to recognize all three targets. When all three targets are evaluated simultaneously, the average power used in the MI calculation is the mean power calculated over all three targets. To make the single target curves comparable with this result, the single target curves were compensated to be relative to the average power of all three targets. This has the effect of penalising targets with lower reflectivity. The maximum MI for all three targets is given by $\log_2(3\times1024)\approx11.584\,963$ bits, which should be reached for high values of SNR.

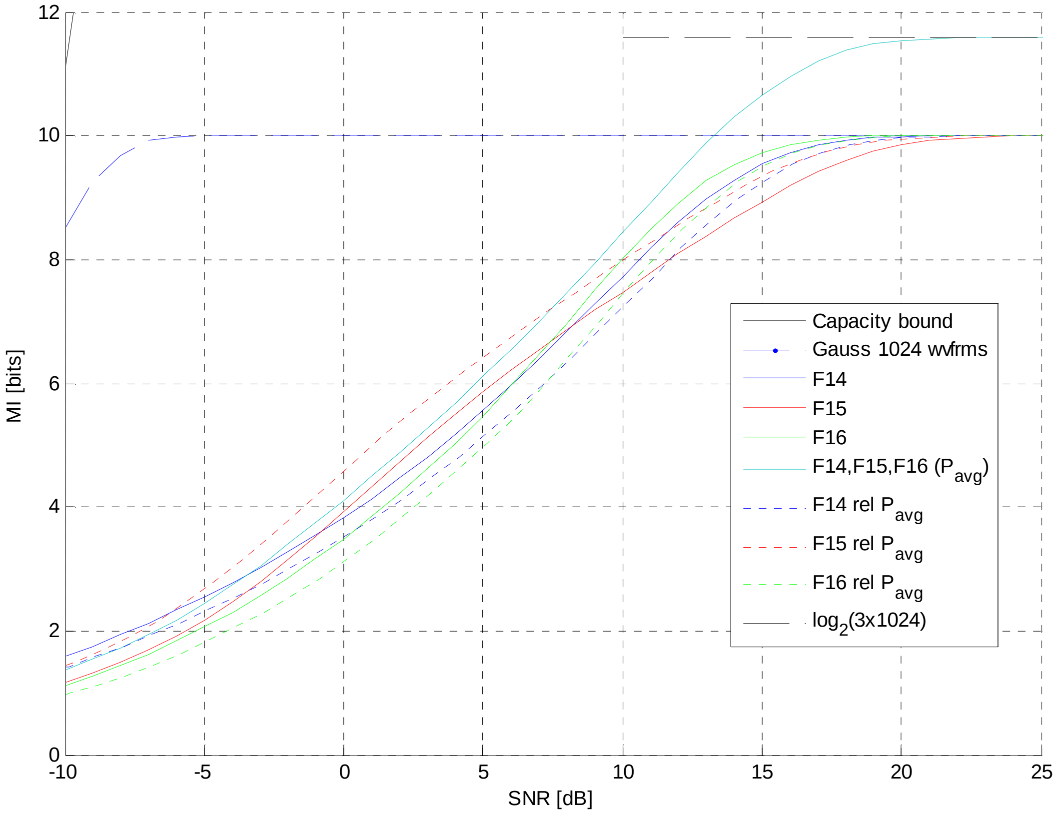


Figure 8.6: Comparison of the MI versus SNR for the F-14, F-15 and F-16 models, with MI for all 3 and MI plots relative to average power over all three targets.

The relative powers for the three targets are given in Table 4. The adjustment of SNR for each of the individual targets can however have the effect of allowing the MI to be higher than the channel capacity at very low SNR's. For these targets, the F-15 suffers from this problem.

It is interesting to note that the performance of the three targets is very similar at high MI values. The point at which the combined set for all three targets reaches its maximum MI value is also very similar to the point where the MI curves, which have been corrected for power, reach their maximum value.

Table 4: Relative power for F-15, F-16 and F-16 targets.

	F-14	F-15	F-16	F-14, F-15, F-16
P_{avg} [W in 1 Ω]	0.1038	0.1959	0.1033	0.1344
Gain relative to P_{avg} [dB]	-1.12	1.64	-1.14	0

This implies that the HRRPs for the three targets do not have a detrimental effect on each other. The Euclidean distances between any HRRP's between different targets are at least as large, or larger, than the Euclidean distances between the HRRP's of a single target.

8.2 Experiment #9: Effect of Restricted Azimuth Sectors

In this experiment, the azimuth sector over which the MI was calculated was restricted. The purpose of this is to evaluate the recognition performance in a radar which can estimate the velocity vector of the target. The recognition subsystem is thus not required to search through the whole database of waveforms, but can restrict its search to a specific azimuth region for the target.

If the aircraft is flying in a cross-wind of 5% of its forward velocity, a heading adjustment of approximately 2.8° is required by the pilot to maintain a given course over the ground. It was thus decided to restrict the azimuth region to 5.6°, which translates to 16 HRRP's out of the 1024 covering the full 360° around the target.

Figure 8.7 below shows the result for the F-14 model. For all three frames the x-axis is the azimuth angle in degrees. The upper frame shows the HRRP's generated using a Chebyshev window with a sidelobe level of -60 dB, the middle frame is the raw frequency domain data, and the bottom frame shows the MI as a function of SNR. Figure 8.8 and Figure 8.9 show the results in the same format for the F-15 and F-16 respectively. The maximum MI for the 5.6° sector for a single target is 4 bits. Figure 8.10 graphs the MI for all three targets, but for the same limited azimuth sector. Note that the maximum MI in this case is $\log_2(3\times16) = 5.584963$ bits.

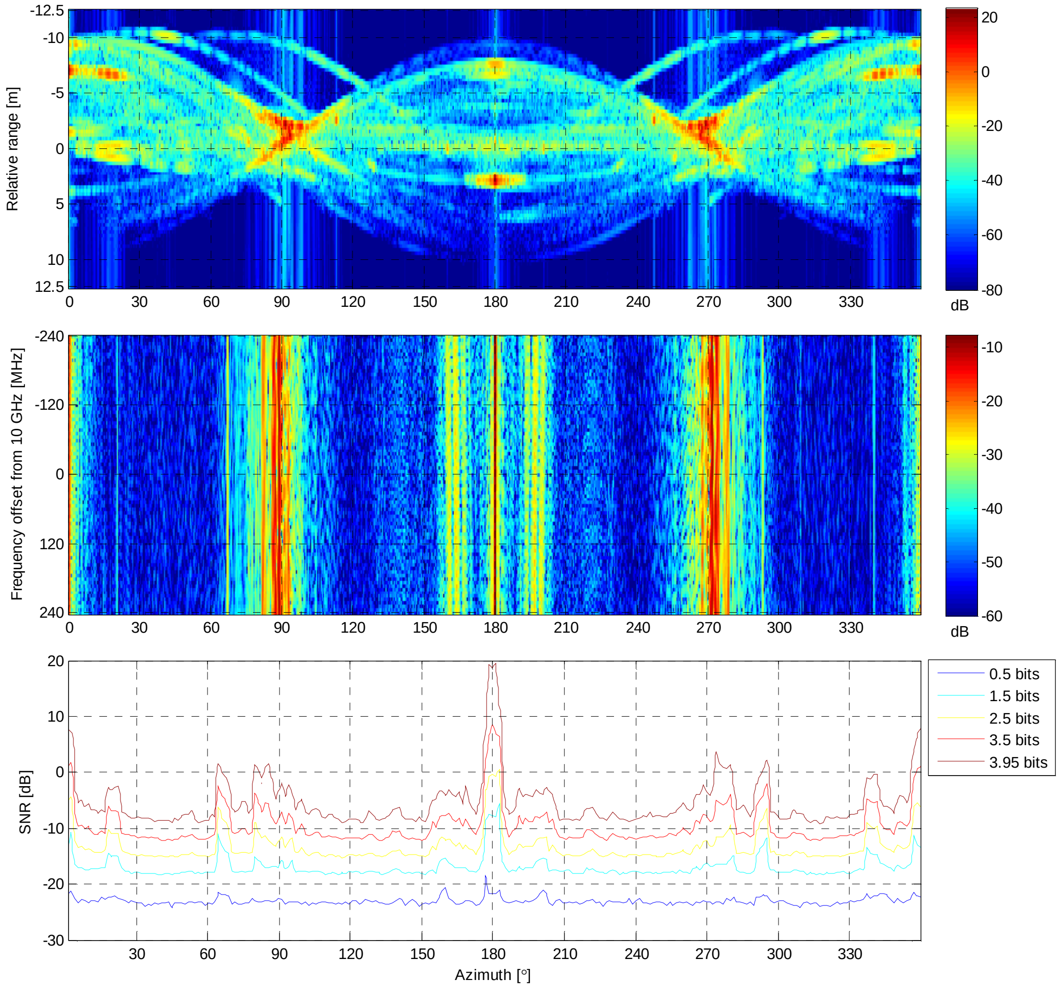


Figure 8.7: F-14 mutual information for azimuth sectors limited to 5.6°.

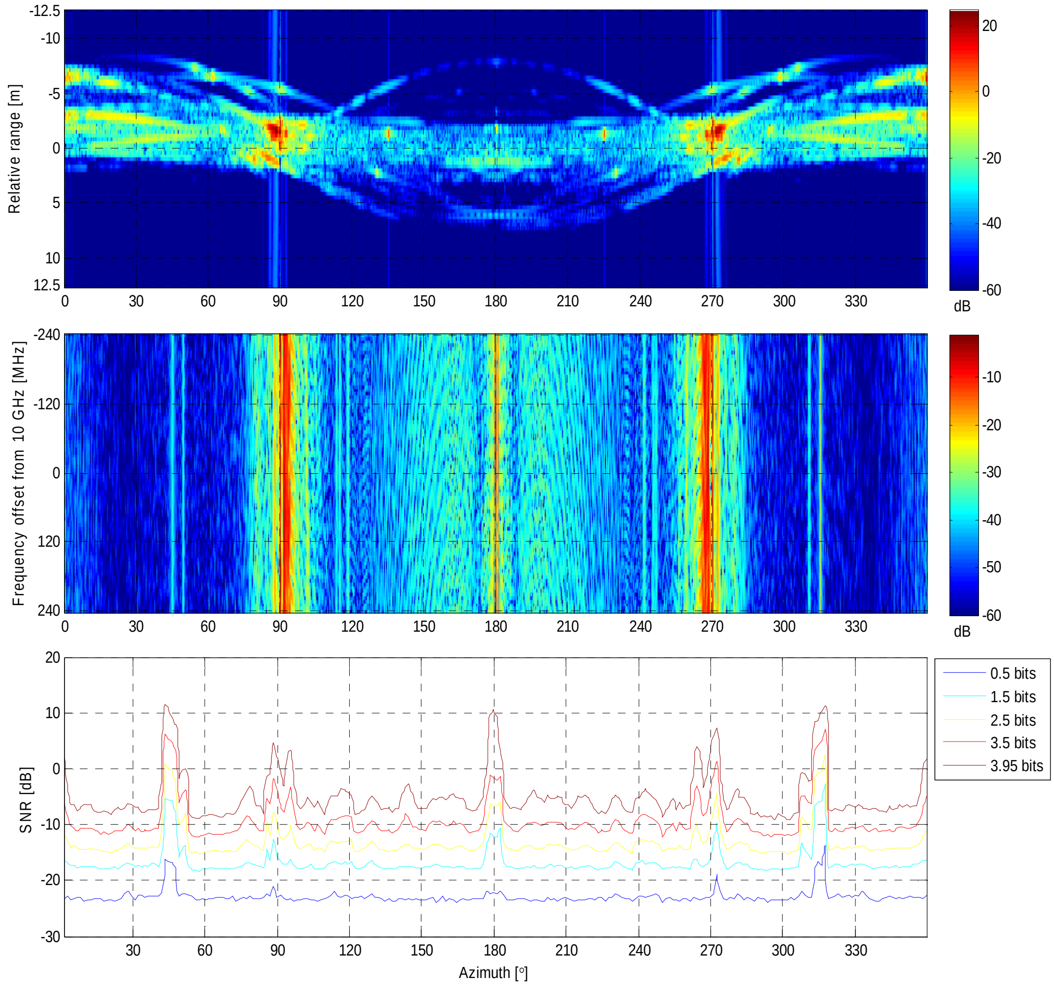


Figure 8.8: F-15 mutual information for azimuth sectors limited to 5.6°.

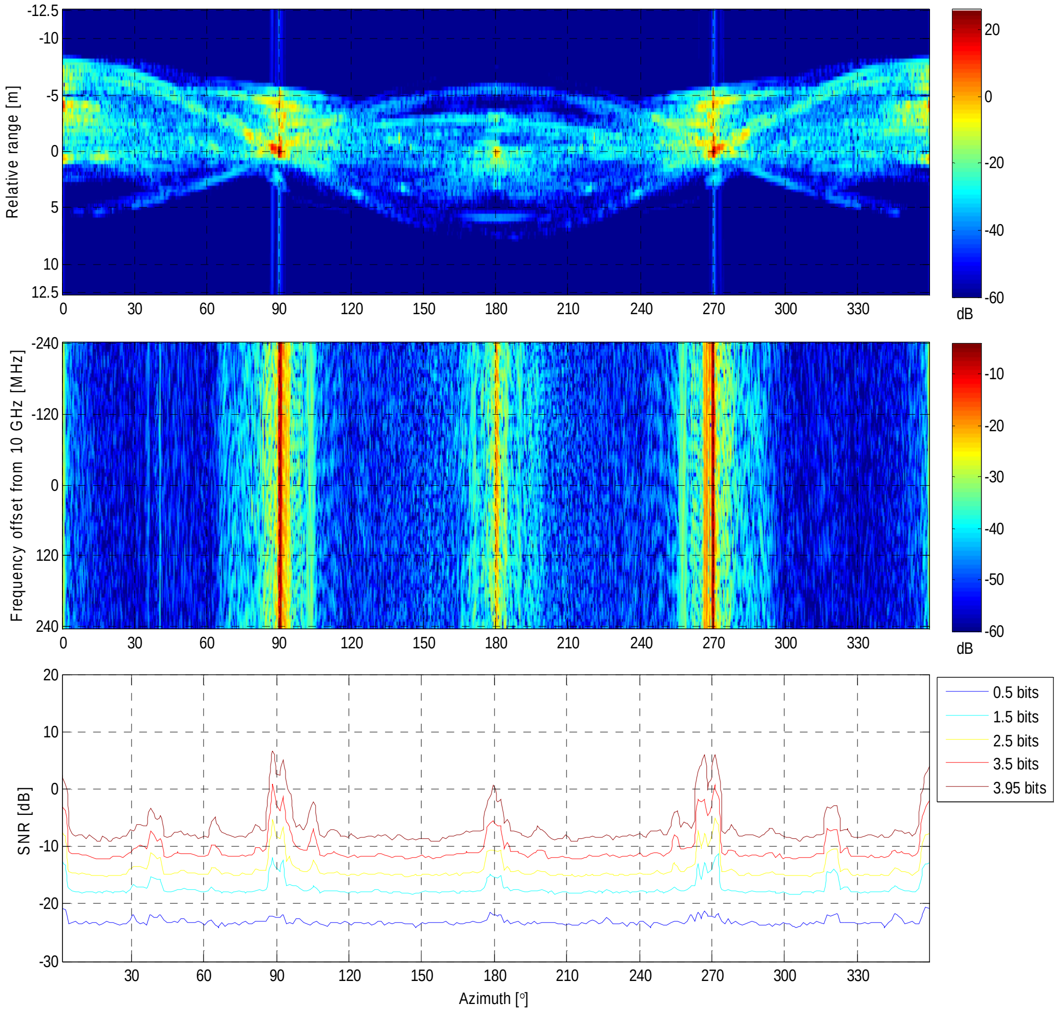


Figure 8.9: F-16 mutual information for azimuth sectors limited to 5.6°.

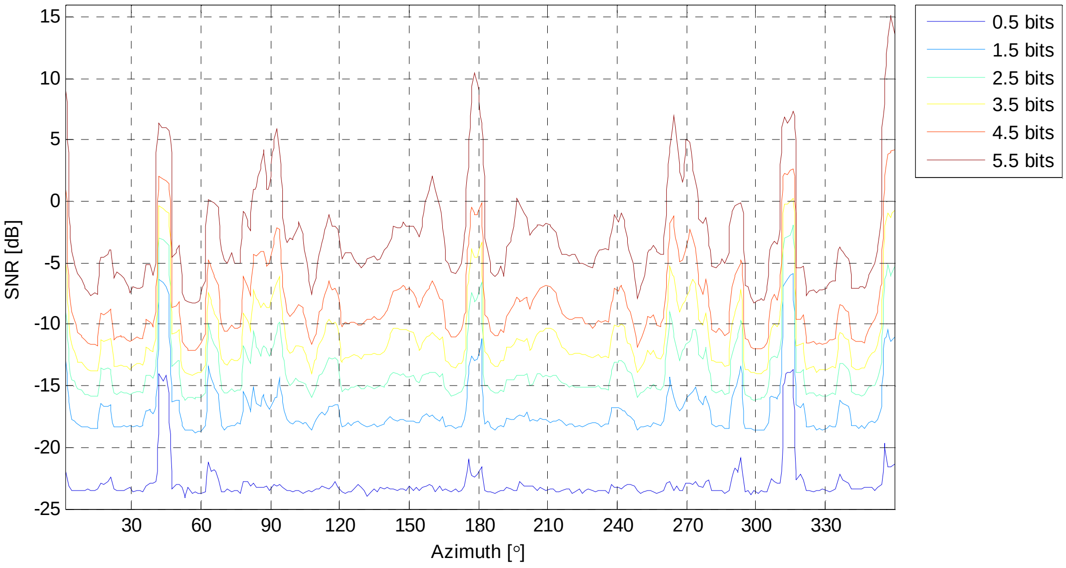


Figure 8.10: Combined F-14, F-15 and F-16 mutual information for azimuth sectors limited to 5.6°.

From the first three figures, for a single target at a time, it can be seen that one of two effects detrimentally effect the transfer of information about the target to the radar. The effects causing a loss of recognition ability are either the high relative power of the received signal, or the fact that the frequency response in the affected regions corresponds to frequency responses which do not exhibit much magnitude fluctuation. In many cases these correspond to HRRPs with few, well resolved, scatterers. In areas where all the scattering effects are approximately of the same magnitude and the HRRP is more noise like, the performance is the best. All three targets have relatively poor performance in the nose-on (0°) and tail-on (180°) aspects as well as the areas around side on (90° and 270°). The F-15 shows an area between approximately 45° and 50°, to either side of the nose, which also exhibits performance which is approximately 17 dB worse than the average performance. This area corresponds to the two vertical lines in the frequency response data, and these lines correspond to the main wing leading edge and the horizontal stabiliser leading edge. These leading edges form long, straight horizontal lines, which cause large single reflections at a well defined range when the radar illuminates the leading edges perpendicularly to their geometry. The F-16 exhibits a similar, but less dramatic effect at approximately 40° to either side of the nose.

The final figure for the performance with all three targets shows that these detrimental effects affect the whole set of waveforms for the specific azimuth sector.

8.3 Experiment #10: High Azimuth Resolution Analysis of F-15

In this section the MI results are presented for the F-15 when HRRP's are generated every 0.01°. This is the required sampling interval at a carrier frequency of 10 GHz, for a target with a cross range dimension of 20 m. The result for the 36 000 waveforms, which span 0° to 359.99°, is given in Figure 8.11 below, along with some of the previous results for comparison. Additionally the MI was also calculated for the first 1024 waveforms (from 0° to 10.23°) of the 36 000 as well as a case with 36 000 waveforms drawn from a Gaussian distribution.

From the graph it can be seen that the 36 000 set of waveforms has a higher information content than the 1024 waveform (over 360°) graph for the F-14, F-15, F-16 and the case for all three targets. The 36 000 waveform graph also exhibits the same shape as the highly oversampled azimuth graph in Figure 8.11 where there is a steep initial increase, followed by a more gradual approach to its final limiting value. The 0 to 10.23° graph also shows this behaviour. This can be explained by considering the Euclidean distances between the waveforms. As the noise reduces the signal points which are far from each other are quicker to become distinct points, leading to the initial increase in MI. The signal points which are closer to each other require a much higher SNR before they become distinct from each other and thus making the MI's approach to its limiting value more gradual. The lower slope of the MI also implies that the waveforms are more correlated with each other in the azimuth direction.

The MI for the 36 000 set of waveforms passes through the 10 bit level at a SNR of 10.91 dB, whereas the 1024 waveform set reaches 9.995 bits at a SNR of 24 dB. This does not necessarily translate into a 13 dB gain as there is uncertainty as to which of the 1024 waveforms of the 36 000 waveforms are correctly identifying the target. There is no guarantee that these are the same 1024 waveforms as the 1024 waveform case.

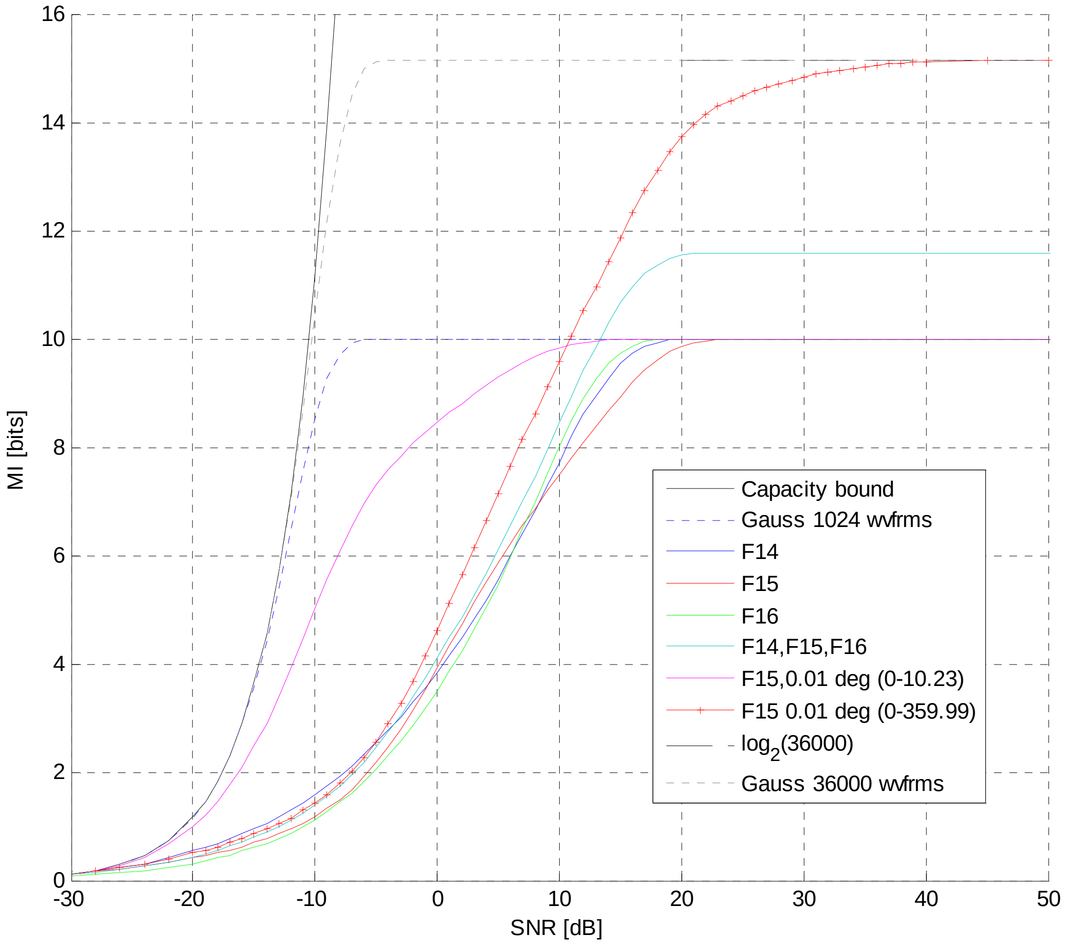


Figure 8.11: Comparison of the MI versus SNR for the F-14, F-15 and F-16 models, with two extra traces for the F-15 where the azimuth spacing was set to 0.01°.

8.4 Conclusion

In this section the MI content of the F-14, F-15 and F-16 have been analysed. The RCS of each target was simulated at 81 frequencies covering 480 MHz around a centre frequency of 10 GHz, for 1024 equally spaced azimuth incident angles, for an elevation angle of 0°. Use was also made of 1024 waveforms drawn from a Gaussian distribution as a benchmark. Use was made of flight simulator CAD models of the three fighters which were downloaded from the internet.

For MI values of 99.95% of the maximum, the error probability is approximately 1 in 4700. At this MI level, the 1024 Gaussian waveforms were approximately 5 dB worse than Shannon bound, but F-14, F-15 and F-16 were between 25 and 30 dB worse than Shannon bound. The F-16 was the most easily recognized, probably due to all the external stores on the model. The F-14 requires 1.7 dB more SNR and the F-15 requires 4.7 dB more SNR than the F-16, for the same recognition performance. The F-15's reduced performance is probably due to the lower quality CAD model.

If the three targets are normalised to have equal average power then the performance was very similar, with the F-16 outperforming the other two by approximately 1 dB. If all three sets of HRRPs are used together then the point at which the combined set of waveforms reach their maximum MI is approximately the same as the performance of the F-14 or F-15 in isolation.

Restricting the azimuth sector over which to perform recognition (for example if the radar's target track information is used) to 5.6° (16 HRRPs) results in an improvement of recognition performance which translates to a SNR gain of approximately 30 dB for relatively broad areas in azimuth. There are some azimuth angles which show performance degradations of between 10 dB and 25 dB. These are caused by areas with extremely high returns that are concentrated in range. The head-on, tail-on and side-on aspects show this type of performance degradation, as well as azimuth angles at which leading edges cause strong reflections at well defined ranges. This was especially the case for the F-15.

When all three targets are to be recognized together, the areas of degraded performance swamp any areas of good performance to degrade the performance for the three target case. The average minimum performance is reduced by approximately 10 dB for the three target case relative to the single target case. The tail-on performance is reduced by 15 dB relative to this and the nose-on performance is 20 dB worse than this.

The HRRPs for the F-15 were also simulated at an azimuth interval of 0.01°, resulting in 36 000 waveforms. The MI performance was compared to the Gaussian waveform case as well as the case where only the first 1024 waveforms (0° to 10.23°) were used. The 1024 waveform case over 360 deg always showed worse performance than the 36 000 waveform case. The MI graph for the 36 000 waveform case also exhibited the same "knee" shape as examples in section 7.1.1 which are oversampled in azimuth.

The azimuth restricted example (0° to 10.23°) exhibited an exceptionally good performance increase (10 dB gain in SNR) up to approximately the 8 bit level, which is the point at which its knee occurs. When it levels out at 10 bits it still has a SNR gain of approximately 8 dB over the F-15 with 1024 waveforms over 360 deg. This again demonstrates the performance gain possible if the angular sector over which recognition has to be performed can be reduced.

9 Interrogation of Radar Targets using Ultra Wideband Waveforms

In this section the mutual information content of the three targets (F-14, F-15 and F-16) is investigated using Ultra Wideband waveforms. The aim was to attempt to improve the recognition performance by forcing the frequency domain data for the targets to become more uncorrelated and Gaussian like. This was achieved by keeping the number of frequencies constant, but making use of large step sizes in frequency. This type of waveform is not generally used in air surveillance radars, but the point of this experiment is to evaluate the recognition performance of such a wideband radar concept. Results were also generated for various frequency bands. The results of the 480 MHz bandwidth waveforms from the previous section will be referred to as wide band (WB) for the purpose of this discussion.

The frequency response for all three targets was simulated for 81 frequencies spaced 100 MHz apart, covering the 10 to 18 GHz band. For the F-14 extra simulations were run covering the 2 to 10 GHz and 2 to 18 GHz bands (200 MHz step size), in both cases making use of 81 Frequency steps. The MI was calculated using the frequency domain data directly. If the data had been converted into the range domain, the spacing of the range ambiguities for the 100 MHz step size would be 1.5 m. The HRRP of a 19 m long aircraft would fold back on itself 12 times. For the 200 MHz step size, the range ambiguities would be every 0.75 m, thus causing the HRRP to fold back on itself 25 times. Most radar design engineers would aim to have a waveform with an ambiguous range that is somewhat longer than the target extent.

9.1 Experiment #11: Comparison of Wideband and Ultra Wideband Waveforms

The following three figures show various versions of the results for the setups described above. Figure 9.1 shows the standard plot of the MI versus SNR. Due to the fact that the traces are relatively close to each other (surprisingly so), the next two plots are processed versions of this data to demonstrate changes in MI and SNR of the UWB waveforms relative to the WB (480 MHz) waveform used in the previous experiments.

In Figure 9.2 the UWB waveform MI results are plotted relative to the WB waveform results. The figure shows the gain (or loss) in MI when using the UWB waveform, relative to using the WB waveform. From Figure 9.2 it can be seen that for SNR’s above -10 dB, the UWB waveforms perform better than the WB waveform. The maximum information increase is 0.78 bits, which implies that nearly double the amount of waveforms can be distinguished in the UWB case. The exception is the F-15, which actually shows a loss in MI of up to 0.1 bits for SNR’s between 10 dB and 20 dB.

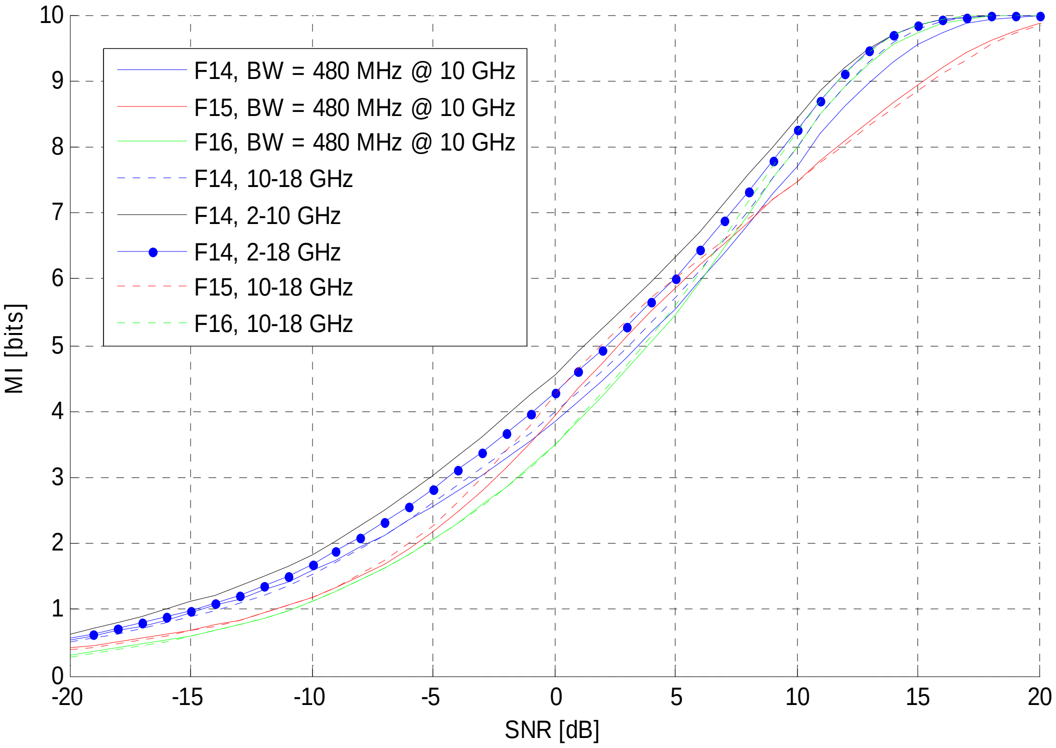


Figure 9.1: Comparative traces of the MI versus SNR for various UWB waveforms for the F-14, F-15 and F-16 targets.

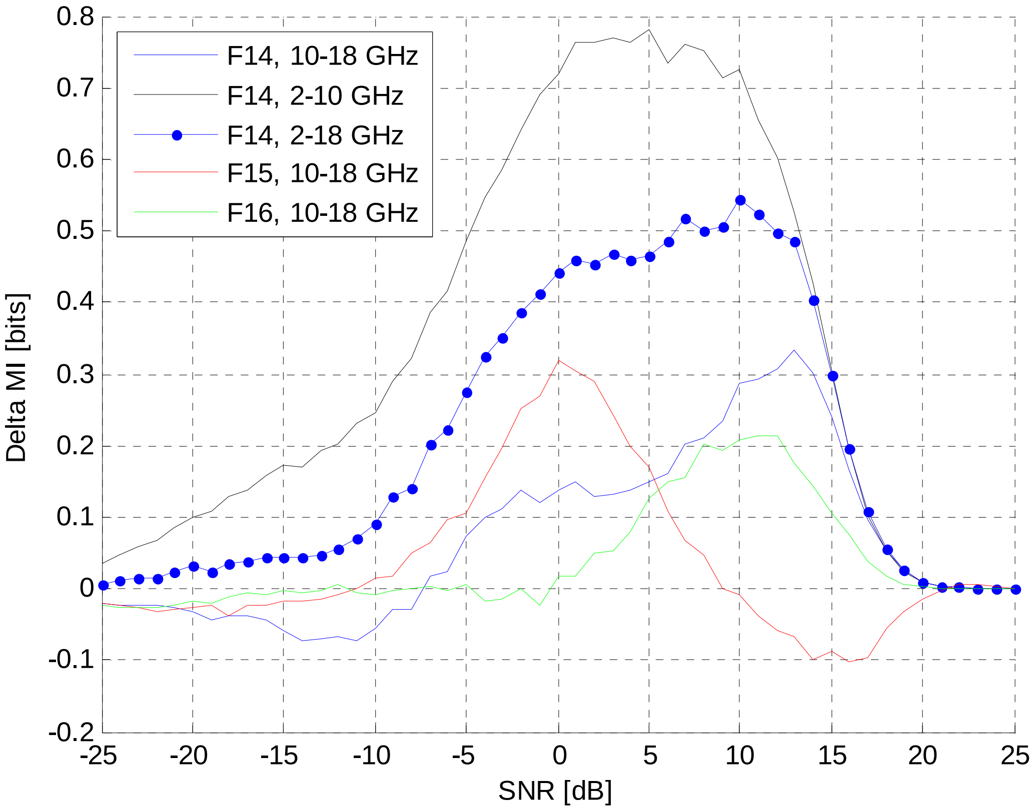


Figure 9.2: Comparison of the change in MI versus SNR for the various UWB waveforms and for the F-14, F-15 and F-16 relative to the WB waveform.

The MI gain is limited to zero at high SNR's due to the use of 1024 waveforms. This means that a maximum of 10 bits of information can be achieved and once the MI curves have reached this value, there is no relative MI gain between the various curves.

It is interesting to note that for the F-14, the 2-10 GHz waveform out performs the 2-18 GHz waveform by approximately 0.25 bits, and that the 10-18 GHz waveform performs the worst of all three. The initial aim of the experiment was to create a more randomized response, so it was theorized that the 10-18 GHz waveform would achieve this better due to the shorter wavelengths inducing more random phase shifts in the data.

Figure 9.3, below, was generated by resampling the MI data sets at fixed MI values and plotting the SNR gain relative to the WB waveform, thus obtaining the SNR gain for each setup.

