OPTIMIZING THE DECISION RULE OF A GPS INTEGRITY MONITORING
SYSTEM FOR IMPROVED AVAILABILITY

A DISSERTATION
SUBMITTED TO THE DEPARTMENT OF MECHANICAL ENGINEERING
AND THE COMMITTEE ON GRADUATE STUDEIS
OF STANFORD UNIVERSITY
IN PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY

Mike Koenig
March 2010

Abstract

The Global Positioning System, which was created by the United States Department of Defense specifically for guiding munitions, has disseminated into many applications of the present. Using a differential, ground based augmentation scheme has enabled GPS guidance to be used for the ultra-safety requisite use of landing aircraft. The civil incantation of this project is called LAAS, for Local Area Augmentation System. The military desired a comparable system for their use, and has instituted the JPALS program, for Joint Precision and Approach Landing System. The object is to deliver an integrity monitor to satisfy the stringent requirements of difficult military operation. Although the distinction of emphasis might seem to be on the inherent danger of physical conflict, since the landing platform is expected to be distant from any combat, it is actually the electro-magnetic conflict which takes precedence in this thesis.

The JPALS land-based, or sea-based, Integrity Monitor must be designed to endure radio frequency interference, RFI. The potential of RFI to undermine the operation of a landing system is quite real. The intrinsic obligation of an Integrity Monitoring System is to bound the errors associated with the guidance being provided. To this end, the system must determine if any of measurements are faulty, and if those fault(s) indicate the presence of a potentially hazardous position error. Hardware and software can be used to mitigate the impact of radio frequency interference to a GPS based landing system. Beam-steering antennas can be used to emphasize the satellites' signals, and to simultaneously deemphasize a jamming source. The focus of this thesis is in the software algorithms. How can we enhance the performance of our system in its robustness to interference, by optimizing the decision rule it uses to exclude satellites, and receivers, thereby maximizing the amount of useable information, and minimizing the error limits that are broadcast to the approaching aircraft.

Acknowledgments

A gracious thank you to my advisor Professor Per Enge. Professor Enge has always fostered my creative thinking and endured my unending questions. He has always been supportive and overly effusive in his praise.

Thank you to Dr. Jason Rife who directed my research and mastered the balance of giving encouragement and setting deadlines. He provided me with much needed guidance in both the professional and personal pursuit of this degree. Also, thank you to Demoz Gebre-Egziabher who had such enthusiasm for my work.

Thank you to the entire Stanford GPS lab, who have been both colleagues and friends. Their collective support helped me to overcome both trying and trivial tribulations.

Thank you to my Defense Committee and Reading Committee: Prof. Cox, Prof. Enge, Prof. Niemeyer, Prof. Rock, Dr. Pullen, and Prof. Kenny.

Thank you to my family, who has never stopped supporting me and who remind me that there is always a place called home.

Sandy Koenig (1991-2007)

The best friend a man has in this world may turn against him and become his enemy. His son or daughter that he has reared with loving care may prove ungrateful. Those who are nearest and dearest to us, those whom we trust with our happiness and our good name, may become traitors to their faith. The money that a man has, he may lose. It flies away from him, perhaps when he needs it the most. A man's reputation may be sacrificed in a moment of ill-considered action. The people who are prone to fall on their knees to do us honor when success is with us may be the first to throw the stone of malice when failure settles its cloud upon our heads. The one absolutely unselfish friend that a man can have in this selfish world, the one that never deserts him and the one that never proves ungrateful or treacherous is his dog.

A man's dog stands by him in prosperity and in poverty, in health and in sickness. He will sleep on the cold ground, where the wintry winds blow and the snow drives fiercely, if only he may be near his master's side. He will kiss the hand that has no food to offer, he will lick the wounds and sores that come in encounters with the roughness of the world. He guards the sleep of his pauper master as if he were a prince. When all other friends desert, he remains. When riches take wings and reputation falls to pieces, he is as constant in his love as the sun in its journey through the heavens.

If fortune drives the master forth an outcast in the world, friendless and homeless, the faithful dog asks no higher privilege than that of accompanying him to guard against danger, to fight against his enemies, and when the last scene of all comes, and death takes the master in its embrace and his body is laid away in the cold ground, no matter if all other friends pursue their way, there by his graveside will the noble dog be found, his head between his paws, his eyes sad but open in alert watchfulness, faithful and true even to death.

- George G. Vest, 1870

Table of Contents

List of Tables

List of Figures

Chapter 1 - Introduction

The rapid transformation of the Global Positioning System from theory to implementation has transformed the modern world of navigation. Various techniques have been developed to improve the performance of GPS, including carrier-smoothing, wide-laning integer ambiguity resolution, Receiver Autonomous Integrity Monitoring (RAIM), dual frequency ionosphere mitigation, etc. The focus of this thesis is to use GPS for aircraft guidance, or more generally for users that cannot tolerate potentially hazardously misleading information. In other words, the navigation system must not be allowed to endanger the lives or equipment of the user. This is the "integrity" criterion of knowing with near certainty that the navigation assistance provided will not jeopardize craft or crew. Specifically, this means that the position error the user aircraft experiences will be less than a stated protection limit. Another objective is to maximize the fraction of time the system can provide navigation assistance with such a guarantee.

This latter requirement introduces the concept of continuity, which is the probability that guidance is continuous after the aircraft approach begins. There must be a balance between the stringency required of integrity and the compliance which enables continuity and practical usage. A system which is always turned off has perfect integrity but no continuity. A system which is always turned on has perfect continuity but suspect integrity. Neither of these extremes is desirable. Therefore, a key system-design trade-off exists – how tight should the thresholds be and how many faults of a particular entity are tolerated before the entire entity is faulted?

When a satellite is being tracked on three independent GPS receivers, the code and carrier-wave measurements are monitored for anomalous trends on each of the receivers. Each receiver can declare a fault on that particular satellite. Collectively that information is interpreted to determine if apparent faults are limited to a single receiver or if it indicates that the satellite itself may be faulted. Two critical questions are 1) what

should the statistical threshold be for the each test, and 2) how many of the three receivers can fault that satellite before the satellite itself is deemed to be faulted?

Using incredibly tight testing thresholds throughout the system would mean that every error would likely be caught but that many times the system would declare an error when none was hazardous. This means that the system would have very high integrity but suffer low continuity. The converse could also be true; the testing thresholds could be loose enough such that almost no measurements were flagged. This would mean the system would almost always provide some kind of navigation assistance, but the integrity of that assistance could be very low. It would be difficult to have any confidence that the navigation assistance would not jeopardize the user aircraft. In addition to meeting an accuracy requirement, when the system is able to meet both its continuity and integrity requirements, then the system has availability. This spectrum of statistical testing thresholds and the effect on availability is shown in Figure 1.1.
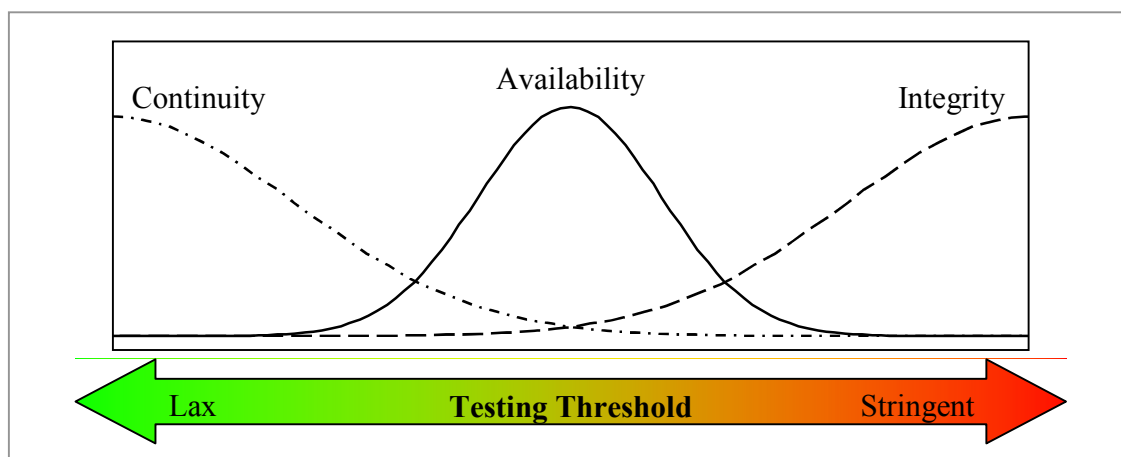
Figure 1.1: How Testing Thresholds Affect Availability

1.1 Motivation for Research

In addition to the creation of the civilian landing system known as the Local Area Augmentation system (LAAS), an analogous landing system for the military has also been developed. This system is called the Joint Precision and Approach Landing System (JPALS). This military system will share similarly strict requirements for safety but has

important operational differences. While LAAS will operate at domestic airports, JPALS must operate anywhere in the world and under very stressing conditions, including the presence of RF interference or jamming [18]. Given the rigidity of the GPS hardware, the ability to continue operation in an interference laden environment must employ intelligent algorithms to avoid excluding useful information via overly conservative testing thresholds and decision rules. Our need is to optimize the efficiency of these algorithms in order to maximize the availability of the landing system.

1.2   Terminology

This thesis contains a variety of acronyms and notations. Although these are always defined prior to their usage, it may not be immediately evident for the reader where that first usage occurred. Therefore, this section acts as a quick reference for the most common terms used throughout this thesis.

Program Level Acronyms: *(Used Throughout)*

GPS	≡	Global Positioning System
DGPS	≡	Differential Global Positioning System
LAAS	≡	Local Area Augmentation System
IMT	≡	Stanford's Integrity Monitoring Testbed
JPALS	≡	Joint Precision and Approach Landing System
JLIM	≡	JPALS Land-based Integrity Monitor
JSIM	≡	JPALS Sea-based Integrity Monitor
JTeP	≡	JPALS Test Platform

System Performance Terms: *(Used Throughout)*

Accuracy	≡	The measure of the navigation output deviation from truth under fault-free conditions, often specified in terms of 95% performance

Integrity   ≡   The ability of a system to provide timely warnings to users when not to use the system for navigation. Integrity risk is the probability of an undetected navigation system error or failure that results in hazardously misleading information onboard the aircraft [30].

Continuity   ≡   The likelihood that the navigation signal-in-space supports accuracy and integrity requirements for the duration of intended operation. Continuity risk is the probability of a detected but unscheduled navigation function interruption after an approach has been initiated.

Availability   ≡   The fraction of time the navigation function is usable (as determined by its compliance with the accuracy, integrity, and continuity requirements) before the approach is initiated [2].

Probability terms: *(Chapters 4, 5)* ***Note the use of lower or uppercase text**

P_{md}	≡	Probability of a Missed Detection, for a given channel/monitor
P_{MD}	≡	Probability of a Missed Detection, for a Satellite or Receiver
P_{ffd}	≡	Probability of a Fault Free Detection, for a given channel/monitor
P_{FFD}	≡	Probability of a Fault Free Detection, for a Satellite or Receiver
$P_{FFD/CH}$	≡	P_{FFD} for a Satellite or Receiver, given a channel fault
MDE	≡	Minimum Detectable Error, given P_{FFD} and P_{MD} requirements

Thesis general terms: *(Used Throughout)*

n_{SV}	≡	Number of Satellites
n_{RX}	≡	Number of Receivers
n_{FQ}	≡	Number of Frequencies

Thesis specific terms: *(Chapters 4, 5)*

$(10 \times 3 \times 2)$	≡	$(n_{SV} \times n_{RX} \times n_{FQ})$; a 10 satellite, 3 receiver, 2 frequency system
VMT	≡	Vector-Matrix-Tensor *(an analytical method)*

M_A	$\equiv$	In EXM, the Averaging Method
M_B	$\equiv$	In EXM, the Boolean Method
M_O	$\equiv$	In EXM, the Oblique Method
M_{OR}	$\equiv$	In EXM, the Oblique Radial Method

1.3   GPS Background

The Global Positioning System typically consists of 24 to 32 operational satellites in six orbits inclined at approximately 55 degrees and an orbital radius of about 26,600 km (Figure 1.2). The system uses time-of-arrival to estimate distance and can achieve roughly 10 m accuracy using the code signal, and 1 cm by tracking the carrier-wave itself [1]. Various differential or relative navigation systems are possible which can further increase the accuracy of the system.
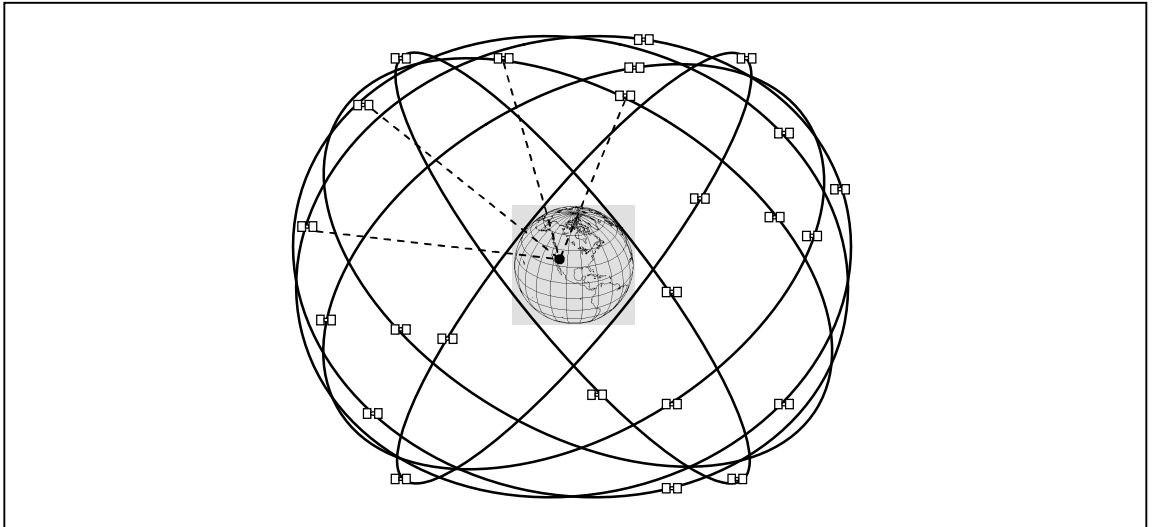
Figure 1.2: The Global Positioning System

The classical GPS range pseudo-range measurement (so called because the actual measurement includes an unknown receiver clock bias) is given by Equation (1.1). Here, ρ is the measured pseudo-range, c is the speed of light, $t_u(t)$ is the arrival time measured by the receiver clock, and $t^s(t)$ is the emission time stamped on the transmitted signal.

$$\rho(t) = c[t_u(t) - t^s(t)] \qquad (1.1)$$

Figure 1.3 depicts the several layers included within the GPS signal. The primary navigation signal is the code modulation. There are two forms, a civil signal with a bit rate of 1.023 million bits per second, and a faster (and encrypted) military signal at 10.23 million bits per second. Equation (1.2) gives the wavelength of the code modulating the civil signal.

$$\lambda_{Code} = \frac{c}{BR_{Code}} \approx \frac{3 \cdot 10^8 \,(m/s)}{1.023 \cdot 10^6 \,(b/s)} \approx 300 \,(m/b) \qquad (1.2)$$

A 50 bits-per-second navigation data signal is also added which provides auxiliary information, including an ephemeris message. Equation (1.3) shows that the length of each bit is then 6,000 km, which is much longer than the code wavelength.

$$\lambda_{ND} = \frac{c}{BR_{ND}} \approx \frac{3 \cdot 10^8 \,(m/s)}{50 \,(b/s)} \approx 6,000 \,(km/b) \qquad (1.3)$$

The ephemeris message gives the satellite location and timing parameters to the user to enable a position fix. This combined GPS signal is then multiplied by the carrier signal of L1 at 1.575 GHz and transmitted. In addition to the signals at L1, there are two more frequencies which carry GPS signals. At the center frequency called L2 (1.2276 GHz), there is a new L2C code which uses the same chipping rate as the code on L1. It has similar ranging abilities as the L1 code, even though the L2C code has some unique properties [16]. There is also an encrypted signal called the P(Y) code at L2 and L1 which is reserved for use by the military. Lastly, there is a forthcoming civil signal called L5, at 1.176 GHz [22].
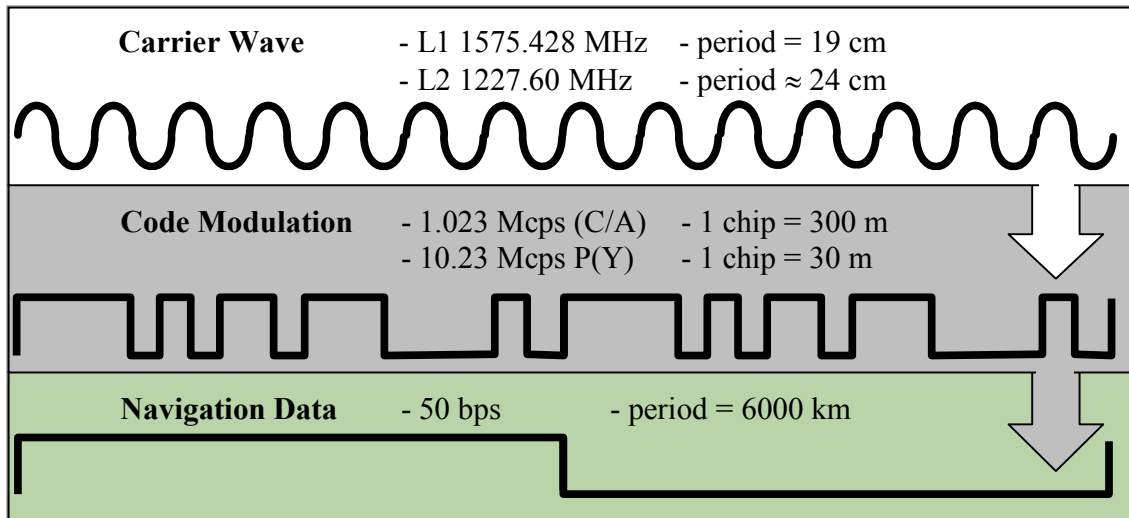
Figure 1.3: GPS Signal Structure

Figure 1.4 shows the frequency spectrum of each of the GPS signals. It shows the broader spectrum of the P(Y) code and also the M-Code, which is an additional signal in development for use by the military.
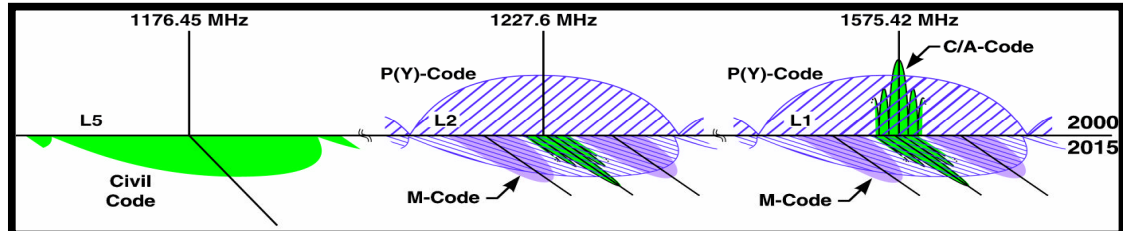
Figure 1.4: GPS Signal Spectrum

The code phase measurement is the inferred pseudo-distance between receiver and satellite by calculating the travel time of the signal. Equation (1.1) shows the fundamental GPS pseudo-range equation. Collecting enough of these equations (from different satellites) enables the receiver to calculate the unknowns of position (x,y,z) and time offset from GPS (τ). With an effective wavelength of 300 (m), the code measurement provides a measure of absolute distance to the satellite, albeit somewhat noisy. A code-phase position solution has a 95% 3-D accuracy of approximately 10 m.