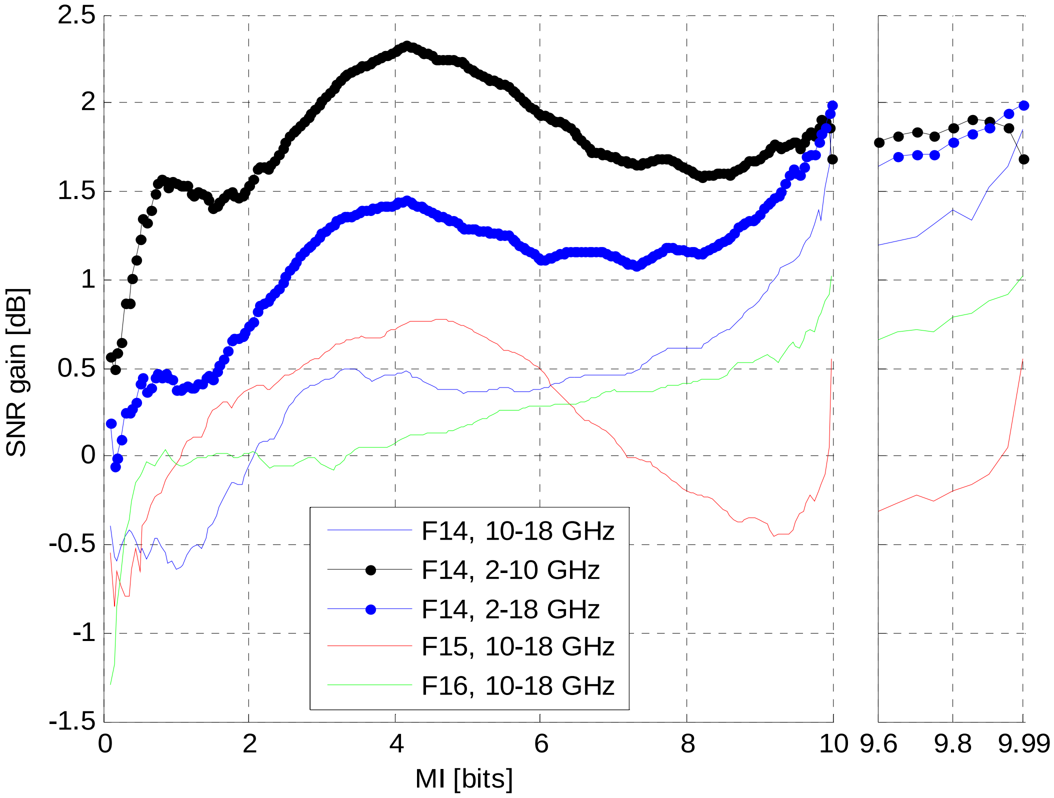


Figure 9.3: Comparison of the gain (loss) in SNR versus MI for the various UWB waveforms for the F-14, F-15 and F-16 with the WB MI as reference. The right panel is a zoom of the high MI region of the left panel.

Once again, all the targets except the F-15 showed a gain in performance for the UWB waveforms over the WB waveform for values of MI above 2 bits (approximately corresponding to a SNR of -10 dB). The F-14 in combination with the 2-10 GHZ waveform showed the highest gain of 2.33 dB, while the F-15 and 10-18 GHz waveform showed a loss of nearly 0.45 dB at a MI value of 9.25 bits. For this experiment the MI value of 9.99 bits was converted to P_e using Fano's inequality, which results in $P_e = 4.43 \times 10^{-4}$, or equivalently, one error in 2257. The SNR gains for each of the experimental setups is tabulated in Table 5 for this value of P_e, where the gains are listed in decreasing order.

Table 5: SNR Gain for UWB waveforms over WB waveforms for F-14, F-15 and F-16 at a probability of error of 1 in 2257 (MI = 9.99 bits).

Target	Waveform	SNR Gain
F-14	2 - 18 GHz	1.99 dB
F-14	10 - 18 GHz	1.85 dB
F-14	2 - 10 GHz	1.69 dB
F-16	10 - 18 GHz	1.02 dB
F-15	10 - 18 GHz	0.55 dB

The reason for the slightly degraded high-SNR performance of the F-14 with the 2-10 GHz waveform is the fact that the frequency responses at each azimuth are more correlated with one another in the low frequency part of the band. This correlation is probably increased by the fact that the F-14 model does not have any ordnance mounted on it. From the head-on and especially tail-on aspect (when the trailing edges of the wings face the radar) this probably reduces the effective cross range dimension of the F-14 as seen by the radar to a value which is less than its physical (i.e. wingtip to wingtip) cross range dimension. A related effect would be that at the lower frequencies each scattering area on the target will have a larger angular region over which it will scatter. From an antenna point of view, the aperture size remains constant so if the frequency is reduced the beamwidth should increase. This effect might distribute extra energy into angular regions where relatively little scattering was present at the higher frequencies. This in turn should increase the separability of the waveforms in azimuth, especially taking into account the fact that the azimuth dimensioned is sampled every 0.352° instead of 0.01° at 10 GHz and 0.055° at 2 GHz (The F-14 has a wingspan of 19.55 m).

Another explanation might be related to the percentage bandwidth of the two waveforms. Percentage bandwidth is defined as the ratio of the bandwidth of a signal to its centre frequency. For the 2-10 GHz waveform the percentage bandwidth is 133%, whereas the 10-18 GHz has a percentage bandwidth of only 57%, even though the two waveforms have the same resultant range resolution. The difference in percentage bandwidth might cause returns from reflection points on the target to have phase angles which are more likely to be random, thus increasing the separability between the waveforms in the signal space.

It is still surprising that the performance of the UWB waveforms is only 0.5 dB to 2 dB better than the WB performance. The increase in radar complexity to support these waveforms would far outweigh the gains.

9.2 Experiment #12: Comparison using Ultra Wideband Waveforms which Span 2-18 GHz

The above set of results were so surprising that the simulations were rerun, using a different approach to ensure that these results are correct. For the second run, the F-14 was simulated over the 2-18 GHz band at 100 MHz steps. This single result was then used as input into the MI calculation, and then sub-sampled to create the various data sets. Due to the fact that data for 161 frequencies was available, the MI for this case was calculated as well. The results are shown in Figure 9.4, Figure 9.5 and Figure 9.6 below.

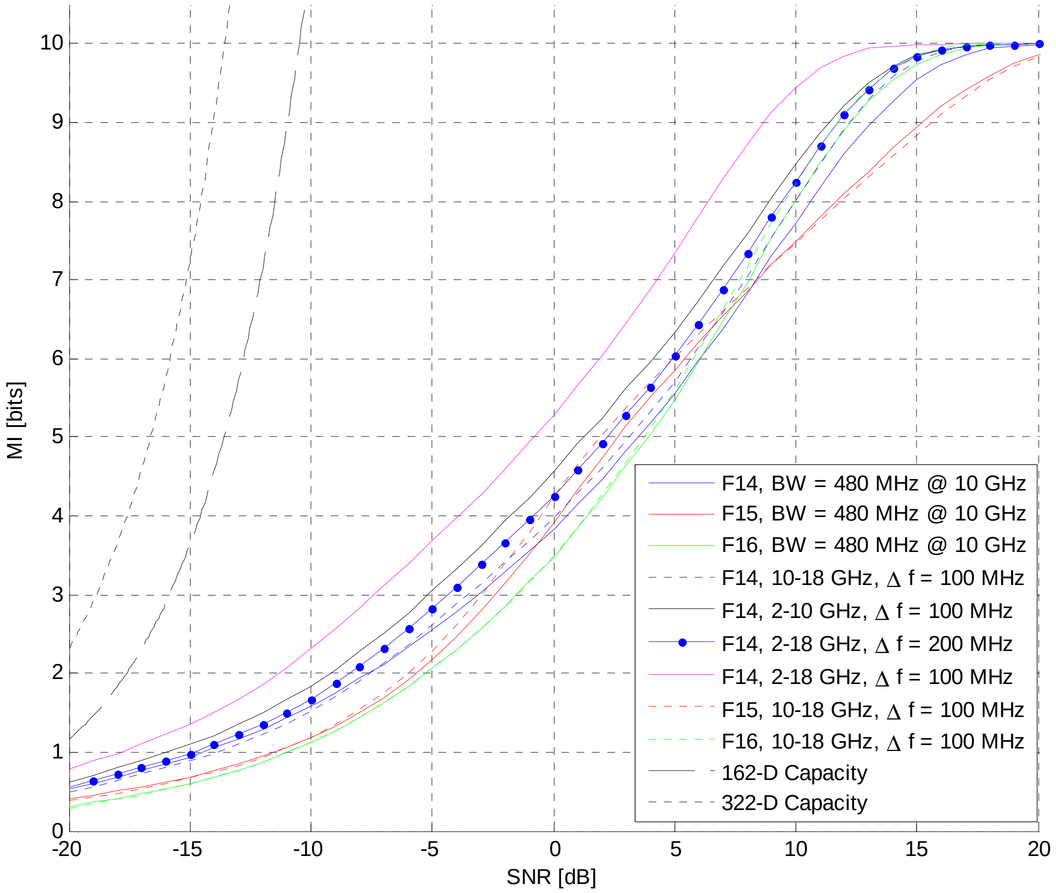


Figure 9.4: Rerun of the comparison of the MI versus SNR for various UWB waveforms for the F-14, F-15 and F-16 targets.

The results in these three figures track the initial set of results very well, thus verifying that no programming errors were made during the data processing steps.

The result for the F-14 using the 100 MHz step size waveform shows that there is nearly an exactly 3 dB increase in performance over the 200 MHz step size case. The last trace in Figure 9.6 shows the difference between the 100 MHz and 200 MHz waveforms.

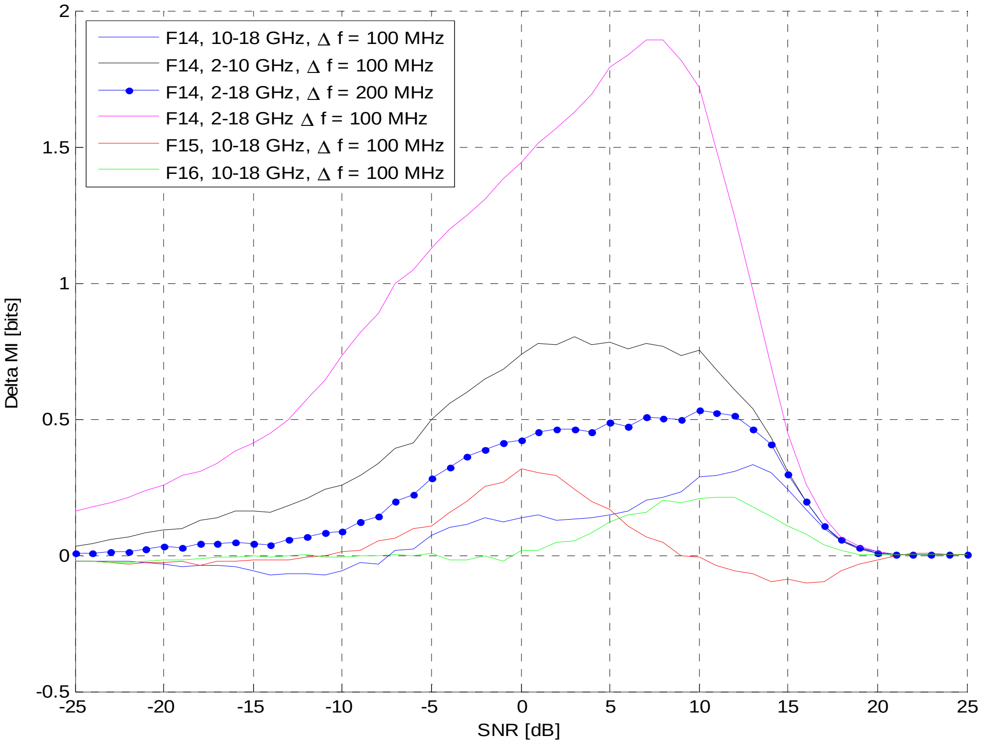


Figure 9.5: Rerun of the comparison of the change in MI versus SNR for the various UWB waveforms for the F-14, F-15 and F-16.

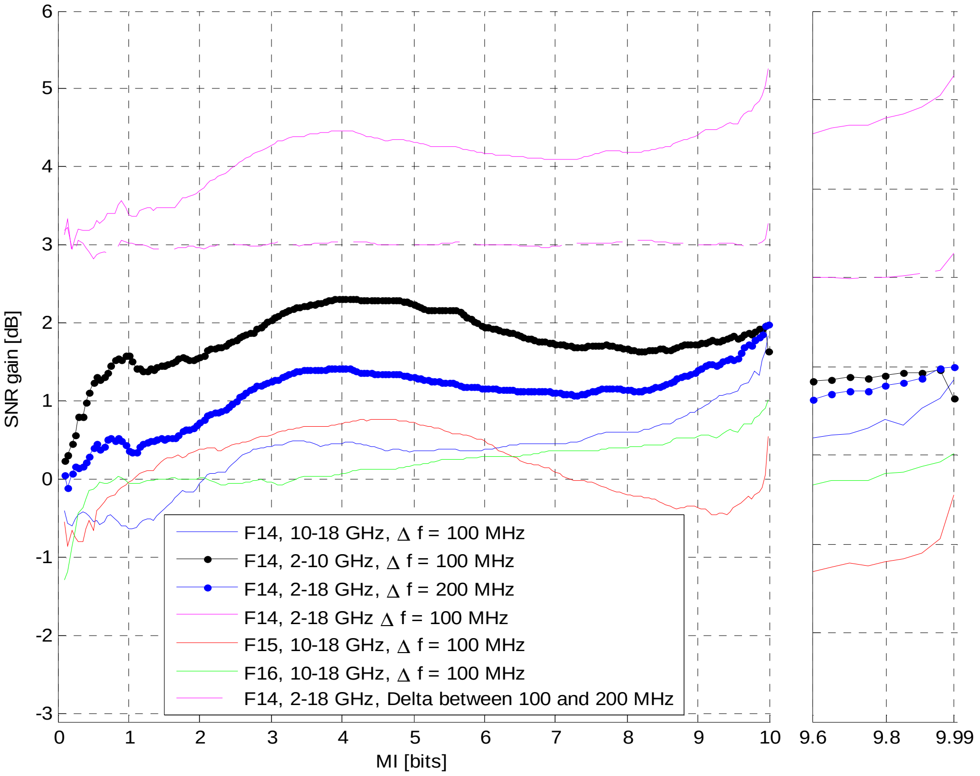


Figure 9.6: Rerun of the comparison of the gain (loss) in SNR versus MI for the UWB waveforms and for the F-14, F-15 and F-16. (Right panel: zoom of the high MI region.)

For the radar to maintain the SNR at each frequency, the radar would have had to transmit double the amount of frequencies for the 100 MHz step case as compared to the 200 MHz step case. This effectively means that the radar increased its power by 3 dB, and a 3 dB recognition performance gain was measured. This means that the radar has not gained any extra information by transmitting the extra frequencies. It has however effectively gained in SNR. This implies that the maximum possible information content was already captured by the waveform spanning 2-18 GHz in 200 MHz steps.

Figure 9.7 shows the histograms of the Euclidean distances between the signal points for the F-14 with various UWB waveforms. This was added out of curiosity and has not led to any extra insight as yet.

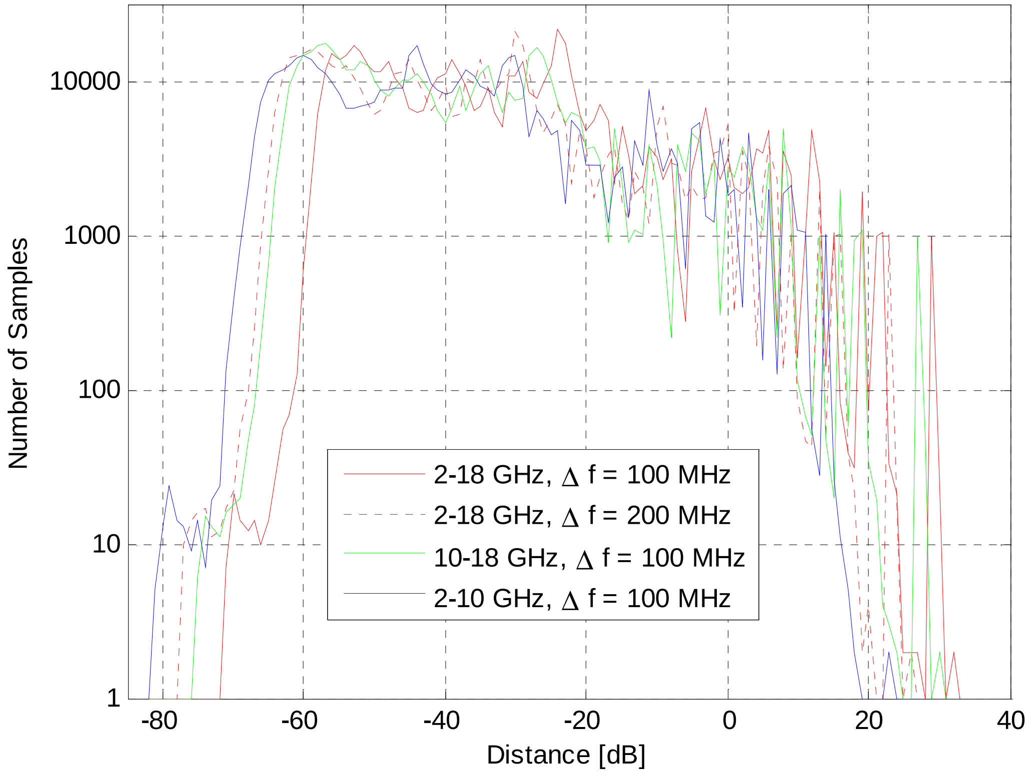


Figure 9.7: Comparison of the histograms of the Euclidean distances between points in the signal space for various UWB waveforms, for the F-14 target.

The following two figures show the data sets used for this section. Figure 9.8 shows the frequency domain data, and Figure 9.9 shows the range domain data. Note that a Chebyschev window with a -40 dB dB sidelobe level was used during the conversion from the frequency to the range domain for Figure 9.9. The MI graphs were calculated using the frequency domain data directly. According to the data processing inequality this will yield the same MI result as the FFT is an invertible function.

From the range domain plot, it can be seen that the data is relatively sparse in this domain, so it is unlikely that the range ambiguity has destroyed large amounts of information.

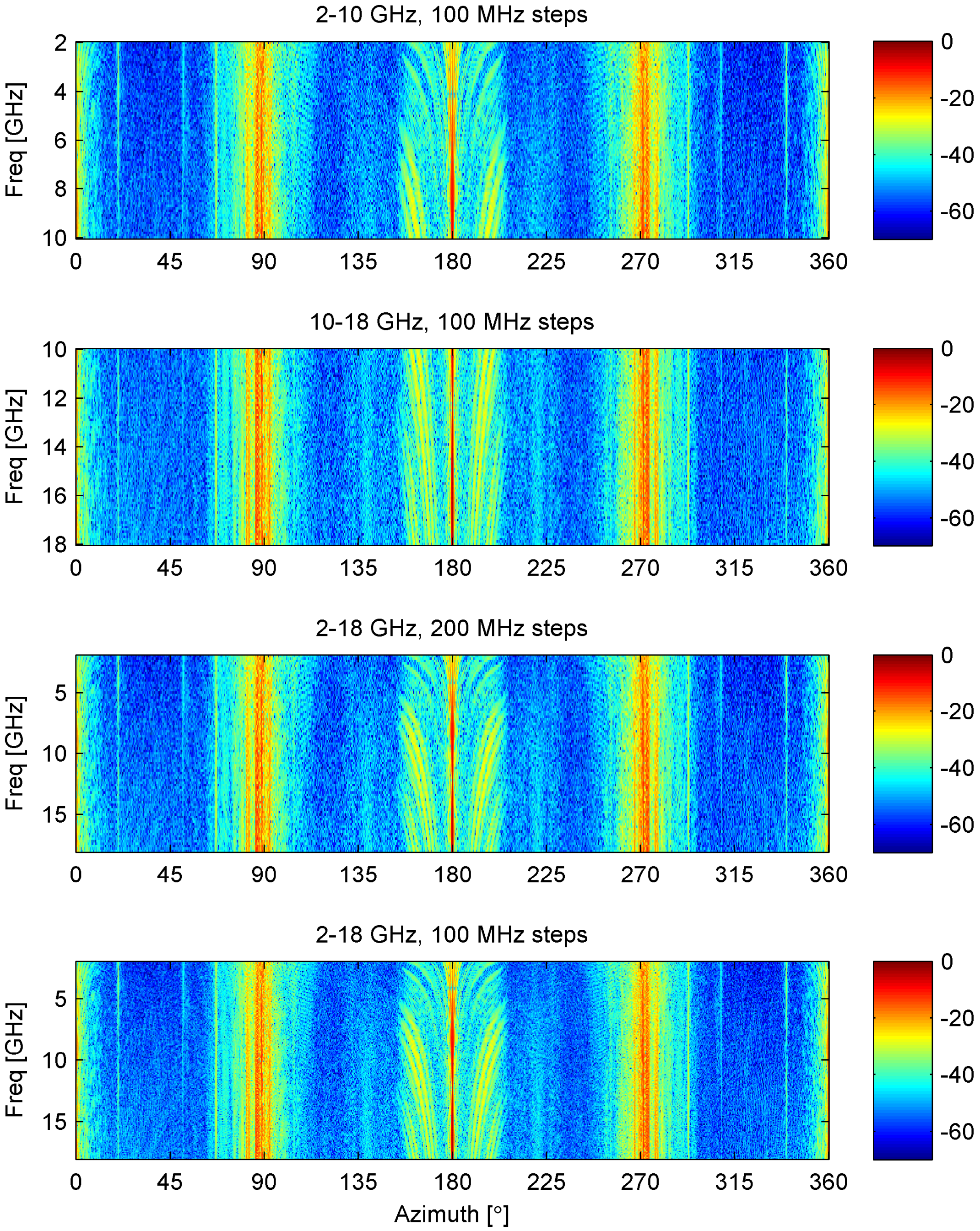


Figure 9.8: UWB frequency domain data sets for the F-14.

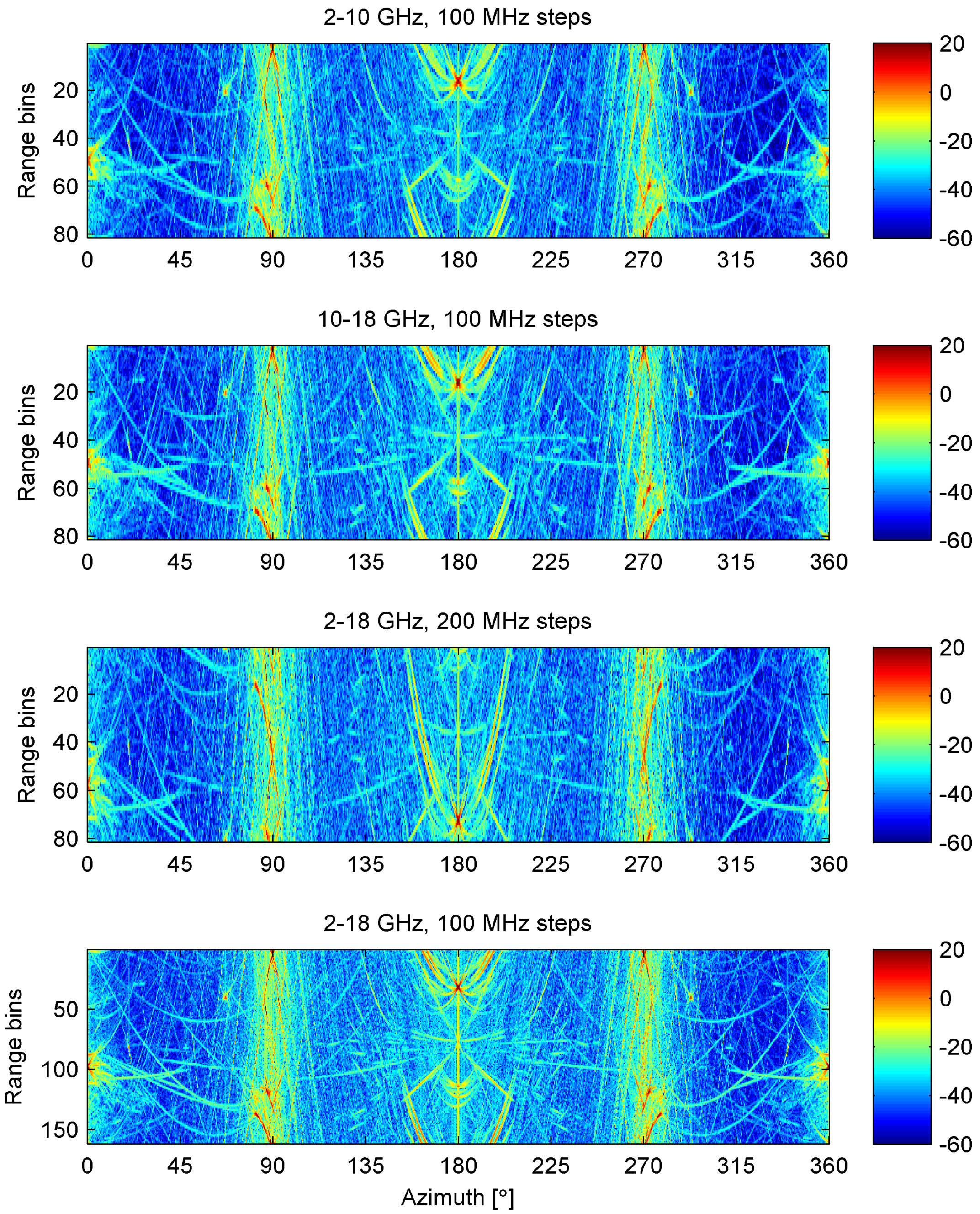


Figure 9.9:UWB data sets for the F-14 after transformation into the range domain.

9.3 Conclusion

The aim of this section was to attempt to improve the recognition performance by forcing the frequency domain data for the target responses to become more uncorrelated and Gaussian like. This was achieved by keeping the number of frequencies constant, but making use of large step sizes in frequency. It was theorized that a 10-18 GHz waveform with large frequency step sizes of 100 MHz would achieve this due to the shorter wavelengths inducing more random phase shifts in the data and the highly range ambiguous nature of the waveforms.

Use was made of the MI results for the wideband waveforms from the previous section (81 frequencies spanning 480 MHz at 10 GHZ) for the F-14, F-15 and F-16 as a baseline. These results were compared to the MI content for waveforms which have 81 frequencies, but span 10-18 GHz. Extra waveforms which span 2-10 GHz at 100 MHz step size and 2-18 GHz at 200 MHz step size were also analysed for the F-14. All the waveforms are highly range ambiguous have range ambiguities of 1.5 m and 0.75 m.

The SNR gain for the 10-18 GHz waveform was between 0.55 dB for the F-15 and 1.85 dB for the F-14 at a probability of error of 4.43×10^{-4} (MI = 9.99 bits). The maximum gain of 2 dB was exhibited by the F-14 for the 2-18 GHz waveform. These are relatively mediocre gains for the implied increase in the complexity of the radar system to support these waveforms.

There might be several effects which are responsible for the observed performance:

1. The highly range ambiguous waveform might be destroying any information gained by the UWB nature of the waveform.
2. The limit of the amount of information contained in the geometry of the target might have been reached.
3. The limit of the amount of information about the target geometry which can be encoded onto the radar waveform might have been reached.

These results were so surprising that the experiment was repeated by using SigmaHat to simulate the whole 2-18 GHz band in 100 MHz steps in a single batch. The various UWB waveforms of the previous section were then extracted and the MI calculated. The results confirmed the results of the previous section to within the tolerance of the MI estimates.

The extra 2-18 GHz at 100 MHz step size waveform showed nearly exactly 3 dB improvement over the 200 MHz step size case. This was to be expected as the radar had transmitted double the number of pulses, thus increasing the amount of energy transmitted by a factor of two. This result invalidated the assumption that the range ambiguous nature of the waveforms was destroying the information content. If this was true then the gain for this comparison should have been more than 3 dB because there would have been 3 dB gain due to the extra transmitted energy, and then some extra gain due to the waveform being less range ambiguous. There was however a slight SNR gain above the 3 dB level of approximately 3.3 dB just at the point where the MI reached 9.99 bits. This would indicate that there might have been some low amplitude effects which were only discernible at high SNR which have been recovered by making the data less range ambiguous.

10 Gain in Information Content by Exploitation of Extra Receiver Channels

This chapter presents simulated and measured results which characterise the potential gain in recognition performance which can be achieved by the addition of extra receiver channels to the radar. Two cases are analysed, the first being the exploitation of polarisation information in the HRRP domain, the second is the use of the difference channel in a monopulse antenna.

10.1 Experiment #13: Gain in information Content by Using Polarisation for the F-18 and F-35

This experiment addresses the prediction of the gain in recognition performance of a High Range Resolution (HRR) X-band radar when the full polarisation matrix is sensed by the radar, and is compared to the performance when only making use of single linear polarisations. The study was conducted for the F-35 and F/A-18F which were simulated using the SigmaHat RCS prediction code [Smit2012a]. The simulations were setup and executed by Johan Smit at CSIR DPSS and the resulting data was made available for this study.

CAD models of a F/A-18F and F-35, were downloaded from the internet and are shown in Figure 10.1 and Figure 10.2 respectively. The outer shell of the aircraft was modelled as a perfectly electrically conducting (PEC) material due to the lack of any information on the materials of these two aircraft. Effects caused by the canopy, radomes and radar absorbing material (RAM) could thus not be modelled.

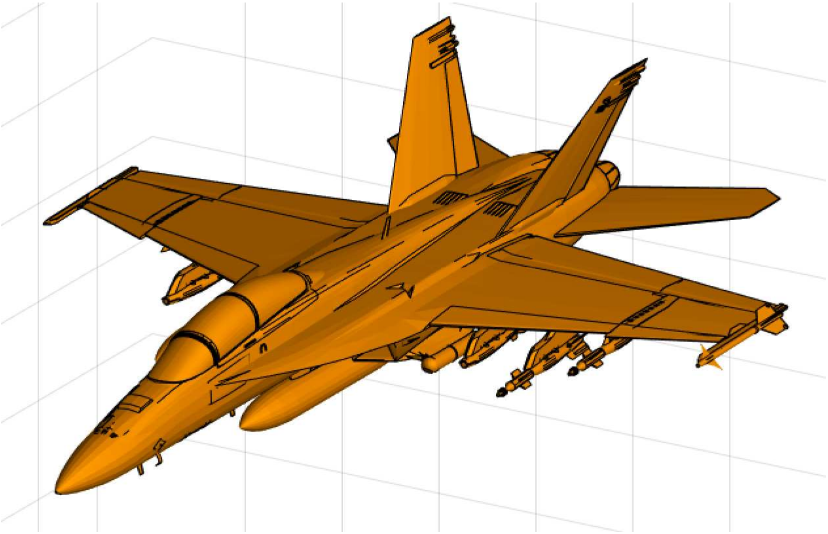


Figure 10.1: CAD model for the FA-18F.

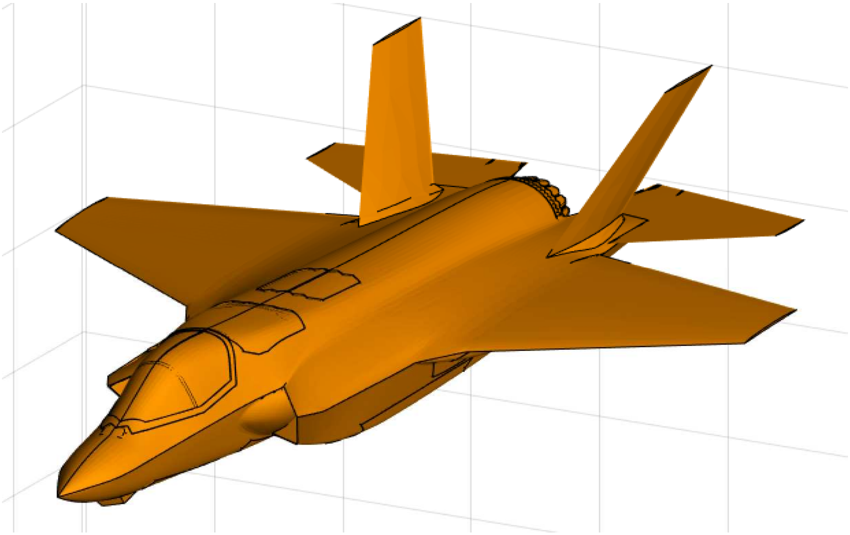


Figure 10.2: CAD model for the F-35 Lightning II.

Also note that the engine models were relatively inaccurate. These models were analysed using an in-house high frequency radar cross section code SigmaHat to obtain their frequency responses at a set of discrete frequencies. SigmaHat makes use of Physical Optics (PO) combined with the shooting and bouncing ray (SBR) techniques to calculate the RCS. Polar plots of the high range resolution profiles (HRRP's) of the targets are given in Figure 10.3 and Figure 10.4.

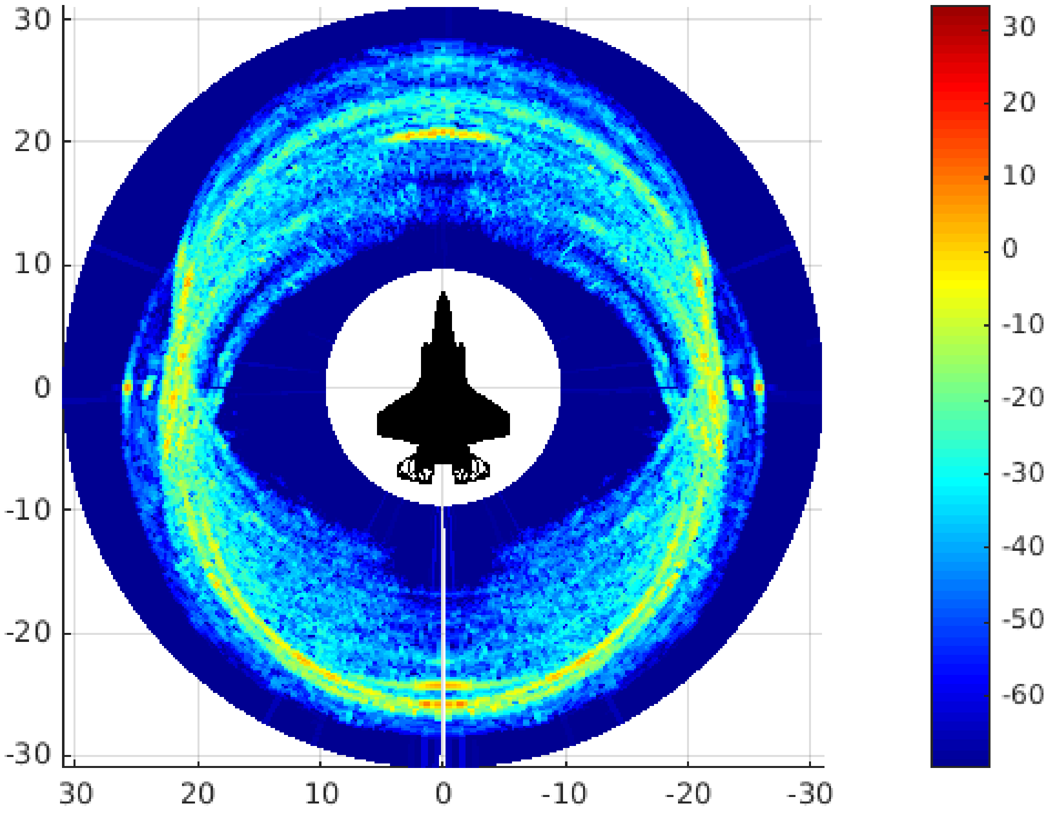


Figure 10.3: HRRP ring plot for the F-35, VV Polarisation.

The simulations were run for all four linear polarisations and each target was illuminated at a set of 116 frequencies centred at 10 GHz in steps of 7 MHz (i.e. 9.6 GHz to 10.405 GHz). This gives a total bandwidth of 805 MHz, resulting in a range resolution of 0.186 m. Each target was simulated at 360 azimuth angles in 1° steps for an elevation of 0°.