The carrier-phase measurement is the difference between the phase of the carrier signal of the GPS satellite and that of the GPS receiver. This phase difference is integrated over time to provide a continuous estimate of the relative change of position between the satellite and receiver. This measurement involves an ambiguity in the absolute number of carrier wave cycles (and as such the absolute distance) as it is just the difference in carrier-phase from one epoch to another. Using blended carrier-code measurements gives a position solution precision of 10 cm, though the 95% 3-D position accuracy is dominated by biases, such as atmospheric signal delays, and is typically 1 m.

Differential GPS can remove the biases by using a reference receiver. This topic is discussed in Section 1.5 and can produce 95% 3-D position accuracies of 1 m using differential code and 1 cm using differential carrier measurements.

1.4 RAIM

RAIM stands for Receiver Autonomous Integrity Monitoring. RAIM means that the GPS receiver is autonomously attempting to detect if any of the measurements are faulty. A position and time solution requires at least four ranging measurements, and RAIM requires at least five because it needs redundancy to determine which of those measurements may be flawed. In fact, RAIM can do fault detection and exclusion, meaning that the GPS receiver attempts to determine if any of the measurements are erroneous and also tries to exclude it. This process requires at least six measurements. When one measurement is excluded, there will be at least five remaining which is sufficient for the fault detection algorithm to run and ensure that none of the remaining measurements are faulty. Testing each measurement for errors and being able to assert that those measurements which pass are error-free is the integrity portion of RAIM.

1.5 Differential GPS

Figure 1.5 shows the fundamental operation of Differential GPS or, more generally, any type of differential navigation system. Differential GPS means that a base

station, which knows its location through a survey, takes ranging measurements from a set of satellites and finds the corrections to those measurements needed to calculate its known position. Those corrections are broadcast to the user because the errors observed by the base station will be highly correlated with the errors experienced by the user provided that the separation is not too great.
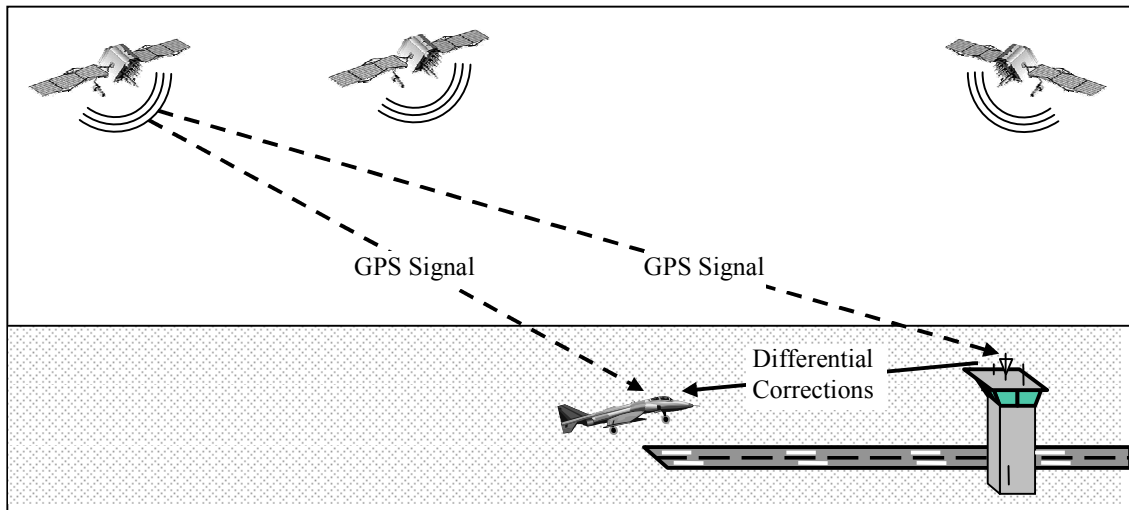
Figure 1.5: Basic Differential Correction Installation

RAIM can detect potentially threatening errors, but it is limited in its ability to mitigate them. That is, it can observe and remove a measurement but it can not necessarily correct it. This would require additional information, specifically, the precise location of the receiver, which is what the Differential Correction Station (DCS) provides. Use of a DCS can reduce positioning errors to sub-meter accuracy. A DCS and RAIM can coexist, given that any commercial receiver which is generating measurements for the DCS will have its own embedded algorithms to validate the measurements. If the concepts of the RAIM integrity monitoring algorithms are applied to the DCS by leveraging the precise location of the installation, the DCS can broadcast the measurement corrections along with a quantification of the expected error on those corrections and the combined system becomes an Integrity Monitoring Station.

1.6   Integrity Monitoring Station

An Integrity Monitoring Station is a system or installation which not only provides differential corrections to improve accuracy, but also attaches an estimate of its confidence in the fidelity of those corrections. However, the ground station cannot be responsible for integrity hazards affecting only the aircraft, such as a airborne GPS receiver failures or airborne multipath. It can only assert that the corrections themselves will not induce hazardously misleading information (HMI). Typically, integrity requirements for precision approaches are on the order of one missed detection per one hundred million operations.

1.7   VPL and Availability

The VPL is a descriptive statistic which stands for "Vertical Protection Level". The VPL is a confidence limit that describes the largest error that may occur given an allowed integrity risk. To ensure safe navigation, the VPL must remain within an envelope called the Vertical Alert Limit (VAL). If the VPL exceeds the VAL, then the user treats the corrections as potentially hazardous and therefore unavailable [5].

Figure 1.6 shows what is commonly referred to as a Triangle Chart [14]. It plots the true vertical error on the horizontal axis and the Vertical Protection Level on the vertical axis. To ensure the safety of the aircraft using the corrections generated by the Integrity Monitoring Station, the Vertical Protection Level must always be greater than the true vertical error. These types of plots are normally constructed by using an external truth reference to estimate the actual altitude of the aircraft. It may be radar, laser ranging, or pre-surveyed position knowledge.

As shown Figure 1.6, if the Vertical Protection Level exceeds the Vertical Alarm Limit, the system is declared unavailable. The system also has an accuracy requirement defined by an accuracy threshold. The accuracy requirement is met when the vertical errors of at least 95% of the data points are below the accuracy threshold.
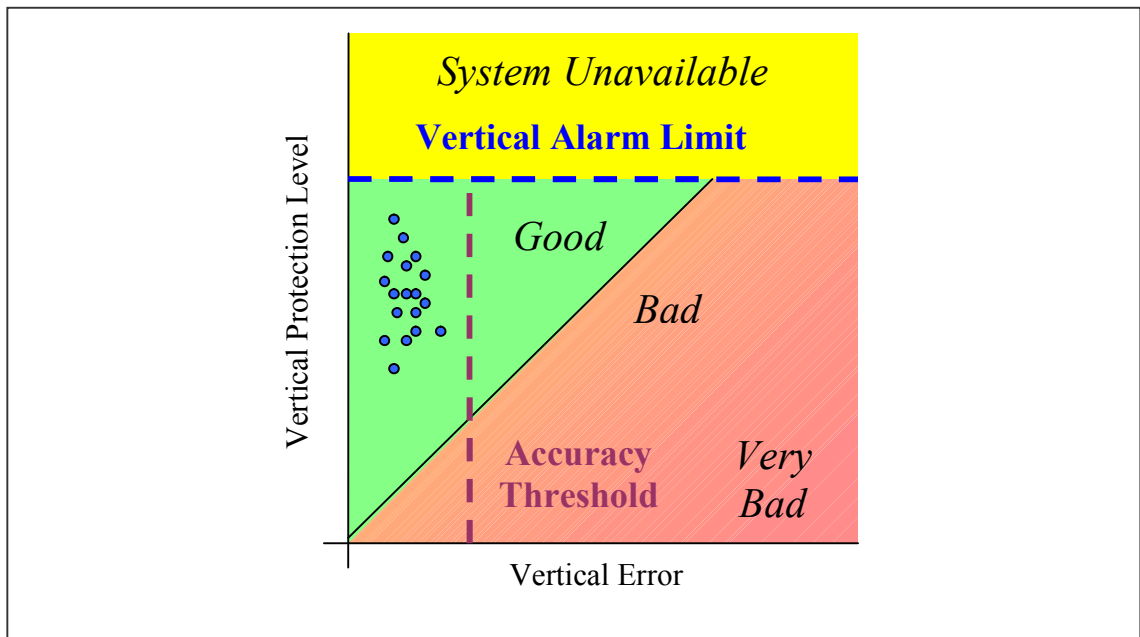
Figure 1.6: Triangle Chart – VPL, VAL, and Availability

1.8   Previous Work

Much work has been done on LAAS to provide high integrity landings at terrestrial airports. JPALS is a program extending the work of LAAS to military applications. JPALS has two variants, land-based and sea-based. This thesis is applicable to all of these, as it concerns decision algorithms for efficient combining of error detection, fault free operation, and environmental robustness.

This section highlights several previous papers that are relevant to this thesis.

1)   M. Luo, S. Pullen, *et. al*. "Development and Testing of the Stanford LAAS Ground Faciilty Prototype," *Proceedings of the ION 2000 National Technical Meeting*. Anaheim, CA., Jan 26-28, 2000, pp. 210-219.

This paper explains the Stanford IMT system architecture and specifically addresses the operation of Measurement Quality Monitoring (MQM), which includes a

set of algorithms using filters to determine if there are any jumps or unusually large accelerations of the received GPS carrier phase. This paper also gives results of two of the MQM test statistics using data with Selective Availability (a deliberate dithering of the satellite clock to degrade performance for non-military users that was turned off by executive order in May of 2000). MQM is discussed in Chapter Two and thoroughly examined in Chapter Three of this thesis and provides the measurements to test the decision logic developed in Chapter Four of this thesis.

2)   S. Pullen, M. Luo, *et. al.*, "GBAS Validation Methodology and Test Results from the Stanford LAAS Integrity Monitor Testbed," *Proceedings of ION GPS 2000*, Salt Lake City, UT, Sept. 19-22, 2000, pp. 1191-1201.

This paper describes the logic of the Executive Monitor (EXM) in the Stanford IMT and how it handles the various scenarios of having multiple faults on either the receivers and/or the satellites. It also details the method of using a small number of test cases with known faults to validate previous detection probability results. EXM is an algorithm which compiles the output of all prior monitors to determine which channels are faulted and if those channel faults indicate the likely presence of a larger fault, such as an unhealthy satellite or receiver. Chapter Five examines EXM in detail to determine if unhealthy satellites and receivers can be detected while minimizing false alarms, which would exclude healthy channels and useable measurements.

3)   G. Xie, S. Pullen, *et. al.*, "Integrity Design and Updated Test Results for the Stanford LAAS Integrity Monitor Testbed," *Proceedings of the ION 57th Annual Meeting*. Albuquerque, NM, Hune 11-13, 2001, pp. 681-693.

This paper follows up on the work of the paper (1) and gives more detail on the operation of the EXM. It also shows MQM results for newer datasets which do not have Selective Availability. The MQM performance detailed here serves as a benchmark for the MQM improvements developed in Chapter Three. The fundamental contribution of

Chapter Three is to increase the sensitivity of the filters described in paper (3) without increasing the detection time.

4)   G. Xie, "Optimal on-Airport Monitoring of the Integrity of GPS-Based Landing System," Doctoral Dissertation to the Electrical Engineering Department of Stanford University. March, 2004.

Dr. Xie's Ph.D. thesis provides a comprehensive reference for the Stanford IMT system and also provides examples of the performance of the MQM test statistics.

5)   Rife, Jason. "Vertical Protection Levels for a Local Airport Monitor for WAAS," *Proceedings of the 61st Annual Meeting*, June 27 - 29, 2005, pp. 745 – 758

The focus of this thesis is to increase the availability of an integrity monitoring system to support landing aircraft, and this paper provides an understanding of the VPL equation and how it affects availability. The error protection levels at the aircraft encompass three fundamental issues: 1) the uncertainty at the ground station, 2) the uncertainty in the propagation medium, and 3) the uncertainty at the aircraft. The contributions of this thesis reduce the uncertainty at the ground station by increasing the sensitivity of several detection algorithms and optimizing the decision logic which interprets the output of those algorithms.

6)   Misra, Pratap, and P. Enge. Global Positioning System: Signals, Measurements, and Performance. Lincoln, MA.: Ganga-Jamuna Press, 2001.

This book serves as a critical reference for understanding the operation of GPS. It details how a GPS receiver can operate in a stand-alone capacity in addition to advanced techniques using multiple GPS receivers, satellites, and frequencies.

1.9   Outline / Contributions

Ultimately, this thesis serves as a design tool specific to the algorithms of detecting satellite faults to protect integrity based on the measurements of multiple channels. The chapter summary and contributions of this thesis are listed here.

Chapter One (this chapter) includes a general introduction to GPS and describes the objective of this work: design and testing algorithms to improve GPS integrity monitoring for aircraft landing systems. More specifically, the objective is to increase the amount of time that system can be used to provide useful and timely navigation assistance for aircraft to land safely. In other words, this research seeks to simultaneously improve integrity, continuity and availability.

Chapter Two explains the operation of a system developed at Stanford called the Integrity Monitoring Testbed (IMT). The IMT is a prototype of the Local Area Augmentation System (LAAS), which is a Differential GPS integrity monitoring system intended to enable civilian aircraft to land at airports using GPS measurements. This thesis was developed under a program called JPALS which is comparable to a military version of LAAS, and the work shown here is part of a prototype called the JPALS Testbed Platform (JTeP) which is comparable to the IMT.

Chapter Three examines one of the monitors in the IMT called MQM and increases its performance for use in the JTeP. This monitor looks for aberrations in a short sequence of the carrier-phase values of a satellite tracked on a particular receiver. The algorithm has been enhanced to increase the ability to detect those aberrations and is presented as Contribution 1.

Contribution 1:

MQM calculates descriptive statistics for the carrier-phase data to determine if any anomalies are present, such as discontinuities or inexplicable rates of change. The methods for calculating those descriptive statistics have been enhanced such that the methods are able to detect smaller anomalies without increasing the detection time.

Chapter Four addresses the operation of the EXM to determine if is possible to alter its decision logic in order to increase its ability to detect faults and reduce the possibility of false alarms. This is the second contribution of this thesis.

Contribution 2:

The EXM decision logic has been modified in order to better detect satellite errors observable across multiple receivers while simultaneously reducing the probability of false detections. For a satellite fault which is observable on each receiver, the previous method does not fully leverage the available information. It is possible to average the estimated statistics across each receiver to characterize a satellite, but such a method means that a substantial fault on one channel may result in the exclusion of the satellite if that faulty channel can not be isolated. Conversely, the old method affords the system great protection against any one channel fault causing the system to erroneously exclude the satellite, but at the loss of detection sensitivity. The method developed in this thesis blends the old method with a method using averaging to produce a Hybrid Method. This Hybrid Method comprehensively outperforms the existing method by being both robust to single-channel faults and more sensitive to the satellite faults deemed hazardous to the operation of the navigation system.

Chapter Five develops a computational method to analyze the effect of using different kinds of decision logic on a multiple receiver, multiple satellite, multiple frequency system. This is the third contribution of this thesis.

Contribution 3:

The decision logic used to handle channel faults and determine whether to remove a satellite or receiver includes rules to exclude a receiver or satellite if there are (m) observed failures over (n) channels. The values of (m) and (n) are variables which are used as input parameters. Typically (m) is 1-3, while (n) can be 2-4 if it represents the

number of receivers or even 4-12 if it represents the number of satellites. The decision logic can quickly get complicated if rules to flag satellites overlap rules to flag receivers, meaning a satellite and receiver can simultaneously be excluded. Instead of calculating conditional probabilities, this computational algorithm efficiently simulates all possible combinations of passing or failing the channels to determine what the effective rate of exclusion will be for that particular set of decision rules. Different rules can be defined to specify how to treat multiple fault configurations on receivers and satellites. A satellite may be tracked on three receivers, and a receiver may track up to 12 satellites. How many faults on a receiver, or how many faults on a satellite, should cause that receiver or satellite to be removed from the solution? The effect of altering these design parameters can be analyzed with the efficient simulation developed here. Additionally, the current implementation of Stanford IMT uses three receivers and tracks up to 12 satellites per receiver on only one frequency, L1, but the certified version of LAAS utilizes four receivers because of the extra level of redundancy it affords. JPALS, as well as potential future versions of LAAS, will also track two or even three frequencies. The method developed here also demonstrates the applicability of the simulation to having additional receivers and frequencies.

Finally, Chapter Six provides a summary of the accomplishments of this thesis and the avenues of future work which may follow.

Chapter 2 - An Integrity Monitoring System

A reference station dramatically enhances the performance of a local or regional navigation system. Ideally, the errors that any user may see will be highly correlated, in space and time, to that of the reference station. By knowing its own position, the reference station can calculate the errors for each ranging signal and then broadcast these errors as corrections to a nearby user. This makes the user utterly dependent upon the reference station for its own position calculation. Such dependency makes the ground station responsible for providing a consistently accurate set of corrections to the user. Whatever frailty the ground station possesses will transfer to the user, and as such, if the ground station had only one antenna and one receiver, then it would be highly susceptible to failure. The use of multiple receivers and antennas isn't focused so much on improving the accuracy of the corrections, but on improving the confidence/integrity of those corrections. Having only one source for corrections requires dependency, having multiple sources allows for a comparison. This introduces the concept of redundancy.

Redundancy

Redundancy provides two critical elements for integrity monitoring, observability and robustness, which directly relate to the ability to detect and isolate faults. For a standalone mobile GPS navigation system, at least four ranging measurements (i.e. four satellites) are needed to determine the position and time of the GPS receiver. With five ranging measurements, the receiver can calculate the root sum-of-squares (RSS) of the residual ranging errors. This value estimates the lack of fit of the solution, and if that RSS value is large it is likely that one (or more) of the ranging measurements is flawed. Having five ranging measurements only allows the receiver to determine if one of the measurements is flawed but it can't easily determine which one of the measurements is flawed. This is because if any one of the measurements is excluded the remaining four measurements will perfectly determine a position and time solution even if one of those

measurements is flawed. With six ranging measurements, the receiver can calculate the RSS residual error to determine if there is a possible flawed measurement. If it is large, the receiver can remove one measurement at a time and recalculate the position and time solution with five measurements. Using five measurements means that the receiver can calculate a meaningful RSS residual error. The solutions with the bad measurement included will have large RSS residual errors, while the one solution with the bad measurement excluded will have a much lower RSS residual error. Having even more ranging measurements (satellites) means that additional detection and isolation can be done, albeit with additional processing.