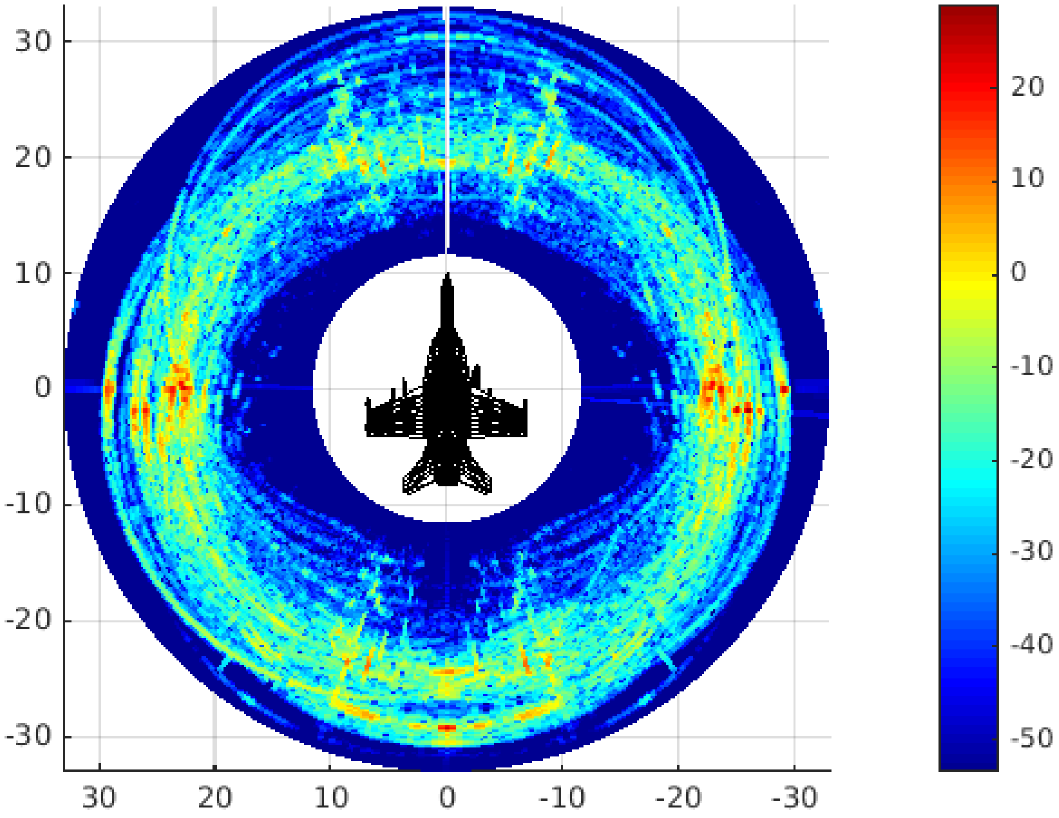


Figure 10.4: Ring plot for the F-18, VV polarisation. Note the external stores protruding from below the wings and the body.

The multidimensional expression for MI was used to evaluate the MI for each of the targets well as the three polarisation options (VV, HH and full polarization) by means of Monte-Carlo integration which was performed over a 232-dimensional Gaussian density for the linear polarisations and over a 696-dimensional space for the full polarisation matrix. For the full polarisation matrix, a vector was formed by concatenating the vectors for the two linear polarisations, followed by the sum of the two cross polarisations for each azimuth angle. The two cross polarisations are added because most radar systems capable of sensing the full polarisation matrix will measure V and H receive channels for V transmit, followed (usually in time) by measurements of V and H for H transmit. In this process, both the VH and HV polarisations have been measured. These measurements are assumed to be equal due to reciprocity and are thus added to obtain a gain in SNR for the co-polarisation channel.

The data sets were not converted to HRRP's, but were processed directly in the frequency domain. According to the data processing inequality this will yield the same MI results as the FFT is an invertible function. The maximum value of the MI is always limited to the entropy of the underlying set of waveforms, which in this case is the number of azimuth angles at which target responses have been simulated. This means that the MI graphs reach a maximum value of 8.492 bits. The MI results for the two targets and various polarisation combinations are shown in Figure 10.5 below.

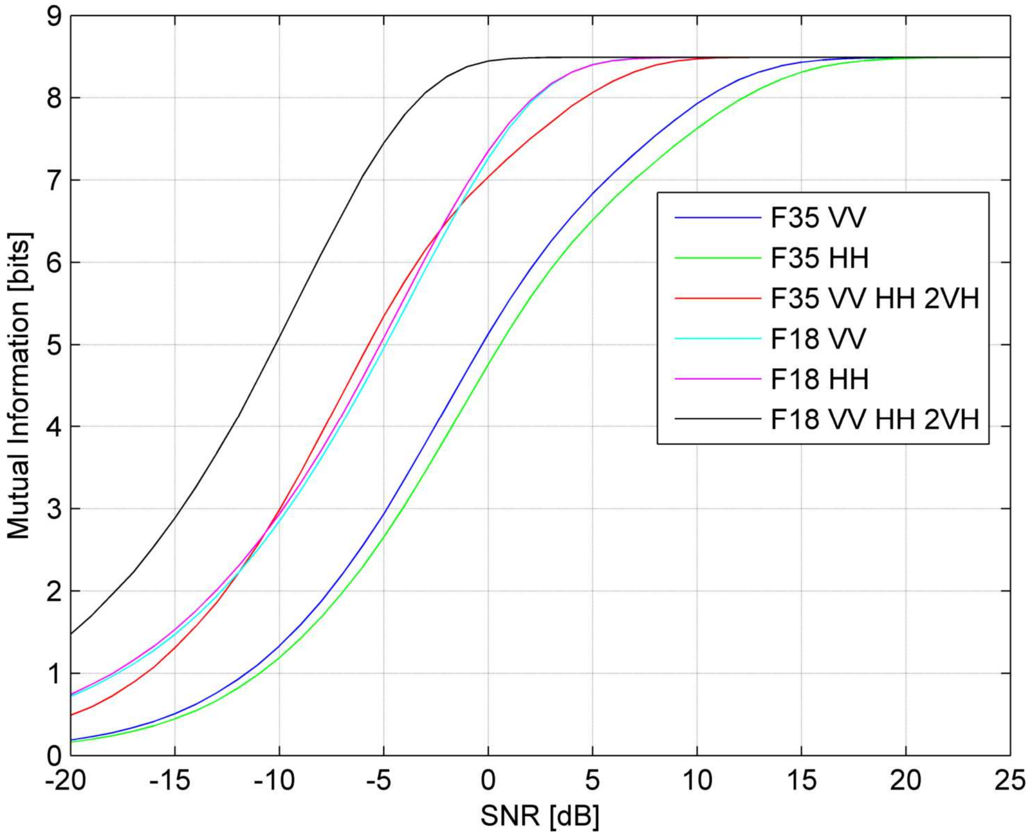


Figure 10.5: Comparative MI results versus SNR for the F-35 and F-18 for three polarisation options, linear VV, linear HH and the full polarisation matrix.

From Figure 10.5 it can be seen that the information contained in the two linear polarisations is approximately equal for F-18. For the F-35 the VV polarisation has a 1-2 dB advantage over the HH polarisation. The F-35 is less easily recognizable than the F-18 and requires approximately 5 dB more SNR to achieve the same recognition performance as the F-18. To further quantify the achievable performance gains, Figure 10.6 shows the differences in SNR between selected curves in Figure 10.5 for equal values of MI.

When both linear polarisations as well as the cross polarisation are used for recognition, the recognition performance of both targets is increased by approximately 5-6 dB. Interestingly, the F-35 is still less recognizable than the F-18 for average to high values of MI by approximately 4-5 dB, but this difference increases to 9 dB at an MI value of 8.48 bits. This illustrates that it will be more difficult to obtain recognition performance close to 100% for the F-35. This is probably due to the fact that for this F-35 CAD model there are no external stores, whereas for the F-18 there are several external stores mounted below the wing and fuselage. The effect of these stores is also evident in the HRRP ring plots where the F-18 has far more discrete scattering centres than the F-35 does. Even using the full polarisation matrix for the F-35 leaves it with approximately a 2.5 dB SNR deficit compared to the two single linear polarizations for the F-18 at an MI value of 8.0 bits (see Figure 10.5).

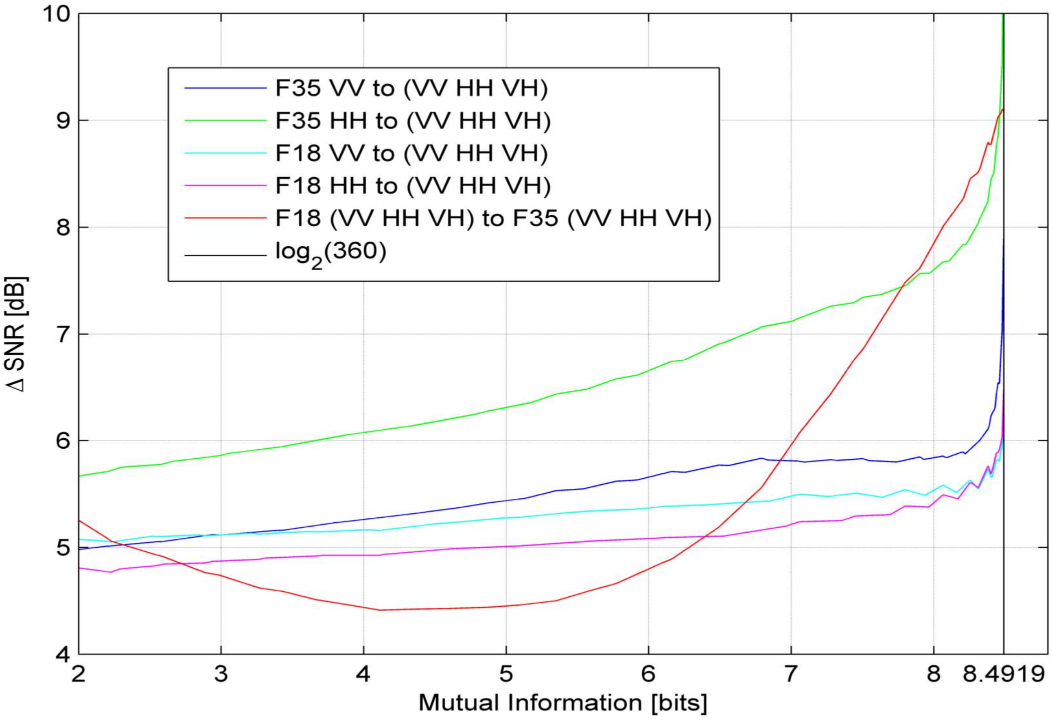


Figure 10.6: Achievable SNR gains versus MI between various polarisations for the F-18 and F-35 targets.

For this problem, a probability of error of 10^{-3} translates to a MI value of 8.472 bits, which is approximately 99.77% of the maximum MI value. At this point, for the F-18, the SNR gain in both VV and HH to full polarization is approximately 6 dB. This means that the average gain going from linear to full polarization is 6 dB for the F-18. For the F-35, the gains going from VV and HH to full polarisation are approximately 7 dB and 9.5 dB respectively, giving an average SNR gain of 8.25 dB.

This experiment has evaluated the increase in recognition performance which can be achieved by making use of a fully polarised radar. For the HRR radar operating in X-band and the two targets analysed, it has been shown that a minimum performance increase of at least 4 dB is possible by using the full polarisation matrix over any linear polarisation. Performance increases of more than 5 dB are possible when operating at high signal to noise ratios. At a probability of error of 10^{-3} (MI value of 8.472 bits, or 99.77% of maximum MI), performance increases of 6 dB for the F-18 and 8.25 dB for the F-35 are possible.

Interestingly, in the case when the full polarisation matrix is used for the recognition of the F-35, its results still show approximately a 2.5 dB loss in SNR relative to either of the linear polarisation results for the F-18.

10.2 Experiment #14: Information Content of Measurements of a 1:25 Scale Boeing 707 Model

A 1:25 scale model of a Boeing 707 was acquired by Prof. Odendaal of the University of Pretoria to serve as one of the scale models for several RCS studies [Pienaar2018, Pianaar2017, Pianaar2017a]. This model had been used for antenna placement studies in the past. Due to this it had several holes in the fuselage which had to be closed with conductive tape. The gaps where the wing roots joined into the fuselage were also closed, as were the inlets and outlets of the engines. The engines were hollow pipes, so it was felt it would be better to close them totally until some form of engine model could be manufactured. The model was mounted on a specially manufactured (by Casper van Zyl) wooden pole in the compact range at the University of Pretoria. The model was aligned to be as level as the mechanics of the mounting would allow. The measurements were conducted for both polarizations by repeating the measurement with the range's feed horns changed from vertical to horizontal polarization. The frequency extent of the measurements was 2-18 GHz and this range was covered by 2001 frequency points, resulting in a step size of 8 MHz. The azimuth rotation of the target was set up to be from -179.8° to +179.8° in steps of 0.2°. This setup resulted in 1799 discrete measurement points in azimuth. Due to the scale of this model, the following table was constructed to give the equivalent dimensions:

Table 6: Scale calculation for 1:25 scale Boeing 707 model.

Length Wingspan	$L = L'p = 2.0 \times 25 = 50m$ $L = L'p = 1.8 \times 25 = 45m$
Frequency	$f = f'/p$ $f' = 2 - 18$ GHz, $f = 80 - 720$ MHz
RCS	$\sigma = \sigma' p^2$

In this table, p is the scale factor of 25, unprimed variables represent full scale dimensions and primed variables represent the scaled down dimensions. Due to the frequency scaling effect these measurements represent full scale frequencies of 80 MHz to 720 MHZ, and the results are thus applicable to VHF and UHF radar designs. Figure 10.7 and Figure 10.8 show the Front view and side view of the Boeing 707 mounted in the compact range. Note that it has been traversed to 180° for the photographs, and is thus effectively facing backwards.

Figure 10.7: Front view of 1:25 scale Boeing 707 model in the University of Pretoria's compact range.

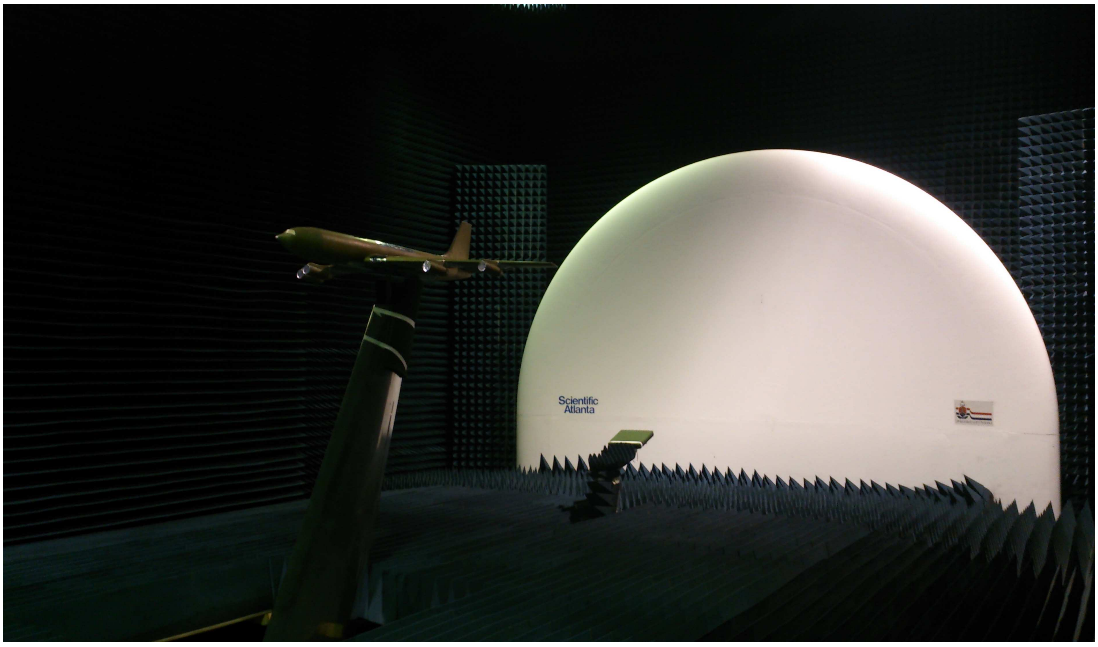


Figure 10.8: Side view of scale Boeing 707 model in the compact range.

The recorded RCS data is shown in Figure 10.9 for both polarizations. This data was used as the input data to the MI calculation. The measured data was converted to HRRP's and the two ring plots of the result are shown in Figure 10.10 for VV polarisation and Figure 10.11 for HH polarisation. The MI results for VV polarization, HH polarization and combined VV and HH polarization are sown in Figure 10.12.

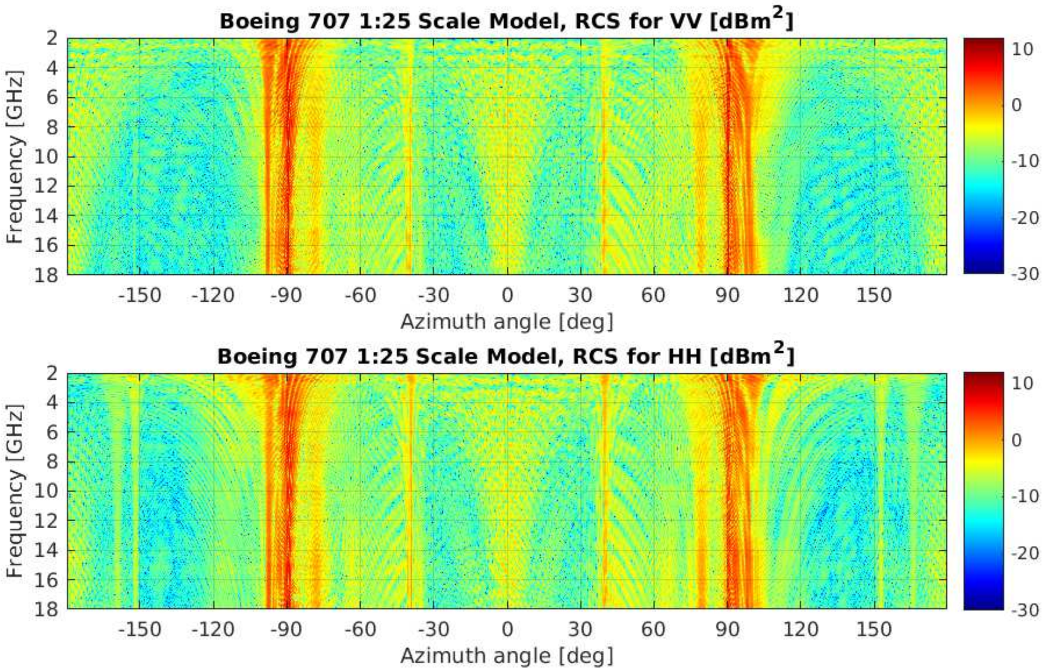


Figure 10.9: Measured RCS of 1:25 scale Boeing 707 vs frequency and azimuth angle for VV (upper plot) and HH (lower plot) polarisation.

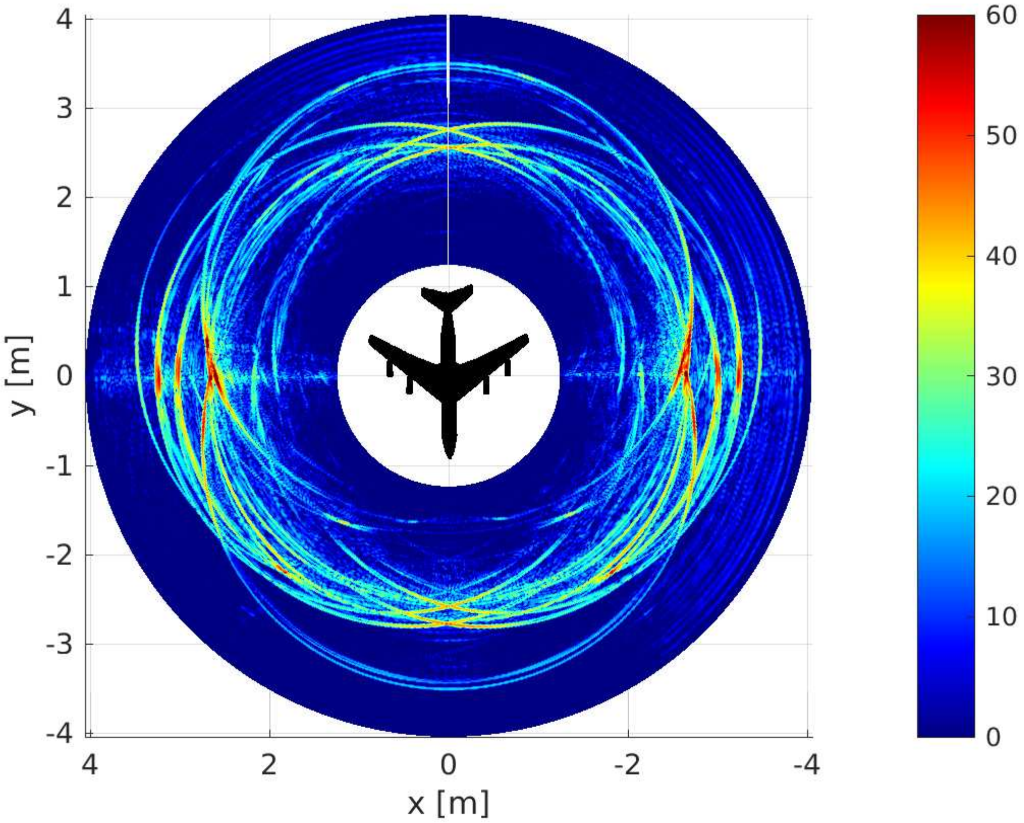


Figure 10.10: Ring plot of the HRRPs for the 1:25 scale Boeing 707 for VV polarisation.

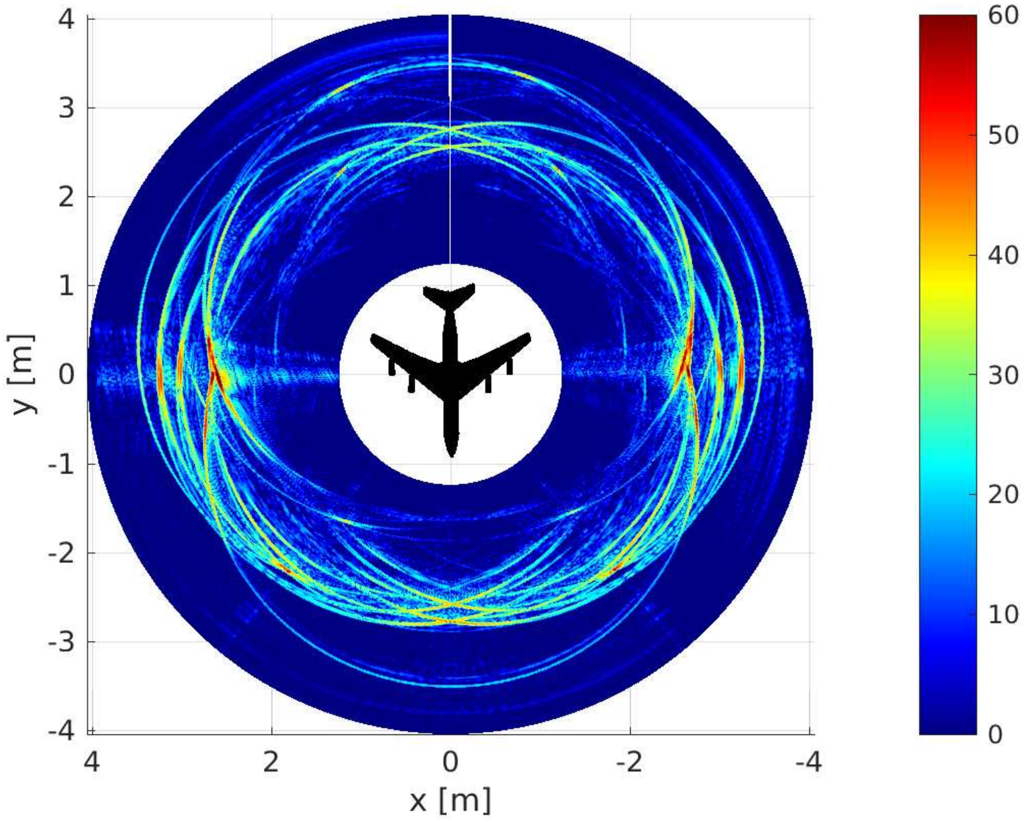


Figure 10.11: Ring plot of the HRRPs for the 1:25 scale Boeing 707 for HH polarisation.

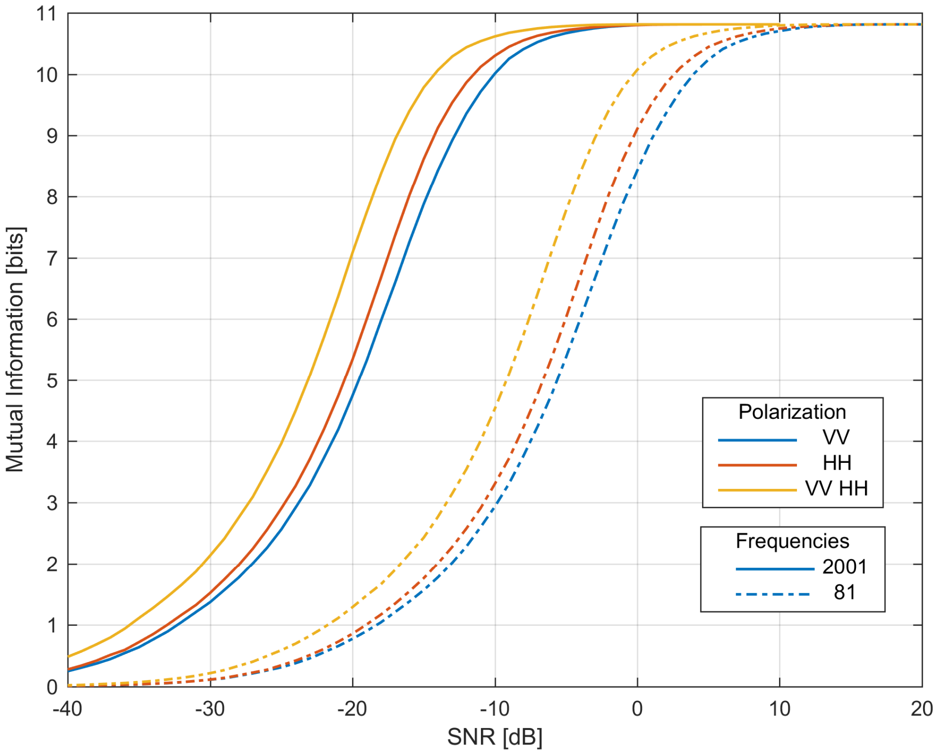


Figure 10.12: MI results versus SNR for the scale model Boeing 707 for various polarizations.

The multidimensional expression for MI was used to evaluate the MI for each of the three polarisation options (VV, HH and dual linear polarisation) by means of Monte-Carlo integration which was performed over a 4002-dimensional Gaussian density for each of the linear polarisations and over a 8004-dimensional space for the dual linear polarisation case. For the dual linear case, a vector was formed by concatenating the vectors for the two linear polarisations for each azimuth angle.

From the results in Figure 10.12 it can be seen that the HH polarization performs approximately 1.5 dB better than the VV Polarization. The dual linear, or HH in combination with VV, exhibits a 4 dB improvement over the VV polarisation and a 2.5 dB improvement over the HH polarisation case. The average performance gain when going from linear to dual linear is thus approximately 3.25 dB, which is approximately equal to the non-coherent sum of the two channels. From this it can be concluded that at these low frequencies polarisation does not lead to the large performance increases seen in the previous section for an X-band radar. The fact that this data set was measured at 2001 frequency points compared to 81 for the simulated data sets for the F-14 allows the MI calculation to exploit extra coherent integration gain. This leads to the substantially improved performance in terms of SNR in Figure 10.12 where the MI .curve reaches its maximum value at a SNR of 0 dB. To further evaluate the difference between this model and the F-14, the 2001 frequency set was sub sampled by a factor of 25, giving exactly 81 frequencies. This reduces the range ambiguity from 18.75 m to 0.75 m, which is less than the length and wingspan of the Boeing 707 model, which are on the order of 1.8 m. The SNR difference between each pair of curves was calculated over the range of the MI values and gave a result between -13.87 dB and -14.1 dB for all three pairs of curves. This corresponds to the SNR loss which is obtained by having decreased the number of samples by a factor of 25, and is given by $10 \times \log_{10}\left(1/25\right) \approx -13.98\ \mathrm{dB}$. This shows that no information has been destroyed in the subsampling process over and above the loss caused by the removal of 24 out of 25 samples.

This experiment demonstrates that the MI calculation technique is stable for very high dimensional signal spaces. In this case the analysis for the dual linear polarisation data set was conducted in 8004-dimensional signal space. The experiment also shows that the MI calculation is stable and valid for measured data, although the results have to be interpreted more carefully. For example, if the target interacts electromagnetically with any objects in the measurement chamber the extra signals induced in the data set will corrupt the final MI values. Interactions which create constant values which add to the data set, such as background reflections from the chamber, will not influence the MI results, but effects such as the target interacting with the support structure, which changes as the target is rotated, will add varying signals to the recorded data which will in turn influence the final MI values which are calculated. The setup of the target in the measurement chamber is thus critical and should be carefully controlled and characterised.

10.3 Experiment #15: Gain in Information Content by Using the Sum and Azimuth Difference Channels in a Monopulse Radar

This section studies the recognition performance gain which is possible through the exploitation of the target signal in the azimuth difference channel and in addition compares the effect of polarization for an X-band HRR monopulse radar. A previous study by Armstrong and Griffiths [Armstrong1991] has shown that the information contained in the monopulse difference channels can be exploited to improve the detection performance of a radar. Borden, [Borden1995a], illustrated by means of simulation that the long term tracking statistics and angle estimations can be combined to form a 3-D pseudo-image of the target being tracked. Tait [Tait2005] also describes in principle how HRRP's on all three monopulse channels can be combined with high accuracy angular measurements to create a pseudo 3-D image of a target [Tait2005, pp. 233]. These techniques could be applied to generate the input to a machine based target classification system.

Standard monopulse processing makes use of two collocated receive antenna patterns to enable the radar to calculate the angular location of a target within the beam of the radar. This technique makes use of a sum and difference beam pattern and associated receive RF paths to generate a ratio of the difference voltage to the sum voltage. The resulting ratio is linearly related to the off-boresight angle of the target being tracked. This processing is usually implemented for narrow instantaneous bandwidths which correspond the radar's native range resolution. When use is made of HRR techniques to increase the range resolution, monopulse processing becomes conceptually cumbersome from a calibration point of view. However, due to the fact that the difference channel senses a different spatial energy distribution over the target's angular extent to the sum beam, the difference channel can be exploited as an additional source of information about the target. This section aims to quantify the amount of information which can be gained from the difference channel without specifying the type of processing or recognition algorithms which will be implemented in the radar. This represents a bound on the maximum performance gain which can be achieved by exploiting the electromagnetic scattering characteristics of the target in the difference channel.

Use was made of the same F-14 CAD model as the previous sections. The outer shell of the aircraft was once again modelled as a perfectly electrically conducting (PEC) material. Effects caused by the canopy, radomes and radar absorbing material (RAM) are thus not modelled accurately. Also note that the engine models were relatively inaccurate. The model was analysed using SigmaHat [Smit2012a], to obtain the F-14's frequency response over a set of discrete frequencies. SigmaHat was set to apply Physical Optics (PO) and shooting and bouncing rays (SBR), but not edge diffraction. The simulations were run to generate the full scattering matrix and the target was illuminated at a set of 116 frequencies centred at 10 GHz in steps of 7 MHz (i.e. 9.6 GHz to 10.405 GHz). This gives a total bandwidth of 805 MHz, resulting in a range resolution of 0.186 m. The target RCS was calculated in 2° steps in azimuth for angles from -180° to +178°, at an elevation angle of 0°. At each azimuth angle the spatial distribution of the RCS was also recorded for each frequency. Use was made of 440 pixels in azimuth and 100 pixels in elevation, and the image spanned 22 m in azimuth and 5 m in elevation. This resulted in square pixels which were 5 cm x 5 cm, over which the scattering values are averaged. Examples of these images are shown in Figure 10.13 and Figure 10.14 for aspect angles of 0° and -10° respectively.

These images were then weighted by the sum and azimuth difference patterns for the monopulse antenna for each frequency. The complex valued pixels in the resultant images were then summed to give a single value for each frequency, thus forming the frequency response of the target for each antenna channel. Note that for this study the antenna patters were taken to be constant in elevation. The monopulse patterns were calculated as a function of frequency using first order approximations for the beamwidth as a function of frequency. The simulated antenna patterns at 10 GHz are shown in Figure 10.15 below.

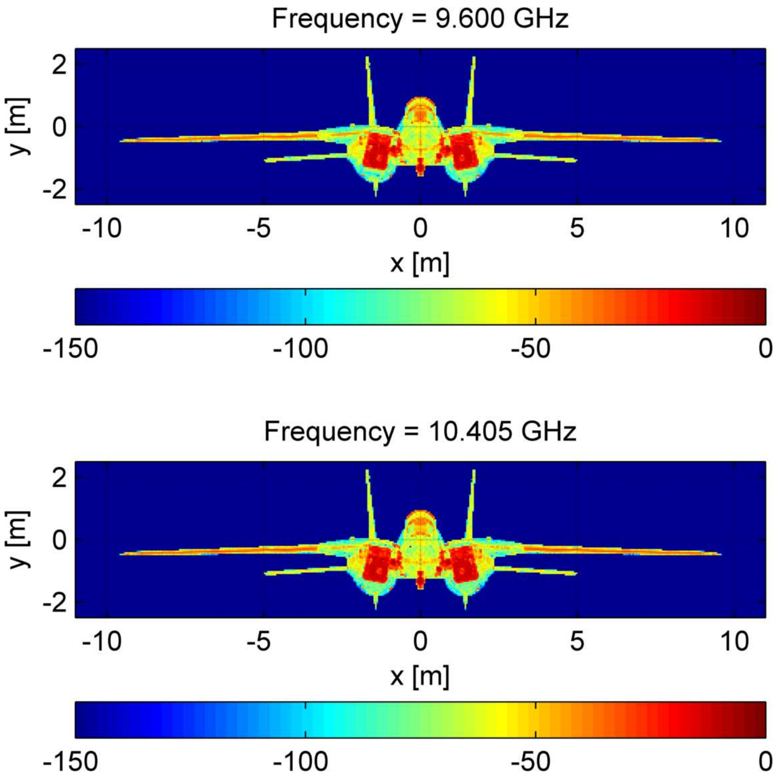


Figure 10.13: Reflectivity image for F-14, VV polarisation, head on, for the minimum and maximum illumination frequencies.

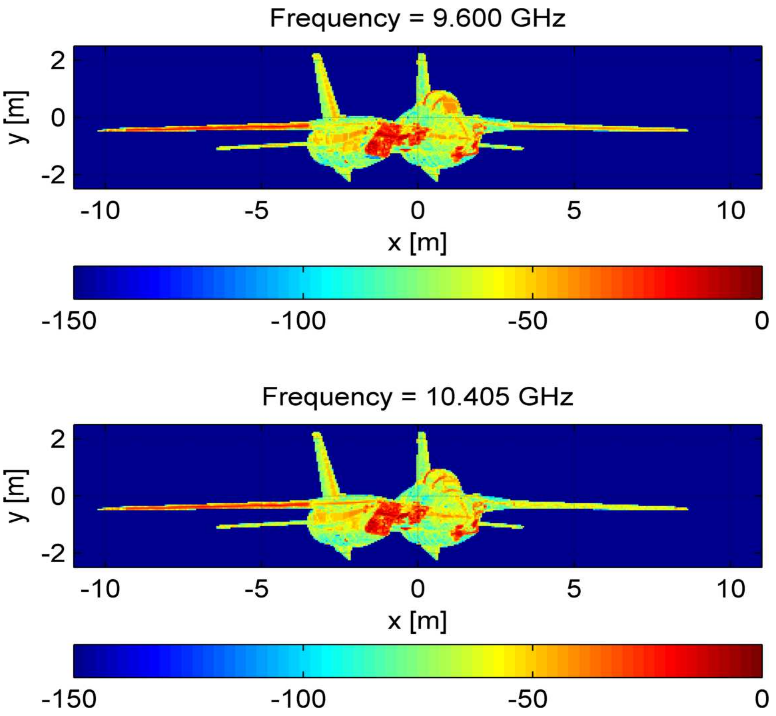


Figure 10.14: Reflectivity image for F-14, on VV polarisation, for an azimuth angle of -10°, for the minimum and maximum illumination frequencies.

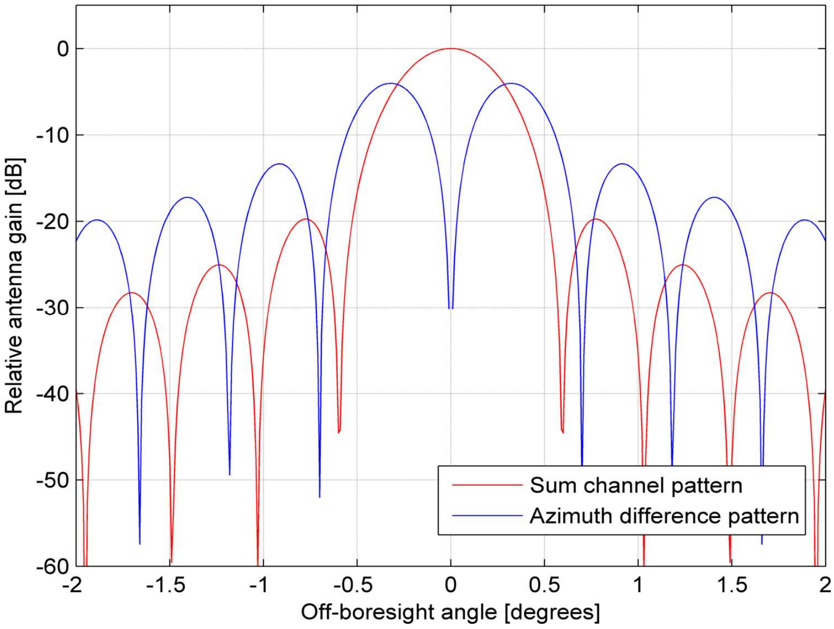


Figure 10.15: Example of the sum and azimuth difference monopulse antenna patterns at 10 GHz.