In the mobile stand alone scenario, the receiver is using measurements to estimate its position and time. Not only does this reduce the available measurements to perform an integrity check, but there is an error associated with the position and time estimates which further obscures the integrity process. When the GPS receiver is stationary and at a known location (through survey) only the receiver clock bias needs to be estimated. Technically, the clock bias estimate is equivalent to a common range error on all of the range measurements. That error dominates other error sources and must be removed at each epoch in order to detect other potential errors, such as severe Ionosphere gradients. This process is still only using one receiver which could result in the following two situations.

1) The receiver has a problem that prevents it from tracking any of the GPS satellites. This means that no ranging measurements are made and the system cannot calculate range corrections to broadcast to the aircraft. Consequently, the aircraft has no navigation assistance and may not be able to perform a landing.

2) The receiver has a problem which affects only one of the ranging measurements. Assuming sufficient measurements exist, and that the integrity algorithms detect the error, this measurement would be excluded even though there was nothing wrong with the signal but instead with the receiver.

To make the integrity monitoring station robust against these two particular scenarios, the GPS integrity monitoring system must utilize multiple GPS receivers. Clearly this would prevent the first scenario. In the second scenario, if the satellite with the receiver induced error is tracked on the other receivers (which do not have induced errors) then the error is observable and the satellite (measurement) on the flawed receiver can be excluded. That satellite can still be used on the other receivers, meaning that the system is not unnecessarily discarding useable ranging information. Using only two receivers means that for a particular satellite, each receiver could calculate a distinct range correction, and it would not be immediately clear which correction is more accurate. Consequently, three or more receivers are typically used in order to quickly identify which receiver may have produced a flawed measurement. Thus also increases confidence of the broadcast correction, meaning less uncertainty is transmitted from the ground station to the aircraft. Figure 2.1 shows a basic ground station using three GPS antennas, each with their own receiver, collecting data from several satellites. The data is processed at the reference station and correction messages are broadcast to any users in the vicinity.
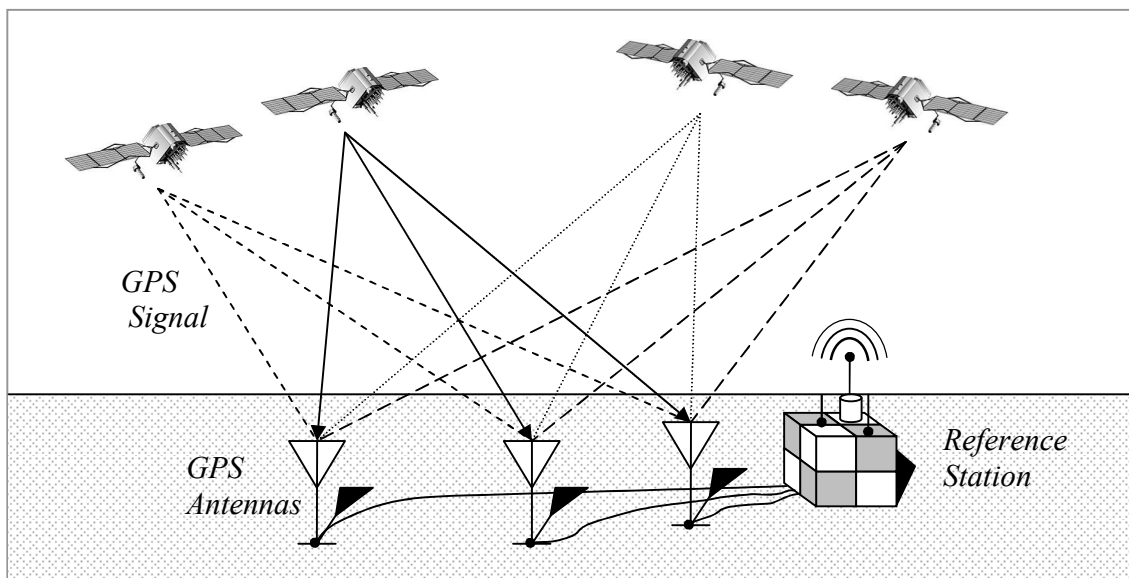
Figure 2.1: Multiple Satellites and Receivers for Error Detection and Isolation

19

2.1 LAAS

LAAS, or the Local Area Augmentation System, is the FAA version of the Ground Based Augmentation System, or GBAS, that has been defined by the International Civil Aviation Organization (ICAO) [39]. LAAS is based on a single GPS reference station facility located on the property of the airport being serviced. This facility has three or more (redundant) reference receivers that independently measure GPS satellite pseudorange and carrier phase and generate differential carrier-smoothed-code corrections that are eventually broadcast to users via a 31.5-kbps VHF data broadcast (in the 108 - 118 MHz band) that also includes safety and approach-geometry information. This information allows users within 45 km of the LAAS ground station to perform GPS-based position fixes with 0.5-meter (95%) accuracy and to perform all civil flight operations up to non-precision approach. Aircraft landing at a LAAS-equipped airport will be able to perform precision approach operations up to at least Category I weather minima. (Stanford LAAS website) A typical LAAS installation would look similar to Figure 2.1 – using multiple GPS receivers at a landing facility and broadcasting GPS corrections and integrity information to the aircraft.

2.2 IMT

The IMT is the Integrity Monitor Testbed, created at Stanford towards the development of LAAS [40]. The IMT uses three GPS receivers, each with their own antennas (shown in Figure 2.2), and evaluates the quality of the GPS measurements including: pseudo-range, carrier phase, ephemeris, tracking loop correlator output, and signal-to-noise ratio.

The three antennas of the IMT are placed atop of the Stanford HEPL (Hansen Experimental Physics Laboratory) building. The antennas are sited in order to minimize the potential for multi-path interference, which is when a signal is received via two or more pathways. This same installation provided the data for the development of the land-based JPALS and its algorithms.
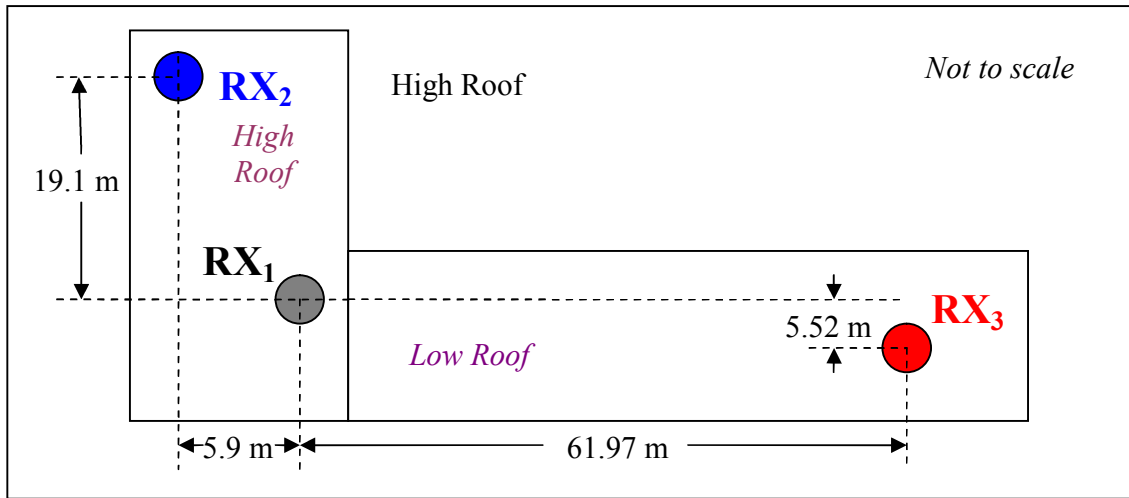
Figure 2.2: The Stanford IMT Antenna Layout (HEPL Rooftop)

Figure 2.3 shows a flowchart of the processes of the IMT. The contributions of this thesis concern the algorithms of the MQM (Measurement Quality Monitoring) and the EXM (Executive Monitoring).
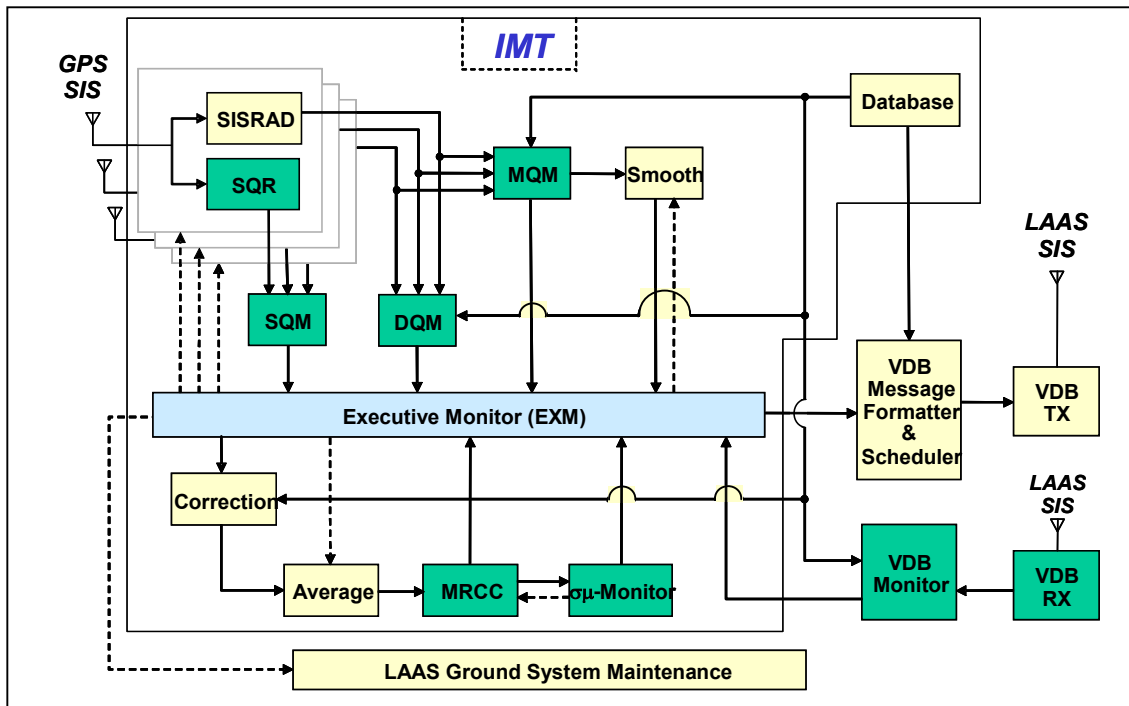
Figure 2.3: Stanford's IMT Flowchart

An overview of the IMT follows. The GPS receiver outputs the tracking loop data to the SQM monitor which determines if the code correlation peak is overly distorted. This monitor is targeted to detect "evil waveforms" similar to the recorded SV19 event (shown later in Figure 2.4) [14]. Concurrently, the DQM monitor evaluates the decoded ephemeris message to determine if it is accurate, and the MQM determines if the carrier phase exhibits any abnormal temporal trends. This can indicate either a system which if off-nominal or a circumstance where the latency between the ground station and the user will enable a hazardously large range error. When the carrier phase for each channel has been passed, it is used to smooth the pseudorange. All of this information is fed into the EXM, which at this stage determines if any of those channels are erroneous and should be excluded from further propagation towards calculating the range corrections. The remaining healthy channels are used to calculate a range correction. Then the average correction (across receivers) is calculated and if the consistency among receivers is sufficient and pass the MRCC, SU-Monitor, and MFRT tests, then that average correction (for each satellite now) is formatted for broadcast to the user along with an estimate in the confidence of each satellites range correction.

Gang Xie's Doctoral Thesis [15] serves as a significant reference for the synopsis of IMT components. The IMT can be divided into three primary sections: (1) nominal processing, (2) integrity monitoring algorithm, and (3) executive monitoring logic.

2.2.1   Nominal Processing

This involves carrier smoothing and calculating differential corrections. Carrier smoothing was mentioned in the first chapter and is a method of blending the GPS carrier phase with the pseudorange to increase the accuracy of the ranging measurements. Equations for this process are given in Section 2.2.2.3.3.

2.2.2   Integrity Monitoring Algorithms

The integrity monitors consist of the blocks labeled Signal Quality Monitoring (SQM), Data Quality Monitoring (DQM), Measurement Quality Monitoring (MQM), Multiple Reference Consistency Check (MRCC), $\sigma\mu$-Monitor, and the Message Field Range Test (MFRT). Each of these monitors is focused on a specific failure mode that is deemed to be potentially hazardous to the operation of an integrity based landing system. The MQM is a critical part of this thesis and is treated further in Chapter Three. The DQM and SQM are given some background here as they can produce flags on channels which affect the decision logic of the EXM, but otherwise are not relevant to the remainder of this thesis. The MRCC, $\sigma\mu$-Monitor, and MFRT algorithms occur after EXM-I and are also not relevant to this thesis, but they are explained here for background. This thesis is focused on the MQM and the EXM-I; the SQM, DQM, MRCC, $\sigma\mu$-Monitor, and MFRT are simply explained to understand the overall treatment of faults.

2.2.2.1 Data Quality Monitoring (DQM)

The DQM is charged with making sure that the satellite navigation data is accurate. It does this by examining the GPS ephemeris and clock data for each satellite which is tracked by the ground station.

When a satellite first becomes visible to the ground station, there is no previous ephemeris for that satellite and the DQM begins to validate the newly received ephemeris. The DQM begins this process when at least two receivers have decoded identical new navigation data.

There are several stages to the validation. The DQM compares the new ephemeris to almanac data for the next six hours at five minute intervals. The ephemeris data are the parameters used to calculate the precise location of the satellite. This is broadcast only by the satellite it describes. However, every satellite also broadcasts almanac data, which is a coarse description of the orbits of each satellite in the GPS constellation. Each sample of the orbit location from the new ephemeris must agree with

the orbit location from the almanac data to within 7,000 meters to validate this stage of the DQM [15].

The GPS navigation messages are typically updated every two hours. The DQM compares satellite positions based on the old and new ephemeris to insure that the two sets of ephemeredes are consistent to within 250 meters over the past two hours and the upcoming two hours. This method uses 70 samples, spaced at 206 second intervals, totaling just over a four hour span. The DQM also compares this new navigation message against the most recent almanac data in the same manner as detailed in the prior paragraph.

The DQM has also incorporated a monitor called the Ye-Te (Yesterday minus Today Ephemeris) test [41]. The purpose of this test is to validate the satellite position of a newly risen satellite, but by comparing it against the most recent validated ephemeris for that satellite instead of comparing it against the almanac derived orbital position.

2.2.2.2 Signal Quality Monitoring (SQM)

The SQM is responsible for detecting anomalies in the GPS ranging signals such as anomalies arising from the GPS satellite itself or from a local interference source, including a reflection of the signal itself. Significant contributions to the SQM have been made by Phelts [17] and Mittelman [18]. Their work involved mitigating the influence of external signals which may corrupt the GPS signal. The SQM is comprised of three tests:

1)   A test of the correlation peak symmetry
2)   A test of the received signal power level
3)   A test of the code-carrier divergence

The correlation peak symmetry test looks at the shape of the GPS signal correlation, which should ideally be a triangle. By sampling this correlation at different points (known as correlator spacings) the SQM can determine if anything has

significantly interfered with the signal. The shape of the ideal correlation triangle compared to a distorted correlation peak is shown in Figure 2.4.
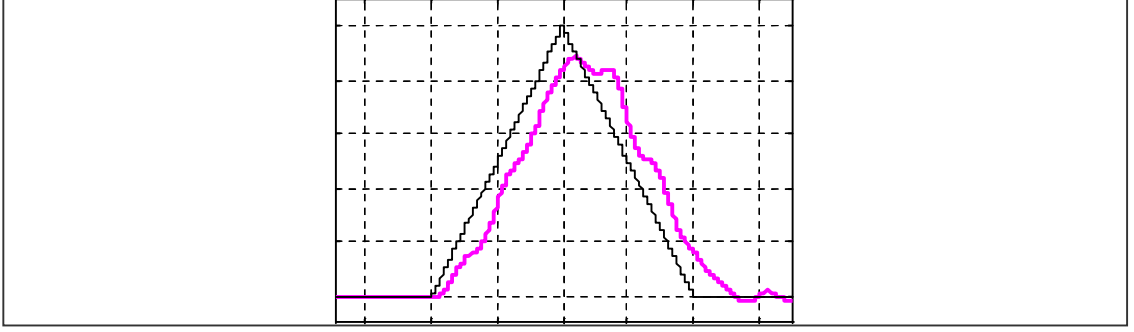
Figure 2.4: Signal Correlation Peak Deformation

The SQM also looks at the received signal power to determine if the signal may be corrupted. The signal power monitor averages the receiver C/N_0 (carrier-to-noise ratio) for each channel at the current (k) and previous ($k-1$) epochs.

$$C/N_{O,Avg}(k) = \frac{1}{2} \cdot \left(C/N_O(k) + C/N_O(k-1) \right) \qquad (2.1)$$

For each installation, the antenna location, antenna gain, cable loses, and receiver signal amplifier are different, thus the averaged C/N_0 is compared against a threshold which is derived for each receiver. Under nominal operating conditions, data for each received is collected and compiled into seven elevation angle bins. The bins are: (0° - 5.5^o), (5° - 15°), (15° - 25°), (25° - 35°), (35° - 45°), (45° - 60°), and (60° - 90°). For each bin, a mean and standard deviation are calculated and a threshold is set to be six-times the standard deviation lesser than the mean. That threshold may be lowered further by an inflation factor to compensate for the fact that the distribution of the signal power level is not well modeled and a six-sigma offset is a heuristic approach.

The test for code-carrier divergence ensures the quality of the carrier-smoothed code. The Ionosphere is the uppermost part of the Earth's atmosphere. It is a dispersive

medium which affects the propagation of electro-magnetic waves. The Ionosphere is a topic of much research, and the effect on the GPS signal is to cause the code phase to experience a delay, while the GPS carrier-wave advances. This results in a bias on the carrier-smoothed code range when the transmitted signal passes through a gradient in the Ionosphere [23][24]. This monitor looks at the geometric moving average of the difference between successive epochs of the code phase minus the carrier phase. The Divergence (D) is expressed in Equations (2.2) and (2.3) as a function of the averaging time (τ_d = 200 seconds), the GPS receiver sampling time (T_S = 0.5 seconds), the code-phase (ρ), and the carrier phase (ϕ).

$$D_k = \frac{\tau_d - T_S}{\tau_d} \cdot D_{k-1} + \frac{T_S}{\tau_d} \cdot dz_k \tag{2.2}$$

$$dz_k = \left(\rho_k - \phi_k \right) - \left(\rho_{k-1} - \phi_{k-1} \right) \tag{2.3}$$

The vale of the estimated Divergence (D) is compared to an empirically derived threshold which is dependent on the satellite elevation angle. Making sure the magnitude of the Divergence is acceptably low further protects the user against a potential threat to the integrity of the broadcast range corrections.

2.2.2.3 Measurement Quality Monitoring (MQM)

The specific tests of the current MQM of the IMT are introduced in some detail here and the contribution to the carrier phase based Acceleration-Step Tests of the new MQM is examined thoroughly in Chapter Three. At present, the MQM has three major components: the Receiver Lock Time Check, the Carrier Acceleration-Step Test, and the Carrier-Smoothed Code (CSC) Innovations Test. The first test supports the operation of the second test. It does not raise a flag, but re-initializes the memory buffer which stores the recent history of carrier phase values that is used by the second test. The Carrier Acceleration-Step Test and the CSC Test can flag channels which exceed the testing thresholds of those tests.

2.2.2.3.1   **The Receiver Lock Time Check**

This test looks at the receiver lock time reported by each receiver to make sure that the carrier phase is being tracked continuously by each receiver. This is important because the Acceleration and Step Tests look at a time series of the carrier phase values, and if there is an unknown jump or discontinuity in time then it will appear to be a discontinuity in the carrier phase. That jump would likely be detected by the MQM Step Test because of its expected large size. This particular type of jump can be regarded as a labeling error, because one epoch of carrier phase is given the wrong time stamp. Even so, they would appear as serious errors in the ranging data and would lead to the unnecessary exclusion of channels. For this reason, the Receiver Lock Time Check does not flag channels but instead resets the memory which is storing the carrier phase values to make sure the spacing between samples in consistent.

2.2.2.3.2   **Acceleration-Step Test**

This section provides an overview of the operation of the Acceleration and Step Tests of the current MQM. The specific equations for the MQM, developed as part of this dissertation, are described in detail in Chapter Three along with the derivation of the new MQM's algorithms.

As presently implemented, the Acceleration and Step Tests look at five seconds worth of carrier phase data, sampled at 2-Hz, for a total of ten epochs of carrier phase data. For each channel, the carrier phase data is modified to compensate for the satellite motion and known satellite clock offset from GPS time by calculating the expected range and satellite clock drift from the satellite navigation data. At this stage the data is called "corrected carrier phase". That data is then compensated for the receiver clock drift. This is done at each epoch on each receiver by averaging the corrected carrier phase across all the satellites tracked on a receiver and subtracting that estimate from the corrected carrier phase. At this stage the data is called "corrected-adjusted carrier phase".

For each channel, a second order polynomial is fit to the ten points of the corrected-adjusted carrier phase using a least-squares method to minimize the lack of fit. Three parameters are directly taken from that model: phase, phase velocity, and phase acceleration. That model is also used to estimate the difference between the current carrier phase reported from the receiver and the corresponding carrier phase predicted by the polynomial model. That difference is the estimate examined by the MQM Step Test. This test is capable of detecting jumps in the satellite clock, but typically it detects carrier phase cycle slips. A cycle slip occurs when the receiver's carrier tracking loop momentarily loses phase lock and unknowingly jumps one cycle of the carrier wave forwards or backwards. This amounts to an error of ± 19 cm for L_1. The Acceleration Test uses the phase acceleration estimate from the model to determine if the carrier phase acceleration is abnormally large. Both the Acceleration and Step Tests compare their estimated values to empirically derived thresholds. Those thresholds are specific to each receiver and are also a function of satellite elevation. They are typically six times the standard deviation of a compilation of data. This is because the false alarm rate for each test is usually set to be on the order of 10^{-8}, which is roughly the exclusion probability at six sigma of a Gaussian distribution. Typically, because the distribution is not perfectly Gaussian the exact threshold is decided empirically by the data. That data is recorded over months to accurately estimate the true characteristics of the data and minimize the impact of any transitory affects which would skew the thresholds and inhibit the test's ability to detect spurious data points.

2.2.2.3.3   CSC Test

The CSC Innovation Test uses both the pseudorange and the carrier phase to determine if the difference in the current pseudorange and the projected smoothed pseudorange is abnormally large.

The smoothed pseudo range is a recursive relation and shown in Equations (2.4-2.6). The smoothing coefficient N_S is variable and is equal to the lesser of the time-in-track, t_T, and the ratio of the time constant, $\tau_S = 100$s, and sample interval, $T_S = 0.5$

seconds [1].   Because the smoothing coefficient is variable, the first value of the smoothed pseudo range is equal to the raw pseudo range.

$$\rho_S(1) = \rho(k) \tag{2.4}$$

$$\rho_S(k) = \frac{1}{N_S} \cdot \rho(k) + \frac{N_S - 1}{N_S} \cdot \left(\rho_S(k-1) + \phi(k) - \phi(k-1)\right) \tag{2.5}$$

$$N_S = \min\left(t_T, \frac{\tau_S}{T_S}\right) \tag{2.6}$$

The Innovation Test statistic is given in Equation (2.7).   It compares the pseudorange of the current epoch with the smoothed pseudorange of the previous epoch advanced by the difference in the carrier phase.

$$I(k) = \rho(k) - \left(\rho_S(k-1) + \phi(k) - \phi(k-1)\right) \tag{2.7}$$

This test also uses an elevation dependent threshold which is approximately six to seven times the standard deviation of the nominal distribution. However, failing this test does not immediately result in a flag for the channel. If the Innovation estimate is above the threshold for at least two out of the last three successive epochs then the channel is flagged. However, if the Innovation estimate only exceeds the threshold for the current epoch then the smoothed pseudo range is recalculated to de-emphasize the current raw pseudorange. Using Equation (2.6), the value of N_S is effectively set to infinity.

2.2.2.4 Multiple Reference Consistency Check (MRCC)

The MRCC uses a test statistic called the B-Value which is a method off expressing the agreement across receivers. B-Values are calculated for every channel for both carrier phase and pseudorange corrections using a common set method [15]. Those B-Values for the pseudorange and carrier phase are compared against their respective thresholds which are both a function of satellite elevation and the number of receivers.

The process for calculating the pseudo range B-values are given in Equation (2.8-2.10). These values are calculated at each epoch k, on receiver M, and satellite N. In Equation (2.8), the smoothed-corrected pseudo range is calculated by removing the range to satellite R, and satellite clock correction, τ. In Equation (2.9), the smoothed-corrected-adjusted pseudo range is calculated by subtracting off the average value of all the smoothed-corrected pseudo ranges for that satellite in the common set, S_C. N_C is the number of satellites in that common set. In Equation (2.10) the B-Value is calculated by taking the difference of the average smoothed-corrected-adjusted pseudo range across all receivers in the common set and the average smoothed-corrected-adjusted pseudo range across the other receivers in the common set. The process for calculating carrier phase B-Values is nearly identical, but the carrier phase difference is used instead of the smoothed pseudo range.

$$\rho_{SC,M,N}(k) = \rho_{S,M,N}(k) - R_{M,N}(k) - \tau_{M,N}(k) \tag{2.8}$$

$$\rho_{SCA,M,N}(k) = \rho_{SC,M,N}(k) - \frac{1}{N_C(k)} \cdot \sum_{j \in S_C(k)} \rho_{SC,M,j}(k) \tag{2.9}$$

$$B_{\rho,M,N}(k) = \frac{1}{M_N(k)} \sum_{i \in S_N(k)} \rho_{SCA,i,N} - (k) - \frac{1}{M_N(k)-1} \cdot \sum_{\substack{i \in S_C(k) \\ i \neq m}} \rho_{SC,i,N}(k) \tag{2.10}$$

This process is a test of variation and is a method to determine if one element in a set differs from the other elements in the set. Equation (2.11) illustrates a simple example of this process applied to a three element set, $\{x_1, x_2, x_3\}$. Only the first B-Value is calculated.

$$\begin{aligned} B_1 &= \frac{1}{3} \cdot (x_1 + x_2 + x_3) - \frac{1}{2} \cdot (x_2 + x_3) \\ &= \frac{1}{6} \cdot (2 \cdot x_1 - x_2 - x_3) \end{aligned} \tag{2.11}$$

2.2.2.5 σμ-Monitor

The σμ-Monitor uses the B-values from the MRCC to help ensure that the distribution of the pseudorange correction error is bounded by a zero-mean Gaussian distribution with a standard deviation equal to the broadcast value. This monitor uses the CUSUM algorithm which is a generalized method to detect a change in process parameters under certain assumptions [19][23].

2.2.2.6 Message Field Range Test (MFRT)

The Message Field Range Test is the final check and makes sure that the average pseudorange correction magnitude is less than 75 meters and the average pseudorange correction rate magnitude is less than 0.8 m/s [20]. It is important to test both the corrections and correction rates because there is latency from when the corrections are computed and when they are received and applied by the aircraft. Testing and bounding the correction rate is a way of providing integrity to the actual correction when it is applied, not just when it is generated.

2.2.3 Executive Monitoring Logic (EXM)

The EXM is divided into two stages, the EXM-I and EXM-II, and interprets the flags generated by all of the integrity monitors in order to isolate and remove any faulty channels.

2.2.3.1 EXM-I

The EXM-I operates on the output of the Quality Monitors in a way which is deliberately conservative towards protecting the user. There are two steps to this section of the EXM, referred to an "exclusion" and "inclusion". The "exclusion" portion deals with how to remove channels which are suspected of being faulted. There are three primary fault scenarios of faulted channels that the EXM handles [15]. The response of the EXM is also included below. Note that a "channel" means one satellite being tracked

on one receiver. Most modern receivers are capable of tracking twelve channels (twelve satellites) on L1.

Fault 1:   There is a flag on a single channel.
Response 1:   The flagged channel is excluded.

Fault 2:   There are multiple flags for one satellite across multiple receivers
Response 2:   The satellite is excluded for all receivers.

Fault 3:   There are multiple flags for one receiver across multiple satellites
Response 3:   The receiver is excluded for all satellite channels.

There is the possibility of having a scenario which is a combination of Fault 2 and Fault 3. In this case, both the satellite and receiver are excluded. Each of the Quality Monitors which produce the flags for the EXM-I have very small fault-free detection probabilities, thus having multiple flags is cause for caution and leads to the overly protective exclusion rules.

The "inclusion" portion of the EXM-I involves which channels should be used to calculate the receiver clock adjustment done in the MQM. Those channels are collectively called a "common set". A common set using three receivers and at least four satellites is preferred. If a common set using three receivers is not possible, then a common set with two receivers and at least four satellites is used. There is the possibility that no common set can be found.

2.2.3.2 EXM-II

The purpose of the EXM-I was to exclude channels with faulty measurements and provide a common set of channels required for subsequent monitors. The EXM-II is responsible for interpreting the output of those subsequent monitors, removing any channels, and finding a new common set to apply to the monitors. Those additional

monitors which can flag channels for the EXM-II to interpret are the MRCC (Multiple Receiver Consistency Check), the σμ-Monitor, and the MFRT (Message Field Range Test). The output of the MRCC feeds into a "pre-screen" of the EXM-II which attempts to resolve which channels are responsible for the source of the problem which manifested as MRCC flags. Similar to the EXM-I, the EXM-II determines a common set of the healthy channels in order to remove the receiver clock bias from the measurements on each receiver. If after four iterations the EXM-II can not exclude the appropriate channels such that no further flags are generated by the MRCC, the σμ-Monitor, or the MFRT, then the systems terminates and all channels are excluded. Otherwise, the pseudo range corrections and pseudo range rate corrections are broadcast to the user. This thesis does not examine the EXM-II. The exact logic of the EXM-II can be found in [4].

In summary, the quality monitors and EXM must work together to identify, isolate, and remove erroneous channels. They do so to minimize the probability that hazardously misleading information (HMI) is contained in the corrections broadcast by LAAS. Lastly, the VHF Data Broadcast (VDB) is a system of hardware and algorithms responsible for broadcasting/transmitting the corrections to airborne users. Its operation is not relevant to the contributions of this thesis.

2.3 JPALS/JLIM

JPALS is the Joint Precision and Approach Landing System being developed for the Military [42]. It incorporates all aspects of a landing system, multi-receiver ground station monitoring, airborne monitoring, inertial sensor integration, and antenna arrays. Two variants are being developed, one for terrestrial landings, the land-based JPALS, quite similar in deployment to LAAS, and a second for landings on Navy aircraft carriers, the sea-based JPALS. The JLIM (JPALS Land-based Integrity Monitor) is a prototype being developed at Stanford to meet the land-based JPALS requirements.

Figure 2.5: U.S.S. Harry Truman, commissioned 1998

2.3.1   The Moving Reference Station

For the sea-based implementation the entire landing platform (the aircraft carrier) is moving. The true position of the antennas upon the ships yard arm is not completely obscure but the motion of the ship does diminish the system's ability to discern errors on the ranging signals, as addressed previously (the distinction between a reference station and an integrity monitoring station). This furthers the necessity to develop highly discriminating statistics and logic to be able to detect any threatening and anomalous signal behavior.

Fortunately, whatever position error would occur at the reference station would be transmitted to the aircraft and the relative error would be ameliorated. Consider the GPS antennas on the aircraft carrier as moving beacons. The aircraft will find its way to the beacon even if that location is not known perfectly in absolute terms. Most importantly, confidence in the corrections being broadcast is diminished because the motion of the ship potentially obscures the ability to detect any systematic errors. My work provides better detection algorithms to reduce that risk.

In addition to the motion of the ship, there is motion of the landing point relative to the GPS antennas. This is an unavoidable consequence of a ship which is over 1000

feet long; there is stretching, twisting, and flexure of the superstructure. The ability to resolve, or bound, that motion is important to the success of such a landing system. It is beyond the scope of this thesis to analyze the impact of that flexure. Figure 2.6 shows just a few of those degrees of freedom and the lever arm which connects the antenna location to the touchdown/landing point.
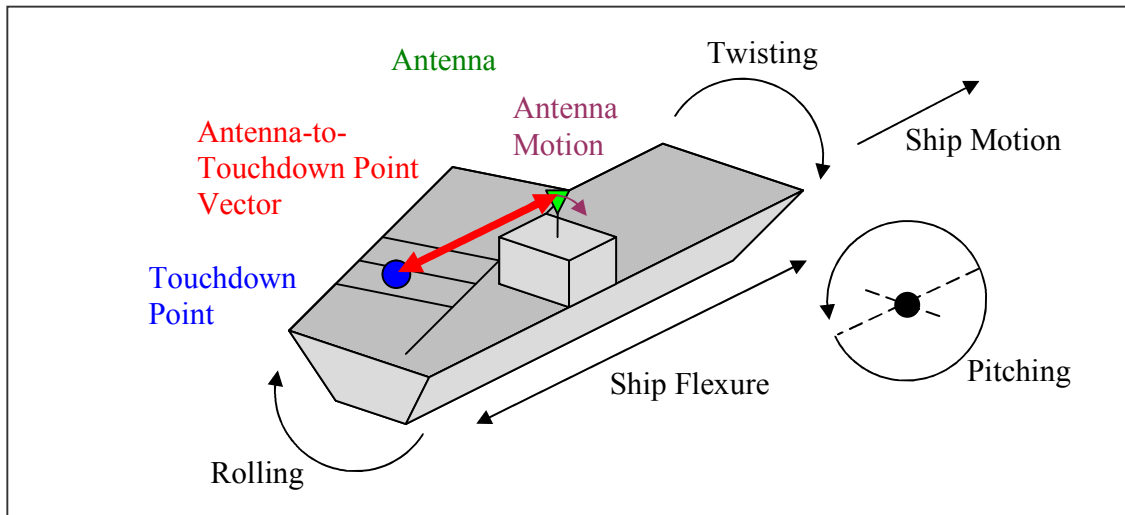
Figure 2.6: Antenna and Ship Motion Relative to the Touchdown Point

My work endeavors to separate potentially hazardous faults from nominal noise. As such, it is applicable to either land-based or ship-based JPALS. JPALS involves many segments, but this thesis work focuses on the algorithm development. The architecture of the JPALS Land-based Integrity Monitor (JLIM) is quite similar to the IMT, except that due to the military's expected difficult operating environment, adaptive antennas are used in conjunction with an RFI estimator. The contributions of this thesis use data collected from the IMT, and so, are not affected by the proposed change of using adaptive antennas. The remaining architecture is largely untouched, save that individual monitors may be refined to enhance their performance. Figure 2.7 shows an obvious influence from the IMT Flowchart of Figure 2.3.
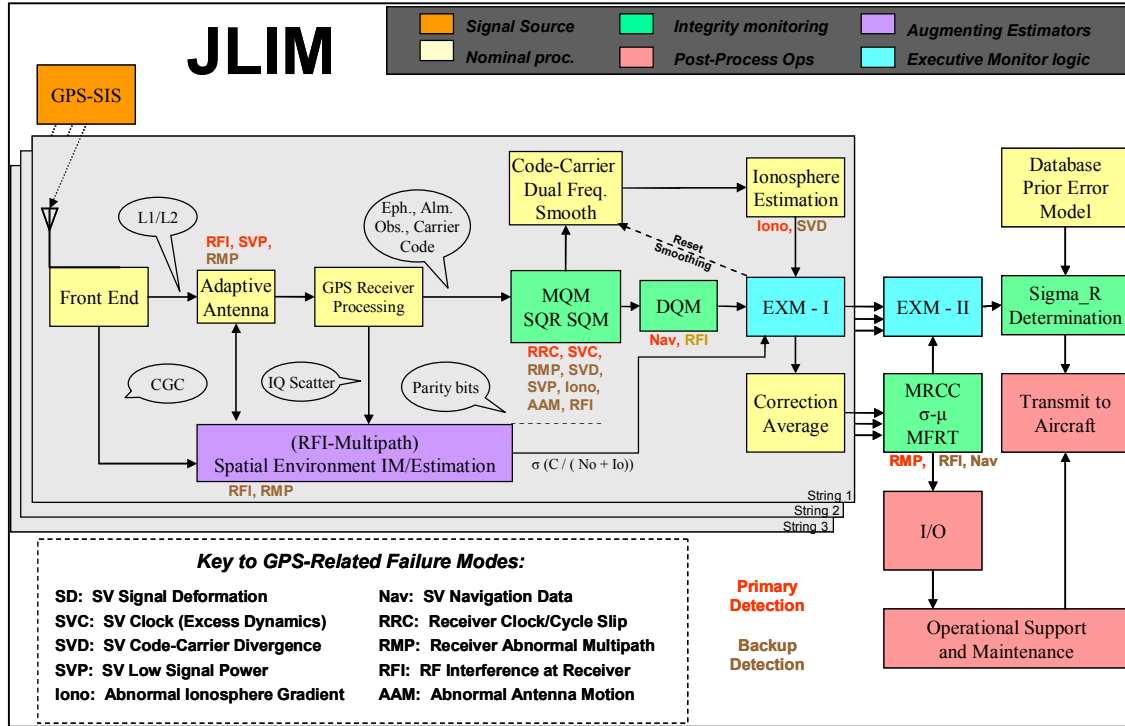
Figure 2.7: Stanford's JLIM Architecture

The JPALS Test-bed Platform (JTeP) is a simplified version of the JLIM. It is a software platform developed specifically for this thesis to enable the creation and testing of the MQM and EXM algorithms. The operations of these two monitors are the focus of this thesis.

2.4   The JPALS' Testbed Platform (JTeP)

The block diagram in Figure 2.4a shows that GPS data is loaded from saved files. The EXM-T assesses the tracking status of each channel, i.e. the presence and continuous age of the pseudorange, carrier phase, CN_0, and ephemeris. At this point, various failures can be injected into the measurements of any particular channel. The JTeP SQM and DQM operate identically to the IMT versions as described previously in Section 2.2.2.1 and 2.2.2.2. There are two new blocks included here, the EXM-T and the EXM-0.