The parameters for the antenna were chosen as follows:

- antenna diameter: 3.5 m,
- antenna illumination: uniform,
- sum channel beamwidth: 0.49°.

The beamwidth of the constituent beams was 0.42° with an offset of 0.19°.

The MI was evaluated by means of Monte-Carlo integration for the F-14 at 6 different ranges from the radar. These ranges were chosen as 1, 2, 5, 10, 20, 50 km, and the unweighted (i.e. no antenna pattern applied) images were used to generate a reference curve, as these correspond to the case where computational electromagnetic (CEM) results are used directly to evaluate a classification algorithm. The reference plots thus required the integration to be performed over a 232-dimensional Gaussian density for the reference images, and over a 464-dimensional Gaussian density for the two channel cases. The data sets were not converted to HRRP's, but were processed directly in the frequency domain. According to the data processing inequality this will produce the maximum MI value. The maximum value of the MI is always limited to the entropy of the underlying set of waveforms, which in this case is the number of azimuth angles at which target responses have been simulated. This means that the MI graphs reach a maximum value of $\log_2(180)$ which is approximately 7.4919 bits. The MI results for each target range are shown in Figure 10.16 for VV polarization and a zoomed in version of this plot is given in Figure 10.17. This aids in assessing the MI close to its maximum value which corresponds to a probability of correct classification of unity.

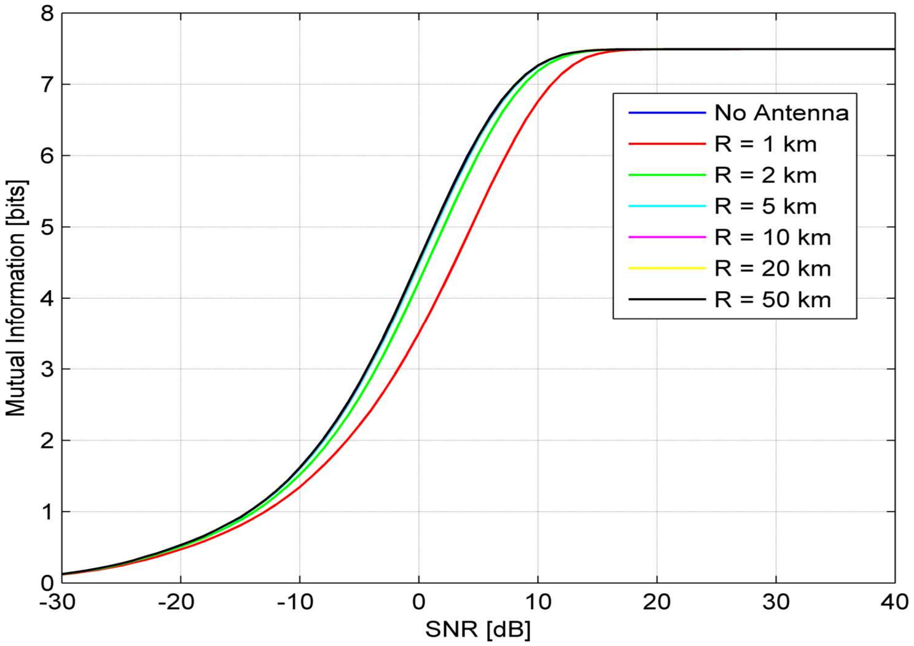


Figure 10.16: Mutual information (MI) for the sum channel as a function of SNR at various target ranges, with the unweighted image as a reference for VV polarisation.

From Figure 10.16 it can be seen that the antenna pattern causes a loss in MI, which is most pronounced at close range. This is due to the extremities of the target being heavily attenuated by the far out regions of the main lobe and thus reducing the SNR of scatterers in these regions. At an MI level of 7.417 bits, which corresponds to 99% of the final MI value (probability of error is 0.0045), there is a loss of approximately 0.7 dB for the 2 km case, and a loss of approximately 1.5 dB for the 1 km case. Note that the angular extent of the F-14 at 1 km is approximately 1.15°, so the target spans the null to null beamwidth of the sum channel.

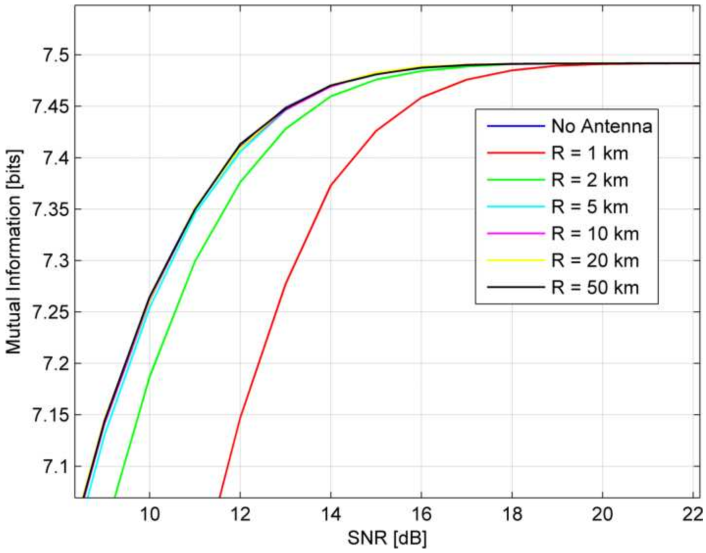


Figure 10.17: Zoomed in version of MI versus SNR at various target ranges, with the unweighted image as a reference for VV polarisation.

It is interesting to note that as the target range decreases there is a loss in SNR for a fixed value of MI. This is due to the effect of the antenna pattern reducing the magnitude of some of the scatterers, especially at close ranges, which effectively reduces the SNR of the target. At longer ranges there is very little loss induced by the beam shape, so the “no antenna” curve and the 5, 10, 20 and 50 km target range curves form a single line in the plot. The target’s angular extent at 5 km is 0.23°, which is approximately 47% of the 3 dB beamwidth

Figure 10.18 and Figure 10.19 give the MI as a function of SNR when the azimuth channel is included in the calculation for VV polarisation. The “No Antenna” curve shows the MI for the sum channel only and acts as a reference curve. Figure 10.19 is again a zoomed in version of Figure 10.18. These graphs show that there is a gain of approximately 3 dB over the “No Antenna” (sum channel only with no antenna weighting) case.

In Figure 10.19, for the high MI cases, the gain over the “No Antenna” case is as much as 4.5 dB for the 1 km and 2 km target ranges. This is due to the fact that the target is close enough to the radar that its angular extent spans the peaks in the difference beam. As the range of the target from the radar increases, the angular extent decreases, and the target return in the azimuth channel is highly attenuated by the null in the azimuth pattern, thus decreasing the magnitude of the scattering in this region, and in turn decreasing the effective SNR in this channel. At long enough ranges the null in the difference channel will suppress the whole target return and it is expected that the MI will reduce to the point where it coincides with the “No Antenna” curve.

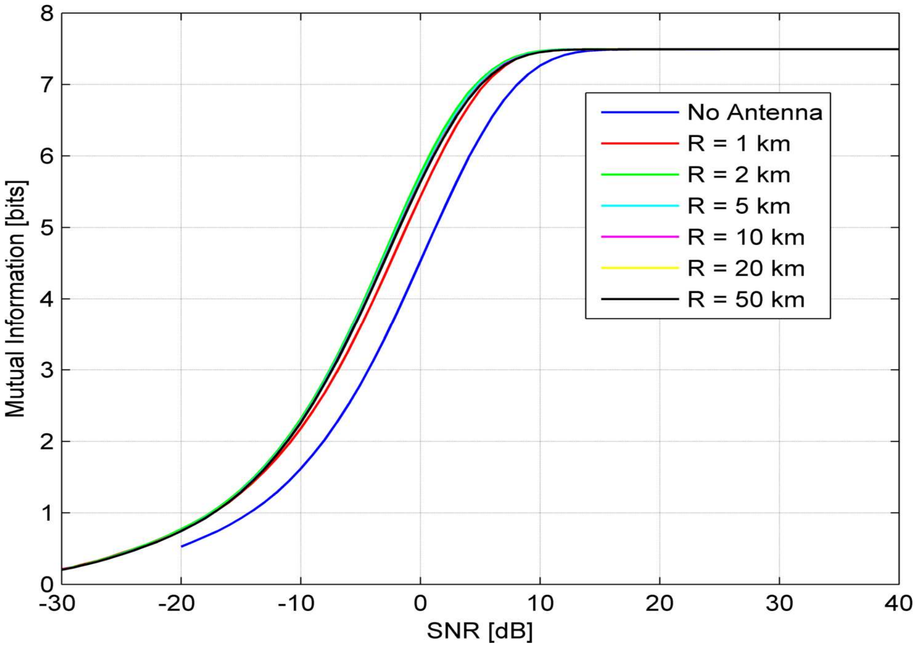


Figure 10.18: Mutual information (MI) for the sum and azimuth channels as a function of SNR at various target ranges, with the unweighted image as a reference for VV polarisation.

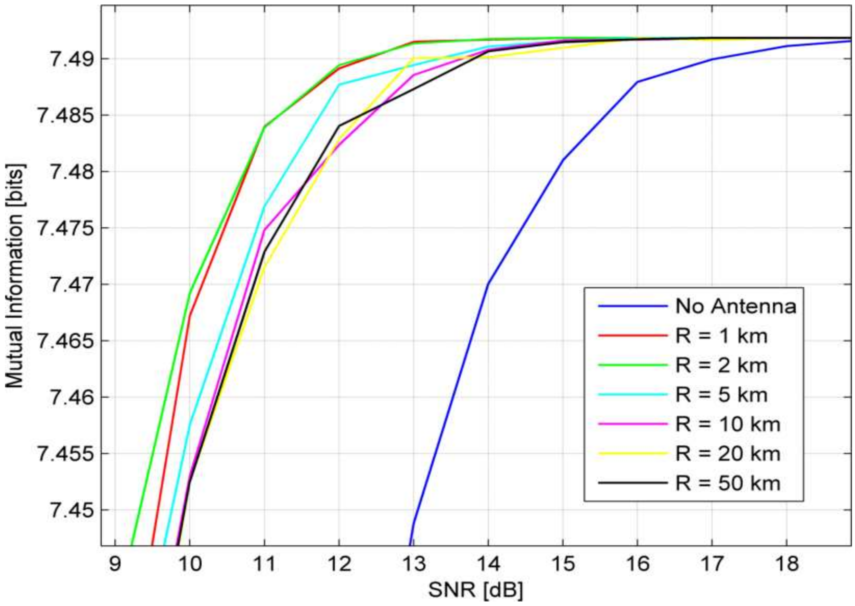


Figure 10.19: Zoomed in version of MI versus SNR for the sum and azimuth channel at various target ranges, with the unweighted image as a reference for VV polarisation.

Figure 10.20 and Figure 10.21 compare the MI when using both the sum and azimuth channel for the VV and HH linear polarizations for target ranges of 2 km and 50 km. Figure 10.21 is once again a zoomed in version of Figure 10.20.

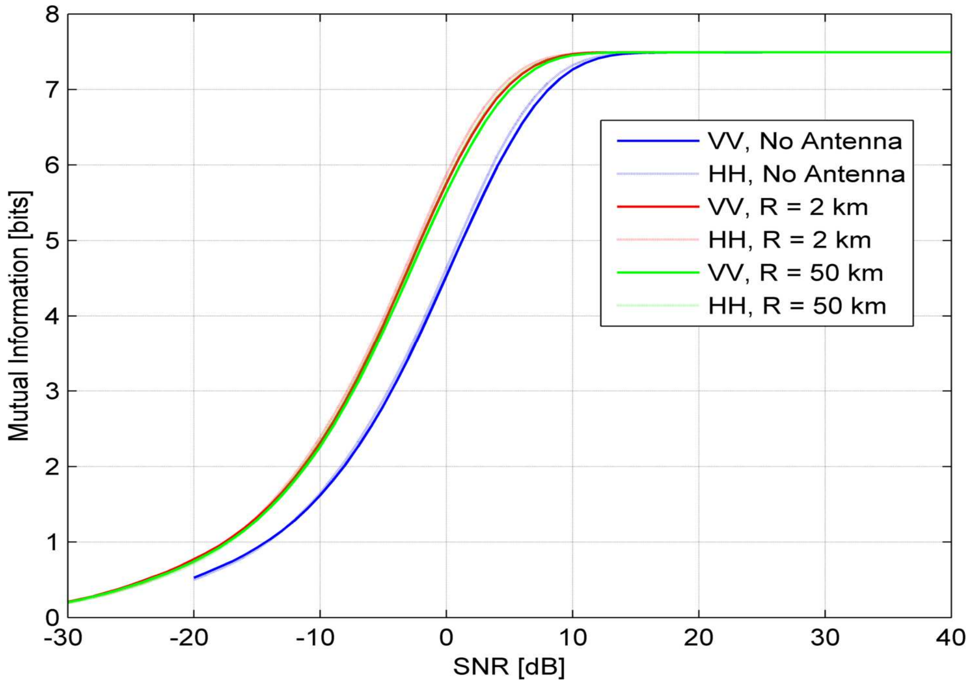


Figure 10.20: Comparison of the effect of VV and HH polarisation on the MI for the sum and azimuth channels as a function of SNR at various target ranges, with the unweighted image as a reference for each polarisation respectively.

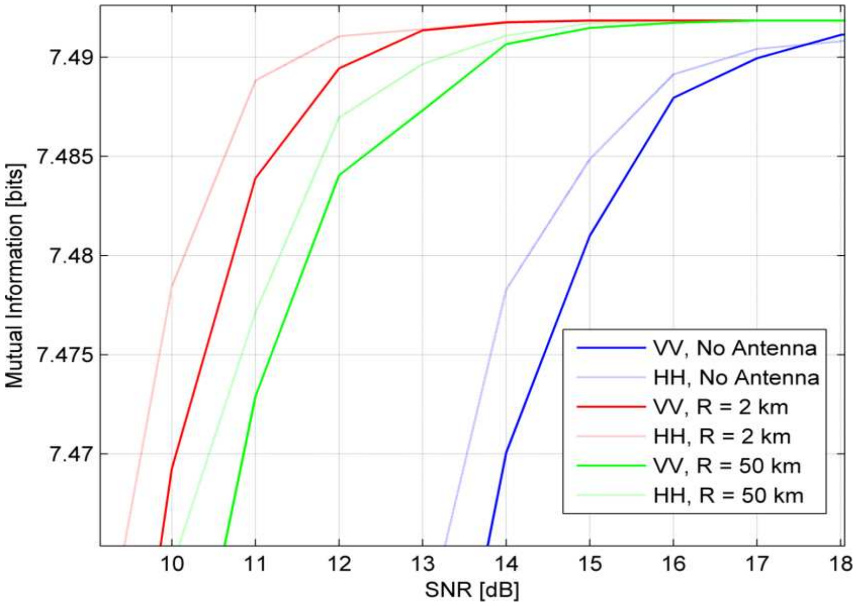


Figure 10.21: Zoomed in version of the comparison of the effect of VV and HH polarisation on the MI for the sum and azimuth channels as a function of SNR at various target ranges.

The 3 dB gain relative to the “No Antenna” curve is evident for both polarisations, and there is a slight gain of approximately 0.5 dB when using HH polarisation over VV polarisation. It is interesting to note that the results presented in section 10.1 showed that VV had a 1.5

dB gain over HH for the F-35 and that the two polarisations were approximately equal for the F-18. The best linear polarisation to use is thus very dependent on the target geometry.

This section has evaluated the increase in recognition performance which can be achieved by making use of the azimuth difference channel signal in a monopulse antenna. It was shown that performance gains of as much as 4.5 dB are achievable at short ranges of 1 km and 2 km, and that gains of 3 dB are achievable at longer ranges up to 50 km for this target and antenna setup. Note that these gains include the SNR advantage due to the longer vector which is available for matched filtering in case where the sum and difference channel signals are used.

The results presented in this section have shown that a 3 dB gain could be achieved with relatively minor changes to an existing monopulse radar. This study also shows that the recognition algorithm should make use of different reference target scattering libraries depending on the range of the target.

10.4 Conclusion

This section has presented results for the achievable information gain from multi-channel radar receivers. Examples were analysed for two polarization cases and a monopulse radar case.

The first experiment analysed the recognition advantage of sensing the full polarization matrix over sensing only one of the linear polarisations for the F-18 and F-35 fighters when illuminated at X-band. The target responses were simulated using SigmaHat for all four linear polarisations and each target was illuminated at a set of 116 frequencies centred at 10 GHz in steps of 7 MHz (i.e. 9.6 GHz to 10.405 GHz). This gives a total bandwidth of 805 MHz, resulting in a range resolution of 0.186 m. Each target was simulated at 360 azimuth angles in 1° steps for an elevation of 0°.

It was shown that the performance for the linear polarizations is very similar for F-18, but for the F-35 the VV polarisation had a 1-2 dB SNR advantage over HH. The achievable gain when using the full polarisation matrix was 5-6 dB for both targets. For a probability of error of 10^{-3}, the average gain going from linear to full polarisation is approximately 6 dB for the F-18 and 8.25 dB for the F-35.

The F-35 requires 9 dB more SNR at the maximum MI as compared to the F-18. Even using full polarisation for the F-35, still requires approximately 2.5 dB more SNR for recognition than the F-18 at high MI values. This is probably due to the stealth design of the F-35 and the fact that the F-18 had stores mounted under the wings whereas the F-35 carries these internally. The F-35 CAD model was also less complex than the F-18 and the F-18 also has more discrete scattering centres in its HRRPs.

The second experiment analysed measurements of a 1:25 scale model of a Boeing 707. These measurements were made in the University of Pretoria’s compact range where the HH and VV response of the target was measured over 2001 frequency points from 2-18 GHz. The MI results showed that average SNR gains when going from linear to dual linear polarisation were approximately 3.25 dB, which is relatively low. The performance gains going from VV and HH to dual linear polarisation were 4 dB and 2.5 dB respectively.

This analysis also demonstrated that the MI calculation is stable for very high dimensional problems, as up to 8004 dimensional data was used as input data for the dual linear case. This section also shows that the MI calculation can be applied to measured data if the experiment is carefully controlled.

The aim of the third experiment was to characterise the gain in information about a target when by exploiting the difference channel of a monopulse radar. The F-14 was used as a target and the spatial distribution of reflected energy over 22 m x 5 m area (horizontal x vertical = 440 x 100 pixels) was simulated using SigmaHat for 116 frequencies centres at 10 GHz in 2° azimuth steps. The pixels were coherently summed in the vertical dimension, and then weighted by the sum and azimuth antenna patterns in the horizontal direction and again coherently summed to form the sum and azimuth channel outputs. A constant antenna pattern in elevation was assumed.

The analysis made use of the case when there is no antenna pattern weighting applied to the image data as the baseline result. A loss at very close ranges 1-2 km of 1.5 to 0.7 dB respectively was observed when using the sum channel only. This was due to the beam shaping loss over target's angular extent. Very little loss was observed at ranges from 5 km to 50 km. When adding the difference channel in to the data set, performance gains of as much as 4.5 dB were achieved at short ranges of 1 km and 2 km. Gains of 3 dB were achieved at longer ranges of up to 50 km. The 3 dB gain could be achieved with relatively minor changes to an existing monopulse radar. This section of the study also highlighted the fact that the target reference libraries used by the recognition algorithm should be a function of the range of the target, especially at ranges at which the antenna pattern cannot be assumed to be constant over the target’s angular extent.

11 Characterisation of the Effect of Envelope Processing on the Information Content of High Range Resolution Profiles (HRRP)

This chapter makes use of the expression which was derived for the mutual information contained in a signal after envelope processing. The expression is repeated below for reference.

$$I\left(a_k, y\right) = \log_2\left(N\right) - \frac{1}{N\ln\left(2\right)} \sum_{k=0}^{N-1}\sum_{j=0}^{N-1} s_{kj}^2 + \frac{1}{N}\sum_{k=0}^{N-1} E\left\{ \log_2\left(\frac{\prod_{j=0}^{N-1} I_0'\left(\frac{r_{kj}s_{kj}}{\sigma^2}\right)}{\sum_{i=0}^{N-1}\prod_{j=0}^{N-1}\exp\left(\frac{s_{kj}^2+2r_{kj}\left(s_{ij}-s_{kj}\right)-s_{ij}^2}{2\sigma^2}\right) I_0'\left(\frac{r_{kj}s_{ij}}{\sigma^2}\right)}\right)\right\} \tag{11.1}$$

Envelope processing is a very standard technique in radar signal processing chains and due to the fact that it is a non-invertible operation on the HRRP data, some of the information content of the complex valued HRRP will be lost due to this processing.

11.1 Experiment #16: Information Content of the Envelope of the HRRP

In this section the MI is calculated for the F-14, F-15 and F-16 for the case where the radar calculates the envelope of the received signal which is then used for target recognition. Due to the fact that this is a non-invertible operation, the signal processing step has effectively discarded any information contained in the phase of the signal. The input data from Section 7 was used for these experiments, thus allowing for the direct comparison of results between the two sections.

Figure 11.1 shows the MI for the coherent as well as non-coherent (labelled “NC” in the legend) case for each of the three targets as well as two bounds for reference. It is interesting to note how well the MI curves for the coherent and non-coherent case track each other for each target. To more accurately quantify the loss of information, the MI was measured at the point where it reaches 99.95% of its final value. These values are compared in Table 7, which shows that there is a loss of between 7.2 dB and 8.7 dB. This loss would require the SNR to be increased by this amount for the NC case to have the same performance as the coherent case.

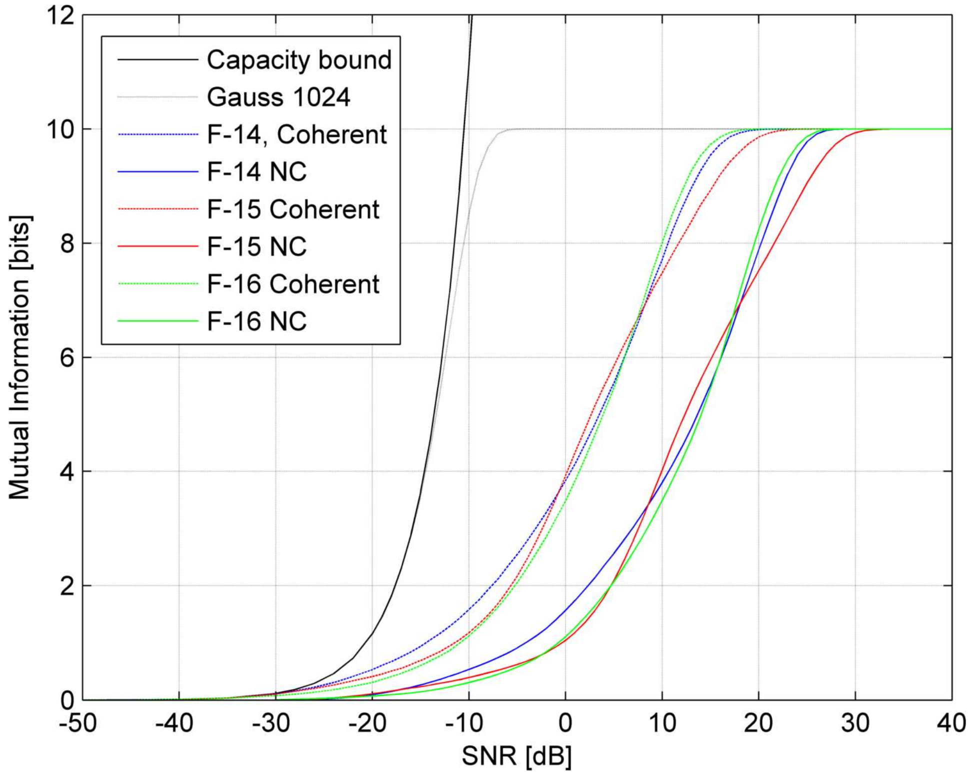


Figure 11.1: Comparison of the F-14, F-15 and F-16 Mutual information versus SNR for the coherent and non-coherent (NC) cases. The input data is the raw measurements per frequency.

To compare the SNR loss over the whole range of MI values, the MI curves were resampled at fixed MI points and the resulting SNR values subtracted for each target. The results of the calculation give the loss when going from coherent processing to non-coherent processing, and are graphed in Figure 11.2. The loss for the F-14 and F-16 track each other remarkably well for most of the range, but separate at MI values above 9 bits. For MI values between 9 bits and 10 bits, which is the range that target recognition algorithms would be interested in, the loss is between 7.5 dB and 9.5 dB.

Table 7: Comparison of SNR as MI reaches 99.95% of its maximum value of 10 bits for the coherent and non-coherent cases.

	Capacity Bound	Gauss 1024 waveforms	F-16	F-14	F-15
MI [bits]	9.995	9.995	9.995	9.995	9.995
SNR [dB], Coherent	-10.5	-5.2	19.3	21.0	24.0
SNR [dB], non-coherent	N/A	N/A	27.8	28.2	32.7
Loss [dB]	N/A	N/A	8.5	7.2	8.7

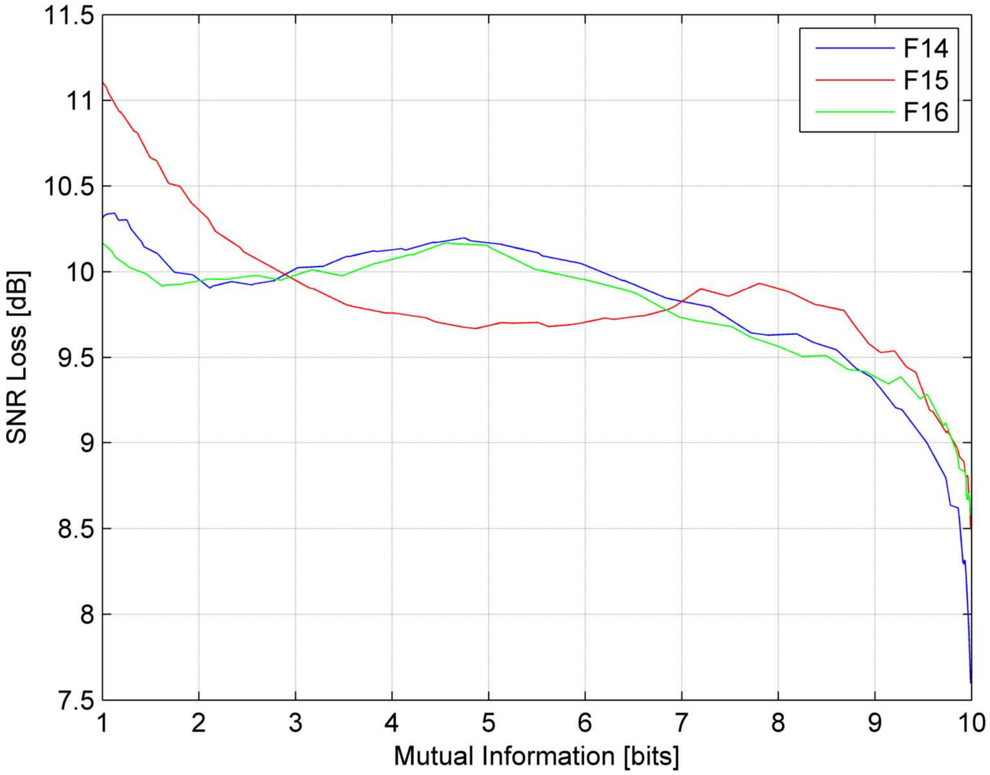


Figure 11.2: SNR loss versus MI, showing the loss caused by non-coherent processing for the F-14, F-15 and F-16.

The fact that the MI curves for the non-coherent (NC) cases in Figure 11.1 still reach the maximum value of 10 bits means that there is still enough information left in the amplitude of the signal to unambiguously identify all 1024 waveforms. This just requires a much higher SNR than for the case where the waveforms are complex valued.

11.2 Experiment #17: Effect of Pre-processing on the Information Content after Envelope Processing

To further investigate and quantify these losses, HRRP processing was executed on the raw data, and then the MI was re-calculated. This was achieved by taking an unwindowed FFT of the raw data. The FFT was not windowed as this would have caused correlation in the noise, which was not included in the derivation of the NC case.

The results for HRRP processing followed by envelope processing are compared to the original performance of the envelope processing on the raw frequency domain data in Figure 11.3 below. Once again the graphs have been resampled to generate the graph in Figure 11.4 which shows the SNR gain achieved using this processing.

This result shows that there is a 2-5 dB SNR gain at MI values below 9 bits due to having executed HRRP processing before discarding the phase of the signal. Unfortunately, this gain drops to zero as the MI approaches its maximum value of 10 bits. In the case of the F-14 the gain becomes a 0.3 dB loss at a MI value of 10 bits.

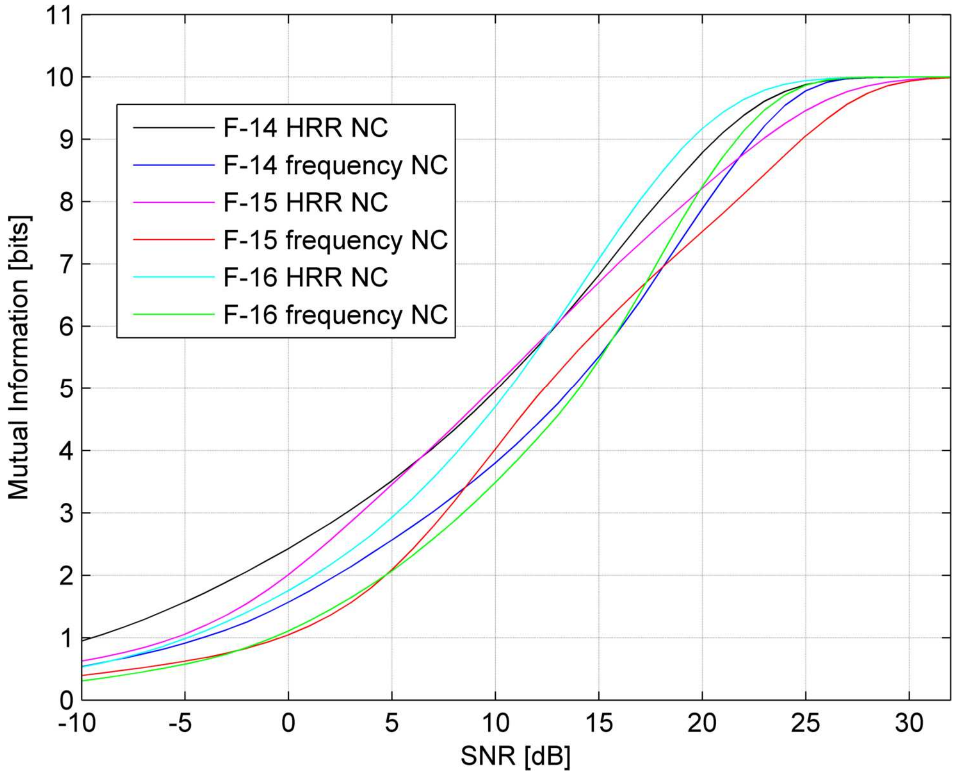


Figure 11.3: Comparison of MI for the three target aircraft when HRRP processing is executed before envelope processing.

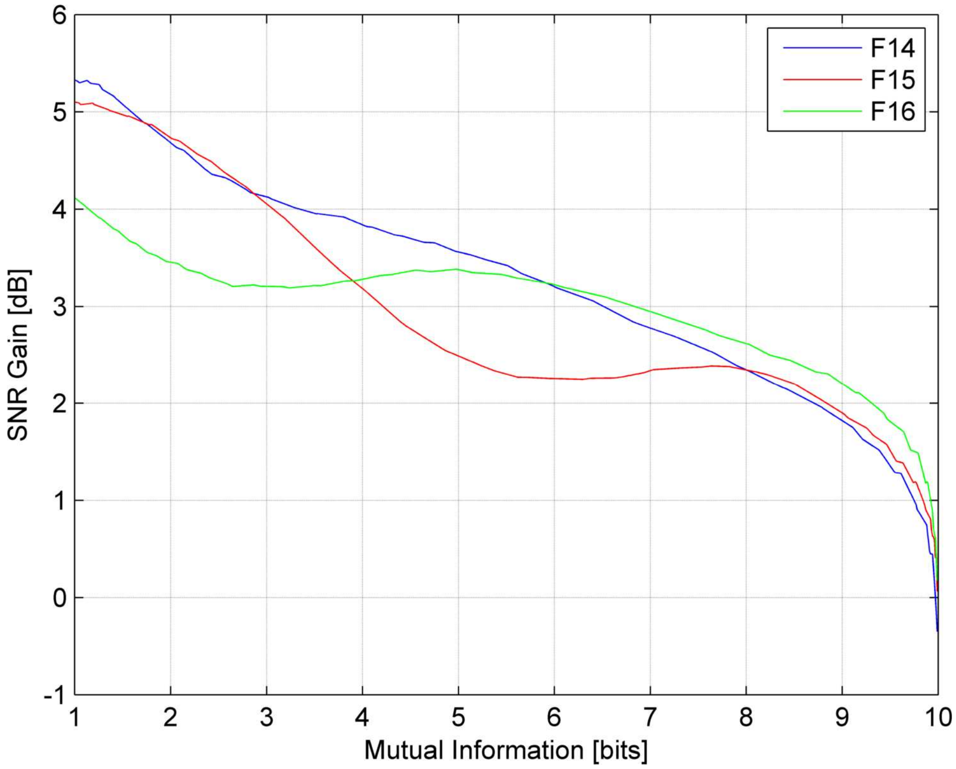


Figure 11.4: SNR gain versus MI for the three target aircraft when HRRP processing is executed before envelope processing.

The implication of this result is that if the recognition algorithm is expected to operate at intermediate values of MI between 4 bits (probability of error of 0.5) and 9.5 bits (probability of error of 0.03) it is worthwhile for the radar signal processor to generate the HRRP as this will give at least a 1.5 dB SNR gain. From a MI value of 9.978 bits (probability of error of 0.001) and upwards it is better not to generate the HRRP of the target and rather use the raw frequency domain data in the recognition process.

To explore whether further gains were possible, the FFT was interpreted as a matrix mapping of the data into a new vector space. This in turn meant that any matrix mapping could be used after which the MI was evaluated. For the next result, use was made of the mapping generated by means of applying principle component analysis (PCA) to the raw frequency data and then applying envelope processing. The result for the F-14 is shown in Figure 11.5 below.

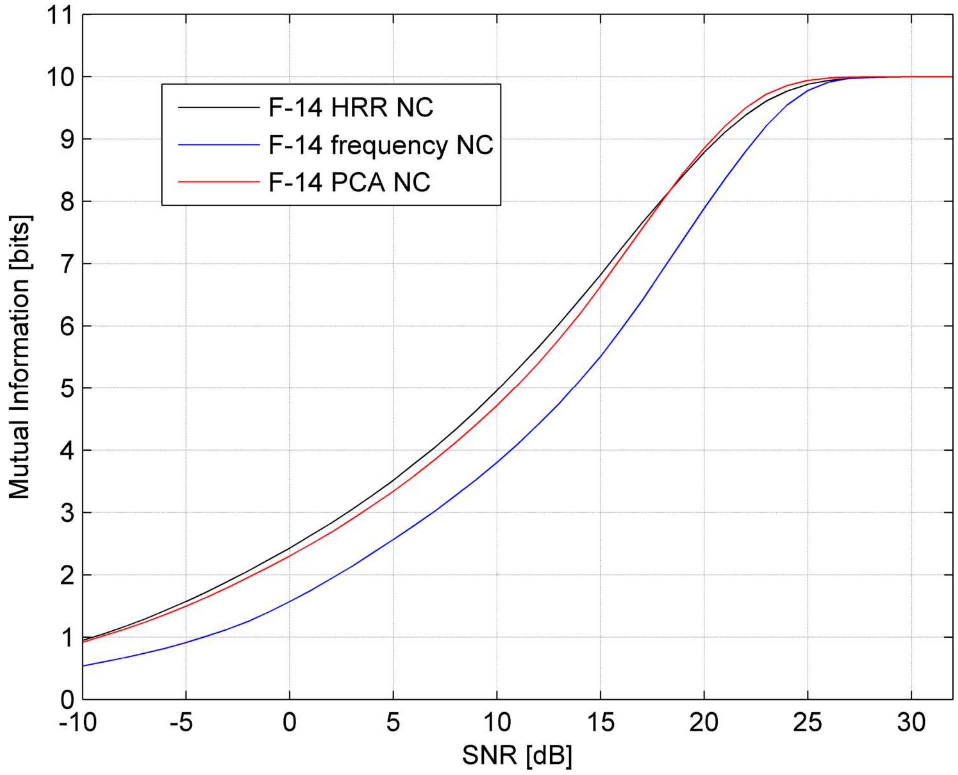


Figure 11.5: Comparison of MI versus SNR for the F-14 for HRRP processing and PCA processing followed by envelope processing.

From the graph it can be seen that the PCA processing causes a slight loss, relative to HRRP processing, over most of the MI range, but that for the high values of MI it exhibits a gain. This was further characterised by plotting the gain when going from the raw frequency domain data to the HRRP processed data and comparing it to the gain when going from the raw frequency domain data to the PCA processed data. This result is given in Figure 11.6. From this figure it can be seen that the PCA processing placed more emphasis on the high MI region of the graph. This led to a gain of more than 1 dB over the

HRRP result, however there is a loss of approximately 0.5 dB for PCA processing below an MI value of 8 bits.

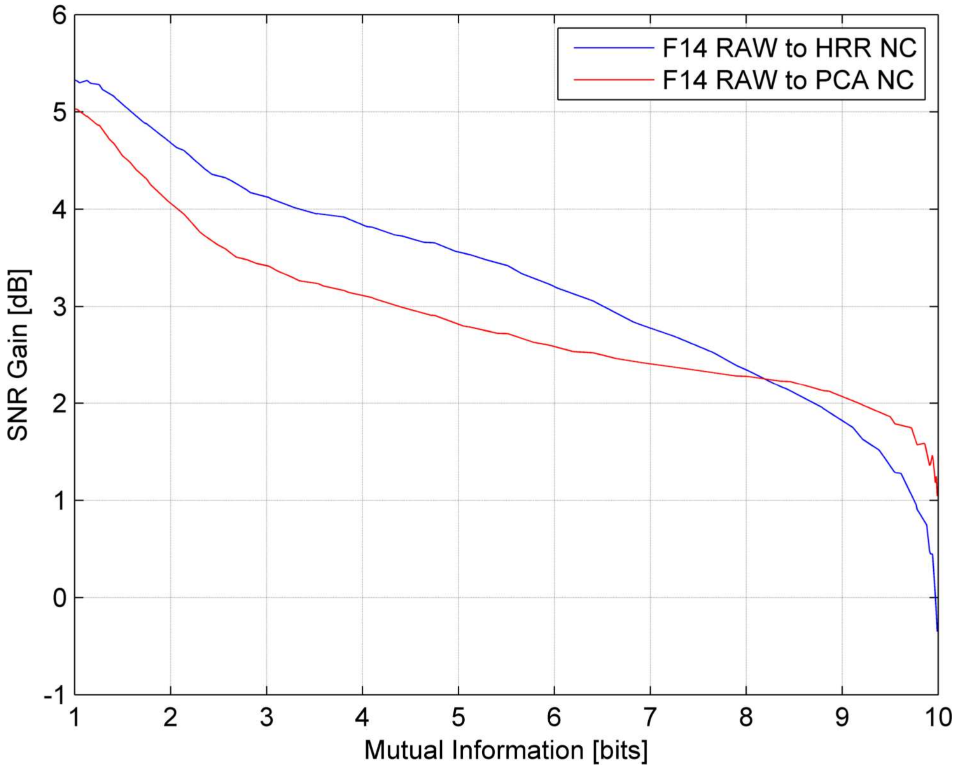


Figure 11.6: Comparison of the SNR gain versus MI for the F-14 between raw frequency domain data and HRRP processing and raw frequency domain data and PCA processing before non-coherent processing.

The results presented in this section have demonstrated the types of analysis and trade-off studies which can be conducted by comparing the performance of coherent waveforms for target recognition with waveforms which have been preceded by various forms of linear processing and then non-coherently detected for recognition purposes.

The most interesting result in this section is that the information contained in the phase of the signal scattered from the target contains enough information to give a 7 dB to 8 dB SNR advantage over recognition algorithms which discard the phase information.

11.3 Conclusion

This section analysis the information lost due to envelope processing of the received signal. To further explore this topic the effects of executing linear processing on the signal before envelope processing are analysed. Two types of linear processing, namely HRRP (FFT) and PCA are analysed by quantifying the loss or gain in mutual information at the output of the envelope processing step.

The first section discusses the loss in MI caused by envelope processing by calculating the MI for the F-14, F-15 and F-16 data sets generated in Section 7. The loss in SNR required to reach an MI level of 99.95% (probability of error = 2.12×10^{-4}) of the maximum of 10 bits was quantified for the three targets. The loss is 7.2 dB for the F-14, 8.5 dB for the F-16 and 8.7 dB for the F-15.

The fact that the MI curves for the non-coherent (NC) cases still reach the maximum value of 10 bits means that there is still enough information left in the amplitude of the signal to unambiguously identify all 1024 waveforms. This just requires a much higher SNR than for the complex valued case. The fact that the phase of the target return contains enough information to give a 7 dB to 8 dB SNR advantage over the envelope only signal, makes research into recognition algorithms which exploit the complex valued target return a worthwhile pursuit.

The next two sections compare the performance of various linear pre-processing algorithms applied to the signal before envelope detection. It is shown that if high levels of performance (low error rates) are expected of the recognition system it is better not to generate the HRRP from the frequency domain data and to rather perform envelope processing directly on the frequency domain data. It was also shown that if use is made of the PCA algorithm rather than the HRRP processing, then the PCA outperforms the HRRP and the raw frequency domain data by approximately 1 dB at high values of MI. At intermediate MI values (below 8 bits) the HRRP processing outperforms the PCA processing by approximately 0.7 dB.

This result also gives insight into the performance of algorithms as a function of SNR. For the processing options analysed, it would be better to make use of HRRP processing at SNRs below 20 dB, and then make use of PCA at SNRs over 20 dB. It is very seldom that designers of radar target recognition algorithms consider switching algorithms as a function of SNR.

12 Effect of Multi-target Scenarios on Information Transfer of a Target

Up until this point the results and discussions have focussed mostly on cases where only a single target is present and the radar is only set up to recognize that target. This type of analysis has been used to quantify and compare the amount of information transferred to the radar about a specific target when various multi-channel measurements are made by the radar or when the radar discards phase information in the received signals.

In this section the scenario will be analysed where the radar has been setup to recognize more than one target. This implies that if points in the signal space belonging to two different targets are close to each other in the signal space, then they will increase the probability of misclassifying both targets. Use is made of the multi-target MI expressions developed in Section 5.6 to execute this analysis. Fano’s inequality will be used to quantify the effect of multiple targets on the probability of error.

12.1 Experiment #18: Multi-target information using a Single Target and Modified Versions of the same Target

For an initial set of tests, use is made of the F-14 dataset of signatures which were calculated for vertical polarization on transmit and receive and the illumination waveform was a set of 81 frequencies centred at 10 GHz in steps of 6 MHz (i.e. 9.76 GHz to 10.24). This gives a total bandwidth of 480 MHz, which in turn translates to a range resolution of 0.3125 m.

To test the multi-target MI formulation use was made of the F-14 dataset for the first target and a slightly modified version of the F-14 dataset for the second target, where Gaussian noise had been added to the second set of F-14 signatures. By adjusting the variance of the Gaussian noise the distance between the two signal sets representing the two targets could be varied. Note that the perturbation was only added once and then the new set of signatures were stored in files for later use in the MI calculation.

The original set of F-14 data consisted of 1024 signatures, thus resulting in a maximum MI value of 10 bits. This is also equal to the entropy of 1024 equally likely events and this entropy defines the maximum value that the MI can reach. For the two target scenario it is assumed that all 2048 waveforms are equally likely, but that the radar receiver is only interested in the 1024 of these which correspond to the target being sensed by the radar. It is informative to note at this point, that due to the symmetry of MI, this situation could be interpreted firstly as a radar which has been configured to recognize a single target but it is possible to observe either of the targets, or secondly, the radar has been configured to recognize both targets but only one is present in the observation

Returning to the first case, the entropy is calculated by assuming all 2048 signatures are equally likely in the input space, but only signatures from one of the two sets of 1024 signatures will occur. This results in an entropy of

$$\begin{aligned} H(X) &= -\sum_{x\in R_X} p(x)\log_2\left(p(x)\right) \\ &= -\frac{1024}{2048}\log_2\left(\frac{1}{2048}\right) = -\frac{1}{2}(-11) = 5.5 \quad \text{bits} \end{aligned} \tag{12.1}$$

Four competing sets of F-14 waveforms were generated with Gaussian noise added at the following levels below the average power of the F-14 waveforms: 30, 50, 80 and 90 dB. The resulting MI curves for this setup are shown in Figure 12.1. The figure shows that the F-14 alone reaches a maximum MI of 10 bits. All the MI curves for the two target case limit at 5.5 bits as expected.

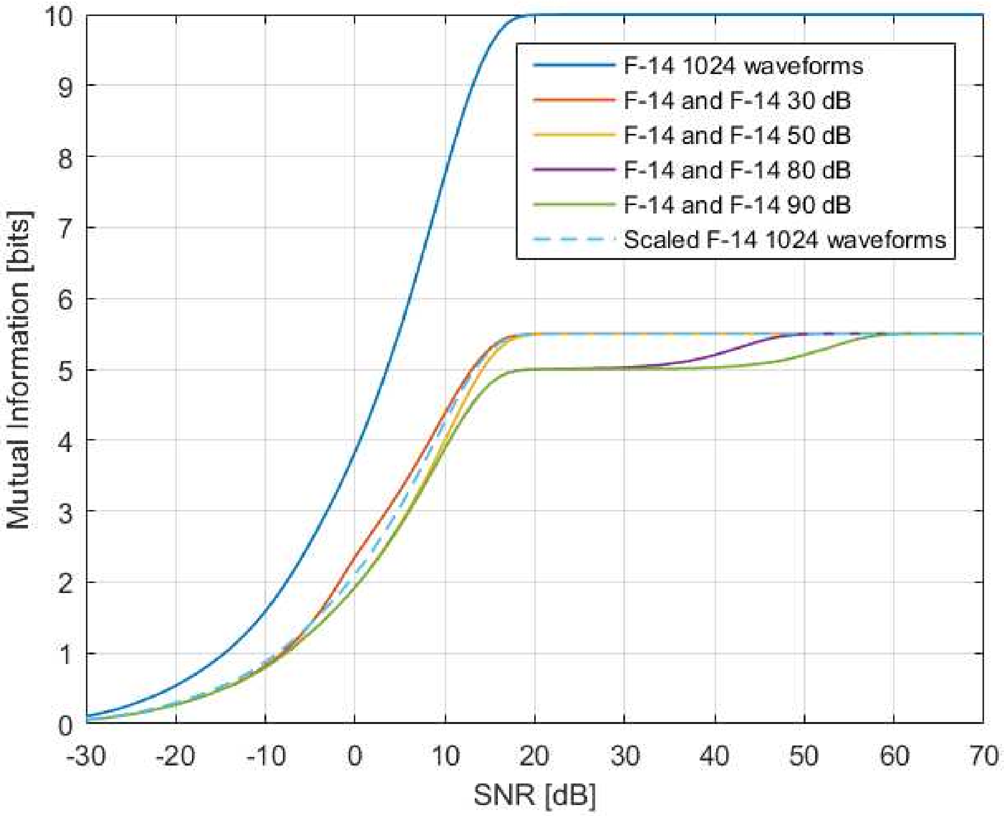


Figure 12.1: Multi-target MI versus SNR test cases using the F-14 for target #1 and the F-14 with Gaussian noise added at 30 dB, 50 dB, 80 dB and 90 dB below the mean power in the F-14 profiles for target #2.

The MI curve for the single F-14 was scaled by a factor of 5.5/10 and plotted as a dotted line for easier comparison with the other curves. For the 90 dB two target curve, the MI first reaches a plateau of 5 bits at a SNR of approximately 20 dB, and then increases to its maximum value of 5.5 bits between an SNR of 40 dB to 60 dB. This behaviour is due to the fact that the signal points for the two sets of waveforms are so close to each other in the signal space that a SNR of 60 dB is required before the two targets can be recognized as separate targets.

In the first plateau region, between 20 dB and 40 dB, each signal point for target #1 is indistinguishable from the same signal point (same azimuth angle) for target #2. This is equivalent to having joined these two events, thus resulting in a probability space containing only 1024 equiprobable events. This gives an entropy of 5 bits. If the second target was an exact copy of the first, then the MI curve would reach a maximum value of 5 bits. For the 80 dB case, it can be seen that the transition section from 5 bits to 5.5 bits has moved to a lower SNR value by approximately 10 dB. This is due to the fact that the smallest distance between the same signal points has been increased and the two targets can thus be distinguished from each other at a lower SNR. For the 30 dB signal set, the transition point is at an SNR of approximately 20 dB. The MI curve thus transitions seamlessly to 5.5 bits, without forming a plateau region. For the 30 dB set of signal points the transition region is at an SNR of approximately 0 dB. It can be seen that the signal points for target #2 are far enough away of those of target #1 that the MI value is increased above that of the single F-14 target. The MI curves for these examples have been normalised to their associated maximum values, so that any MI curve can only reach a maximum value of unity, and have been plotted in Figure 12.2.

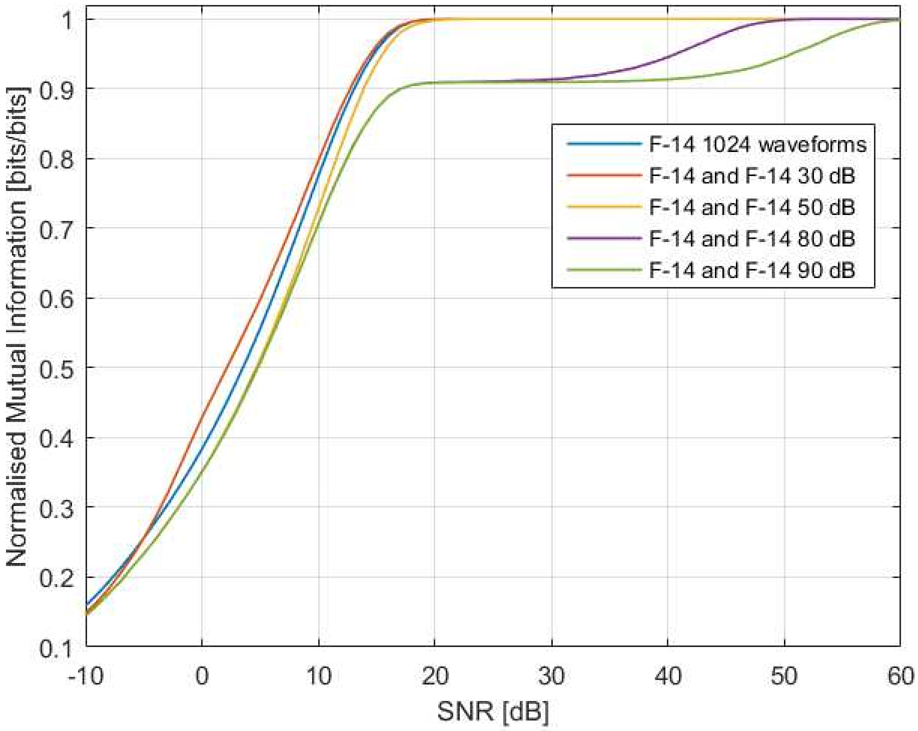


Figure 12.2: Multi-target Normalised MI test cases using the F-14 for target #1 and the F-14 with Gaussian noise added at 30 dB, 50 dB, 80 dB and 90 dB below the mean power in the F-14 profiles for target #2.

The set of waveforms for each target of interest can thus be analysed against the sets of waveforms associated with competing targets. If any of the signal points for two competing targets are close to each other in the signal space, then the MI will be reduced which will lead to an increase in the probability of misclassification.

The principles at play have been demonstrated in this section. These principles will be applied to a more realistic problem in the next section.

12.2 Experiment #19: Multi-target Information and Probability of Error for the F-14, F-15 and F-16

This experiment was conducted using the F-15 and F-16 data sets with the same radar waveform parameters as the F-14 data set. The MI for the individual targets as well as the multi-target MI is plotted in Figure 12.3. The entropy for the multi-target case can be calculated as:

$$\begin{aligned} H(X) &= -\sum_{x\in R_X} p(x)\log_2\left(p(x)\right) \\ &= -\frac{1024}{3072}\log_2\left(\frac{1}{3072}\right) \approx -\frac{1}{3}(-11.585) = 3.862 \quad \text{bits}, \end{aligned} \tag{12.2}$$

which agrees with the limiting value for the MI on the three lower traces.

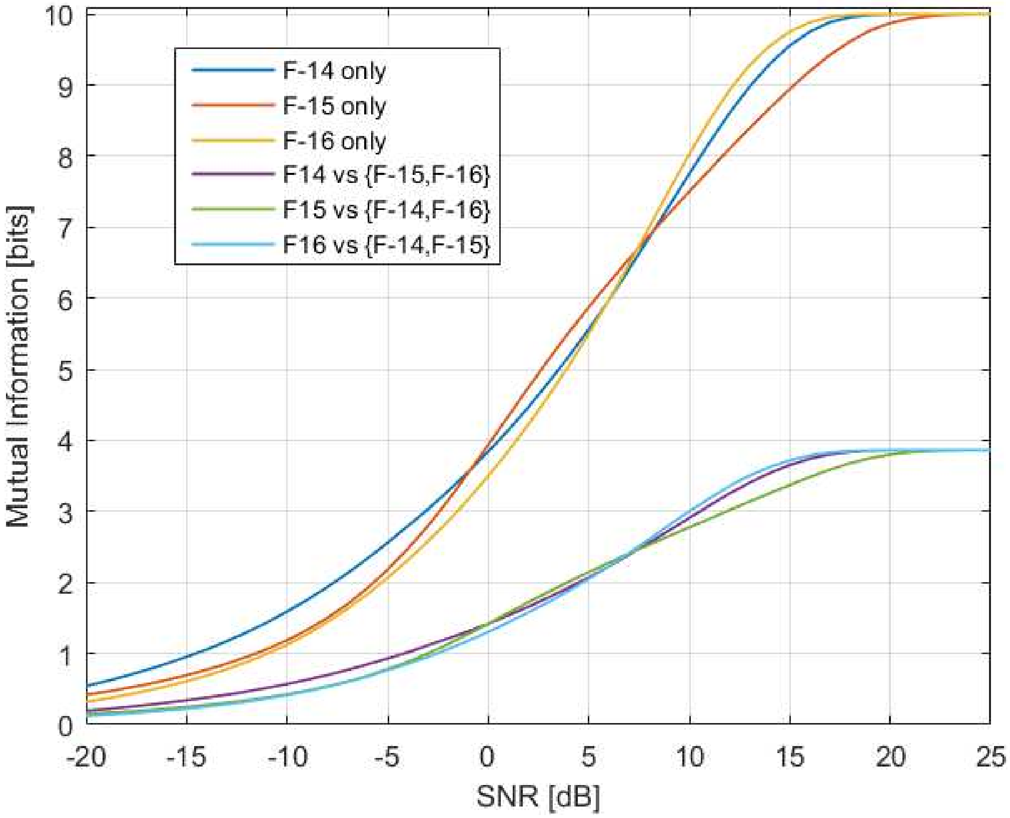


Figure 12.3: MI and multi-target MI versus SNR for the F-14, F-15 and F-16.

These MI curves were then converted to normalised MI, and the results are plotted in Figure 12.4. From this set of graphs it can be seen that all three the multi-target MI curves have suffered a SNR loss relative to their single target counterparts.

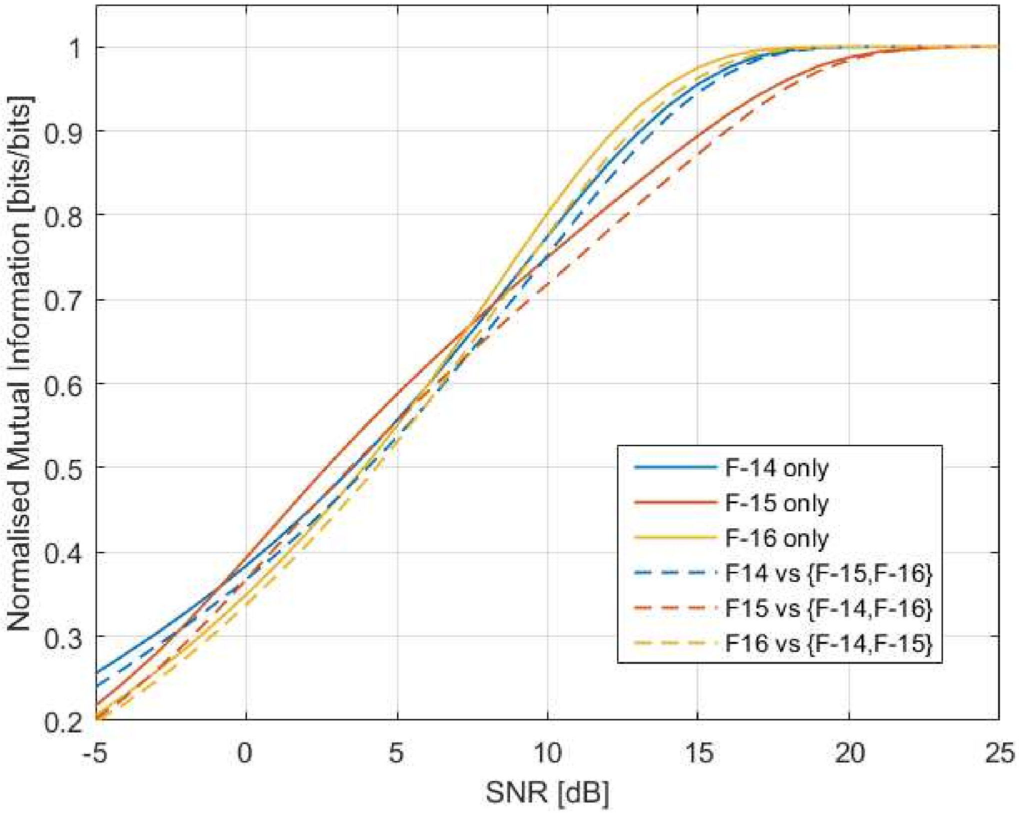


Figure 12.4: Normalised MI and normalised multi-target MI versus SNR for the F-14, F-15 and F-16.

The SNR loss between each MI curve for the single target and the corresponding multi-target case has been calculated and are plotted in the Figure 12.5 below.

The losses at high normalised MI values (above 0.9 bits) range between 0.1 dB and 0.8 dB. The average loss over the whole normalised MI range is approximately 0.6 dB, which is relatively benign. This result demonstrates the use of the multi-target MI for the calculation of SNR losses due to the radar being set up to recognise multiple targets.

Although the loss calculated for this example is relatively low, this loss will be higher in a practical system for the following reasons:

1. As extra targets with their associated sets of signatures are added, the signal space becomes more and more densely packed, thus reducing the Euclidean distance between adjacent signal points.
2. The addition of extra signatures for each target as a function of the elevation angle (relative to the target) will also increase the density of the points in the signal space, again reducing the Euclidean distance between signal points.

3. Depending on the application of the radar system it might be necessary to add multiple versions of each aircraft type corresponding to each possible configuration of the aircraft that will have an impact on the signature. This should at least include external stores and the undercarriage in its stowed and extended positions.

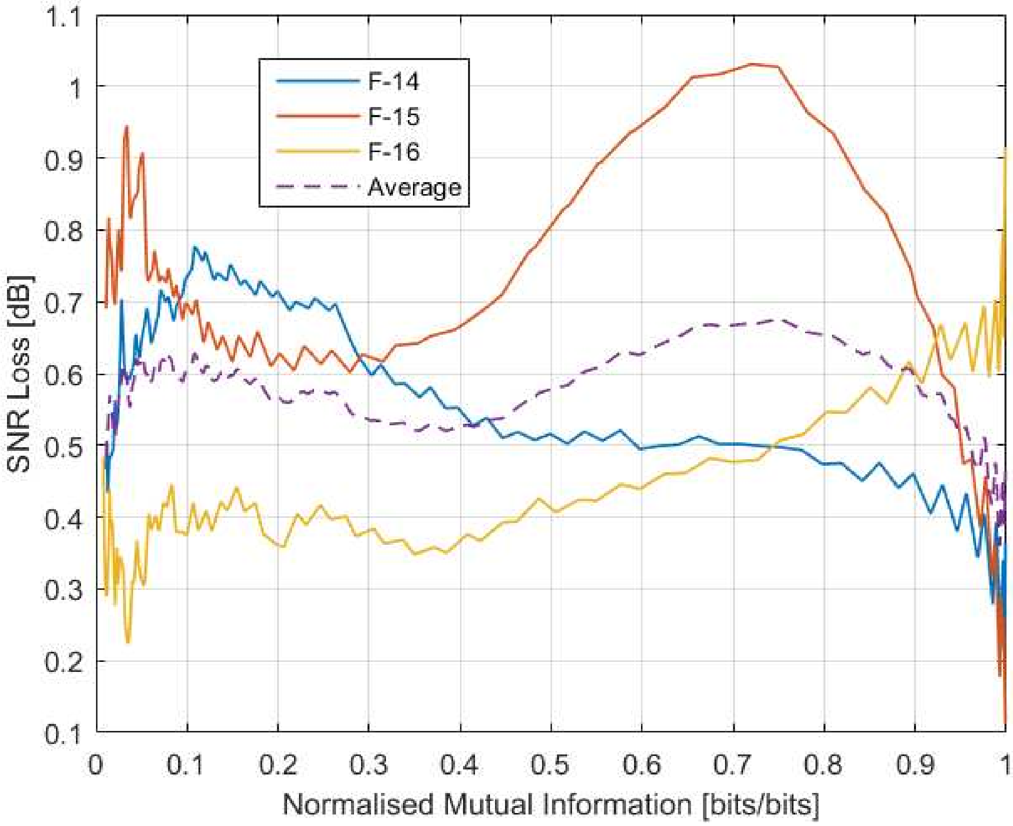


Figure 12.5: SNR loss versus MI between the single target case and the multi-target case for the F-14, F-15 and F-16.

A practical radar system can counter the first two of these effects by exploiting the fact that the target's track is usually estimated by the radar, and this knowledge can be used to reduce the signal set to a specified angular volume around the estimated orientation of the radar's line of sight relative to the target. Variables such as the airspeed and altitude of the aircraft can be used to limit the number of aircraft that the recognition system has to match against.

The MI graphs were converted to probability of error graphs by making use of Fano's inequality. The results are plotted in Figure 12.6. A graph of the normalised MI has been included for reference. For the range of P(error) plotted, use was only made of normalised MI values above 0.7. The loss of accuracy in the P(error) graph at high SNR's is due to very slight variations at values of the normalised MI which are very close to unity.

A zoomed in version of the figure is given in Figure 12.7 to show the cause of this inaccuracy. For these graphs the standard deviation goal for the MI in the Monte Carlo integration process was set to 0.002, and the accuracy goal for the multi-target runs was set to 0.005. This means that if the radar designer is interested in low levels of error probability, then accuracy specification for the Monte Carlo integration will have to be set to a much lower level, and the SNR axis will have to be sampled in smaller step sizes.

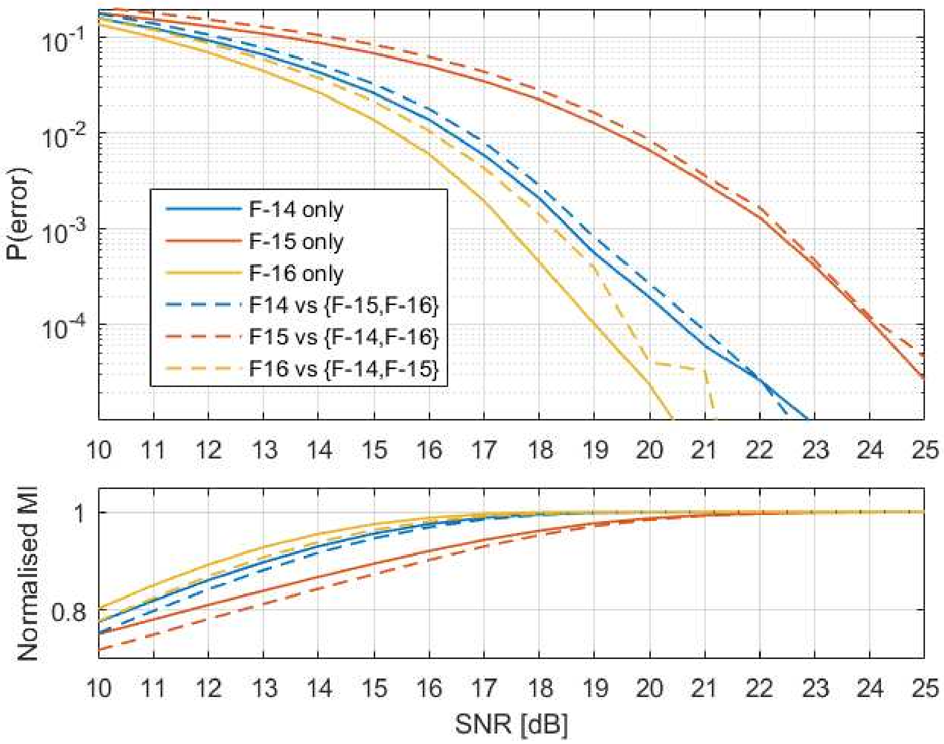


Figure 12.6: Probability of error versus SNR graphs for the multi-target setup, along with normalised MI for reference.

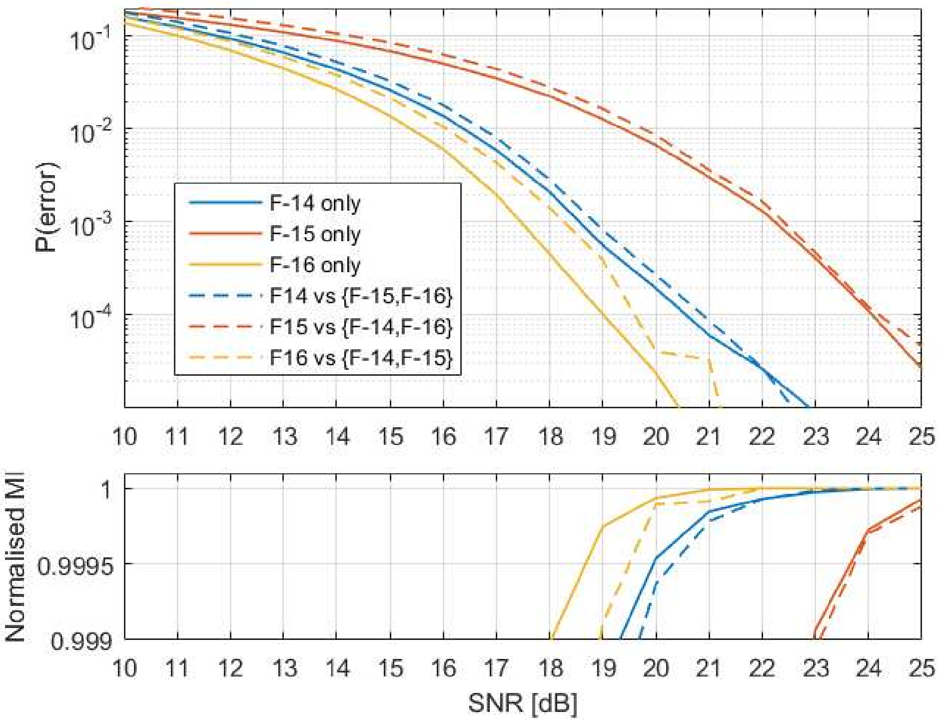


Figure 12.7: Probability of error versus SNR graphs for the multi-target setup, along with normalised MI for reference, zoomed to show detail of normalised MI.

An interesting side effect of the Monte Carlo MI calculation is that it becomes more efficient at high SNR values. This means that the can be optimised by first running a calculation over a wide SNR range to determine the trend, and then running a high accuracy calculation over the region where the Normalised MI approaches unity. The extended P(error) graph is given in Figure 12.7, even though it is inaccurate, it still gives an idea of the SNR's required to reach P(error) values below 10^{-6}.

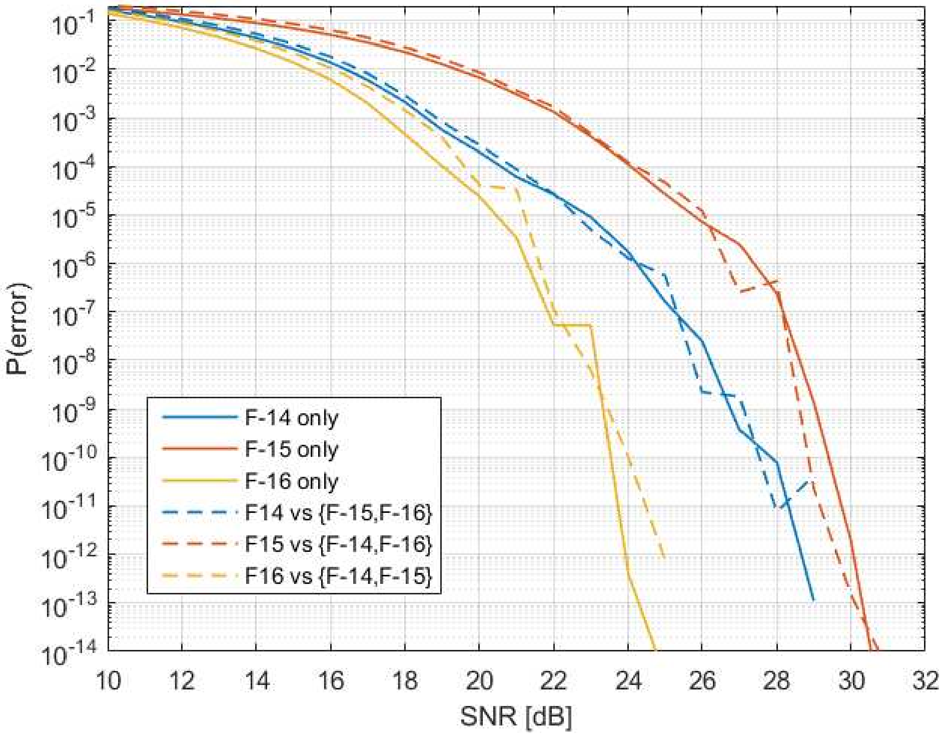


Figure 12.8: Probability of error versus SNR graphs for the multi-target setup.

12.3 Conclusion

In this section the effect of multiple targets on the achievable recognition performance was analysed. Use was made of the multi-target MI expressions developed in Section 5.6 to calculate the mutual information. Fano's inequality was then applied for the calculation of the probability of error from the mutual information result.

This was achieved by first making use of a single F-14 target and then modifying its signature set by adding Gaussian noise at different levels. This allowed the similarity between the waveform sets to be varied at will, which in turn allowed for the explanation of the possible effects on the MI curves.

The next experiment made use of multi-target MI to study the loss in a single target's recognition performance when the radar's recognition sub-system has to cater for multiple targets. For the three target scenario, making use of the F-14, F-15 and F-16 X-band data sets, it was shown that an average loss of 0.6 dB was incurred over the whole range of MI values. The highest loss for high MI values was for the F-16, which suffered a loss of 0.9 dB. This loss can be clearly seen on the probability of error graphs for the F-16.

The losses reported in this section were relatively benign, which implies that for this radar waveform, these three targets have large inter-target Euclidean distances, and are thus easily separable.

These losses will however increase as more targets, more target configurations, and more aspect angles per target are added to the radar's data base of targets for recognition.