In the IMT, the EXM-I created a Tracking Matrix which was a Boolean matrix to record which channels are currently tracked. The JTeP EXM-T formalizes this operation and moves it to the beginning. Deciding which channels are being tracked is something which is implicitly done in the IMT; it is just given a formal function in the JTeP so that all subsequent function can directly reference this tracking matrix.

The JTeP EXM-0 provides an important function not present in the IMT. As seen in Figure 2.8, in the IMT the output of the DQM and SQM are not considerations for the MQM. The MQM uses the measurements from all of the satellites tracked on a receiver in order to estimate and remove the receiver clock bias from those measurements for further processing. An SQM flag on a given channel will cause that channel to be excluded by the EXM-I algorithm of the IMT, yet that channel was still allowed to influence the operation of the MQM since the DQM, SQM, and MQM flags are evaluated simultaneously by the EXM-I. The JTeP EXM-0 looks at the SQM and DQM flags to determine if any channels should immediately be excluded. This means that the MQM is no longer obscured by a channel already flagged by another monitor and can operate more efficiently.

The JTeP MQM provides the same fundamental operation described in Section 2.2.2.3, the only difference is that it tests various filters to increase the detection probability of the algorithm. In summary, it looks at the temporal trends on the carrier phase. The JTeP EXM-I determines which channels to exclude and foremost, which satellites or receivers to also exclude, just like the IMT EXM-I. In the JTeP, the operation of the EXM-I is enhanced so that it evaluates more than just the Boolean output of the MQM. It does this in order to better detect failures which are present across all receivers for the same satellite. The JTeP EXM-I can also assume the task of determining which common set will produce the lowest VPL (Vertical Protection Level), discussed in Chapter 1. Selecting a common set was addressed in Section 2.2.3.1. In the IMT it merely involves deciding between potentially several common sets based on which set has the largest number of receivers, and then by the largest number of satellites.

That selection process can be changed to instead select a common set which results in the lowest VPL. Calculating the VPL is discussed in [5].
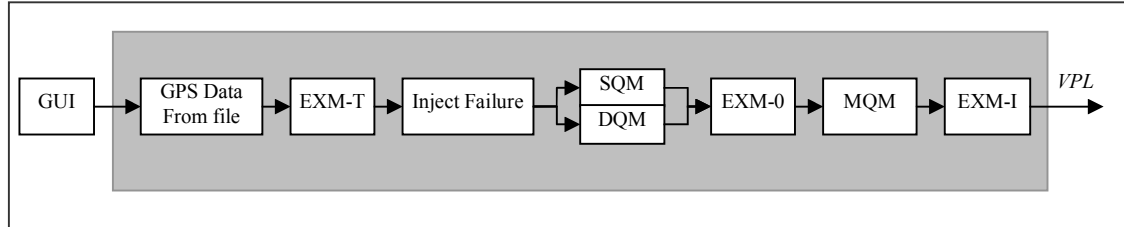
Figure 2.8: JTeP Overview

JTeP in Matlab

The JTeP is coded on the Matlab platform, a style similar to C-code. One of the most useful aspects of this creation is the use of a Graphical User Interface (GUI). While an interface could be constructed using C-Code, the Matlab platform code enables ease-of-use in manipulating input and output files, and certainly so when applying the mathematical abilities of Matlab upon those output files. The use of Matlab significantly increases the ability to design the detect algorithms presented in this thesis. The JTeP GUI is shown in Figure 2.9 below. This construction provides the means to develop the contributions to the MQM in Chapter Three, and the EXM in Chapter Four.

The JTeP GUI is comprised of four primary sub-structures.

1.   File Input (Top): The system can run on multiple input file types. Typically, GPS files are in the RINEX (Receiver Independent Exchange) Format. This is a commonly used format for all GPS receivers. The system has been designed to also operate on something called "packets". Packets are one-hour blocks of GPS data which have been pre-processed so that measurements are loaded as a large matrix instead of being read in line by line as would be with the ASCII text or binary style RINEX files. Loading a large matrix saves significant time because the hard-drive does not need to be accessed every epoch.

2.   Receiver Parameters (Top Middle): Each GPS receiver displays its antenna coordinates stored in both XYZ (Earth Centered, Earth Fixed) coordinates and LLA (Latitude, Longitude, and Altitude). The display also shows the Start and Stop time of measurement data and the First time of ephemeris data for each receiver.

3   Run-Time Parameters (Lower Left): This section allows the user some flexibility in running the file. The user can specify how many epochs the system should run for. The GUI displays the progress of running the file (it is sometimes several hours). The user can also enter a filename for logging output data from the system.

4.   Logging Parameters (Lower Right): This is perhaps the most useful section of the code. Under the old IMT system, the user would have to open the (.C) files and write in code for the software to write certain parameters to file. The IMT project for that (.C) would then have to be recompiled and then run. Under the JTeP Matlab code, every parameter to be logged can simply be chosen from the drop down menus. Those parameters can also be logged for a specific channel, or for all receivers, satellites, and frequencies which have that parameter. The output files generated are conveniently saved as Matlab structure which can be indexed by their receiver, satellite, and frequency number.

Figure 2.9: The JTeP in Operation

Chapter 3 - The Measurement Quality Monitor

The MQM is a Quality Monitor which examines the carrier phase on each channel for abnormal temporal trends. It applies a second-order polynomial fitting to a 10 epoch segment of continuous two hertz carrier-phase data to determine if there are any jumps (which might indicate a cycle slip) or overly large accelerations (which may cause an error on the corrected carrier-phase broadcast to the user). The reason for monitoring temporal trends on the carrier-phase are twofold: 1) an unusually high value would indicate the system is not performing nominally and decrease the confidence in the performance because all thresholds are determined under nominal conditions, and 2) there is a latency between the time a correction is calculated at the ground station and the time it is applied by the aircraft. This means that the aircraft could be using an aged and inaccurate correction. This is similar in principle to the operation of the MFRT which monitors the size of the range correction rate in addition to the size of the correction itself.

The JTeP was designed to be able to run in real time or post process data. All of the analysis of the MQM in this chapter and EXM in the next chapter was done by post processing saved data files.

There are three major sections to this chapter. Section 3.1 explains how the raw carrier-phase measurements are processed to remove the receiver clock bias. Section 3.2 contains the bulk of the chapter and explains how each of the filters of the IMT MQM are derived, then continues by deriving better performing acceleration and velocity filters. Section 3.3 succinctly summarizes the chapter with recommendations for the new JTeP MQM.

3.1   Data Reduction in the JTeP

The data reduction which leads up to the MQM occurs as follows. First the carrier-phase data is pulled from the data file. Then if that channel has adequate elevation, tracking times, and CNo values, it will pass those carrier-phase values to the MQM to be processed. The carrier-phase values for the first two hours of the data set are shown in Figures 3.1 and 3.2. Each color is a different satellite. Notice that the data for the highest carrier phase values there are frequent outages. Lower elevation satellites are actually farther away, thus resulting in the larger carrier phase values and also the cause of the outages.
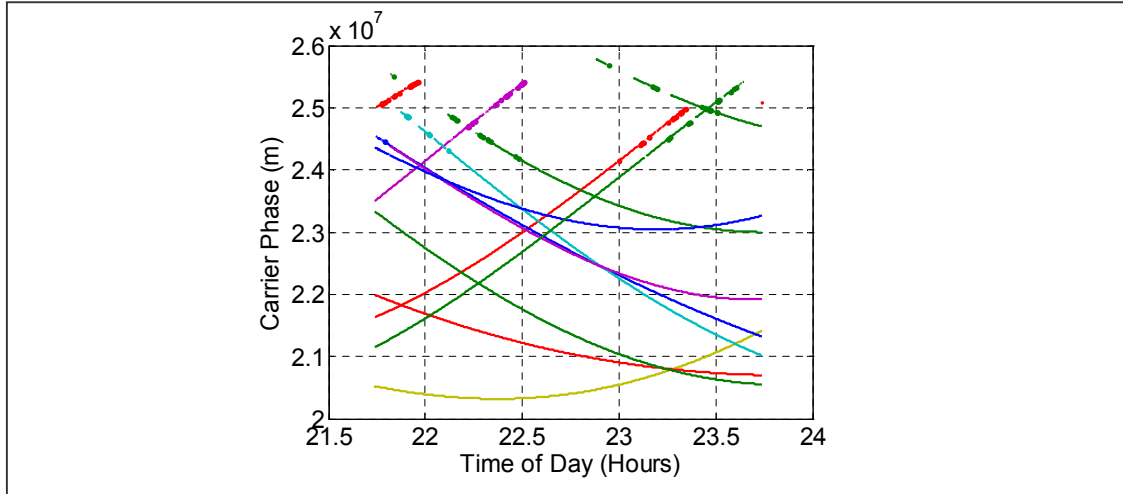
Figure 3.1: Raw CPH (Φ) Data for Receiver 1

The expected range and satellite clock correction are subtracted from the carrier phase. This is specific to each channel, i.e. receiver/satellite. Equation (3.1) shows the corrected carrier phase (ϕ_C) calculated from the carrier phase (ϕ), the range to the satellite (R), the satellite clock correction (τ).

$$\phi_C^{(RX,SV)} = \phi^{(RX,SV)} - (R^{(RX,SV)} + \tau^{(RX,SV)}) \tag{3.1}$$

42
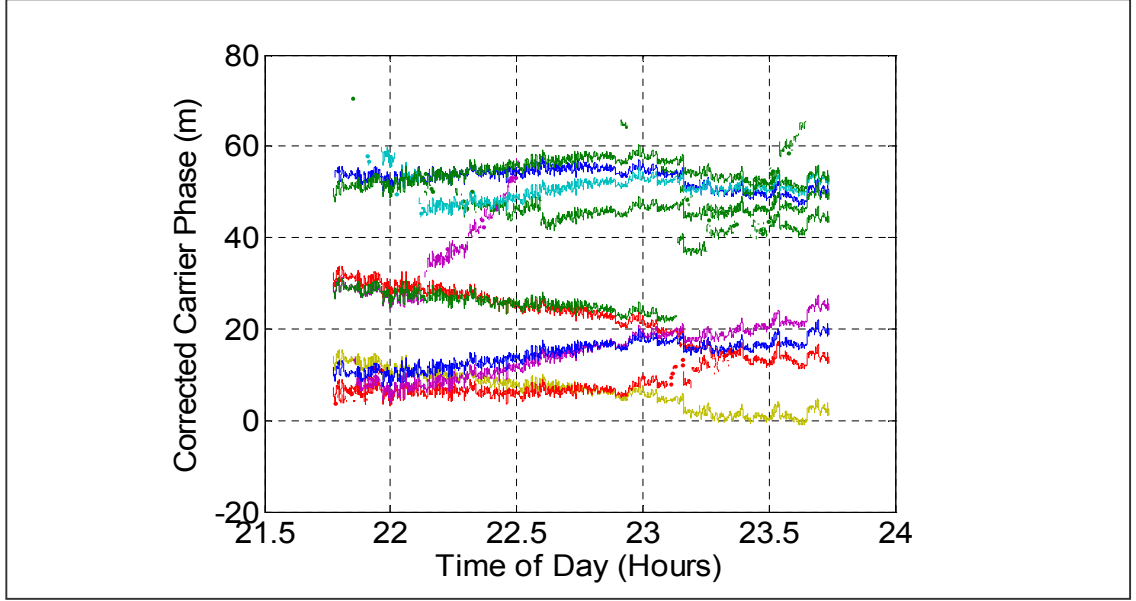
Figure 3.2: Corrected CPH (Φ_C) Data for Receiver 1

As shown in Figure 3.2, all of the satellites exhibit similar, but offset, curves. The offset is due to different ionosphere delays, while the similarity is due to the receiver clock bias at each epoch. In order to determine the temporal trend of each channel this bias is removed by subtracting the average Φ_C values for each respective receiver at each epoch.

Now, all of the corrected carrier-phase values are assembled and a common set is determined. The common set means that for each receiver in the set, there are the same satellites. This enables the receiver clock bias to be subtracted for each receiver without introducing a satellite bias. N_C is the number of satellites in the common set. It is hopeful that a common set can be found across all three receivers, but if only two can be used then the third receiver is immediately excluded.

$$\phi_{CA}^{(RX,SV)} = \phi_C^{(RX,SV)} - \frac{1}{N_C}\sum_{i=1}^{N_C}\phi_C^{(RX,i)} \qquad (3.2)$$

43
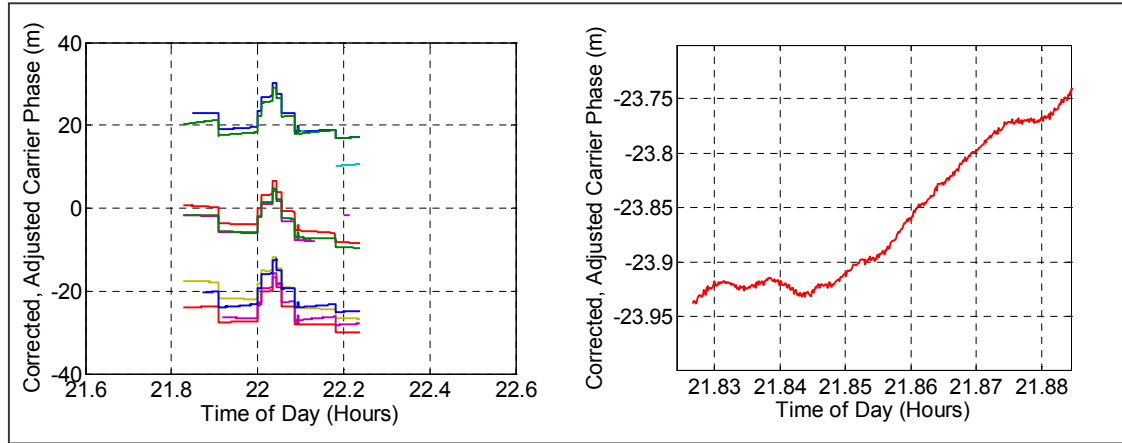
Figure 3.3: a) Corrected-Adjusted CPH (Φ_{CA}) Data for Receiver 1, b) Zoomed

Figure 3.3a above shows that by removing the receiver clock bias the curves become much more stable. There are obvious jumps in the curves, but this is due to changing the common set which estimates the receiver bias. There is a buffer which stores the last ten epochs of carrier phase data, thus when there is a change in the common set, it can be applied to all the carrier phase data in the buffer. This means that those jumps only appear in the plot and do not appear as jumps in the carrier phase to the MQM. Figure 3.3b is a zoom in which shows the millimeter level phase noise atop the lower frequency variation attributable to the ephemeris and satellite clock.

3.2 Polynomial Fitting

The IMT MQM is tested for overly large carrier-phase discontinuities (Step-Test) and carrier-phase accelerations (Acc-Test). Including a test for carrier-phase velocity (Vel-Test) was considered for the IMT MQM, but was not incorporated because the test statistic proved to have too much noise to be useful. The MQM developed as part of this thesis uses the Step-Test from the IMT MQM but alters the Acc-Test to increase its performance. The Vel-Test is also altered in order to dramatically reduce the noise on the test statistic so that it can be considered for inclusion in a future monitoring system. The velocity seen on the carrier-phase is affected significantly more by the Ionosphere than the satellite clock. As such, adopting the Vel-Test might interfere with the algorithms monitoring the Ionosphere. This requires careful consideration and is a topic for future

44

work.   At present the IMT MQM uses a finite length second order polynomial fitting which given in Equation (3.3)

$$\phi_{CA}^{(RX,SV)} = \beta_0 + \beta_1 \cdot n + \tfrac{1}{2}\beta_2 \cdot n^2 \qquad (3.3)$$

The Beta terms are the common representation of polynomial coefficients. β_0 is the phase estimate, β_1 is the ramp estimate and β_2 is the acceleration estimate. The impulse estimate arises by comparing the carrier-phase of the current epoch to that of the model. The (RX, SV) superscript has been dropped for simplification. The equation for estimating the carrier phase impulse is given in Equation (3.4), which says the impulse is the difference between the actual corrected-adjusted carrier phase and the corrected-adjusted carrier phase predicted from the model derived from Equation (3.3).

$$I = \phi_{CA}(n) - \hat{\phi}_{CA}(n) \qquad (3.4)$$

Whether or not to include a step detector is discussed later. Continuing with the calculation of velocity and acceleration, the method for estimating the coefficients is simply the matrix pseudo-inversion. The estimates of β are a result of the measured Φ_{CA} values, and the indices matrix, M. The superscript $\dagger$ is the notation for the pseudo-inverse of a matrix. It is expressed in Equation (3.6).

$$\begin{pmatrix} \phi_{CA}(1) \\ \vdots \\ \phi_{CA}(10) \end{pmatrix} = \begin{pmatrix} 1 & 1 & \tfrac{1}{2}1^2 \\ \vdots & \vdots & \vdots \\ 1 & n & \tfrac{1}{2}n^2 \end{pmatrix} \cdot \begin{pmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \end{pmatrix} = M \cdot B \qquad (3.5)$$

$$\begin{aligned} \hat{B} &= \left(M^T \cdot M \right)^{-1} \cdot M^T \cdot \Phi \\ &= M^{\dagger} \cdot \Phi \end{aligned} \qquad (3.6)$$

The coefficients of each correlator which result from this method are shown below. The Ramp correlator is a straight line (this is not actually the case for the IMT

and will be addressed in another section of this chapter), the Acceleration correlator is a parabola, but the odd shape of the Impulse correlator comes about because it is looking at the discrepancy of the first epoch to that of the other nine. The x-axis represents the order of the coefficients such that epoch one is applied to the current measurement and epoch ten is applied to the measurement which is already nine epochs old.
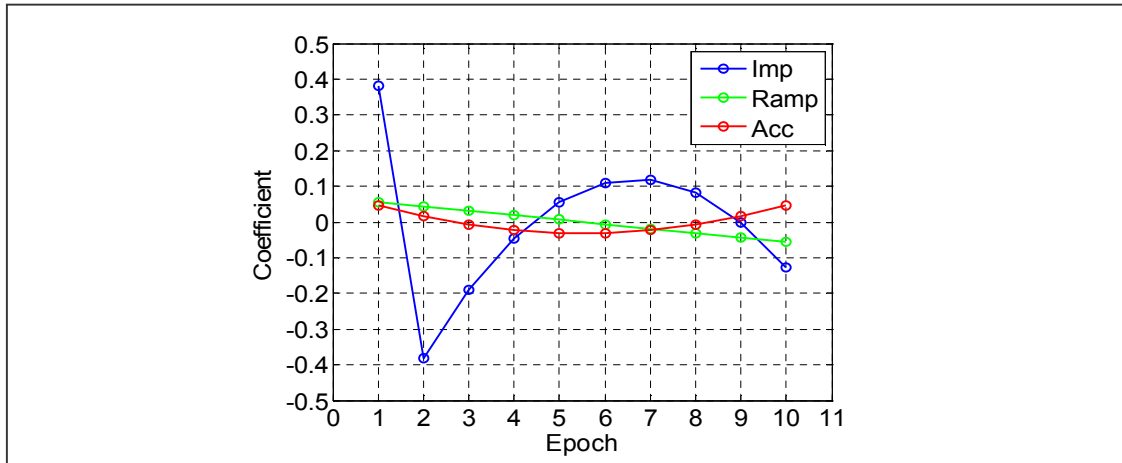
Figure 3.4: The MQM Impulse, Ramp, and Acceleration Correlator Curves

The MQM process involves fitting a polynomial, but because this is done using carrier phase samples equally spaced in time, it is equivalent to a finite impulse response filter, or FIR. No polynomial fitting must actually be done, because the coefficients associated with these FIRs need only be calculated once, before the system ever runs. The filters are time-invariant. Those FIRs are then applied continuously to the data coming in, accounting of course for any resets in the carrier phase due to loss of tracking continuity or channel resets due to faults. This is essentially convolving a particular curve against a finite segment of data and seeing if there is a response. This produces the polynomial "correlator" curves shown in Figure 3.4 above.