13 Conclusions and Future Work

This chapter summarises the achievements of this research along with highlighting some of the more insightful results and discusses the possibilities for future research in the area of information theory applied to radar target recognition.

13.1 Conclusion

The topic of non-cooperative target recognition (NCTR) for radar systems was introduced by means of a high level description of the problem and an overview of pattern recognition, which is the usual means by which such problems are solved. The next chapter introduced the effects which can be exploited by a radar designer to recognise targets. This focussed on results from the theory of electromagnetism which describe the interaction between the radar signal and the target. This highlighted the fact that the interaction between the target and the EM wave is extremely complex, but can still be analyzed as a linear system in most cases.

A literature overview was presented, which highlighted the various approaches which have been exploited by radar researchers to recognize targets. This review showed that there is very little work published in the field of predicting the upper bound on the performance of target recognition in a radar system.

The next section gave an overview of the theoretical development of information theory as well as some worked examples from open literature, where this theory has been applied to radar problems. Two examples were constructed to illustrate the superiority of mutual information over correlation coefficients as a measure of dependence between two random variables. This was followed by detailed derivations pertaining to the calculation of mutual information for a known set of waveforms, which are transmitted over an additive white Gaussian noise (AWGN) channel. This section also showed that expressions for mutual information can be derived for envelope and envelope squared sensing, as well as for the case where multiple targets are to be recognized by the radar system. Validation results showing that the theoretical derivations agree with known results from the literature were also presented. Most results were also compared to Shannon's channel capacity bound to prove that this fundamental limit is not exceeded.

The results of various MI calculations are presented over the next six chapters, where each chapter focuses on a different type of mutual information analysis. These chapters describe various detailed investigations into the use of mutual information for the analysis and comparison of various approaches to radar target recognition.

Initially use was made of point scatterer models to demonstrate that increasing the radar's range resolution resulted in an increase in recognition performance. These results also showed that the mutual information calculation inherently took correlation in the range dimension as well as the azimuth dimension into account. A set of 16 random targets were also generated, and it was demonstrated that these targets could be recognized as a set, and that the amount of mutual information contained in the set was higher than that of a single target for the same SNR. This result also showed that it is possible to differentiate between all 16 targets if the SNR is high enough. The next section compared recognition performance using ISAR imaging to HRRP processing. It was shown that the MI for the unwindowed ISAR image compared to the MI for the raw HRRP's resulted in exactly the same value of MI for a given SNR, thus resulting in the same recognition performance. This result also demonstrates the practical application of the data processing theorem, which states that when data is processed, the information content either remains constant or is reduced.

The next five chapters concentrated on more realistic targets. Use was made of an in-house CEM software package to calculate the HRRP's for a F-14, F-15 and F-16 using Physical Optics (PO) and Shooting and Bouncing Ray (SBR) techniques. Initially results were generated for a 480 MHz bandwidth (81 frequency steps) centred at 10 GHz. The results showed a 5 dB spread in the SNR required to achieve 99.95 % of the maximum mutual information for the three targets. It was also shown that an optimum set of waveforms (i.e. if the radar designer could choose the target signatures to be optimum from an information transfer point of view) would outperform the true target waveforms by approximately 25 dB. Next it was shown that recognition performance could be increased by exploiting the radar's knowledge of the target's bearing. If the search space was reduced from 360° to 5.6°, maximum information transfer could occur at SNR levels between than 0 and 10 dB, instead of 20 dB and above. A 36000 azimuth sample (0.01° steps over 360°) waveform set for the F-15 was compared to the 1024 waveform set, and it was shown that the MI curve suffered a reduced slope at high SNRs which was caused by correlation between the profiles in the azimuth dimension.

The next set of experiments attempted to make use of much wider bandwidths, and HRRP's that were ambiguous in range in an attempt to create waveform sets which were more like Gaussian noise. To this end waveform sets for 2-10 GHz, 10-18 GHz and 2-18 GHz bands were generated, each still only using 81 frequencies. Comparison of the 2-10 GHz, 10-18 GHz and 2-18 GHz bands showed that slightly higher recognition performance was available in the 2-10 GHz band. This counter-intuitive insight demonstrated the usefulness of the MI approach during the conceptualisation and design of radar based target recognition systems. Another unexpected result was that although the waveforms are ambiguous in range, the information content is not significantly altered. To complete this section, the F-14 was simulated over the 2-18 GHz band and results for 100 MHz and 200 MHz frequency step sizes were extracted. It was shown that the performance for the 100 MHz step size was nearly exactly 3 dB better than that for the 200 MHz step size. This is however due to the fact that the radar transmitted 3 dB worth of extra power, so in this case no extra information was acquired about the target, even though the amount of measurement frequencies were doubled. Reducing the frequency step size also increased the range ambiguity. If any information was gained in this process then the SNR gain would have been more than 3 dB. The fact that only a 3 dB gain in SNR was measured implies that no information was being destroyed by the more ambiguous waveform. There was however a 0.5 dB increase in performance at an MI level of 99.9% of the maximum information content, which implies that there will be slightly increased performance when very low probabilities of error are required.

The next section focused on the extra information which could be extracted when making use of multiple receiver channels in the radar. The information gain when using a fully polarimetric radar as well as a monopulse radar were analysed. The first experiment made use of an F-18 and an F-35 CAD model to calculate the fully polarimetric target signature. It was shown that the F-35 requires between 5 and 9 dB more SNR to reach the same recognition performance as the F-18 when use is made of the full polarisation response of the targets. In general the full polarisation showed approximately a 5 dB improvement over any single linear polarisation. This was followed by the analysis of measured data from the University of Pretoria’s compact range. A 1:25 scale model of a Boeing 707 was measured over a 2-18 GHz frequency span. The scaling makes this data set applicable to the 80-720 MHz band for a full-scale 707. It was shown that a gain of approximately 3.25 dB could be achieved when moving from a single linear polarisation to dual linear polarisation. Importantly this demonstrated the validity of the mutual information calculation on measured data sets if the measurements are carefully set up and made in a controlled environment. It also demonstrated that the calculation was stable for very high input signal dimensionalities. The highest dimension required during this analysis was 8004 dimensions. The monopulse radar case was analysed next. The amount of gain which can be extracted by including the azimuth channel in a monopulse radar was shown to be approximately 3 dB for the F-14 at long ranges (50 km) and up to 4.5 dB at shorter ranges (2 km). This section also highlighted the important fact that the reference signatures stored in the radar’s database should be compensated for the range of the target and the effect of the antenna pattern over the target’s angular extent.

The next chapter analysed the amount of information which was lost when envelope processing is applied to the receive signal. This effectively discards the phase information in the signal. The amount of information which is lost when a radar discards phase information when classifying targets was shown to be in the region of 7 dB over the three fighter targets (F-14, F-15 and F-16) at X-band. This was calculated by deriving the expression for MI after envelope processing and then running Monte-Carlo numerical integration on the expression. It was also shown that some information gain was possible if linear pre-processing was applied to the signal before the envelope processing. Examples for FFT processing and PCA processing showed that the pre-processing gave a SNR gain of approximately 3 dB over most of the MI range. At high MI values, the PCA outperformed the FFT by approximately 1 dB.

The final results chapter made use of the multi-target MI expression in conjunction with Fano's inequality to convert the MI result to a probability of error for the case where the radar is programmed to identify three targets, but only one of the targets is interrogated by the radar. This allows for the case where the radar can incorrectly identify the target. The analysis of this scenario demonstrated that the recognition suffers a loss of approximately 0.6 dB. This is due to the fact that the recognition algorithm has to allow for more target signatures, and that as the number of signatures increase, the likelihood of mis-classification events increases. Use was once again made of the three jet targets at X-band to generate this result.

This research has thus demonstrated that information theory, and specifically the concept of mutual information can be used to compare disparate approaches to designing a radar concept which is capable of target recognition without resorting to choosing specific feature extraction and classification algorithms. The analysis of the various radar concepts and techniques was compared quantitatively, which allowed the differences between targets, waveforms, radar receiver architectures and signal processing techniques, as well as combinations of these, to be characterised in terms of SNR gains and/or losses. From the mutual information result, the probability of error versus SNR could also be calculated. It was also shown that these results could be calculated efficiently for high input dimensionality signals. This research has also shown that it is vital to consider the performance as a function of SNR as the performance gains and losses vary as a function of SNR, often in a non-intuitive manner.

Importantly, the focus of this research has been to develop techniques to predict the maximum achievable radar target recognition performance within a set of radar system constraints. The mutual information represents the fundamental limit on the maximum amount of information which can be extracted by the radar for a given scenario (combination of target, radar waveforms, radar architecture and signal processing). The data processing inequality shows that it is impossible to achieve better performance by any clever processing or manipulation of the input data. In fact the data processing inequality proves that information can only be destroyed or, in the best case, remain constant when data is passed through a chain of signal processing algorithms.

It is hoped that this study will aid designers of radar systems in making design choices in the early stages of radar designs to allow for, and optimize, target recognition functionality in future radar systems. The next section gives an overview of research areas that would compliment or extend the work presented in this thesis.

13.2 Future Work

This research has also shown that some unexpected insights can be obtained through the use of mutual information and information theory in general as an analysis tool for predicting the maximum achievable target recognition performance of a radar concept. Many of these insights or ideas emanating from them could become topics for dedicated studies in their own right. This study has focussed on the calculation of the mutual information contained in radar returns from a target and there are many topics which could build directly on this research. These topics range from understanding the lifting of the assumptions that were made during this study, through extending the analysis to more targets, waveforms and radar architectures, to attempting to use mutual information as a cost function in a recursive optimization of a radar design for target recognition.

The following limitations exist regarding the targets and the interaction of the EM wave with the targets. The target is constructed from perfectly electrically conducting material (PEC) and the fields impinging on the target and being sensed by the radar are far fields. Due to the large electrical size of the targets of interest, the interaction of the EM field and the target was predicted using high frequency approximations to Maxwell's equations. Use had to be made of freely available CAD models of the targets, and it is uncertain how close these are to the physical aircraft. This also means that the engines were not modelled accurately and were stationary during the analysis. Radomes for the radar as well as other antennas were modelled as PEC material. The cockpit canopy was also modelled as a PEC material. Any research into improving the electromagnetic modelling of the target will thus improve the fidelity of the MI results.

There are several ways in which the target space for this analysis could be extended. More targets and more varied types of targets would be the first step. The expansion of the spatial sampling of the targets from just a "water line" cut at zero elevation to full spherical coverage would give a much better idea of the true achievable recognition performance. Each target could also be extended to include various variations of the same target. This should include the effect of variations of external stores, the effect of flight control surfaces, the effect of landing gear and could even include the effect of variations in the target geometry caused by different loading of the aircraft and manoeuvre induced flexing of the airframe. A far more difficult problem would be the analysis of battle damage to an aircraft. What amount of damage could be sustained before the radar would not be able to recognize the target correctly? How sensitive would a recognition algorithm be to small variations in geometry such as dents and single bullet holes in the aircraft skin?

The current formulation of the MI approach does not allow for the use of an "unknown" target, that is, that the received signal point is far enough away from the known targets in the signal space that it can be safely declared that the received signal does not belong to any of the targets. Extending the MI approach to allow for unknown targets would thus be an important contribution to this analysis technique. Equally important would be an analysis of the effect of storing an approximation of the target response in the radar, instead of a perfect copy. For example, if a set of scattering centres for a target was extracted and stored in the radar's target data base, how accurate would this representation have to be so that the probability of error is not unduly affected?

This study made the assumptions that the radar has perfect timing and phase information. This effectively means that the radar has knowledge of the exact range to the target. The characterisation of the performance loss if these assumptions are lifted would give a better estimate of the recognition performance that could be obtained by a practical system. For this study the only interfering signal was additive white Gaussian noise (AWGN), so research into the effects induced by the propagation environment such as clutter and multipath as well as external signals, such as different types of jamming signals would be of interest. This would also open the door to the analysis of synthetic aperture radar (SAR) targets, as these are inherently imbedded in clutter and multipath. The analysis of SAR targets would require a much richer set of intra-target models. This is due to the fact that land vehicles occur in far more configurations than aircraft. For example a main battle tank would have to be characterised for all combinations of cannon elevations and turret positions, not even to mention all the variability in the equipment which can be attached to the outside of the tank.

Given the above discussion, a set of standard test targets and corresponding data sets should be developed along with published mutual information results for these data sets. These would then serve as absolute performance bounds against which researchers could benchmark their target recognition algorithms.

This research has studied mutual information over the whole range of SNR values. A high performance NCTR system would only require the mutual information to be calculated in the region from 90% to 100% of the maximum MI value. Efficient algorithms for the location of this region could be developed and there might be optimizations to the actual MI calculation which can be made. This study focussed on the use of high range resolution profiles (HRRP) and the equivalent frequency domain data as inputs to the MI calculation. The extension of the MI calculation to other radar observables such as propeller modulation, jet engine modulation and helicopter blade modulation would lead to some very interesting results. The combination of these techniques with HRRP would start to give an idea of the maximum achievable performance of NCTR algorithms if all information sources that the radar can sense about the target are fused.

The effects of various limitations in the radar's hardware, such as phase noise, spurious signals caused by mixing stages in the RF, amplitude discretisation by the analogue to digital converter and various signal processing algorithms could also be framed as a mutual information analysis. This type of analysis will help radar designers develop detailed specifications for radar hardware that is capable of supporting a given level of target recognition performance. The gain in recognition performance for multichannel systems such as collocated MIMO and multistatic radar systems would also be a worthwhile pursuit, along with the specific limitations induced by this type of architecture.

Finally, it might be possible to use MI as a cost function for the optimization of a NCTR radar. The waveform parameters (carrier frequency, bandwidth, etc.) could be seen as free parameters for the optimization algorithm. These would then have to be used to evaluate the EM response of the set of targets for this waveform and then the MI could be calculated. The MI value, possibly as a function of SNR, could then be used to adapt the waveform to improve the recognition performance. This whole process could be automated as a recursive loop to optimize the waveform design and hence the radar design for optimal target recognition performance.

14 Bibliography

[Abromowitz1972] M. Abromowitz & I.A. Stegun, Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables, 10th Printing, Dover Publications, **1972**.

[Aldhubaib2009] F.F. Aldhubaib & N.V. Shuley, *Characteristic polarization states estimation in an ultrawideband context: A frequency approach,* IEEE Transactions on Geoscience and Remote Sensing, **2009**, vol. 47, no. 8, pp. 2808-2817.

[Aldhubaib2010] F. Aldhubaib & N.V. Shuley, *Radar target recognition based on modified characteristic polarization states,* IEEE Transactions on Aerospace and Electronic Systems, **2010**, vol. 46, no. 4, pp. 1921-1933.

[Angell2003] T. Angell, G. Hsiao & L. Wen, *On the two-dimensional inverse scattering in electromagnetics,* Applicable Analysis, **2003**, vol. 82, no. 5, pp. 483-497.

[Armstrong1991] Armstrong, B.C.; Griffiths, H.D., *Use of difference channel information for detection in monopulse radars*, Radar and Signal Processing, IEE Proceedings F , vol.138, no.3, pp.199-210, June **1991**.

[Astanin1997] Astanin & A. Kostylev, Ultrawideband Radar Measurements, Analysis and Processing, IET, **1997**.

[Augustyn1992] K. Augustyn, *A new approach to automatic target recognition,* IEEE Transactions on Aerospace and Electronic Systems, **1992**, vol. 28, no. 1, pp. 105-114.

[Baum1994] C.E. Baum, *Signature-based target identification and pattern recognition,* IEEE Antennas and Propagation Magazine, **1994**, vol. 36, no. 3, pp. 44-51.

[Bell1988] M. Bell, Information Theory and Radar: Mutual Information and the Design and Analysis of Waveforms and Radar Systems, PhD Thesis, Caltech, **1988**.

[Bell2003] A.J. Bell, *The co-information lattice*, 4th International Symposium on Independent Component Analysis and Blind Source Separation, Nara, Japan, **2003**.

[Bhalla1996] R. Bhalla & H. Ling, *Three-dimensional scattering centre extraction using the shooting and bouncing ray technique,* IEEE Transactions on Antennas and Propagation, **1996**, vol. 44, no. 11, pp. 1445-1453.

[Bhalla1997] R. Bhalla, J. Moore & H. Ling, *A global scattering centre representation of complex targets using the shooting and bouncing ray technique,* IEEE Transactions on Antennas and Propagation, **1997**, vol. 45, no. 12, pp. 1850-1856.

[Bhanu1986] B. Bhanu, *Automatic target recognition: State of the art survey,* IEEE Trans. Aerosp. Electron. Syst., **1986**, vol. AES-22, pp. 364–377.

[BlacknellGriffiths2013] D. Blacknell and H. Griffiths, Radar Automatic Target Recognition (ATR) and Non-cooperative Target Recognition (NCTR), IET Radar, Sonar and Navigation, Series 33, **2013**.

[Borden1986] B.H. Borden, *High-frequency statistical classification of complex targets using severely aspect-limited data,* Antennas and Propagation, IEEE Transactions on, IEEE, **1986**, vol. 34, no. 12, pp. 1455-1459.

[Borden1995a] B.H. Borden, *Enhanced range profiles for radar-based target classification using monopulse tracking statistics,* IEEE Transactions on Antennas and Propagation, **1995**, vol. 43, no. 8, pp. 759 - 765.

[Borden2002] B. Borden, *Mathematical problems in radar inverse scattering,* Inverse Problems, **2002**, vol. 18, no. 1, pp. R1-R28.

[Briles1993] S. Briles, The theory for, and Demonstration of, Information Theory Applied to Radar Target Identification, PhD Thesis, Kansas State University, **1993**.

[Briles1993] S.D. Briles, *Information-theoretic performance bounding of Bayesian identifiers,* SPIE Automatic Object Recognition III, **1993**, pp. 255 - 266.

[Butman1973a] S. Butman & M. Klass, *Capacity of noncoherent channels*, Jet Propulsion Lab., Pasadena, CA, JPL, **1973**, pp. 85-93.

[Butman1973b] S. Butman & B. Levitt, Capacity for Noncoherent, Soft-Decision MFSK Signaling, The Deep Space Network Progress Report, **1973**, vol. 15, pp. 146-155.

[Challa2001] S. Challa & G.W. Pulford, *Joint target tracking and classification using radar and ESM sensors,* IEEE Transactions on Aerospace and Electronic Systems, **2001**, vol. 37, no. 3, pp. 1039-1055.

[Cilliers2002] J.E. Cilliers, Multi-dimensional Lattice Equaliser for Q2PSK, Masters Dissertation, University of Pretoria, **2002**.

[Cilliers2007] J.E. Cilliers & J. Smit, *Pulse Compression Sidelobe Reduction by Minimization of* L_p*-Norms,* IEEE Transactions on Aerospace and Electronic Systems, **2007**, vol. 43, no. 3, pp. 1238-1247.

[Cilliers2008] A. Cilliers & W. Nel, *Helicopter parameter extraction using joint time-frequency and tomographic techniques*, 2008 International Conference on Radar, Adelaide, Australia, **2008**, pp. 598-603.

[Cilliers2010] A. Cilliers, Helicopter Blade Parameter Extraction for Purposes of Radar Target Identification, MSc Dissertation, Department of Electrical Engineering, University of Cape Town, **2010**.

[Cilliers2014] J. E. Cilliers, J. M. Steyn, J. C. Smit, C. Pienaar and M. Pienaar, *Considering CAD model accuracy for Radar Cross Section and signature calculations of electrically large complex targets*, 2014 International Radar Conference, Lille, **2014**, pp. 1-6.

[Cilliers2015a] J.E. Cilliers, J.C. Smit, C.J. Baker & K. Woodbridge, *On the gain in recognition performance due to the addition of polarisation in an X-band High Range Resolution radar evaluated for F-18 and F-35 targets using asymptotic EM techniques*, in Radar Conference (RadarCon), 2015 IEEE , vol., no., pp.1296-1299, 10-15 May **2015**.

[Che2000] W. Che & E.K.-n. Yung, *Research on double-channel variable polarization system with short switching time,* IEEE Antennas and Propagation Society International Symposium, **2000**, pp. 2134-2137.

[Cowen1991a] M.N. Cohen, *An overview of radar-based, automatic, noncooperative target recognition techniques,* IEEE International Conference on Systems Engineering, **1991.**

[Cover2006] T. Cover & J.A. Thomas, Elements of Information Theory, Wiley, **2006**.

[Cover1991] T. Cover & J.A. Thomas, Elements of Information Theory, Wiley, **1991**.

[DeVore2002] M. DeVore, R. Chamberlain, G. Engel, J. O'Sullivan & M. Franklin, *Tradeoffs between quality of results and resource consumption in a recognition system,* The IEEE International Conference on Application-Specific Systems, Architectures and Processors, **2002**, pp. 391- 402.

[DeVore2004] M. DeVore & J. O'Sullivan, *Quantitative statistical assessment of conditional models for synthetic aperture radar,* Image Processing, IEEE Transactions on, **2004**, vol. 13, no. 2, pp. 113- 125.

[DeVore2000] M. DeVore, N. Schmid & J. O'Sullivan, *Analytical and experimental performance-complexity tradeoffs in ATR,* Conference Record of the Thirty-Fourth Asilomar Conference on Signals, Systems and Computers, **2000**, vol. 2, pp. Pages:1519-1523.

[Devroye1996] L. Devroye, L. Gyorfi & G. Lugosi, A Probabilistic Theory of Recognition, Springer-Verlag, **1996**.

[Duda2001] R.O. Duda, P.E. Hart and D.G. Stork, Pattern Classification, Second Edition*,* Wiley, **2001**.

[Dudgeon2000] D.E. Dudgeon, *ATR Performance Modelling and Estimation,* Digital Signal Processing, **2000**, vol. 10, no. 4, pp. 269–285.

[Erdogmus2004] D. Erdogmus and J.C. Principe, *Lower and upper bounds for misclassification probability based on renyi's information*, Journal of VLSI signal processing systems for signal, image and video technology, vol. 37, pp. 305-317, June **2004**.

[Fan2002] Y. Fan, T. Jiang & D. Evans, *The parallel genetic algorithm for electromagnetic inverse scattering of a conductor,* Journal of Computer Mathematics, **2002**, vol. 79, no. 5, pp. 573-586.

[Fano1961] Robert M. Fano, Transmission of Information: A Statistical Theory of Communications, M.I.T. Press, **1961**.

[Fortuny1998] J. Fortuny, *An efficient 3-D near-field ISAR algorithm,* IEEE Transactions on Aerospace and Electronic Systems, **1998**, vol. 34, no. 4, pp. 1261-1270.

[Fukunaga1990] K. Fukunaga, Introduction to Statistical Pattern Recognition, 2nd ed., Academic Press, **1990**.

[Gallager1968] R.G. Gallager, Information Theory and Reliable Comminication, MIT, John Wiley and Sons, **1968**.

[Garber1997] F. Garber & E. Zelnio, *On some simple estimates of ATR performance, and initial comparisons for a small data set,* SPIE Algorithms for Synthetic Aperture Radar IV, **1997**, vol. 3070, pp. 150–161.

[Giuli1986] D. Giuli, *Polarization diversity in radars,* Proceedings of the IEEE, **1986**, vol. 74, no. 2, pp. 245-269.

[Gross1999] D. Gross, *A neural network ATR for high range resolution radar signature recognition of moving ground targets,* Asilomar Conference on Signals, Systems, and Computers, **1999**, pp. 1235-1239.

[Herman2002] S. Herman & P. Moulin, *A particle filtering approach to FM-band passive radar tracking and automatic target recognition,* Proceedings, IEEE Aerospace Conference, **2002**, vol. 4, pp. 1789-1808.

[Huang2007] C. Huang, C. Liu, C. Chiu, Y. Wu, T.Wysocki & B. Wysocki, *Electromagnetic transverse electric-wave inverse scattering of a two-dimensional object by genetic algorithm,* Electromagnetics, **2007**, vol. 27, no. 5, pp. 241-251.

[Jacobs1997] S. Jacobs & J. O'Sullivan, *High resolution radar models for joint tracking and recognition,* Proceedings of the 1997 IEEE National Radar Conference, **1997**, pp. 99-104.

[Jacobs2000] S. Jacobs & J. O'Sullivan, *Automatic target recognition using sequences of high resolution radar range-profiles,* IEEE Transactions on Aerospace and Electronic Systems, **2000**, vol. 36, no. 2, pp. 364-381.

[Jain2002] A. Jain, P. Moulin, M. Miller & K. Ramchandran, *Information-theoretic bounds on target recognition performance based on degraded image data,* IEEE Transactions of Pattern Analysis and Machine Intelligence, **2002**, vol. 24, no. 9, pp. 1153-1166.

[Jakulin2004] Aleks Jakulin & Ivan Bratko, *Quantifying and visualizing attribute interactions: An approach based on entropy,* arXiv:cs.AI/0308002v3, **2004**.

[Jenn2005] D. Jenn, Radar and Laser Cross section Engineering, American Institute of Aeronautics and Astronautics, **2005**.

[Jin2007] J. Jin, Q. Wang & Y. Shen. Jenn, *Registering multiple medical images using the shared chain mutual information,* Chineses Optical Letters, **2007**, vol. 5, no. 7, July 10.

[Kennaugh1965] E. Kennaugh & D. Moffatt, *Transient and impulse response approximations,* Proceedings of the IEEE, Ohio State University, **1965**, vol. 53, no. 8, pp. 893-901.

[Kidera2010] S. Kidera, T. Sakamoto & T. Sato, *Accurate UWB radar three-dimensional imaging algorithm for a complex boundary without range point connections,* IEEE Transactions on Geoscience and Remote Sensing, **2010**, vol. 48, no. 4, pp. 1993-2004.

[Killen1989] G.A. Killen, *Advanced Technology MMW Seeker Testbed: a multi-technology demonstration sensor,* Proceedings of the 1989 IEEE National Radar Conference, **1989**, pp. 35-41.

[Kim2005] K.-T. Kim, D.-K. Seo & H.-T. Kim, *Efficient classification of ISAR images,* IEEE Transactions on Antennas and Propagation, **2005**, vol. 53, no. 5, pp. 1611-1621.

[Knott1992], E.F. Knott, Radar Observables, in Tactical Missile Aerodynamics: General Topics, Vol. 141, **1992**, Chap. 4, pp. 107-194, American Institute of Aeronautics and Astronautics, Washington, DC..

[Knott2004], E. F. Knott, J. F. Shaeffer, M. T. Tuley, Radar Cross Section, 2nd Edition, Scitech Publishing, Inc., **2004**.

[Ksienski1975] A.A. Ksienski, Y.-T. Lin & L.J. White, *Low-frequency approach to target identification,* Proceedings of the IEEE, **1975**, vol. 63, no. 12, pp. 1651-1660.

[Lewis1969] R.M. Lewis, *Physical optics inverse diffraction,* IEEE Transactions on Antennas and Propagation, **1969**, vol. 17, no. 3, pp. 308-314.

[Li1993] H.-J. Li & S.-H. Yang, *Using range profiles as feature vectors to identify aerospace objects,* IEEE Transactions on Antennas and Propagation, **1993**, vol. 41, no. 3, pp. 261-268.

[Li1996a] H.-J. Li, Y.-D. Wang & L.-H. Wang ,*Matching score properties between range profiles of high-resolution radar targets,* IEEE Transactions on Antennas and Propagation, **1996**, vol. 44, no. 4, pp. 444-452.

[Li1998] Q. Li, P. Ilavarasan, J.E. Ross, E.J. Rothwell, D.P. Nyquist & K.-M. Chen, *Radar target identification using a combined early-time/late-time E-pulse technique,* IEEE Transactions on Antennas and Propagation, **1998**, vol. 46, no. 9, pp. 1272-1278.

[Lin1981] H. Lin & A.A. Ksienski, *Optimum frequencies for aircraft classification,* IEEE Transactions on Aerospace and Electronic Systems, **1981**, vol. AES-17, no. 5, pp. 656-665.

[Maasdorp2013] F. Maasdorp, J. Cilliers, M. Inggs & C. Tong, *Simulation and measurement of propeller modulation using FM broadcast band commensal radar*, IET Electronics Letters, **2013**, vol. 49, no. 23, pp. 1481-1482

[Malas2015] J.A. Malas, J.A. Cortese and P. Ryan, *Uncertainty propagation and the Fano based information theoretic method,* IEEE Radar Conference (RadarCon), **2015**, pp. 1638-1643.

[Martorella2011] M. Martorella, E. Giusti, L. Demi, Z. Zhou, A. Cacciamano, F. Berizzi & B. Bates, *Target recognition by means of polarimetric ISAR images,* IEEE Transactions on Aerospace and Electronic Systems, **2011**, vol. 47, no. 1, pp. 225-239.

[Morgan1988] M.A. Morgan & B.W. McDaniel, *Transient electromagnetic scattering: data acquisition and signal processing,* IEEE Transactions on Instrumentation and Measurement, **1988**, vol. 37, no. 2, pp. 263-267.

[Nebabin1995] V.G. Nebabin, Methods and Techniques of Radar Target Recognition, Artech House, **1995**.

[Novak1991] L.M. Novak, *A comparison of 1-D and 2-D algorithms for radar target classification,* IEEE International Conference on Systems Engineering, **1991**, pp. 6-12.

[O'Sullivan1998] J. O'Sullivan, R. Blahut & S.D.L., *Information theoretic image formation,* Invited IEEE Transactions on Information Theory, **1998**, vol. 44, no. 6, pp. 2094-2123.

[Pearl1988] J. Pearl, Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference, Morgan Kaufmann, **1988**.

[Pienaar2017] C Pienaar, JW Odendaal, J Joubert, JC Smit & JE Cilliers, *RCS validation of asymptotic techniques using measured data of an electrically large complex model airframe*, Applied Computational Electromagnetics Society Journal, **2017**, vol. 32, No. 1.

[Pienaar2017a] M Pienaar, JW Odendaal, J Joubert, C Pienaar & JC Smit, *Bistatic RCS measurements in a compact range*, 2017 International Conference on Electromagnetics in Advanced Applications (ICEAA), pp. 1199-1202, **2017**.

[Pienaar2018] C Pienaar, JW Odendaal, JC Smit, J Joubert & JE Cilliers, *RCS results for an electrically large realistic model airframe*, Applied Computational Electromagnetics Society Journal, **2018**, Vol. 33, No. 1.

[Proakis1989] J.G. Proakis, Digital Communications, McGraw-Hill, **1989**.

[Renyi1961] A. Renyi, *On measures of entropy and information,* Proceeding of the fourth Berkeley Symposium on Mathematics, Statistics and Probability, pp. 547-561, **1961**.

[Rihaczek1996] A.W. Rihaczek & S.J. Hershkowitz, Radar Resolution and Complex-image Analysis, Artech House, **1996**.

[Saha1989] D. Saha & T. Birdsall, *Quadrature-quadrature phase-shift keying,* IEEE Transactions on Communications, vol. 37, no. 5, pp. 437-448, **1989**.

[Seybold1996] J.S. Seybold & S.J. Bishop, *Three-dimensional ISAR imaging using a conventional high-range resolution radar,* Proceedings of the 1996 IEEE National Radar Conference, IEEE, **1996**, pp. 309-314.

[Shannon1948] C. Shannon, *A Mathematical Theory of Communication,* Bell System Technical J., **1948**, vol. 27, pp. 379-423 and pp. 623-656.

[Shirman2002] Y.D. Shirman, Computer Simulation of Aerial Target Radar Scattering, Recognition, Detection and Tracking, Artech House, **2002**.

[Smit2012] J.C. Smit & J.E. Cilliers, *Insights into factors contributing to the observability of a submarine at periscope depth by modern radar: Part 2 - EM simulation of mast RCS in a realistic sea surface environment,* 2012 IEEE-APS Topical Conference on Antennas and Propagation in Wireless Communications (APWC), IEEE, **2012**, pp. 989-992.

[Smit2012a] J. Smit, E. Burger & J. Cilliers, *Comparison of MLFMM, PO and SBR for RCS investigations in radar applications,* IET International Conference on Radar Systems (Radar 2012), Institution of Engineering and Technology, **2012**, pp. 22-22.

[Stausberger1992] D. Stausberger, E. Garber, N. Chamberlain & E. Walton, *Modelling and performance of HF/OTH radar target classification systems,* IEEE Transactions on Aerospace and Electronic Systems, **1992**, vol. 28, no. 2, pp. 396-403.

[Steedly1991] W.M. Steedly & R.L. Moses, *High resolution exponential modelling of fully polarized radar returns,* IEEE Transactions on Aerospace and Electronic Systems, **1991**, vol. 27, no. 3, pp. 459-469.

[Steinberg1988] B.D. Steinberg, *Microwave imaging of aircraft,* Proceedings of the IEEE, **1988**, vol. 76, no. 12, pp. 1578-1592.

[Stuff2002], M. Stuff, *Derivation and estimation of Euclidean invariants of far field range data,* University of Michigan, **2002**.

[Stuff2003] M. Stuff, P. Sanchez & M. Biancalana, *Extraction of three-dimensional motion and geometric invariants from range dependant signals,* Multidimensional Systems and Signal Processing, **2003**, vol. 14, pp. 161-181.

[Tait2005], P. Tait, Introduction to Radar Target Recognition, IEE Radar Sonar and Navigations series, **2005**.

[Ungerboeck1982], G. Ungerboeck, Channel *Channel coding with multilevel/phase signals,* IEEE transactions on information theory, **1982**, vol. 28, no. 1, pp. 55-67.

[Urkowitz1983], H. Urkowitz, Signal theory and random processes, Artech House, **1983**.

[Wang2010a], D.-w. Wang, X.-y. Ma, A.-L. Chen & Y. Su, *High-resolution imaging using a wideband MIMO radar system with two distributed arrays,* IEEE transactions on image processing, **2010**, vol. 19, no. 5, pp. 1280-1289.

[Wang2010] D.-w. Wang, X.-y. Ma & Y. Su, *Two-dimensional imaging via a narrowband MIMO radar system with two perpendicular linear arrays,* IEEE transactions on image processing, **2010**, vol. 19, no. 5, pp. 1269-79.

[Watanabe1960] S. Watanabe, *Information theoretical analysis of multivariate correlation,* IBM Journal of Research and Development, Vol. 4, pp. 66-82, **1960**.

[Williams2000] R. Williams, J. Westerkamp, D. Gross & A. Palomino, *Automatic target recognition,* IEEE Aerospace and Electronic Systems Magazine, **2000**, vol. 15, no. 4, pp. 37-43.

[Woodward1953] P. Woodward, Probability and Information Theory, with Applications to Radar, Artech House, **1953**.

[Xu2001] X. Xu & R.M. Narayanan, *Three-dimensional interferometric ISAR imaging for target scattering diagnosis and modelling,* IEEE transactions on image processing, **2001**, vol. 10, no. 7, pp. 1094-102.

[Zembower1998] A. Zembower, *Emerging trends in sensor data management,* 15th Annual AESS/IEEE Dayton Section Symposium. Sensing the World: Analog Sensors and Systems Across the Spectrum (Cat. No.98EX178), IEEE, **1998**, pp. 71-78.

[Zhu2011] Y. Zhu & Y. Su, *A type of* M^2*-transmitter* N^2*-receiver MIMO radar array and 3D imaging theory,* Science China Information Sciences, **2011**, vol. 54, no. 10, pp. 2147-2157.

[Zwart2003] J. Portegies Zwart, R. van der Heiden, S. Gelsema & F. Groen, *Fast translation invariant classification of HRR range profiles in a zero phase representation,* IEE Proceedings - Radar, Sonar and Navigation, **2003**, vol. 150, no. 6, pp. 411.

[Zyweck1994] A. Zyweck & R. Bogner, *High-resolution radar imagery of the Mirage III aircraft,* IEEE Transactions on Antennas and Propagation, **1994**, vol. 42, no. 9, pp. 1356-1360.

[Zyweck1996] A. Zyweck & R. Bogner, *Radar target classification of commercial aircraft,* IEEE Transactions on Aerospace and Electronic Systems, **1996**, vol. 32, no. 2, pp. 598-606.

Appendix A Derivation of Multidimensional Mutual Information

In this section a detailed derivation of the Mutual information for a discrete input continuous output channel is given. The approach taken was to keep the pdf's as $p(\bullet)$ for as long as possible in the derivation, and then only substitute the multi-dimensional pdf's towards the end of the derivation.