3.2.1   Signal to Noise Ratio, Γ

As with any system which tries to distinguish between a signal and the nominal noise, a monitor or filter which can maximize such a ratio is ideal. The SNR (signal to

noise ratio) relates the power of each source, whereas here it is ratio of the amplitude of each source which is of interest. The amplitude of the noise is its standard deviation, and the amplitude of the signal is just its response to the filter.

$$SNR = \frac{P_{Signal}}{P_{Noise}} \tag{3.7}$$

$$\Gamma = \frac{A_{Signal}}{A_{Noise}} \tag{3.8}$$

A two-point mean filter can be readily compared to a three-point mean filter, assuming the input noise is zero-mean Gaussian. Equations (3.10) and (3.12) give the steady-state noise amplitude.

$$\mu_2(n) = \tfrac{1}{2}\big[x(n) + x(n-1)\big] \tag{3.9}$$

$$A_{2,Noise} = \sigma_2 = \tfrac{1}{\sqrt{2}} \cdot \sigma_x \tag{3.10}$$

$$\mu_3(n) = \tfrac{1}{3}\big[x(n) + x(n-1) + x(n-2)\big] \tag{3.11}$$

$$A_{3,Noise} = \sigma_3 = \tfrac{1}{\sqrt{3}} \cdot \sigma_x \tag{3.12}$$

The responses of each filter are shown in the first plot of Figure 3.5a and 3.5b. The responses when scaled for their associated noise factors are shown in the second plot. The second plot indicates that the response of the three point filter will generally lag behind that of the two-point filter, but ultimately reach a higher Gamma value over a long enough period. This would suggest that fast detection is accomplished by the shorter filter, whereas more reliable detection is accomplished by the longer filter. This isn't necessarily the case though if the signal was to rapidly fluctuate. The longer filter may detect a smaller signal, but only if the signal is present for a sufficient amount of time. This is important to recognize otherwise it would seem that a longer filter may always be better.
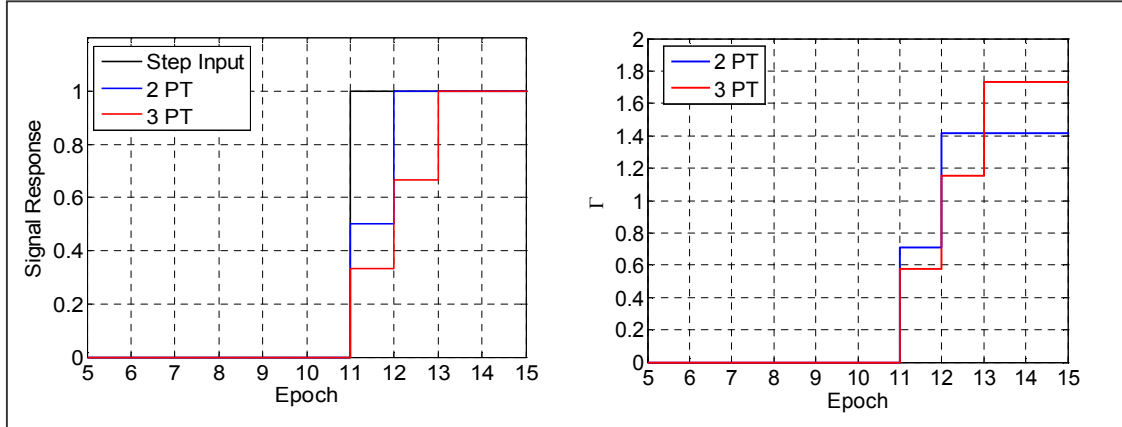
Figure 3.5: 2-Pt, and 3-Pt Averaging Filters, a) Step Response, b) Γ Response

3.2.2 Discrete/Frequency Domain

The MQM uses Finite Impulse Response Filters, or FIRs. The well known concept of the discrete domain and Z-Transforms can be used to analyze the performance of these filters. This method also allows an analysis in the frequency domain which allows for a quick solution of of noise throughput of multiple filters. When some variable y is a linear combination of an input time series x, such as below, the Z-Transform creates a simple transfer function between the input x and output y.

$$y_n = \alpha_0 \cdot x_n + \alpha_1 \cdot x_{n-1} + ...\alpha_N \cdot x_{n-N} = \sum_{i=0}^{N} \alpha_i \cdot x_{n-i} \tag{3.13}$$

The Z-Domain representation is.

$$Y(Z) = \alpha_0 \cdot X(Z) + \alpha_1 \cdot X(Z) \cdot z^{-1} + ...\alpha_{N-1} \cdot X(Z) \cdot z^{1-N} + \alpha_N \cdot X(Z) \cdot z^{-N} \tag{3.14}$$

The transfer function can be then written.

$$\frac{Y(Z)}{X(Z)} = \frac{\alpha_0 \cdot z^N + \alpha_1 \cdot z^{N-1} + ...\alpha_{N-1} \cdot z + \alpha_N}{z^N} \tag{3.15}$$

When the *X* data values are independent, the following relations hold between the output noise, input noise, coefficients, and frequency spectrum.

$$\sigma_Y^2 = \sigma_X^2 \cdot \sum_{i=0}^{N} \left(\alpha_i^2 \right) \qquad (3.16)$$

$$\sigma_Y^2 = \sigma_X^2 \cdot \int_{\Omega=0}^{1} \left| H(-j \cdot 2 \cdot \pi \cdot \Omega) \right|^2 \cdot dz \qquad (3.17)$$

The frequency responses of the two mean filters are shown in Figure 3.6 below. The amount of power each filter contributes is the integral of the square of the amplitude. The power of the three-point mean filter is less than that of the two-point mean filter as is evident in Figure 3.6. Roughly speaking, the less area under the curve, the better the filter is at rejecting noise.
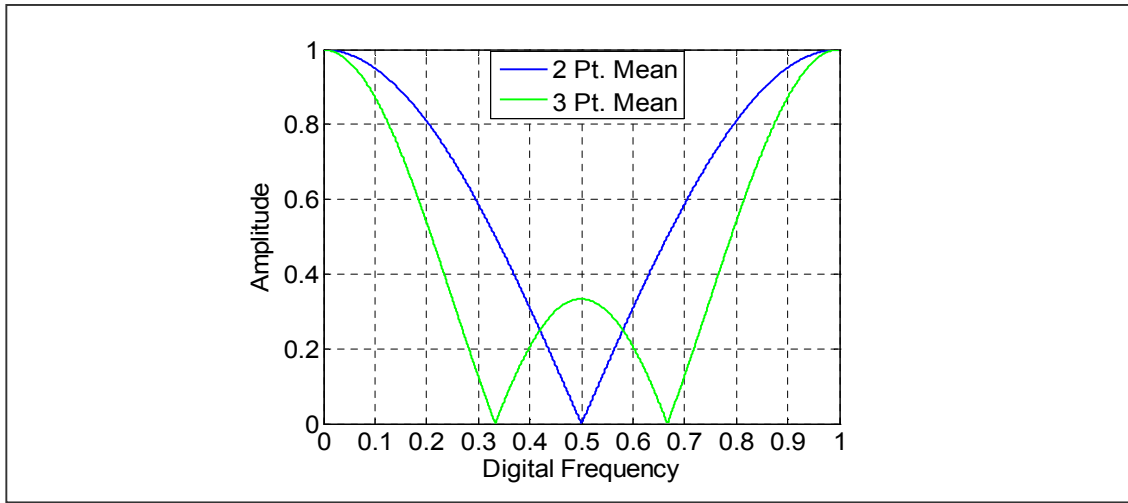
Figure 3.6: Frequency Responses of the 2-Pt and a 3-Pt. Mean Filters

3.2.3 Correlation

Data Correlation

When the input data is correlated there is a prominent effect on the output of the MQM filters. The frequency domain facilitates a simple calculation of the output noise.

The correlation of the input noise is defined with a single parameter, τ. Using a vector notation of Equation (3.13), let, $y = \bar{\alpha} \cdot \bar{x}$, with

$$E[x_n \cdot x_{n-1}] = \tau \cdot \sigma_x^2 \tag{3.18}$$

The resulting covariance of y is given by the following,

$$
\begin{aligned}
\sigma_y^2 &= (\bar{\alpha} \cdot \bar{x}) \cdot (\bar{\alpha} \cdot \bar{x})^T \\
&= \bar{\alpha} \cdot \Sigma_x \cdot \bar{\alpha}^T \\
&= \bar{\alpha} \cdot
\begin{pmatrix}
1 & \tau & \cdots & \tau^{N-1} \\
\tau & 1 & \tau & \vdots \\
\vdots & \tau & 1 & \tau \\
\tau^{N-1} & \cdots & \tau & 1
\end{pmatrix}
\cdot \bar{\alpha}^T \cdot \sigma_x^2
\end{aligned}
\tag{3.19}
$$

The above method is very useful for the relatively short FIR filters of the IMT MQM. Some of the candidate filters described later in this chapter are not FIR and the calculation of the covariance can be done through an integration of the frequency domain describe previously. The Z-Domain representation of a correlated input is given in Equation (3.20). The end multiplier keeps the correlated noise to a fixed variance (see App E). When white noise data is run through a first order filter, the amount of power that is transmitted through the filter is a function of the correlation coefficient, τ. The last term of Equation (3.20) keeps the power transmitted through the filter constant, regardless of the correlation coefficient.

$$\frac{Y(Z)}{X(Z)} = \left(\frac{\alpha_0 \cdot z^N + \alpha_1 \cdot z^{N-1} + \dots \alpha_{N-1} \cdot z + \alpha_N}{z^N} \right) \cdot \frac{z}{z - \tau} \cdot \sqrt{1 - \tau^2} \tag{3.20}$$

Filter Autocorrelation

The single parameter, τ, can capture the essence of the input data because it is very well conditioned. There are no innately exotic circumstances which correlate the

50

data. The filters themselves are somewhat different. It is less appropriate to attribute a single parameter to describe the correlation that occurs within the filter. In this case, the full autocorrelation is used to describe the filters. The autocorrelation function of the two-point and three-point mean filters are shown in Figure 3.7. The autocorrelation of a filter uses the standard definition of autocorrelation operating on the impulse response of the filter. This is expressed in Equation (3.21), where the alpha terms are the coefficients of the FIR.

$$A_{\Delta N} = \frac{1}{\|\vec{\alpha}\|} \cdot \sum_{i=0}^{n-|\Delta N|} \left(\alpha_i \cdot \alpha_{i+|\Delta N|} \right) \tag{3.21}$$
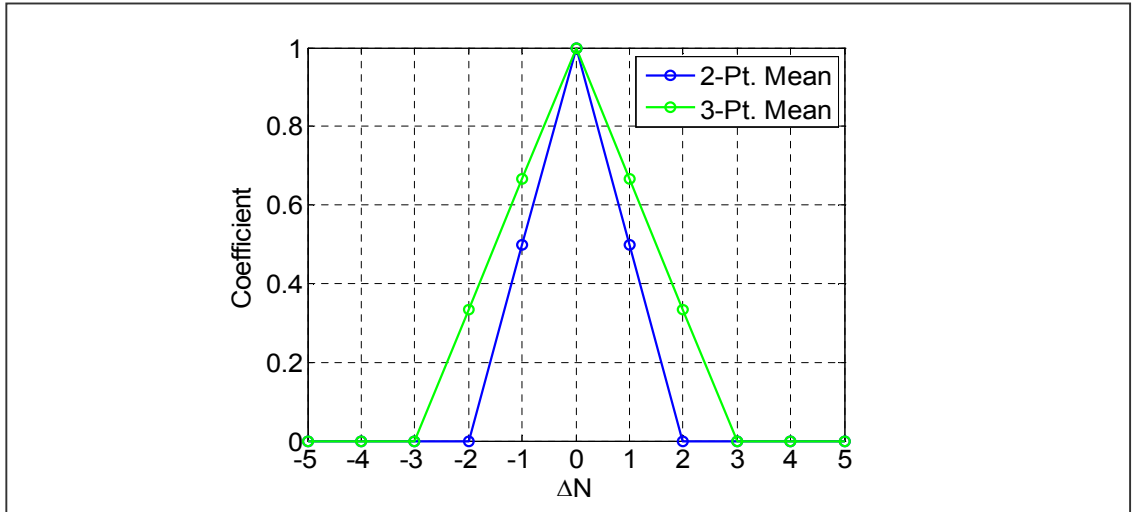
Figure 3.7: 2-Point and 3-Point Mean Autocorrelations

3.2.4 Impulse and Step Detectors

Impulse Detector

The polynomial coefficients from Equation (3.6) are used to estimate the discrepancy between the current data point, and the model suggested by the n most recent epochs. It is a bit of a misnomer because although this test is called a step test in the IMT it is really detecting an impulse. The reason it was given the name of Step Test is

because the intended fault mode was a cycle slip, which looks like a step in the carrier-phase measurement. For the JTeP, the proper terminology will be used. If the IMT step test is being referred to, it will be clearly stated. The Impulse estimate comes from Equation (3.22).

$$I = \phi(n) - \hat{\phi}(n) = \phi(n) - \left(\beta_0 + n\beta_1 + \tfrac{1}{2}n^2\beta_2\right) \qquad (3.22)$$

This detector is always sufficient to detect carrier-phase cycle slips and even half cycle slips because the reference antennas of the JTeP are stationary.

Step Detector

A "step" is a sustained shift in the mean. The question is whether this shift is most detectable at its instantiation or after some sustained period. The Impulse Filter served well to detect steps, typically cycle slips, and it is actually the case that designing a filter to detect a prolonged step causes a high correlation to the ramp detector. The ramp detector suffers from being susceptible to measuring the ionosphere gradient and causes there to be large testing thresholds, and this permeates to the step detector as well. Currently, the impulse detector provides adequate detection of steps to subside the necessity of an additional detector. Figure 3.8 shows a) the impulse detector, b) a pure step detector. Figure 3.9 shows the pure detector when orthogonalized to a ramp. Notice that the scale on Figure 3.9 has doubled, indicating a larger noise presence. At present, the impulse detector precludes a necessity for designing a "step" specific monitor.
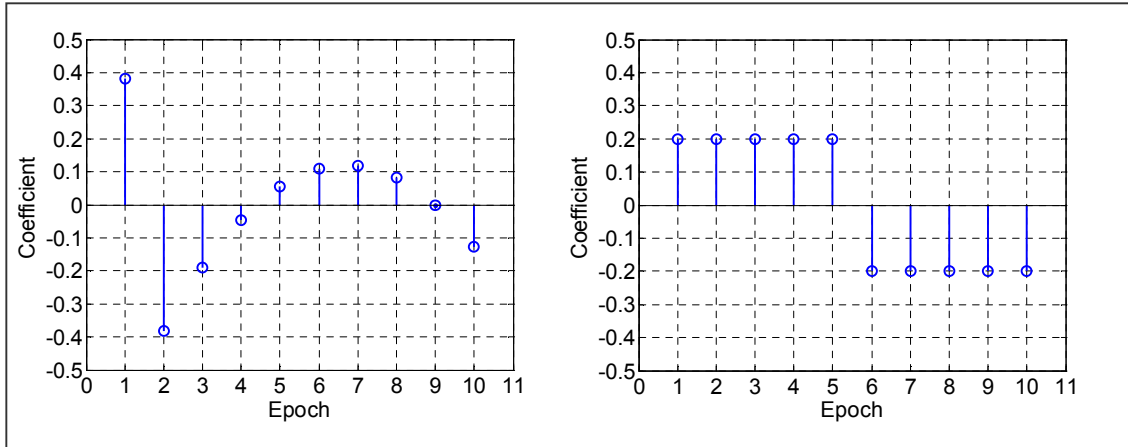
Figure 3.8: a) IMT Step Test Correlator, b) Basic Step Correlator
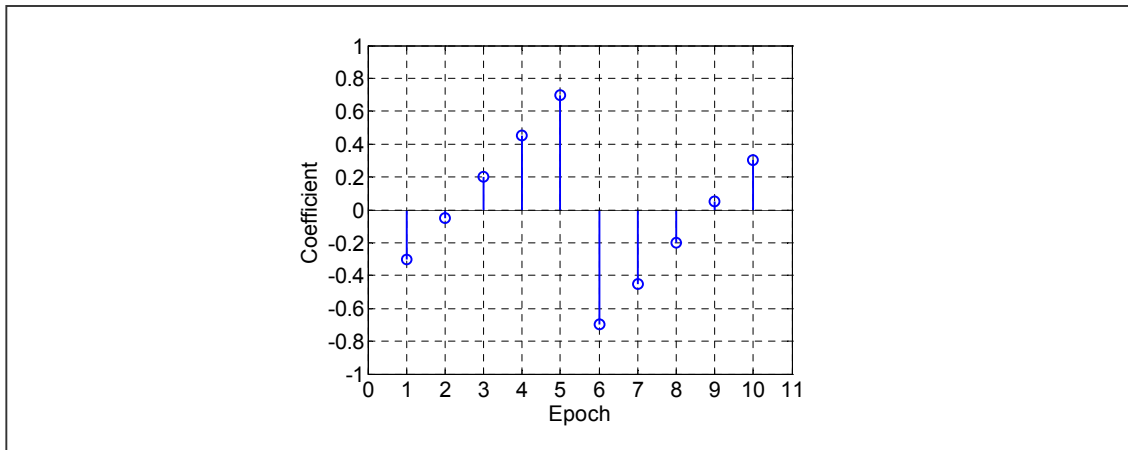
Figure 3.9: Step Correlator Orthogonal to a Ramp Input

The next four sections demonstrate how to improve the velocity and acceleration estimator currently in use by the IMT. Here is a quick summary of the sections to come:

3.2.5   The IMT Velocity Estimation (old method)

3.2.6   The JTeP Velocity Estimation (new method developed in this thesis)

3.2.7   The IMT Acceleration Estimation (old method)

3.2.8   The JTeP Acceleration Estimation (new method developed in this thesis)

This chapter ends with a summary comparison of the new and old methods for velocity and acceleration estimation of the carrier phase.