Starting with the input-output relationship of the channel:

$$y_n = a_n + z_n \tag{A.1}$$

The expression for the mutual information can be simplified as:

$$\begin{aligned}
I(a_k, y) &= \sum_{k=0}^{N-1} P(k) \int_{-\infty}^{\infty} p(y|a_k) \cdot \log_2 \left\{ \frac{p(y|a_k)}{\sum_{i=0}^{N-1} P(i)\, p(y|a_i)} \right\} dy \\
&= \sum_{k=0}^{N-1} \frac{1}{N} \int_{-\infty}^{\infty} p(y|a_k) \cdot \log_2 \left\{ \frac{Np(y|a_k)}{\sum_{i=0}^{N-1} p(y|a_i)} \right\} dy \\
&= \frac{1}{N} \sum_{k=0}^{N-1} \int_{-\infty}^{\infty} p(y|a_k) \cdot \left[\log_2(N) + \log_2 \left\{ \frac{p(y|a_k)}{\sum_{i=0}^{N-1} p(y|a_i)} \right\} \right] dy \\
&= \frac{1}{N} \sum_{k=0}^{N-1} \int_{-\infty}^{\infty} p(y|a_k) \cdot \log_2(N)\, dy + \frac{1}{N} \sum_{k=0}^{N-1} \int_{-\infty}^{\infty} p(y|a_k) \cdot \log_2 \left\{ \frac{p(y|a_k)}{\sum_{i=0}^{N-1} p(y|a_i)} \right\} dy \\
&= \frac{\log_2(N)}{N} \sum_{k=0}^{N-1} \int_{-\infty}^{\infty} p(y|a_k) dy + \frac{1}{N} \sum_{k=0}^{N-1} \int_{-\infty}^{\infty} p(y|a_k) \cdot \log_2 \left\{ \frac{p(y|a_k)}{\sum_{i=0}^{N-1} p(y|a_i)} \right\} dy \\
&= \frac{\log_2(N)}{N} \sum_{k=0}^{N-1} 1 + \frac{1}{N} \sum_{k=0}^{N-1} \int_{-\infty}^{\infty} p(y|a_k) \cdot \log_2 \left\{ \frac{p(y|a_k)}{\sum_{i=0}^{N-1} p(y|a_i)} \right\} dy \\
&= \log_2(N) + \frac{1}{N} \sum_{k=0}^{N-1} \int_{-\infty}^{\infty} p(y|a_k) \cdot \log_2 \left\{ \frac{p(y|a_k)}{\sum_{i=0}^{N-1} p(y|a_i)} \right\} dy
\end{aligned} \tag{A.2}$$

where use was made of the fact that $p(y|a_k) = p(z)$, and then making a change of variable using $y = a_k + z$ which gives $dy = dz$. The y, a_i and a_k are vectors. They each have M real elements, or $L = M/2$ complex elements. The noise on each real variable component of the vector is independent, so the joint pdf can be written as the product of the pdfs in each dimension.

$$\begin{aligned} p\left(y|a_i\right) &= \prod_{j=0}^{M-1} \frac{1}{\sqrt{2\pi\sigma}} \exp\left(\frac{-\left(y_j - a_{i,j}\right)^2}{2\sigma^2}\right) \\ &= \left(\frac{1}{\sqrt{2\pi\sigma}}\right)^M \prod_{j=0}^{M-1} \exp\left(\frac{-\left(y_j - a_{i,j}\right)^2}{2\sigma^2}\right) \\ &= \left(2\pi\sigma^2\right)^{-\frac{M}{2}} \prod_{j=0}^{M-1} \exp\left(\frac{-\left(a_{k,j} + z_j - a_{i,j}\right)^2}{2\sigma^2}\right) \end{aligned} \tag{A.3}$$

Where $a_{i,j}$ is the *j*th element in the a_i vector. Remember that $y_j = a_{k,j} + z_j$, and that a_k is fixed externally to the above expression. Now, substitute the pdfs into the MI equation:

$$
\begin{aligned}
I\left(a_k, y\right) &= \log_2\left(N\right) + \frac{1}{N}\sum_{k=0}^{N-1}\int_{-\infty}^{\infty} p\left(y\middle|a_k\right)\cdot\log_2\left\{\frac{\left(2\pi\sigma^2\right)^{-\frac{M}{2}}\prod_{j=0}^{M-1}\exp\left(\frac{-\left(a_{k,j}+z_j-a_{k,j}\right)^2}{2\sigma^2}\right)}{\sum_{i=0}^{N-1}\left(2\pi\sigma^2\right)^{-\frac{M}{2}}\prod_{j=0}^{M-1}\exp\left(\frac{-\left(a_{k,j}+z_j-a_{i,j}\right)^2}{2\sigma^2}\right)}\right\}dy \\
&= \log_2\left(N\right) + \frac{1}{N}\sum_{k=0}^{N-1}\int_{-\infty}^{\infty} p\left(y\middle|a_k\right)\cdot\log_2\left\{\frac{\prod_{j=0}^{M-1}\exp\left(\frac{-\left(z_j\right)^2}{2\sigma^2}\right)}{\sum_{i=0}^{N-1}\prod_{j=0}^{M-1}\exp\left(\frac{-\left(a_{k,j}+z_j-a_{i,j}\right)^2}{2\sigma^2}\right)}\right\}dy \\
&= \log_2\left(N\right) + \frac{1}{N}\sum_{k=0}^{N-1}\int_{-\infty}^{\infty} p\left(y\middle|a_k\right)\cdot\log_2\left\{\frac{1}{\left(\prod_{j=0}^{M-1}\exp\left(\frac{-\left(z_j\right)^2}{2\sigma^2}\right)\right)^{-1}\sum_{i=0}^{N-1}\prod_{j=0}^{M-1}\exp\left(\frac{-\left(a_{k,j}+z_j-a_{i,j}\right)^2}{2\sigma^2}\right)}\right\}dy \\
&= \log_2\left(N\right) + \frac{1}{N}\sum_{k=0}^{N-1}\int_{-\infty}^{\infty} p\left(y\middle|a_k\right)\cdot\log_2\left\{\left[\left(\prod_{j=0}^{M-1}\exp\left(\frac{-\left(z_j\right)^2}{2\sigma^2}\right)\right)^{-1}\sum_{i=0}^{N-1}\prod_{j=0}^{M-1}\exp\left(\frac{-\left(a_{k,j}+z_j-a_{i,j}\right)^2}{2\sigma^2}\right)\right]^{-1}\right\}dy \\
&= \log_2\left(N\right) - \frac{1}{N}\sum_{k=0}^{N-1}\int_{-\infty}^{\infty} p\left(y\middle|a_k\right)\cdot\log_2\left\{\prod_{j=0}^{M-1}\exp\left(\frac{\left(z_j\right)^2}{2\sigma^2}\right)\sum_{i=0}^{N-1}\prod_{j=0}^{M-1}\exp\left(\frac{-\left(a_{k,j}+z_j-a_{i,j}\right)^2}{2\sigma^2}\right)\right\}dy \\
&= \log_2\left(N\right) - \frac{1}{N}\sum_{k=0}^{N-1}\int_{-\infty}^{\infty} p\left(y\middle|a_k\right)\cdot\log_2\left\{\sum_{i=0}^{N-1}\left(\prod_{j=0}^{M-1}\exp\left(\frac{\left(z_j\right)^2}{2\sigma^2}\right)\prod_{j=0}^{M-1}\exp\left(\frac{-\left(a_{k,j}+z_j-a_{i,j}\right)^2}{2\sigma^2}\right)\right)\right\}dy \tag{A.4} \\
&= \log_2\left(N\right) - \frac{1}{N}\sum_{k=0}^{N-1}\int_{-\infty}^{\infty} p\left(y\middle|a_k\right)\cdot\log_2\left\{\sum_{i=0}^{N-1}\left[\prod_{j=0}^{M-1}\left(\exp\left(\frac{-\left(a_{k,j}+z_j-a_{i,j}\right)^2}{2\sigma^2}\right)\exp\left(\frac{\left(z_j\right)^2}{2\sigma^2}\right)\right)\right]\right\}dy \\
&= \log_2\left(N\right) - \frac{1}{N}\sum_{k=0}^{N-1}\int_{-\infty}^{\infty} p\left(y\middle|a_k\right)\cdot\log_2\left\{\sum_{i=0}^{N-1}\left[\prod_{j=0}^{M-1}\exp\left(\frac{-\left(a_{k,j}+z_j-a_{i,j}\right)^2+\left(z_j\right)^2}{2\sigma^2}\right)\right]\right\}dy \\
&= \log_2\left(N\right) - \frac{1}{N}\sum_{k=0}^{N-1}\int_{-\infty}^{\infty} p\left(y\middle|a_k\right)\cdot\log_2\left\{\sum_{i=0}^{N-1}\left[\exp\left(\sum_{j=0}^{M-1}\frac{-\left(a_{k,j}+z_j-a_{i,j}\right)^2+\left(z_j\right)^2}{2\sigma^2}\right)\right]\right\}dy \\
&= \log_2\left(N\right) - \frac{1}{N}\sum_{k=0}^{N-1}\int_{-\infty}^{\infty} p\left(y\middle|a_k\right)\cdot\log_2\left\{\sum_{i=0}^{N-1}\left[\exp\left(\frac{1}{2\sigma^2}\sum_{j=0}^{M-1}\left(-\left(a_{k,j}+z_j-a_{i,j}\right)^2+\left(z_j\right)^2\right)\right)\right]\right\}dy \\
&= \log_2\left(N\right) - \frac{1}{N}\sum_{k=0}^{N-1}\int_{-\infty}^{\infty} p\left(y\middle|a_k\right)\cdot\log_2\left\{\sum_{i=0}^{N-1}\left[\exp\left(\frac{1}{2\sigma^2}\left(-\sum_{j=0}^{M-1}\left(a_{k,j}+z_j-a_{i,j}\right)^2+\sum_{j=0}^{M-1}\left(z_j\right)^2\right)\right)\right]\right\}dy \\
&= \log_2\left(N\right) - \frac{1}{N}\sum_{k=0}^{N-1}\int_{-\infty}^{\infty} p\left(y\middle|a_k\right)\cdot\log_2\left\{\sum_{i=0}^{N-1}\left[\exp\left(\frac{1}{2\sigma^2}\left(-\left|a_k+z-a_i\right|^2+\left|z\right|^2\right)\right)\right]\right\}dy \\
&= \log_2\left(N\right) - \frac{1}{N}\sum_{k=0}^{N-1}E\left\{\log_2\left[\sum_{i=0}^{N-1}\exp\left(\frac{1}{2\sigma^2}\left(-\left|a_k+z-a_i\right|^2+\left|z\right|^2\right)\right)\right]\right\} \\
&= \log_2\left(N\right) - \frac{1}{N}\sum_{k=0}^{N-1}E\left\{\log_2\left[\sum_{i=0}^{N-1}\exp\left(-\frac{\left|a_k+z-a_i\right|^2-\left|z\right|^2}{2\sigma^2}\right)\right]\right\}
\end{aligned}
$$

Once again, this expression agrees with that given by Ungerboeck [Ungerboeck1982], but the derivation has been conducted rigorously to ensure that this expression can be applied to multi-dimensional problems.

Appendix B Derivation of Envelope Based Expressions for MI

This Appendix gives the detailed derivations for the amplitude squared case and amplitude case to obtain expressions for the mutual information.

B.1 Derivation for Amplitude Squared HRR MI

The general form for MI with discrete inputs and continuous outputs is given by:

$$\begin{aligned} I(a_k, y) &= \sum_{k=0}^{N-1} P(k) \int_{-\infty}^{\infty} p(y|a_k) \cdot \log_2 \left\{ \frac{p(y|a_k)}{\sum_{i=0}^{N-1} P(i)\, p(y|a_i)} \right\} dy \\ &= \log_2(N) + \frac{1}{N} \sum_{k=0}^{N-1} \int_{-\infty}^{\infty} p(y|a_k) \cdot \log_2 \left\{ \frac{p(y|a_k)}{\sum_{i=0}^{N-1} p(y|a_i)} \right\} dy \end{aligned} \tag{B.1}$$

For the envelope only case, each range bin is processed by:

$$Y = I^2 + Q^2 . \tag{B.2}$$

This means that the pdf of the envelope squared of a range bin is given by a non-central chi-square pdf with two degrees of freedom. The generalized non-central chi-square pdf is given by:

$$p(y) = \frac{1}{2\sigma^2} \left(\frac{y}{s^2} \right)^{(n-2)/4} \exp\left(-\frac{s^2 + y}{2\sigma^2} \right) I_{n/2-1} \left(\sqrt{y} \frac{s}{\sigma^2} \right) \tag{B.3}$$

Which simplifies to

$$p(y) = \frac{1}{2\sigma^2} \exp\left(-\frac{s^2 + y}{2\sigma^2} \right) I_0 \left(\sqrt{y} \frac{s}{\sigma^2} \right) \tag{B.4}$$

For n=2 and has a non-centrality parameter of

$$s^2 = m_1^2 + m_2^2 . \tag{B.5}$$

The pdf can be reduced further if the non-centrality parameter is zero

$$p(y) = \frac{1}{2\sigma^2} \exp\left(-\frac{y}{2\sigma^2} \right) \tag{B.6}$$

If each range bin is assumed to be independent, then:

$$\begin{aligned}
p\left(y \middle| a_k\right) &= \prod_{j=0}^{N-1} p\left(y_j \middle| a_{kj}\right) \\
&= \prod_{j=0}^{N-1} \frac{1}{2\sigma^2} \exp\left(-\frac{s_{kj}^2 + y_{kj}}{2\sigma^2}\right) I_0\left(\sqrt{y_{kj}}\,\frac{s_{kj}}{\sigma^2}\right) \\
&= \prod_{j=0}^{N-1} \frac{1}{2\sigma^2} \exp\left(-\frac{a_{Ikj}^2 + a_{Qkj}^2 + \left(a_{Ikj} + z_{Ij}\right)^2 + \left(a_{Qkj} + z_{Qj}\right)^2}{2\sigma^2}\right) \times \\
&\quad I_0\left(\sqrt{\left(a_{Ikj} + z_{Ij}\right)^2 + \left(a_{Qkj} + z_{Qj}\right)^2}\,\frac{\sqrt{a_{Ikj}^2 + a_{Qkj}^2}}{\sigma^2}\right)
\end{aligned} \tag{B.7}$$

Note that the dy does not have to be converted to dz because samples from $p\left(y \middle| x\right)$ will be generated directly.

$$\begin{aligned}
p\left(y \middle| a_i\right) &= \prod_{j=0}^{N-1} p\left(y_j \middle| a_{ij}\right) \\
&= \prod_{j=0}^{N-1} \frac{1}{2\sigma^2} \exp\left(-\frac{s_{ij}^2 + y_{kj}}{2\sigma^2}\right) I_0\left(\sqrt{y_{kj}}\,\frac{s_{ij}}{\sigma^2}\right) \\
&= \prod_{j=0}^{N-1} \frac{1}{2\sigma^2} \exp\left(-\frac{a_{Iij}^2 + a_{Qij}^2 + \left(a_{Ikj} + z_{Ij}\right)^2 + \left(a_{Qkj} + z_{Qj}\right)^2}{2\sigma^2}\right) \times \\
&\quad I_0\left(\sqrt{\left(a_{Ikj} + z_{Ij}\right)^2 + \left(a_{Qkj} + z_{Qj}\right)^2}\,\frac{\sqrt{a_{Iij}^2 + a_{Qij}^2}}{\sigma^2}\right)
\end{aligned} \tag{B.8}$$

The inner probability ratio is then given by

$$\begin{aligned}
\frac{p\left(y|a_k\right)}{\sum_{i=0}^{N-1} p\left(y|a_i\right)} &= \frac{\prod_{j=0}^{N-1} \frac{1}{2\sigma^2} \exp\left(-\frac{s_{kj}^2 + y_{kj}}{2\sigma^2}\right) I_0\left(\sqrt{y_{kj}}\,\frac{s_{kj}}{\sigma^2}\right)}{\sum_{i=0}^{N-1} \prod_{j=0}^{N-1} \frac{1}{2\sigma^2} \exp\left(-\frac{s_{ij}^2 + y_{kj}}{2\sigma^2}\right) I_0\left(\sqrt{y_{kj}}\,\frac{s_{ij}}{\sigma^2}\right)} \\
&= \frac{\left(\frac{1}{2\sigma^2}\right)^N \prod_{j=0}^{N-1} I_0\left(\sqrt{y_{kj}}\,\frac{s_{kj}}{\sigma^2}\right)}{\left(\frac{1}{2\sigma^2}\right)^N \sum_{i=0}^{N-1} \prod_{j=0}^{N-1} \exp\left(\frac{s_{kj}^2 + y_{kj}}{2\sigma^2}\right) \exp\left(-\frac{s_{ij}^2 + y_{kj}}{2\sigma^2}\right) I_0\left(\sqrt{y_{kj}}\,\frac{s_{ij}}{\sigma^2}\right)} \\
&= \frac{\prod_{j=0}^{N-1} I_0\left(\sqrt{y_{kj}}\,\frac{s_{kj}}{\sigma^2}\right)}{\sum_{i=0}^{N-1} \prod_{j=0}^{N-1} \exp\left(\frac{s_{kj}^2 + y_{kj} - s_{ij}^2 - y_{kj}}{2\sigma^2}\right) I_0\left(\sqrt{y_{kj}}\,\frac{s_{ij}}{\sigma^2}\right)} \\
&= \frac{\prod_{j=0}^{N-1} I_0\left(\sqrt{y_{kj}}\,\frac{s_{kj}}{\sigma^2}\right)}{\sum_{i=0}^{N-1} \prod_{j=0}^{N-1} \exp\left(\frac{s_{kj}^2 - s_{ij}^2}{2\sigma^2}\right) I_0\left(\sqrt{y_{kj}}\,\frac{s_{ij}}{\sigma^2}\right)} \\
&= \frac{\prod_{j=0}^{N-1} I_0\left(\sqrt{y_{kj}}\,\frac{s_{kj}}{\sigma^2}\right)}{\sum_{i=0}^{N-1} \left(\prod_{j=0}^{N-1} \exp\left(\frac{s_{kj}^2 - s_{ij}^2}{2\sigma^2}\right) \prod_{j=0}^{N-1} I_0\left(\sqrt{y_{kj}}\,\frac{s_{ij}}{\sigma^2}\right)\right)} \\
&= \frac{\prod_{j=0}^{N-1} I_0\left(\sqrt{y_{kj}}\,\frac{s_{kj}}{\sigma^2}\right)}{\sum_{i=0}^{N-1} \left(\exp\left(\frac{1}{2\sigma^2} \sum_{j=0}^{N-1} \left(s_{kj}^2 - s_{ij}^2\right)\right) \prod_{j=0}^{N-1} I_0\left(\sqrt{y_{kj}}\,\frac{s_{ij}}{\sigma^2}\right)\right)}
\end{aligned} \tag{B.9}$$

Substituting this into the expression for MI:

$$\begin{aligned}
I\left(a_k, y\right) &= \sum_{k=0}^{N-1} P\left(k\right) \int_{-\infty}^{\infty} p\left(y|a_k\right) \cdot \log_2 \left\{ \frac{p\left(y|a_k\right)}{\sum_{i=0}^{N-1} P\left(i\right) p\left(y|a_i\right)} \right\} dy \\
&= \log_2\left(N\right) + \frac{1}{N} \sum_{k=0}^{N-1} \int_{-\infty}^{\infty} p\left(y|a_k\right) \cdot \log_2 \left\{ \frac{\prod_{j=0}^{N-1} I_0\left(\sqrt{y_{kj}}\,\frac{s_{kj}}{\sigma^2}\right)}{\sum_{i=0}^{N-1} \left(\exp\left(\frac{1}{2\sigma^2} \sum_{j=0}^{N-1} \left(s_{kj}^2 - s_{ij}^2\right)\right) \prod_{j=0}^{N-1} I_0\left(\sqrt{y_{kj}}\,\frac{s_{ij}}{\sigma^2}\right)\right)} \right\} dy \\
&= \log_2\left(N\right) + \frac{1}{N} \sum_{k=0}^{N-1} E \left\{ \log_2 \left[\frac{\prod_{j=0}^{N-1} I_0\left(\sqrt{y_{kj}}\,\frac{s_{kj}}{\sigma^2}\right)}{\sum_{i=0}^{N-1} \left(\exp\left(\frac{1}{2\sigma^2} \sum_{j=0}^{N-1} \left(s_{kj}^2 - s_{ij}^2\right)\right) \prod_{j=0}^{N-1} I_0\left(\sqrt{y_{kj}}\,\frac{s_{ij}}{\sigma^2}\right)\right)} \right] \right\}
\end{aligned} \tag{B.10}$$

Where

$$\begin{aligned} s_{ij} &= \sqrt{a_{Iij}^2 + a_{Qij}^2} \\ s_{kj} &= \sqrt{a_{Ikj}^2 + a_{Qkj}^2} \\ y_{kj} &= \left(a_{Ikj} + z_{Ij}\right)^2 + \left(a_{Qkj} + z_{Qj}\right)^2 \end{aligned} \tag{B.11}$$

B.2 Derivation for Envelope HRR MI

This approach was analysed as a check on the envelope squared HRR derivation. One would expect the two to approaches to give the same MI.

The general form for MI with discrete inputs and continuous outputs is given by:

$$\begin{aligned} I\left(a_k, y\right) &= \sum_{k=0}^{N-1} P\left(k\right) \int_{-\infty}^{\infty} p\left(y \middle| a_k\right) \cdot \log_2 \left\{ \frac{p\left(y \middle| a_k\right)}{\sum_{i=0}^{N-1} P\left(i\right) p\left(y \middle| a_i\right)} \right\} dy \\ &= \log_2\left(N\right) + \frac{1}{N} \sum_{k=0}^{N-1} \int_{-\infty}^{\infty} p\left(y \middle| a_k\right) \cdot \log_2 \left\{ \frac{p\left(y \middle| a_k\right)}{\sum_{i=0}^{N-1} p\left(y \middle| a_i\right)} \right\} dy \end{aligned} \tag{B.12}$$

For the envelope only case, each range bin is processed by:

$$Y = \sqrt{I^2 + Q^2} \; . \tag{B.13}$$

This means that the pdf of the envelope squared of a range bin is given by a Ricean pdf. The Ricean distribution is obtained by a change of variable induced by the function $R = \sqrt{Y}$, performed on a non-central Chi-squared, 2 degrees of freedom RV, and is given by:

$$p\left(y\right) = \frac{r}{\sigma^2} \exp\left(-\frac{r^2 + s^2}{2\sigma^2}\right) I_0\left(\frac{rs}{\sigma^2}\right) \tag{B.14}$$

where

$$s^2 = m_1^2 + m_2^2 \tag{B.15}$$

If each range bin is assumed to be independent, then:

$$
\begin{aligned}
p\left(y \mid a_k\right) &= \prod_{j=0}^{N-1} p\left(y_j \mid a_{kj}\right) \\
&= \prod_{j=0}^{N-1} \frac{r_{kj}}{2\sigma^2} \exp\left(-\frac{s_{kj}^2 + r_{kj}^2}{2\sigma^2}\right) I_0\left(\frac{r_{kj} s_{kj}}{\sigma^2}\right) \\
&= \prod_{j=0}^{N-1} \left[\frac{\sqrt{\left(a_{Ikj} + z_{Ij}\right)^2 + \left(a_{Qkj} + z_{Qj}\right)^2}}{2\sigma^2} \times \right. \\
&\quad \exp\left(-\frac{a_{Ikj}^2 + a_{Qkj}^2 + \left(a_{Ikj} + z_{Ij}\right)^2 + \left(a_{Qkj} + z_{Qj}\right)^2}{2\sigma^2}\right) \times \\
&\quad \left. I_0\left(\sqrt{\left(a_{Ikj} + z_{Ij}\right)^2 + \left(a_{Qkj} + z_{Qj}\right)^2} \, \frac{\sqrt{a_{Ikj}^2 + a_{Qkj}^2}}{\sigma^2}\right) \right]
\end{aligned}
\tag{B.16}
$$

Note that the dy does not have to be converted to dz because samples from $p\left(y \mid x\right)$ will be generated directly. The inner probability ratio is then given by

$$\begin{aligned}
\frac{p\left(y|a_k\right)}{\sum\limits_{i=0}^{N-1} p\left(y|a_i\right)} &= \frac{\prod\limits_{j=0}^{N-1} \frac{r_{kj}}{2\sigma^2} \exp\left(-\frac{s_{kj}^2 + r_{kj}^2}{2\sigma^2}\right) I_0\left(\frac{r_{kj} s_{kj}}{\sigma^2}\right)}{\sum\limits_{i=0}^{N-1} \prod\limits_{j=0}^{N-1} \frac{r_{kj}}{2\sigma^2} \exp\left(-\frac{s_{ij}^2 + r_{kj}^2}{2\sigma^2}\right) I_0\left(\frac{r_{kj} s_{ij}}{\sigma^2}\right)} \\
&= \frac{\prod\limits_{j=0}^{N-1} \exp\left(-\frac{s_{kj}^2 + r_{kj}^2}{2\sigma^2}\right) I_0\left(\frac{r_{kj} s_{kj}}{\sigma^2}\right)}{\sum\limits_{i=0}^{N-1} \prod\limits_{j=0}^{N-1} \exp\left(-\frac{s_{ij}^2 + r_{kj}^2}{2\sigma^2}\right) I_0\left(\frac{r_{kj} s_{ij}}{\sigma^2}\right)} \\
&= \frac{\prod\limits_{j=0}^{N-1} I_0\left(\frac{r_{kj} s_{kj}}{\sigma^2}\right)}{\sum\limits_{i=0}^{N-1} \prod\limits_{j=0}^{N-1} \exp\left(\frac{s_{kj}^2 + r_{kj}^2}{2\sigma^2}\right) \exp\left(-\frac{s_{ij}^2 + r_{kj}^2}{2\sigma^2}\right) I_0\left(\frac{r_{kj} s_{ij}}{\sigma^2}\right)} \\
&= \frac{\prod\limits_{j=0}^{N-1} I_0\left(\frac{r_{kj} s_{kj}}{\sigma^2}\right)}{\sum\limits_{i=0}^{N-1} \prod\limits_{j=0}^{N-1} \exp\left(\frac{s_{kj}^2 + r_{kj}^2}{2\sigma^2} - \frac{s_{ij}^2 + r_{kj}^2}{2\sigma^2}\right) I_0\left(\frac{r_{kj} s_{ij}}{\sigma^2}\right)} \\
&= \frac{\prod\limits_{j=0}^{N-1} I_0\left(\frac{r_{kj} s_{kj}}{\sigma^2}\right)}{\sum\limits_{i=0}^{N-1} \prod\limits_{j=0}^{N-1} \exp\left(\frac{s_{kj}^2 - s_{ij}^2}{2\sigma^2}\right) I_0\left(\frac{r_{kj} s_{ij}}{\sigma^2}\right)} \\
&= \frac{\prod\limits_{j=0}^{N-1} I_0\left(\frac{r_{kj} s_{kj}}{\sigma^2}\right)}{\sum\limits_{i=0}^{N-1} \prod\limits_{j=0}^{N-1} \exp\left(\frac{s_{kj}^2 - s_{ij}^2}{2\sigma^2}\right) \prod\limits_{j=0}^{N-1} I_0\left(\frac{r_{kj} s_{ij}}{\sigma^2}\right)} \\
&= \frac{\prod\limits_{j=0}^{N-1} I_0\left(\frac{r_{kj} s_{kj}}{\sigma^2}\right)}{\sum\limits_{i=0}^{N-1} \left(\exp\left(\frac{1}{2\sigma^2} \sum\limits_{j=0}^{N-1} \left(s_{kj}^2 - s_{ij}^2\right)\right) \prod\limits_{j=0}^{N-1} I_0\left(\frac{r_{kj} s_{ij}}{\sigma^2}\right)\right)}
\end{aligned} \tag{B.17}$$

Substituting this into the expression for MI:

$$
\begin{aligned}
I\left(a_k, y\right) &= \sum_{k=0}^{N-1} P\left(k\right)\int_{-\infty}^{\infty} p\left(y\middle|a_k\right)\cdot\log_2\left\{\frac{p\left(y\middle|a_k\right)}{\sum_{i=0}^{N-1} P\left(i\right)p\left(y\middle|a_i\right)}\right\}dy \\
&= \log_2\left(N\right)+\frac{1}{N}\sum_{k=0}^{N-1}\int_{-\infty}^{\infty} p\left(y\middle|a_k\right)\cdot\log_2\left\{\frac{\prod_{j=0}^{N-1} I_0\left(\frac{r_{kj}s_{kj}}{\sigma^2}\right)}{\sum_{i=0}^{N-1}\left(\exp\left(\frac{1}{2\sigma^2}\sum_{j=0}^{N-1}\left(s_{kj}^2-s_{ij}^2\right)\right)\prod_{j=0}^{N-1} I_0\left(\frac{r_{kj}s_{ij}}{\sigma^2}\right)\right)}\right\}dy \qquad \text{(B.18)} \\
&= \log_2\left(N\right)+\frac{1}{N}\sum_{k=0}^{N-1} E\left\{\log_2\left[\frac{\prod_{j=0}^{N-1} I_0\left(\frac{r_{kj}s_{kj}}{\sigma^2}\right)}{\sum_{i=0}^{N-1}\left(\exp\left(\frac{1}{2\sigma^2}\sum_{j=0}^{N-1}\left(s_{kj}^2-s_{ij}^2\right)\right)\prod_{j=0}^{N-1} I_0\left(\frac{r_{kj}s_{ij}}{\sigma^2}\right)\right)}\right]\right\}
\end{aligned}
$$

where

$$
\begin{aligned}
s_{ij} &= \sqrt{a_{Iij}^2 + a_{Qij}^2} \\
s_{kj} &= \sqrt{a_{Ikj}^2 + a_{Qkj}^2} \qquad \text{(B.19)} \\
r_{kj} &= \sqrt{\left(a_{Ikj} + z_{Ij}\right)^2 + \left(a_{Qkj} + z_{Qj}\right)^2}
\end{aligned}
$$

Testing for further simplification:

$$
\begin{aligned}
\frac{\prod_{j=0}^{N-1} I_0\left(\frac{r_{kj}s_{kj}}{\sigma^2}\right)}{\sum_{i=0}^{N-1}\prod_{j=0}^{N-1}\exp\left(\frac{s_{kj}^2-s_{ij}^2}{2\sigma^2}\right)I_0\left(\frac{r_{kj}s_{ij}}{\sigma^2}\right)} &= \frac{\prod_{j=0}^{N-1}\exp\left(-\frac{s_{kj}^2}{2\sigma^2}\right)\prod_{j=0}^{N-1} I_0\left(\frac{r_{kj}s_{kj}}{\sigma^2}\right)}{\sum_{i=0}^{N-1}\prod_{j=0}^{N-1}\exp\left(-\frac{s_{ij}^2}{2\sigma^2}\right)I_0\left(\frac{r_{kj}s_{ij}}{\sigma^2}\right)} \\
&= \frac{\exp(N/2\sigma^2)\prod_{j=0}^{N-1}\exp\left(-s_{kj}^2\right)\prod_{j=0}^{N-1} I_0\left(\frac{r_{kj}s_{kj}}{\sigma^2}\right)}{\exp(N/2\sigma^2)\sum_{i=0}^{N-1}\prod_{j=0}^{N-1}\exp\left(-s_{ij}^2\right)I_0\left(\frac{r_{kj}s_{ij}}{\sigma^2}\right)} \qquad \text{(B.20)} \\
&= \frac{\exp\left(-\sum_{j=0}^{N-1}s_{kj}^2\right)\prod_{j=0}^{N-1} I_0\left(\frac{r_{kj}s_{kj}}{\sigma^2}\right)}{\sum_{i=0}^{N-1}\left(\exp\left(-\sum_{j=0}^{N-1}s_{ij}^2\right)\prod_{j=0}^{N-1} I_0\left(\frac{r_{kj}s_{ij}}{\sigma^2}\right)\right)}
\end{aligned}
$$

Taking the logarithm to the base 2 gives:

$$
\begin{aligned}
&\log_2\left(\frac{\exp\left(-\sum_{j=0}^{N-1}s_{kj}^2\right)\prod_{j=0}^{N-1}I_0\left(\frac{r_{kj}s_{kj}}{\sigma^2}\right)}{\sum_{i=0}^{N-1}\left(\exp\left(-\sum_{j=0}^{N-1}s_{ij}^2\right)\prod_{j=0}^{N-1}I_0\left(\frac{r_{kj}s_{ij}}{\sigma^2}\right)\right)}\right)\\
&=\log_2\left(\exp\left(-\sum_{j=0}^{N-1}s_{kj}^2\right)\right)+\log_2\left(\frac{\prod_{j=0}^{N-1}I_0\left(\frac{r_{kj}s_{kj}}{\sigma^2}\right)}{\sum_{i=0}^{N-1}\left(\exp\left(-\sum_{j=0}^{N-1}s_{ij}^2\right)\prod_{j=0}^{N-1}I_0\left(\frac{r_{kj}s_{ij}}{\sigma^2}\right)\right)}\right)\\
&=-\frac{\sum_{j=0}^{N-1}s_{kj}^2}{\ln(2)}+\log_2\left(\frac{\prod_{j=0}^{N-1}I_0\left(\frac{r_{kj}s_{kj}}{\sigma^2}\right)}{\sum_{i=0}^{N-1}\left(\exp\left(-\sum_{j=0}^{N-1}s_{ij}^2\right)\prod_{j=0}^{N-1}I_0\left(\frac{r_{kj}s_{ij}}{\sigma^2}\right)\right)}\right)
\end{aligned}
\tag{B.21}
$$

Substituting into the expression for MI gives:

$$
\begin{aligned}
I(a_k,y)&=\log_2(N)+\frac{1}{N}\sum_{k=0}^{N-1}E\left\{-\frac{\sum_{j=0}^{N-1}s_{kj}^2}{\ln(2)}+\log_2\left(\frac{\prod_{j=0}^{N-1}I_0\left(\frac{r_{kj}s_{kj}}{\sigma^2}\right)}{\sum_{i=0}^{N-1}\left(\exp\left(-\sum_{j=0}^{N-1}s_{ij}^2\right)\prod_{j=0}^{N-1}I_0\left(\frac{r_{kj}s_{ij}}{\sigma^2}\right)\right)}\right)\right\}\\
&=\log_2(N)+\frac{1}{N}\sum_{k=0}^{N-1}\left\{-\frac{\sum_{j=0}^{N-1}s_{kj}^2}{\ln(2)}+E\left\{\log_2\left(\frac{\prod_{j=0}^{N-1}I_0\left(\frac{r_{kj}s_{kj}}{\sigma^2}\right)}{\sum_{i=0}^{N-1}\left(\exp\left(-\sum_{j=0}^{N-1}s_{ij}^2\right)\prod_{j=0}^{N-1}I_0\left(\frac{r_{kj}s_{ij}}{\sigma^2}\right)\right)}\right)\right\}\right\}\\
&=\log_2(N)-\frac{1}{N\ln(2)}\sum_{k=0}^{N-1}\sum_{j=0}^{N-1}s_{kj}^2+\frac{1}{N}\sum_{k=0}^{N-1}E\left\{\log_2\left(\frac{\prod_{j=0}^{N-1}I_0\left(\frac{r_{kj}s_{kj}}{\sigma^2}\right)}{\sum_{i=0}^{N-1}\left(\exp\left(-\sum_{j=0}^{N-1}s_{ij}^2\right)\prod_{j=0}^{N-1}I_0\left(\frac{r_{kj}s_{ij}}{\sigma^2}\right)\right)}\right)\right\}
\end{aligned}
\tag{B.22}
$$

Appendix C Derivation of MI with Correlated Noise

The expression for MI is in terms of real variables, so first the relationship between the covariance matrix of a complex RV and the covariance of the equivalent real covariance matrix has to be derived - see Urkowitz, pp 296 [Urkowitz1983].

Definition of a Gaussian RV vector (actually Gaussian random vector variable)(Urkowitz pp. 313):

$$\begin{aligned} p(\mathbf{x}) &= p(x_1, x_2, \ldots, x_D) \\ &= \frac{1}{(2\pi)^{D/2}\sqrt{\mathbf{K}}} \exp\left(-\frac{1}{2}\sum_{i=1}^{D}\sum_{k=1}^{D}(x_i - m_i)(x_k - m_k)[\mathbf{K}^{-1}]_{ik}\right) \\ &= \frac{1}{(2\pi)^{D/2}\sqrt{\mathbf{K}}} \exp\left(-\frac{1}{2}(\mathbf{x}-\mathbf{m})^T \mathbf{K}^{-1}(\mathbf{x}-\mathbf{m})\right) \end{aligned} \tag{C.1}$$

Notes

$\mathbf{x}$ is a column vector.