3.2.5 Velocity Estimation – IMT (Old Method)

The estimate of the velocity of the carrier phase is the β_1 term from the polynomial fitting of Equation (3.3). The problem the IMT faces is that this estimate of the velocity is a backwards extrapolation due to the choice of regression indices. The following figure shows the response of the IMT's velocity estimator to both a ramp input and an acceleration input.
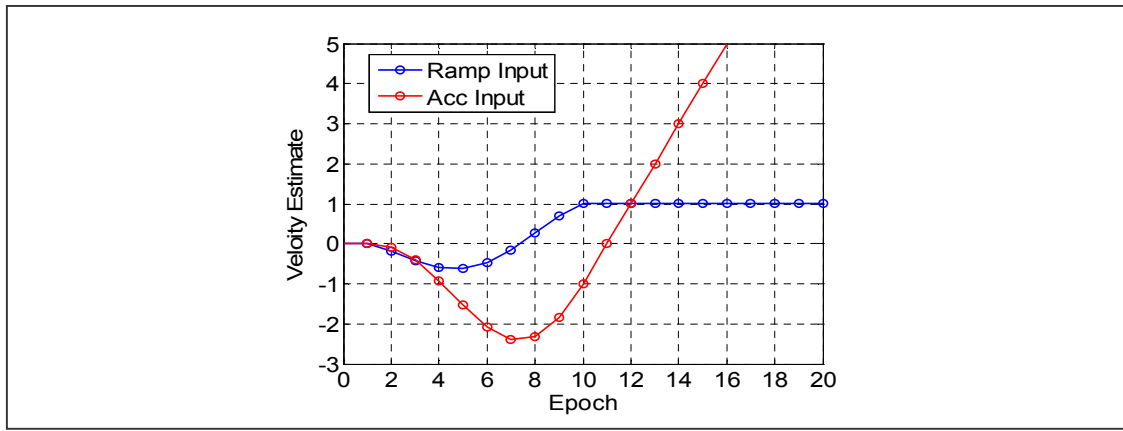
Figure 3.10: IMT MQM Velocity Estimate of a Ramp and Acceleration Input

The velocity onset starts at epoch 1, but that input value is zero, and thus so is the output. What is alarming is the negative response to a velocity input as well as the acc input. The cause for this undesired performance is simply the location of the regression indices. What is tacitly occurring is that the lower order estimates are a function of their indices, or placement on the horizontal axis.

$$
\begin{aligned}
y^A &= \tfrac{1}{2}\hat{\beta}_2^A \cdot x^2 + \hat{\beta}_1^A \cdot x + \hat{\beta}_0^A \\
&= \tfrac{1}{2}\hat{\beta}_2^A \cdot (x-k)^2 + \hat{\beta}_1^A \cdot (x-k) + \hat{\beta}_0^A \\
&= \tfrac{1}{2}\hat{\beta}_2^A \cdot x^2 + (\hat{\beta}_1^A - \hat{\beta}_2^A k)\cdot x + (\hat{\beta}_0^A - \hat{\beta}_1^A k + \tfrac{1}{2}\hat{\beta}_2^A k^2) \\
&= \tfrac{1}{2}\hat{\beta}_2^B \cdot x^2 + \hat{\beta}_1^B \cdot x + \hat{\beta}_0^B
\end{aligned}
\tag{3.23}
$$

$$
\hat{\beta}_1^B = (\hat{\beta}_1^A - \hat{\beta}_2^A k)
\tag{3.24}
$$

By shifting the indices backwards, the method for estimating velocity (shown in Figure 3.10 will: 1) be a function of $\hat{\beta}_2^A$ and k, 2) potentially be negative of $\widehat{\beta}_1^A$ if k is large, and 3) have an increased variance of $\sigma_{\beta_1}^2 = \sigma_{\beta_1}^2 + k^2 \sigma_{\beta_2}^2$.

Figure 3.11 below shows that the IMT is actually extrapolating backwards to estimate the velocity. The minimum error of a polynomial coefficient will occur at the center of the regressed indices (assuming consistent spacing), so the IMT is not only increasing the noise by extrapolating, but it is delaying the response by estimating the velocity at some previous epoch.
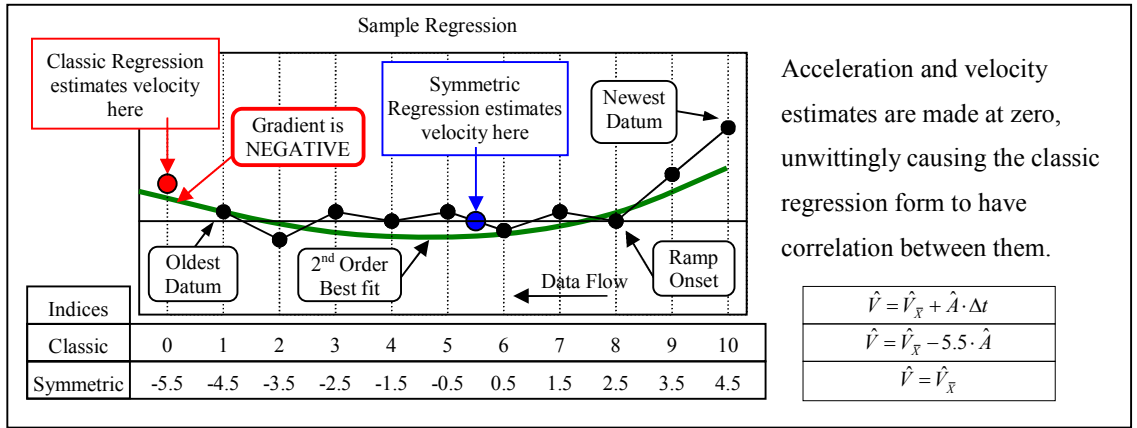
Figure 3.11: Symmetric Indices Decouple Velocity Estimate from Acceleration

Figure 3.11 compares the ramp and acceleration responses using the IMT's current backward predicting ramp detector to a proposed forward predicting ramp detector. Figure 3.12 shows the response curves as a function of the epoch of estimation using the IMT {1-10} indices. The square root of the noise power of each filter is {0.49, 0.24, 0.11, 0.24, 0.49} moving from the top curve to the bottom in Figure 3.12b. Thus for such a set of indices, the epoch of 5.5 is actually the center point, which consequently yields the lowest sigma.
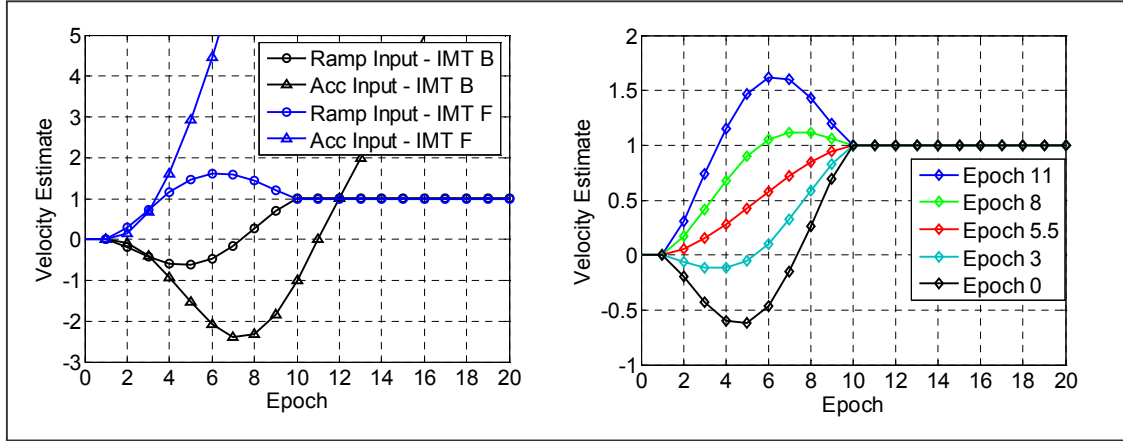
Figure 3.12: a) IMT Forward Predicting Velocity, b) Sample Epochs

Correlation to Acceleration

When the velocity is predicted away from the central point, referred to as $\hat{\beta}_1^*$, it becomes a function of the estimated acceleration.

$$\hat{\beta}_1^* = \hat{\beta}_1 + \Delta n \cdot \hat{\beta}_2 \tag{3.25}$$

When the non-symmetric indices are used the resulting covariance matrix of the polynomial estimates is:

$$\Sigma_\beta = \begin{pmatrix} 1.3833 & -0.5250 & 0.0833 \\ -0.5250 & 0.2413 & -0.0416 \\ 0.0833 & -0.0416 & 0.0075 \end{pmatrix} \tag{3.26}$$

When the above covariance matrix is normalized, the result is:

$$\Sigma_\beta^N = \begin{pmatrix} 1 & -0.908 & 0.814 \\ -0.908 & 1 & -0.975 \\ 0.814 & -0.975 & 1 \end{pmatrix} \tag{3.27}$$

56

There is obviously a large coupling between the velocity and acceleration. In a second order polynomial fitting, the position correlator is very similar to the acceleration correlator and that is the reason for the connection between velocity and position. The affect of the correlation is not bi-directional however. Lower order derivatives are affected by higher order derivatives, but not vice-versa. That is, the position estimate is affected by the true velocity and acceleration, and the velocity estimate is affected by the true acceleration. This is the flaw in the IMT velocity estimate. To complete the claim, the acceleration estimate is not affected by the true velocity or position, and the velocity estimate is not affected by the true position.

3.2.6 Velocity Estimation – JTeP (New Method)

The JTeP uses a modified version of the IMT's MQM; it uses the symmetric indices instead. The covariance matrix that results from this simple change is:

$$\Sigma = \begin{pmatrix} 0.2289 & 0 & -0.0313 \\ 0 & 0.0121 & 0 \\ -0.0313 & 0 & 0.0075 \end{pmatrix} \tag{3.28}$$

Normalized the above covariance matrix gives:

$$\Sigma_N = \begin{pmatrix} 1 & 0 & -0.7504 \\ 0 & 1 & 0 \\ -0.7504 & 0 & 1 \end{pmatrix} \tag{3.29}$$

The difference is immediate. The velocity and acceleration estimates are now independent, and aptly the variance/noise value of the JTeP ramp estimator is drastically reduced. The covariance matrices showed a reduction from 0.2413 to 0.0121. That's a ratio of almost 20, meaning the sigmas/thresholds can be approximately ($\sqrt{20} = 4.47$)

times tighter. That is under a white noise input assumption. Figure 3.13 shows the velocity detector responses using real GPS data with an injected ramp error.
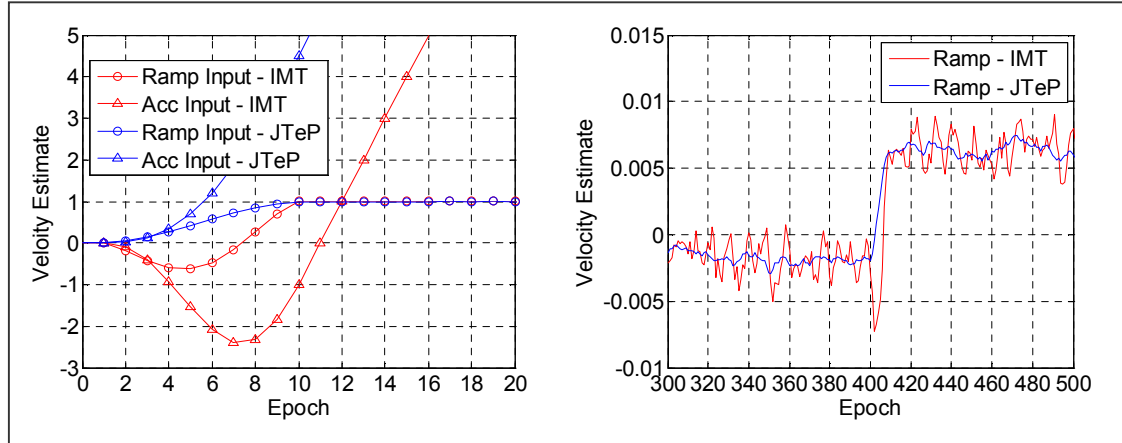
Figure 3.13: Vel. Estimate for a) Ramp and Acc. Input, and b) Ramp Input with Noise

Actually realizing the benefit of the ramp estimator is difficult because it requires knowledge of the ionosphere gradient. The ionosphere is a source of error in the GPS range measurements. As the satellite moves through the sky the amount of that error changes. The acceleration estimator and the velocity estimator will both register the respective rates of change of the ionosphere. However, the actual "acceleration" of the ionosphere error is much smaller than the amount of noise seen in the filter, but this is not true for the velocity estimator. In fact, the dominant signal in the velocity estimator is the ionosphere error gradient. The IMT attempts to subtract off the expected ionosphere gradient, but this is only done based on elevation, and it is different for every satellite with a different trajectory. The theory is that whether by ionosphere or by satellite clock, there is still a carrier phase velocity. The velocity due to the satellite clock affects both the ground station and the aircraft. The velocity due to the ionosphere is highly correlated between the ground station and aircraft because of the spatial proximity. This last issue is one of great interest and involves modeling of maximum ionosphere gradients [26][27]. The difficulty in doing so means that relatively high thresholds are used on the ramp estimates which may obscure a ramp in the satellite clock. This causes the ramp estimator to be an ineffective monitor.

3.2.7　Acceleration Estimation – IMT (Old Method)

The acceleration estimate is not affected by the placement of indices because it is the highest polynomial estimated, thus there is no propagation/extrapolation effect. The acceleration response to an acceleration input for the IMT's 10-point filter is shown in the Figure 3.14.
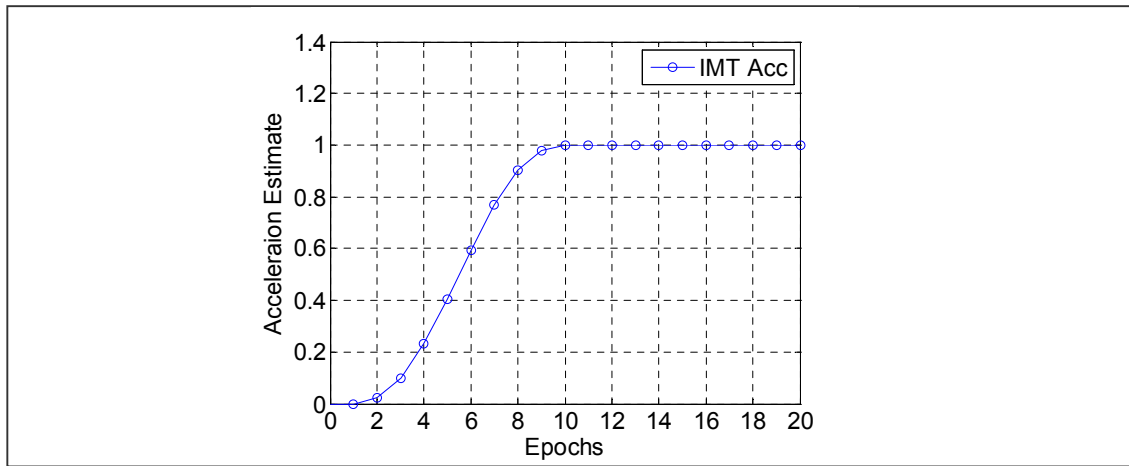
Figure 3.14: IMT Acceleration Response

Figure 3.14 shows that the filter responds slowly at first, then hits its maximum value at 10 epochs, the length of the filter. Using this response curve and the associated noise value, the goal for the JTeP MQM is to design a filter which will outperform this basic construction.

3.2.8　Acceleration Estimation – JTeP (New Method)

The objective of refining the acceleration monitor is to facilitate better detection, which can involve either less false detection or more positive detections. By reducing the nominal noise of the acceleration estimate, there is flexibility to achieve either. Instead of the simple ten-point polynomial fit used for the IMT, a longer or shorter filter can be used. The trade-offs are responsiveness and noise. The compromise is to attain noise reduction with equal response time. There are several areas to investigate to achieve this.

3.2.8.1 Filter Length

The number of regression points can be altered. Using the solution for β_2 from Equation (3.3), the covariance matrix as a function of the number of regression points, n, is calculated. Leaving the equations as a function of (n) and solving for the variance yields:

$$\sigma^2_{A(n)} = \frac{720}{n^5 - 5n^3 + 4n} \tag{3.30}$$

When there is correlated noise, the equation does not easily collapse into a form like Equation (3.30). Instead, it becomes significantly larger because it involves the convolution of the FIR acceleration estimation filter and the IIR correlation filter. For more complicated filters, the output noise is calculated using a numerical integration in the frequency domain by specifying n and τ., referred to in Equation (3.17). The effects of filter length and correlation are shown in Figure 3.15a and Figure 3.15b. What is evident is that long filters (20 or more points) will suffer an increase in the noise value as correlation increases to a certain point, although it will still be lower than the noise of a shorter filter. The "CL" is used to describe the "classic" acceleration filter design of the IMT that preceded the current work. This serves as a basis for comparison with the candidate acceleration filters examined in this section.
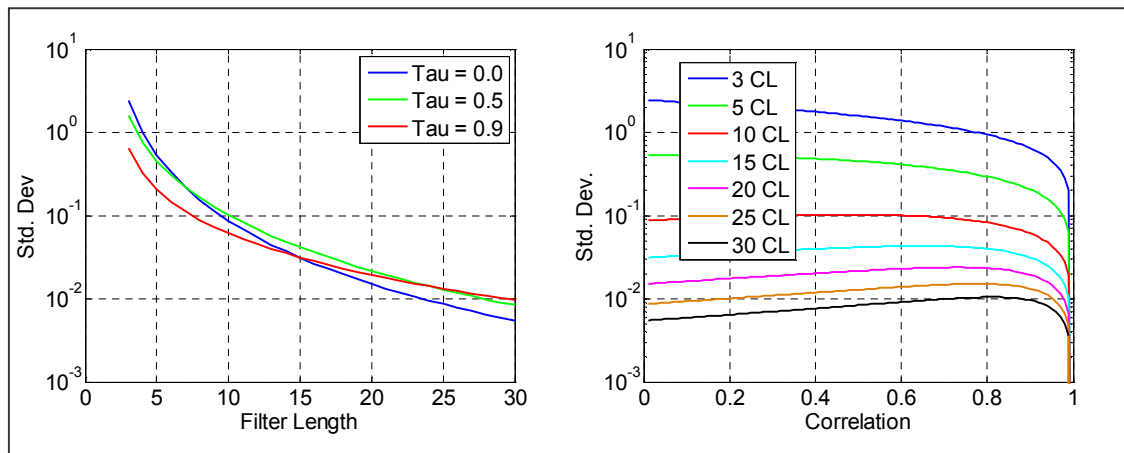
Figure 3.15: The Effects of a) Filter Length, and b) Correlation, on MQM ACC