$\mathbf{x}$ is the independent variable.

If $\mathbf{z}$ is a n-vector of independent Gaussians, then $\mathbf{x} = \mathbf{Az} + \mathbf{u}$.

$\mathbf{A}$ is k x n, and covariance matrix is $\mathbf{AA}^{\mathrm{T}}$, which is k x k, but has a rank of n.

Principle axes of ellipsoids are given by the eigenvectors of the covariance matrix.

Squared relative lengths of the principle axes are given by the corresponding eigenvalues.

For noise, the mean is zero, giving:

$$p(\mathbf{x}) = \frac{1}{(2\pi)^{D/2}\sqrt{\mathbf{K}}} \exp\left(-\frac{1}{2}\mathbf{x}^T\mathbf{K}^{-1}\mathbf{x}\right) \tag{C.2}$$

Each element of the covariance matrix has the following structure:

$$\begin{aligned} K_{ij} &= \rho_{ik}\sigma_i\sigma_k \\ K_{ij} &= K_{ji} \\ \rho_{ik} &= \rho_{ki} \end{aligned} \tag{C.3}$$

For the case at hand, all the noise powers are at the same level, so

$$K_{ij} = \rho_{ik}\sigma^2 \tag{C.4}$$

and the noise power can be factored out of the covariance matrix:

$$\begin{aligned}\mathbf{K} &= \sigma^2\mathbf{K}' \\ \mathbf{K}^{-1} &= \frac{1}{\sigma^2}\mathbf{K}'^{-1}\end{aligned} \tag{C.5}$$

The mutual information for a discrete memoryless channel with continuous outputs is given by:

$$\begin{aligned}I(a_k, y) &= \sum_{k=0}^{N-1} P(k)\int_{-\infty}^{\infty} p(y|a_k)\cdot\log_2\left\{\frac{p(y|a_k)}{\sum\limits_{i=0}^{N-1} P(i)\,p(y|a_i)}\right\}dy \\ &= \frac{1}{N}\sum_{k=0}^{N-1}\int_{-\infty}^{\infty} p(y|a_k)\cdot\log_2\left\{\frac{p(y|a_k)}{\frac{1}{N}\sum\limits_{i=0}^{N-1} p(y|a_i)}\right\}dy \\ &= \frac{1}{N}\sum_{k=0}^{N-1}\int_{-\infty}^{\infty} p(y|a_k)\cdot\log_2(N)\,dy + \frac{1}{N}\sum_{k=0}^{N-1}\int_{-\infty}^{\infty} p(y|a_k)\cdot\log_2\left\{\frac{p(y|a_k)}{\sum\limits_{i=0}^{N-1} p(y|a_i)}\right\}dy \qquad \text{(C.6)} \\ &= \frac{1}{N}N\log_2(N) + \frac{1}{N}\sum_{k=0}^{N-1}\int_{-\infty}^{\infty} p(y|a_k)\cdot\log_2\left\{\frac{p(y|a_k)}{\sum\limits_{i=0}^{N-1} p(y|a_i)}\right\}dy \\ &= \log_2(N) + \frac{1}{N}\sum_{k=0}^{N-1}\int_{-\infty}^{\infty} p(y|a_k)\cdot\log_2\left\{\frac{p(y|a_k)}{\sum\limits_{i=0}^{N-1} p(y|a_i)}\right\}dy\end{aligned}$$

Where N is the number of symbols and P() is the a-priori probability of the symbols and a_k is a vector.

The interference is correlated, but additive: i.e. $y = a_k + z$, where z is from a correlated Gaussian noise source, with a known K.

$$
\begin{aligned}
p\left(y\,|\,\mathbf{a}_k\right) &= \frac{1}{\left(2\pi\right)^{D/2}\sqrt{|\mathbf{K}|}}\exp\left(-\frac{1}{2}\left(y-\mathbf{a}_k\right)^T\mathbf{K}^{-1}\left(y-\mathbf{a}_k\right)\right)\\
&= \frac{1}{\left(2\pi\right)^{D/2}\sqrt{|\mathbf{K}|}}\exp\left(-\frac{1}{2}\left(\mathbf{a}_k+z-\mathbf{a}_k\right)^T\mathbf{K}^{-1}\left(\mathbf{a}_k+z-\mathbf{a}_k\right)\right)\\
&= \frac{1}{\left(2\pi\right)^{D/2}\sqrt{|\mathbf{K}|}}\exp\left(-\frac{1}{2}\left(z\right)^T\mathbf{K}^{-1}\left(z\right)\right)\\
&= \frac{1}{\left(2\pi\right)^{D/2}\sqrt{|\mathbf{K}|}}\exp\left(-\frac{1}{2}z^T\mathbf{K}^{-1}\mathrm{z}\right)
\end{aligned}
\qquad \text{(C.7)}
$$

Now, for the $p\left(y|a_i\right)$ term:

$$
\begin{aligned}
p\left(y\,|\,\mathbf{a}_i\right) &= \frac{1}{\left(2\pi\right)^{D/2}\sqrt{|\mathbf{K}|}}\exp\left(-\frac{1}{2}\left(y-\mathbf{a}_i\right)^T\mathbf{K}^{-1}\left(y-\mathbf{a}_i\right)\right)\\
&= \frac{1}{\left(2\pi\right)^{D/2}\sqrt{|\mathbf{K}|}}\exp\left(-\frac{1}{2}\left(\mathbf{a}_k+z-\mathbf{a}_i\right)^T\mathbf{K}^{-1}\left(\mathbf{a}_k+z-\mathbf{a}_i\right)\right)
\end{aligned}
\qquad \text{(C.8)}
$$

Now substitute into the argument of the $\log_2\left(\bullet\right)$, assuming P(i) are equal:

$$
\begin{aligned}
\frac{p\left(y|a_k\right)}{\sum_{i=0}^{N-1}p\left(y|a_i\right)} &= \frac{\frac{1}{\left(2\pi\right)^{D/2}\sqrt{|\mathbf{K}|}}\exp\left(-\frac{1}{2}z^T\mathbf{K}^{-1}\mathrm{z}\right)}{\sum_{i=0}^{N-1}\frac{1}{\left(2\pi\right)^{D/2}\sqrt{|\mathbf{K}|}}\exp\left(-\frac{1}{2}\left(\mathbf{a}_k+\mathrm{z}-\mathbf{a}_i\right)^T\mathbf{K}^{-1}\left(\mathbf{a}_k+\mathrm{z}-\mathbf{a}_i\right)\right)}\\
&= \frac{\exp\left(-\frac{1}{2}z^T\mathbf{K}^{-1}\mathrm{z}\right)}{\sum_{i=0}^{N-1}\exp\left(-\frac{1}{2}\left(\mathbf{a}_k+z-\mathbf{a}_i\right)^T\mathbf{K}^{-1}\left(\mathbf{a}_k+z-\mathbf{a}_i\right)\right)}\\
&= \left[\exp\left(\frac{1}{2}z^T\mathbf{K}^{-1}z\right)\sum_{i=0}^{N-1}\exp\left(-\frac{1}{2}\left(\mathbf{a}_k+z-\mathbf{a}_i\right)^T\mathbf{K}^{-1}\left(\mathbf{a}_k+z-\mathbf{a}_i\right)\right)\right]^{-1}\\
&= \left[\sum_{i=0}^{N-1}\exp\left(\frac{1}{2}z^T\mathbf{K}^{-1}\mathrm{z}\right)\exp\left(-\frac{1}{2}\left(\mathbf{a}_k+z-\mathbf{a}_i\right)^T\mathbf{K}^{-1}\left(\mathbf{a}_k+z-\mathbf{a}_i\right)\right)\right]^{-1}\\
&= \left[\sum_{i=0}^{N-1}\exp\left(\frac{1}{2}z^T\mathbf{K}^{-1}\mathrm{z}\right)\exp\left(-\frac{1}{2}\left(\mathbf{a}_k+z-\mathbf{a}_i\right)^T\mathbf{K}^{-1}\left(\mathbf{a}_k+z-\mathbf{a}_i\right)\right)\right]^{-1}\\
&= \left[\sum_{i=0}^{N-1}\exp\left(-\frac{1}{2}\left(\left(\mathbf{a}_k+z-\mathbf{a}_i\right)^T\mathbf{K}^{-1}\left(\mathbf{a}_k+z-\mathbf{a}_i\right)-z^T\mathbf{K}^{-1}\mathrm{z}\right)\right)\right]^{-1}
\end{aligned}
\qquad \text{(C.9)}
$$

Now, multiplying out the second $\exp(\bullet)$ argument to obtain a $\mathbf{z}^{\mathbf{T}}\mathbf{K}\mathbf{z}$:

$$\begin{aligned}
&\left(\mathbf{a}_k + z - \mathbf{a}_i\right)^T \mathbf{K}^{-1}\left(\mathbf{a}_k + z - \mathbf{a}_i\right) \\
&= \left(\left(\left(\mathbf{a}_k - \mathbf{a}_i\right)^T + z^T\right)\mathbf{K}^{-1}\right)\left(\left(\mathbf{a}_k - \mathbf{a}_i\right) + z\right) \\
&= \left(\left(\mathbf{a}_k - \mathbf{a}_i\right)^T \mathbf{K}^{-1} + z^T\mathbf{K}^{-1}\right)\left(\left(\mathbf{a}_k - \mathbf{a}_i\right) + z\right) \\
&= \left(\left(\mathbf{a}_k - \mathbf{a}_i\right)^T \mathbf{K}^{-1} + z^T\mathbf{K}^{-1}\right)\left(\mathbf{a}_k - \mathbf{a}_i\right) + \left(\left(\mathbf{a}_k - \mathbf{a}_i\right)^T \mathbf{K}^{-1} + z^T\mathbf{K}^{-1}\right)z \\
&= \left(\mathbf{a}_k - \mathbf{a}_i\right)^T \mathbf{K}^{-1}\left(\mathbf{a}_k - \mathbf{a}_i\right) + z^T\mathbf{K}^{-1}\left(\mathbf{a}_k - \mathbf{a}_i\right) + \left(\mathbf{a}_k - \mathbf{a}_i\right)^T \mathbf{K}^{-1}z + z^T\mathbf{K}^{-1}z
\end{aligned} \tag{C.10}$$

Cancelling the zKz term gives:

$$\begin{aligned}
&\left(\mathbf{a}_k - \mathbf{a}_i\right)^T \mathbf{K}^{-1}\left(\mathbf{a}_k - \mathbf{a}_i\right) + z^T\mathbf{K}^{-1}\left(\mathbf{a}_k - \mathbf{a}_i\right) + \left(\mathbf{a}_k - \mathbf{a}_i\right)^T \mathbf{K}^{-1}z \\
&= \mathbf{d}_{ki}^T\mathbf{K}^{-1}\mathbf{d}_{ki} + z^T\mathbf{K}^{-1}\mathbf{d}_{ki} + \mathbf{d}_{ki}^T\mathbf{K}^{-1}z \\
&\text{OR} \\
&= \left(\mathbf{a}_k - \mathbf{a}_i\right)^T \mathbf{K}^{-1}\left(\mathbf{a}_k - \mathbf{a}_i + z\right) + z^T\mathbf{K}^{-1}\left(\mathbf{a}_k - \mathbf{a}_i\right) \\
&= \mathbf{d}_{ki}^T\mathbf{K}^{-1}\left(\mathbf{d}_{ki} + z\right) + z^T\mathbf{K}^{-1}\mathbf{d}_{ki}
\end{aligned} \tag{C.11}$$

The $\mathbf{K}^{-1}$ might whiten the $\mathbf{z}$, but lack of the handy E{} means we can't substitute $\sigma\mathbf{I}$.

The first option requires two dot products and a scalar addition, whereas the second requires a vector addition and two dot products.

Using the 2nd line of the first option, we can simplify further:

$$\begin{aligned}
&\mathbf{d}_{ki}^T\mathbf{K}^{-1}\mathbf{d}_{ki} + z^T\mathbf{K}^{-1}\mathbf{d}_{ki} + \mathbf{d}_{ki}^T\mathbf{K}^{-1}z \\
&= \mathbf{d}_{ki}^T\mathbf{K}^{-1}\mathbf{d}_{ki} + z^T\mathbf{K}^{-1}\mathbf{d}_{ki} + \left(\left(\mathbf{d}_{ki}^T\mathbf{K}^{-1}z\right)^T\right)^T \\
&= \mathbf{d}_{ki}^T\mathbf{K}^{-1}\mathbf{d}_{ki} + z^T\mathbf{K}^{-1}\mathbf{d}_{ki} + \left(z^T\mathbf{K}^{-1}\mathbf{d}_{ki}\right)^T
\end{aligned} \tag{C.12}$$

But the factor inside the brackets is a scalar, so:

$$\begin{aligned}
&\mathbf{d}_{ki}^T\mathbf{K}^{-1}\mathbf{d}_{ki} + z^T\mathbf{K}^{-1}\mathbf{d}_{ki} + \left(z^T\mathbf{K}^{-1}\mathbf{d}_{ki}\right)^T \\
&= \mathbf{d}_{ki}^T\mathbf{K}^{-1}\mathbf{d}_{ki} + 2z^T\mathbf{K}^{-1}\mathbf{d}_{ki} \\
&= \left(\mathbf{d}_{ki}^T + 2z^T\right)\mathbf{K}^{-1}\mathbf{d}_{ki}
\end{aligned} \tag{C.13}$$

Substituting back using the second line (i.e. first option), we have:

$$\begin{aligned}\frac{p(y|a_k)}{\sum_{i=0}^{N-1} p(y|a_i)} &= \left[\sum_{i=0}^{N-1}\exp\left(-\frac{1}{2}\left((\mathbf{a}_k+z-\mathbf{a}_i)^T\mathbf{K}^{-1}(\mathbf{a}_k+z-\mathbf{a}_i)-z^T\mathbf{K}^{-1}z\right)\right)\right]^{-1}\\ &=\left[\sum_{i=0}^{N-1}\exp\left(-\frac{1}{2}(\mathbf{d}_{ki}^T+2z^T)\mathbf{K}^{-1}\mathbf{d}_{ki}\right)\right]^{-1}\end{aligned} \tag{C.14}$$

Now substituting this back we have:

$$\begin{aligned}I(a_k,y) &= \log_2(N)+\frac{1}{N}\sum_{k=0}^{N-1}\int_{-\infty}^{\infty} p(y|a_k)\cdot\log_2\left\{\frac{p(y|a_k)}{\sum_{i=0}^{N-1}p(y|a_i)}\right\}dy\\ &= \log_2(N)+\frac{1}{N}\sum_{k=0}^{N-1}\int_{-\infty}^{\infty} p(y|a_k)\cdot\log_2\left\{\left[\sum_{i=0}^{N-1}\exp\left(-\frac{1}{2}(\mathbf{d}_{ki}^T+2z^T)\mathbf{K}^{-1}\mathbf{d}_{ki}\right)\right]^{-1}\right\}dy\\ &= \log_2(N)-\frac{1}{N}\sum_{k=0}^{N-1}\int_{-\infty}^{\infty} p(y|a_k)\cdot\log_2\left\{\sum_{i=0}^{N-1}\exp\left(-\frac{1}{2}(\mathbf{d}_{ki}^T+2z^T)\mathbf{K}^{-1}\mathbf{d}_{ki}\right)\right\}dy\\ &= \log_2(N)-\frac{1}{N}\sum_{k=0}^{N-1}E\left\{\log_2\left[\sum_{i=0}^{N-1}\exp\left(-\frac{1}{2}(\mathbf{d}_{ki}^T+2z^T)\mathbf{K}^{-1}\mathbf{d}_{ki}\right)\right]\right\}\\ &= \log_2(N)-\frac{1}{N}\sum_{k=0}^{N-1}E\left\{\log_2\left[\sum_{i=0}^{N-1}\exp\left(-\frac{1}{2}(\mathbf{d}_{ki}^T+2z^T)\mathbf{K}^{-1}\mathbf{d}_{ki}\right)\right]\right\}\end{aligned} \tag{C.15}$$

Then using:

$$\begin{aligned}\mathbf{K} &= \sigma^2\mathbf{K}'\\ \mathbf{K}^{-1} &= \frac{1}{\sigma^2}\mathbf{K}'^{-1}\end{aligned} \tag{C.16}$$

Gives

$$I(a_k,y)=\log_2(N)-\frac{1}{N}\sum_{k=0}^{N-1}E\left\{\log_2\left[\sum_{i=0}^{N-1}\exp\left(-\frac{1}{2\sigma^2}(\mathbf{d}_{ki}+2z)^T\mathbf{K}'^{-1}\mathbf{d}_{ki}\right)\right]\right\} \tag{C.17}$$

Which allows the SNR to be manipulated more readily.

Define SNR as the SNR before windowing, but make use of windowed waveforms on the reception side.

C.1 Verification of MI Expression for Correlated Noise when there is no Correlation

Attempt to prove that the final expression from the previous section is equivalent to the "standard" MI expression when the noise is not correlated.

Standard Expression:

$$I\left(a_k, y\right) = \log_2\left(N\right) - \frac{1}{N}\sum_{k=0}^{N-1} E\left\{\log_2\left[\sum_{i=0}^{N-1}\exp\left(-\frac{\left|a_k + z - a_i\right|^2 - \left|z\right|^2}{2\sigma^2}\right)\right]\right\} \tag{C.18}$$

Correlated noise expression:

$$I\left(a_k, y\right) = \log_2\left(N\right) - \frac{1}{N}\sum_{k=0}^{N-1} E\left\{\log_2\left[\sum_{i=0}^{N-1}\exp\left(-\frac{1}{2}\left(\mathbf{d}_{ki}^T + 2z^T\right)\mathbf{K}^{-1}\mathbf{d}_{ki}\right)\right]\right\} \tag{C.19}$$

The two expressions above have the same form, except for the argument of the $\exp\left(\bullet\right)$. Now, given that the noise power can be factored out of the covariance matrix, and the noise is uncorrelated, the following expression can be derived and substituted into the correlated noise MI expression:

$$\begin{aligned}\mathbf{K} &= \sigma^2\mathbf{K}' \\ \mathbf{K}^{-1} &= \frac{1}{\sigma^2}\mathbf{K}'^{-1} \\ &= \frac{1}{\sigma^2}\mathbf{I}^{-1} \\ &= \frac{1}{\sigma^2}\mathbf{I}\end{aligned} \tag{C.20}$$

Substituting this result in the correlated noise MI expression gives:

$$\begin{aligned}I\left(a_k, y\right) &= \log_2\left(N\right) - \frac{1}{N}\sum_{k=0}^{N-1} E\left\{\log_2\left[\sum_{i=0}^{N-1}\exp\left(-\frac{1}{2}\left(\mathbf{d}_{ki}^T + 2z^T\right)\mathbf{K}^{-1}\mathbf{d}_{ki}\right)\right]\right\} \\ &= \log_2\left(N\right) - \frac{1}{N}\sum_{k=0}^{N-1} E\left\{\log_2\left[\sum_{i=0}^{N-1}\exp\left(-\frac{1}{2}\left(\mathbf{d}_{ki}^T + 2z^T\right)\left(\frac{1}{\sigma^2}\mathbf{I}\right)\mathbf{d}_{ki}\right)\right]\right\} \\ &= \log_2\left(N\right) - \frac{1}{N}\sum_{k=0}^{N-1} E\left\{\log_2\left[\sum_{i=0}^{N-1}\exp\left(-\frac{1}{2\sigma^2}\left(\mathbf{d}_{ki}^T + 2z^T\right)\mathbf{d}_{ki}\right)\right]\right\} \\ &= \log_2\left(N\right) - \frac{1}{N}\sum_{k=0}^{N-1} E\left\{\log_2\left[\sum_{i=0}^{N-1}\exp\left(-\frac{1}{2\sigma^2}\left(\left(\mathbf{a}_k - \mathbf{a}_i\right)^T + 2z^T\right)\left(\mathbf{a}_k - \mathbf{a}_i\right)\right)\right]\right\}\end{aligned} \tag{C.21}$$

Expanding the inner expression gives:

$$\begin{aligned}&\left((\mathbf{a}_k-\mathbf{a}_i)^T+2z^T\right)(\mathbf{a}_k-\mathbf{a}_i)\\&=\left(\mathbf{a}_k^T-\mathbf{a}_i^T+2z^T\right)\left(\mathbf{a}_k-\mathbf{a}_i\right)\\&=\mathbf{a}_k^T\mathbf{a}_k-\mathbf{a}_i^T\mathbf{a}_k+2z^T\mathbf{a}_k-\mathbf{a}_k^T\mathbf{a}_i+\mathbf{a}_i^T\mathbf{a}_i-2z^T\mathbf{a}_i\end{aligned} \tag{C.22}$$

Now, using the fact that inner products are commutative, that is:

$$\begin{aligned}z^T\mathbf{a}_k&=\mathbf{a}_k^T z\\z^T\mathbf{a}_i&=\mathbf{a}_i^T z\end{aligned} \tag{C.23}$$

for this case and expanding the inner expression for the uncorrelated MI expression gives:

$$\begin{aligned}&\left|\mathbf{a}_k-\mathbf{a}_i+z\right|^2-\left|z\right|^2\\&=\left(\mathbf{a}_k-\mathbf{a}_i+z\right)^T\left(\mathbf{a}_k-\mathbf{a}_i+z\right)-z^Tz\\&=\mathbf{a}_k^T\mathbf{a}_k-\mathbf{a}_i^T\mathbf{a}_k+z^T\mathbf{a}_k-\mathbf{a}_k^T\mathbf{a}_i+\mathbf{a}_i^T\mathbf{a}_i-z^T\mathbf{a}_i+\mathbf{a}_k^Tz-\mathbf{a}_i^Tz+z^Tz-z^Tz\\&=\mathbf{a}_k^T\mathbf{a}_k-\mathbf{a}_i^T\mathbf{a}_k+z^T\mathbf{a}_k-\mathbf{a}_k^T\mathbf{a}_i+\mathbf{a}_i^T\mathbf{a}_i-z^T\mathbf{a}_i+\mathbf{a}_k^Tz-\mathbf{a}_i^Tz\\&=\mathbf{a}_k^T\mathbf{a}_k-\mathbf{a}_i^T\mathbf{a}_k+2z^T\mathbf{a}_k-\mathbf{a}_k^T\mathbf{a}_i+\mathbf{a}_i^T\mathbf{a}_i-2z^T\mathbf{a}_i\end{aligned} \tag{C.24}$$

which shows that the expression for the correlated case simplifies to the expression for the uncorrelated case.

C.2 Derivation of Covariance Matrix Induced by a Windowed FFT

This section describes the derivation of the complex covariance matrix of the output of a windowed FFT.

Starting with a correlated multidimensional Gauss vector with covariance matrix K.

$$y = Ax \tag{C.25}$$

Where y and x are column vectors, and x is a zero mean unity variance vector of Gaussian random variables. (All variables could be complex if required.)

$$K = AA^H \tag{C.26}$$

Is the Cholesky decomposition of K. (in Matlab use chol() function with 'lower' parameter.)

Now let z be a correlated (due to windowing) sample of the M-D Gaussian vector having covariance matrix K.

$$\begin{aligned} z &= \operatorname{diag}\left(\left(Ax\right)b\right) \\ &= B\left(Ax\right) \end{aligned} \tag{C.27}$$

Where b is a row vector containing the window function and B is a square matrix, with the window function placed on the diagonal and zero otherwise. The function diag( ) extracts the diagonal of its argument as a column vector.

Now let u be the FFT of z:

$$\begin{aligned} u &= Wz \\ &= WBAx \end{aligned} \tag{C.28}$$

The covariance matrix of the output can be calculated as follows:

$$\begin{aligned} K_{FFT} &= E\left\{uu^H\right\} \\ &= E\left\{\left(WBAx\right)\left(WBAx\right)^H\right\} \\ &= E\left\{WBAxx^H A^H B^H W^H\right\} \\ &= WBE\left\{Axx^H A^H\right\} B^H W^H \\ &= WBKB^H W^H \\ &= WQW^H \end{aligned} \tag{C.29}$$

where

$$Q = \left(b^H b\right) \circ K \tag{C.30}$$

and $\circ$ is the Hadamard product (i.e. elementwise product)

Note that if the DFT matrix is unitary (scaled by $1/\sqrt{N}$) then:

$$W_u^{-1} = W_u^H \tag{C.31}$$

This should have been derived starting from the expression below:

$$W_u^{-1} = W_u^* \tag{C.32}$$

but

$$\begin{aligned} W_u &= W_u^T \\ W_u^* &= W_u^H \end{aligned} \tag{C.33}$$

so

$$W_u^{-1} = W_u^H \tag{C.34}$$

And

$$\begin{aligned} W_u^H &= W_u^{-1} \\ \left(W_u^H\right)^{-1} &= \left(W_u^{-1}\right)^{-1} = W_u \end{aligned} \tag{C.35}$$

If there is no transform A, then:

$$\begin{aligned} K &= E\left\{uu^H\right\} \\ &= E\left\{\left(WBx\right)\left(WBx\right)^H\right\} \\ &= E\left\{WBxx^H B^H W^H\right\} \\ &= WBE\left\{xx^H\right\} B^H W^H \\ &= 2\sigma^2 WBIB^H W^H \\ &= 2\sigma^2 WB^2 W^H \end{aligned} \tag{C.36}$$

Then the inverse of covariance matrix can be calculated as:

$$\begin{aligned} K^{-1} &= \left(2\sigma^2 W B^2 W^H\right)^{-1} \\ &= \left(2\sigma^2\right)^{-1}\left(W^H\right)^{-1}\left(WB^2\right)^{-1} \\ &= \left(2\sigma^2\right)^{-1}\left(W^H\right)^{-1}\left(B^2\right)^{-1}\left(W\right)^{-1} \\ &= \left(2\sigma^2\right)^{-1} W B^{-2} W^H \end{aligned} \tag{C.37}$$

Best approach seems to be to take the mean of the upper triangular matrix and conjugate transpose of the lower triangular, and then copy this as the lower and upper triangular matrix. Then use the real part of the diagonal as the diagonal followed by the Choleskey inverse.

The same seems to be the case for the real matrix, although the averaging idea still needs to be tested.

Appendix D Multi-variate Information Measures

This appendix describes the difference between multi-information, co-information and conditional information. This section is a summary based on the following sources: [Jakulin2004, Jin2007 and Bell2003].

The notation and nomenclature is summarised as follows:

Attribute – input RV to a “system”. (independent but mutually exclusive events)

Label – output RV from a system. (Also an attribute)

Interaction – dependence in a whole set of events, but not any subset.

Extra attributes cannot change existing interactions.

Interactions are unambiguous, symmetric and undirected. Interactions which connect a set of attributes together.

D.1 Conditional Independence

Events are independent if and only if

$$P(A,B) = P(A)P(B) \tag{D.1}$$

This implies that neither event has any influence over the other event:

$$P(A|B) = P(A) \tag{D.2}$$

and

$$P(B|A) = P(B) \tag{D.3}$$

But A and B might become dependent if C is observed – “flickering dependencies”

Conditional independence proposed as a solution to this problem:

$$P(A,B|C) = P(A|C)P(B|C) \tag{D.4}$$

or equivalently:

$$P(A|B,C) = P(A|C) \tag{D.5}$$

This concept should be verified for the case: NOT A and NOT B (both independent and conditionally. independent)

If A and B are independent, they do not 2-interact.

If C affects the dependence between A and B, then A, B and C 3-interact.

Negative interaction implies redundance, and positive interaction implies synergy.

D.2 Entropy

$$H(A) = \sum_{a \in A} P(a) \log_2 \left(P(a) \right) \tag{D.6}$$

The higher the entropy, the less reliable predictions that can be made about the random variable.

Entropy for two attributes: Conditional entropy – uncertainty about A having knowledge of B.

$$\begin{aligned} H(A \mid B) &= \sum_{a \in A, b \in B} P(a,b) \log_2 \left(P(a \mid b) \right) \\ &= H(A,B) - H(B) \end{aligned} \tag{D.7}$$

D.3 Mutual Information

$$I(A;B) = \sum_{a \in A, b \in B} P(a,b) \log_2 \left(\frac{P(a,b)}{P(a)P(b)} \right) \tag{D.8}$$

$$\begin{aligned} I(A;B) &= H(A) + H(B) - H(A,B) \\ &= H(A) - H(A \mid B) \\ &= H(B) - H(B \mid A) \\ &= I(B;A) \end{aligned} \tag{D.9}$$

This is a measure of correlation between attributes or alternatively the reduction in uncertainty.

MI always zero or positive, and is zero if and only if the attributes are independent.

I(Attribute;Label) is the information provided by A about L, and is called the “information gain”.

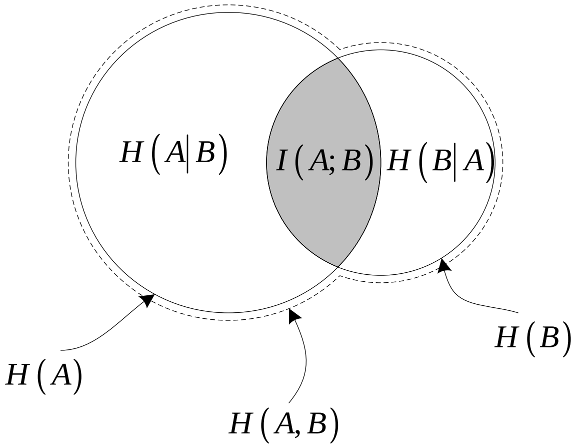


Figure D.1: Venn diagram for two attribute interactions.

Examples of joint probability distributions to aid understanding of MI:

D.4 Entropy of Three Attributes

How much uncertainty remains about A after having observed B and C ?

$$H(A|B,C) = H(A,B,C) - H(B,C) \tag{D.10}$$

Conditional mutual information: Effect of C on interaction between A and B.

$$I(A;B|C) = \sum_{A,B,C} P(a,b,c) \log_2 \left(\frac{P(a,b|c)}{P(a|c)P(b|c)} \right) \tag{D.11}$$

To derive this in terms of entropies:

$$\begin{aligned} I(A;B|C) &= H(A|C) + H(B|C) - H(A,B|C) \\ &= H(A|C) - H(A|B,C) \\ &= H(A,C) + H(B,C) - H(C) - H(A,B,C) \end{aligned} \tag{D.12}$$

Conditional MI: Always positive, and zero iff A and B are independent given knowledge of C. (i.e. C completely explains the association between A and B.) This implies that C can be predicted from A and B.

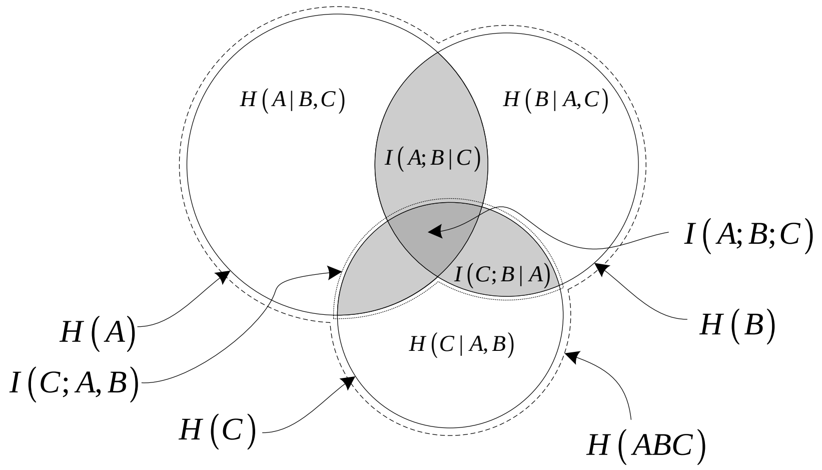


Figure D.2 : Three attribute interactions and entropies.

Note that care needs to be taken when using this figure as some of the areas represent negative values.

D.5 Co-Information

Nomenclature:

> Information shared by all K RV's :
>
> Interaction information [McGill1954], - multivariate generalisation of Shannon's mutual information.

Multiple mutual information or Interaction information or co-information [Bell2003] is a measure of the intersection of all three attributes.

$$I(A;B;C) \triangleq I(A;B|C) - I(A;B) \qquad \text{(D.13)}$$

Three forms

$$\begin{aligned} I(A;B;C) &= I(A;B|C) - I(A;B) \\ &= I(A;C|B) - I(A;C) \\ &= I(B;C|A) - I(B;C) \end{aligned} \qquad \text{(D.14)}$$

Also, (from [Jakulin2004]):

$$\begin{aligned} &I(A;B;C) = I(A,B;C) - I(A,C) - I(B,C) \\ &= H(A,B) + H(B,C) + H(A,C) - H(A) - H(B) - H(C) - H(A,B,C) \end{aligned} \qquad \text{(D.15)}$$

Interaction information is symmetric (same as mutual information)

$$\begin{aligned} I\left(A;B;C\right) &= I\left(A;C;B\right) \\ &= I\left(B;A;C\right) = I\left(B;C;A\right) \\ &= I\left(C;B;A\right) = I\left(C;A;B\right) \end{aligned} \tag{D.16}$$

Interaction information can be negative.

Interaction magnitude is the absolute value of the interaction information.

Total Correlation was first described in [Watanabe1960]. It describes the total amount of redundancy or dependency among the attributes:

$$\begin{aligned} C\left(A,B,C\right) &\triangleq H\left(A\right)+H\left(B\right)+H\left(C\right)-H\left(A,B,C\right) \\ &= I\left(A;B\right)+I\left(B;C\right)+I\left(A;C\right)+I\left(A;B;C\right) \end{aligned} \tag{D.17}$$

It is always positive, and zero if and only if all attributes are independent.

It will be greater than zero even if only a pair of attributes are dependent.

$$P\left(a,b,c\right)=P\left(a\right)P\left(b\right)P\left(c\right) \tag{D.18}$$

It will not be zero even if only a pair of attributes are dependent, e.g.

$$P\left(a,b,c\right)=P\left(a,b\right)P\left(c\right) \tag{D.19}$$

Appendix E Achievable Pulse Compression Sidelobe Levels

To obtain a realistic value for the lowest level for the sidelobes, use was made of a L-p Norm optimization routine based on [Cilliers2007], but where the transmit and receive filter remained matched. The pulse was constrained in that the real and imaginary samples were bounded not to exceed a value of 1.5. An example for a time-bandwidth product of 50 is shown in the following figure, for P=10 and P=408.

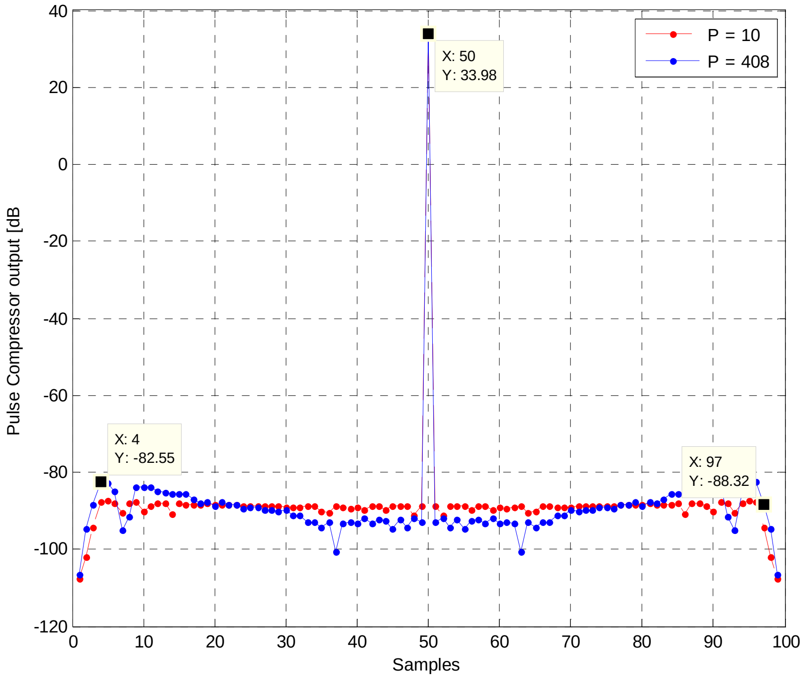


Figure E.3: Example of the pulse compression waveform for a time-bandwidth product of 50.

This result shows that relatively flat sidelobes are achievable for a peak to sidelobe level of approximately 121 dB. Given this result, the sidelobe structure could be simulated from a level of 150 dB, and the sidelobe structure can be modelled as a flat line.