Figure 3.16a and Figure 3.16b give an important illustration of the spectrum of colored noise, the spectrum of the classic acceleration filter, and the product of these two entities.
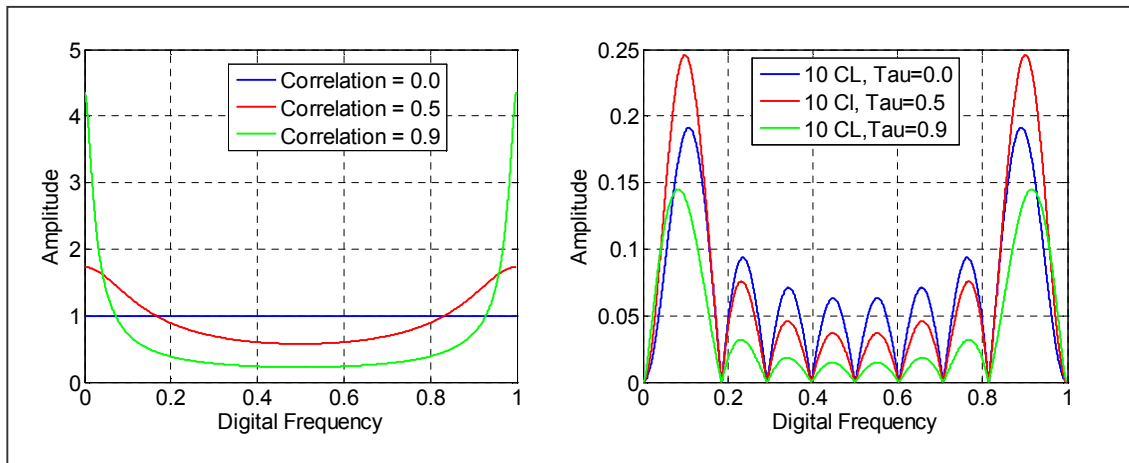
Figure 3.16: a) Correlated Noise in the Digital Domain, b) the Effect on 10 CL

The primary issue is that the longer filter will be slower to respond, and there is an expectation that this will be compensated for by a smaller sigma, which may not be the case under high correlation. The response vs. filter length is shown in Figure 3.17.
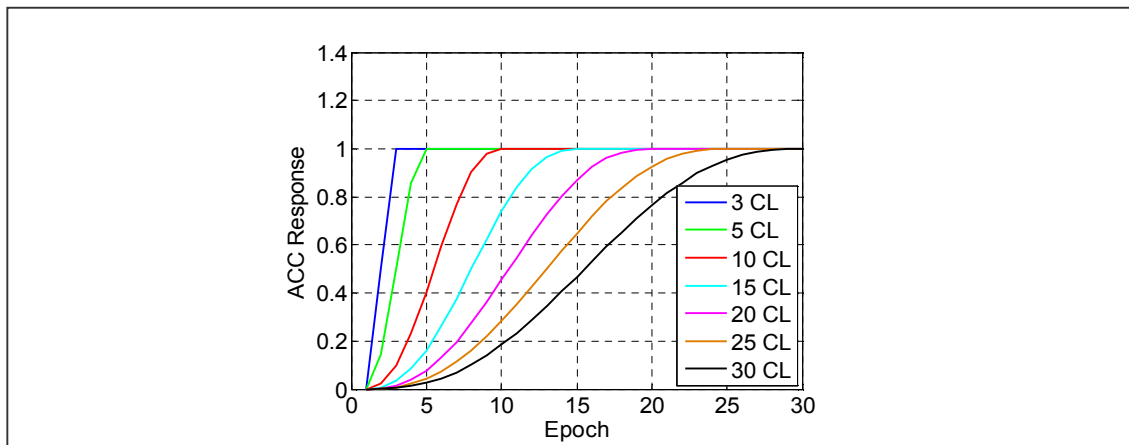
Figure 3.17: Acceleration Responses

$$\Gamma = \frac{\text{Response}}{\sigma_{Noise}} \qquad (3.31)$$

The responses of the longer filters are slower as expected, but when scaled for their sigmas, yielding their gammas (see Equation 3.31), the distinction is hardly as clear. The purpose for using this scaling ratio is that the thresholds of the test statistic will be a function of the standard deviation of the noise that the filter transmits. For example, in Chapter Two, it was repeatedly mentioned that various monitors use thresholds of 6-7σ. This means that if the new filter design can maintain the same response to a signal input while reducing the size of the response to just noise, then the sensitivity of the filter has increased and it enables the detection of smaller faults because the detection thresholds are reduced.

Figure 3.18 shows that, with regard to the responsiveness, the longest filter provides an effective bound on the responses of the shorter filters. It has essentially the same short term gamma, but while also having the largest long-term gamma, indicating its superior sensitivity.
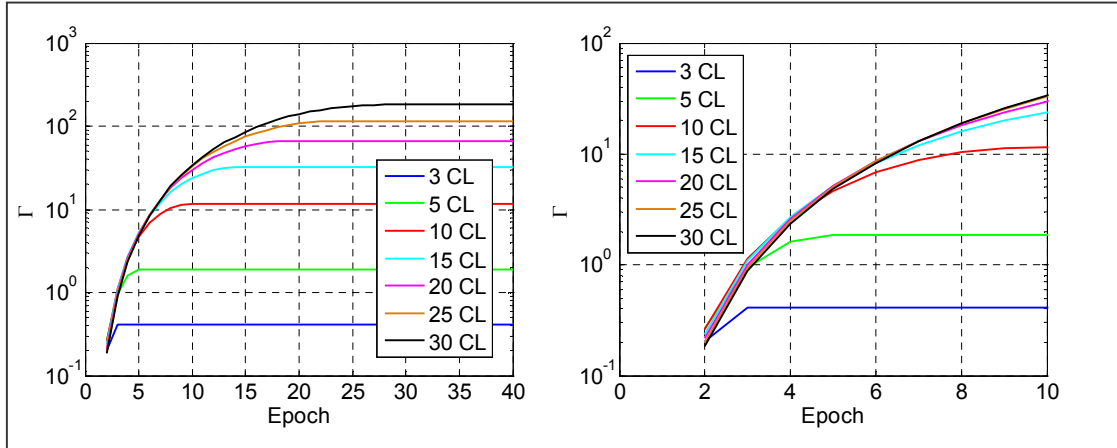
Figure 3.18: Response a) when Scaled for Sigma, b) Zoomed

When the input noise is correlated, the bounding property of the long filter is lost. Figure 3.19 shows that when $\tau = 0.9$, for the short-term, or quick detection, the shorter filters produce a higher detection ratio, Γ. The 30 point filter, while still the most

62

sensitive, doesn't even match the 10-point filter response until 9 epochs into the acceleration event. This should instill hesitation towards adopting a longer filter, and the subsequent analysis of the GPS data indicates that there is an appreciable correlation on the noise of the Φ_{CA} data.
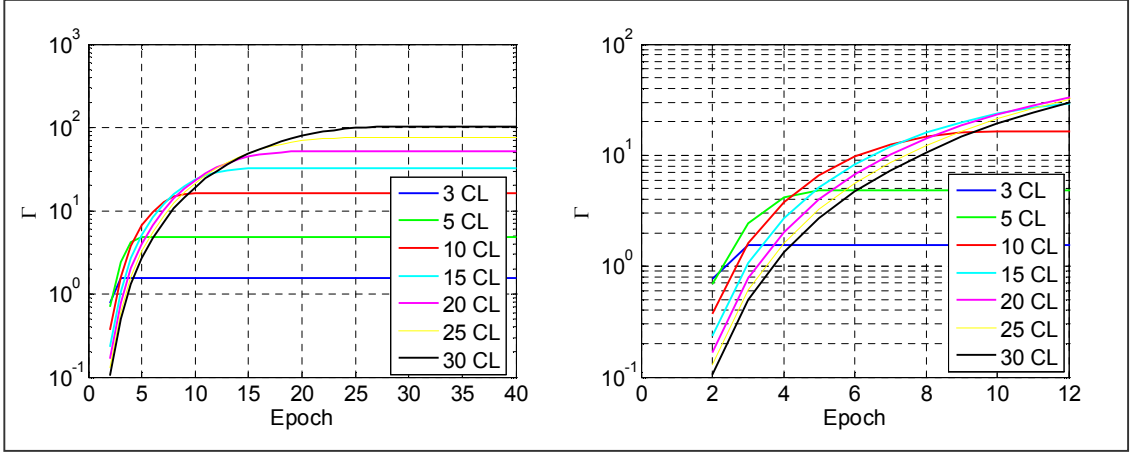
Figure 3.19: Response a) when Scaled for Sigma $\tau = 0.9$, b) Zoomed

The exact length of the optimal filter will be dependent upon the characteristics of the underlying noise. Additionally, there are several methods which may improve the performance beyond that of the Classic Filter. Consequently, these methods are investigated along the way to determine what is the desired filter length and shape, comparing a short filter of around 3 points, to a longer filter of around 10 points (as the IMT currently uses).

3.2.8.2 Spliced Regression

Spliced regression is just a term given to the idea of a piecewise polynomial fit (or splines) where two polynomials are used and their derivatives are forced to be equal at a common point [29]. The motivation for this came from the analysis of Section 3.2.8.1. The idea was to increase the number of points in the filter to reduce the amount of noise but in a way which didn't slow the responsiveness of the filter.

For the JTeP, two ten point filters are splicing together and forcing their zero and first derivatives to be equal, the result is a 19 point FIR filter. Figure 3.20 shows an acceleration beginning at epoch 10 (black), then with additional noise (blue), and finally how a spliced model will precisely mold to this situation (red).
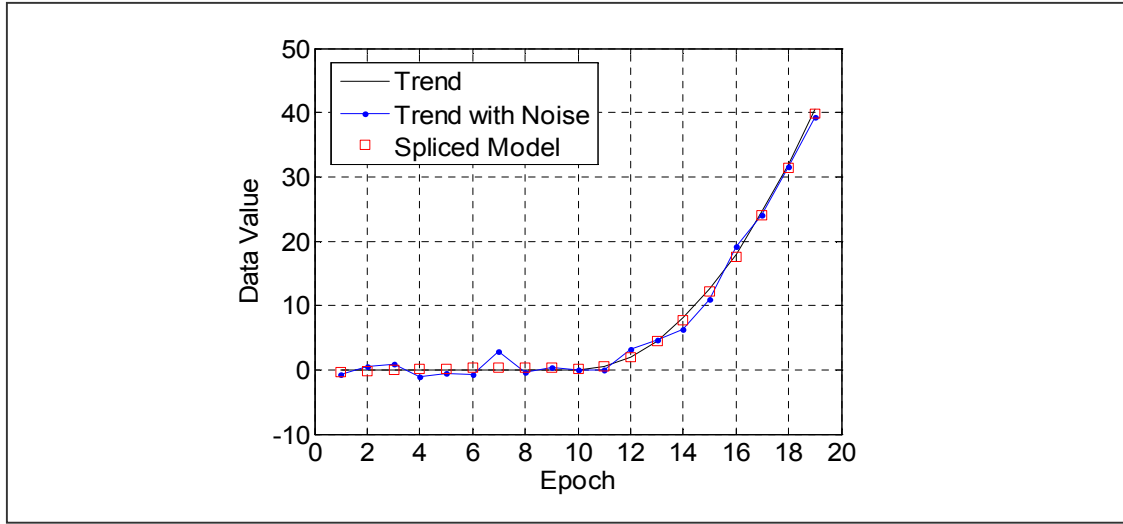
Figure 3.20: How Spliced Regression Works

This is the initial matrix for spliced regression, assuming an odd number of regression points. An even number would shift the values slightly, but not fundamentally change the matrix.

$$M_N = \begin{pmatrix} 1 & -(1-n) & -\frac{1}{2}(1-n)^2 & 0 \\ \vdots & \vdots & \vdots & \vdots \\ 1 & -1 & -\frac{1}{2}1^2 & 0 \\ 1 & 0 & 0 & 0 \\ 1 & 1 & 0 & \frac{1}{2}1^2 \\ \vdots & \vdots & \vdots & \vdots \\ 1 & (n-1) & 0 & \frac{1}{2}(n-1)^2 \end{pmatrix}, or\, M_{10} = \begin{pmatrix} 1 & -9 & -\frac{1}{2}9^2 & 0 \\ \vdots & \vdots & \vdots & \vdots \\ 1 & -1 & -\frac{1}{2}1^2 & 0 \\ 1 & 0 & 0 & 0 \\ 1 & 1 & 0 & \frac{1}{2}1^2 \\ \vdots & \vdots & \vdots & \vdots \\ 1 & 9 & 0 & \frac{1}{2}9^2 \end{pmatrix} \qquad (3.32)$$

The figures in this chapter use the abbreviations "CL" for classic filter, "SP" for spliced filter, and "HY" for hybrid filter, which is the linear combination of two spliced filters. The 10 point spliced filter (10 SP) response is strictly greater than the 10 point

64

classic filter (10 CL) response when there is no correlation on the input noise. This is not true when there is a high correlation. Figure 3.21 shows that when $\tau = 0.9$, the 10 point classic filter response is initially faster than the 10 point spliced filter. The 10 point spliced filter ultimately has a higher gamma due to its sensitivity though. In order to make a robust argument for a better acceleration detector, the objective is to design a filter which performs better over the entire span of τ correlation values.
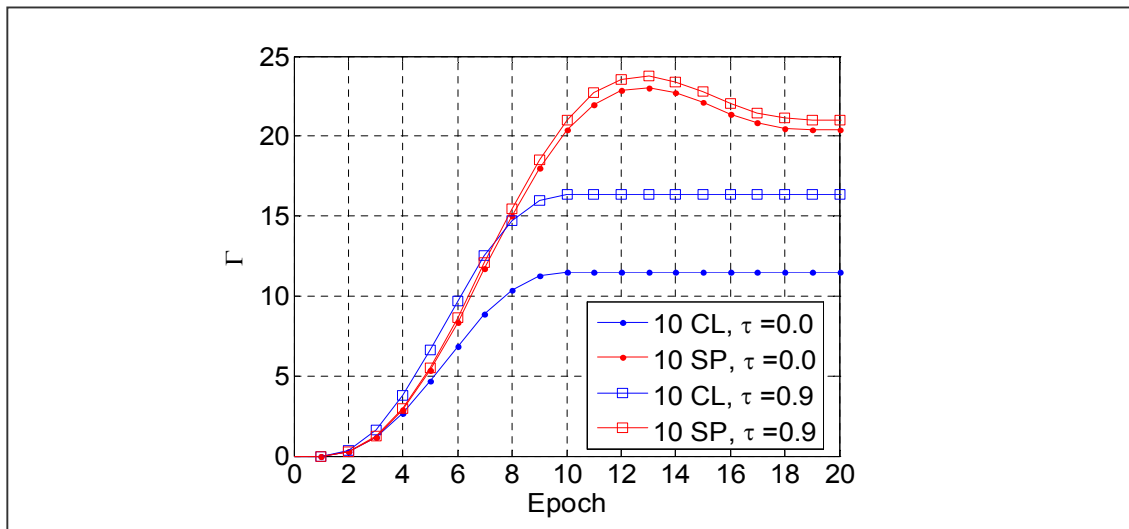
Figure 3.21: Spliced Filter Response Lags the Classic Filter when $\tau = 0.9$

3.2.8.3 Offsetting the Spliced Regression

One of the fundamental notions of data filtering is that increasing the number of samples will slow the response time. The intent of the spliced regression was to allow the first acceleration filter to respond quickly and let the second reduce the noise. This type of filter conforms to the above rule when the input noise is highly correlated. However, it needn't be so. By offsetting the number of points of the acceleration filters, this rule can be broken. The rapid response filter is shortened and the noise attenuating filter is lengthened. So instead of two 10 points filter, a 9 point and 11 point are spliced, or an 8 point and 12 point, etc. These would use the notation 10/0 (or 10 SP), 10/1, and

10/2 respectively to indicate the nominal length of each filter and the offset of the splicing point from the center.

$$M_{N/S} = \begin{pmatrix} 1 & -(1-n+s) & -\frac{1}{2}(1-n+s)^2 & 0 \\ \vdots & \vdots & \vdots & \vdots \\ 1 & -1 & -\frac{1}{2}1^2 & 0 \\ 1 & 0 & 0 & 0 \\ 1 & 1 & 0 & \frac{1}{2}1^2 \\ \vdots & \vdots & \vdots & \vdots \\ 1 & (n-1+s) & 0 & \frac{1}{2}(n-1+s)^2 \end{pmatrix}, or\; M_{10/2} = \begin{pmatrix} 1 & -7 & -\frac{1}{2}7^2 & 0 \\ \vdots & \vdots & \vdots & \vdots \\ 1 & -1 & -\frac{1}{2}1^2 & 0 \\ 1 & 0 & 0 & 0 \\ 1 & 1 & 0 & \frac{1}{2}1^2 \\ \vdots & \vdots & \vdots & \vdots \\ 1 & 11 & 0 & \frac{1}{2}11^2 \end{pmatrix} \quad (3.33)$$

The design objective for this thesis was to design a nominal response curve which was strictly better than the 10 point classic filter response, and then verify that the sigma was lower for all correlation values between zero and unity. Because these acceleration estimators are FIRs, they can be linearly combined to extract the optimal balance of their properties. Different combinations were tried using this blending, and the best performing among them is represented as the "Hybrid" Filter.

Long Filters

These are the response curves comparing the standard classic filter, two spliced filters, and two offset spliced filters. The 10 point spliced filter matches the response at the filter length 10 epochs, but it is slower to respond prior. A 7 point spliced filter is also shown and demonstrates a faster response than the classic filter with nearly identical noise performance, as shown in the table below. Using the eight epoch response time as a target, the 10/2 and 9/1 offset filter responses are also shown. The 9/1 offset filter is always faster than the 10 point classic filter and the 10/2 offset filter response is marginally less than the 10 point classic filter. The responses of several candidate "long" filters are shown in Figure 3.22a, along with a direct difference from the 10 point classic filter in Figure 3.22b. Table 3.1 shows the standard deviation of the noise transmitted by the filters.

66
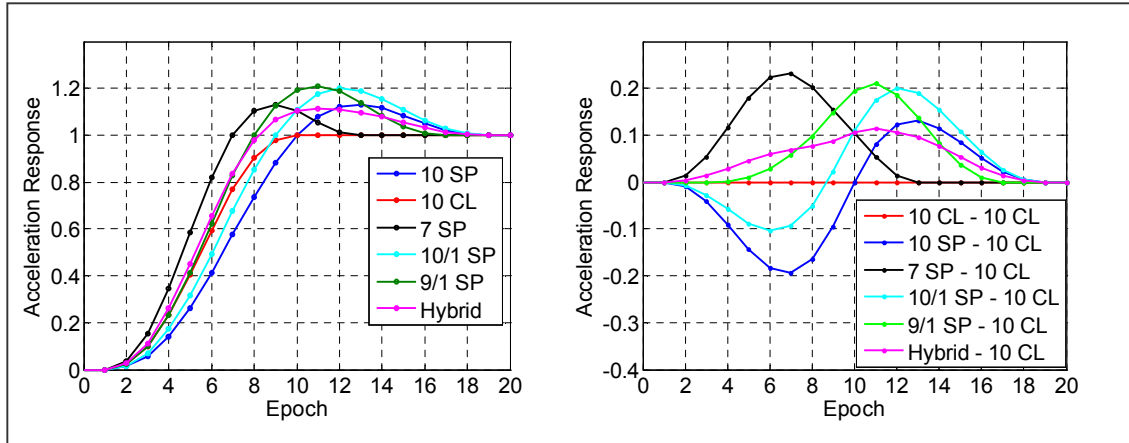
Figure 3.22: a) Response of Acceleration Filters, b) Response Differenced from 10 CL

Method	10 CL	10 SP	8 SP	7 SP	10/1	10/2	9/1	$7_{30}+10/1_{70}$
Std. Dev	0.0870	0.0491	0.0893	0.1285	0.0592	0.0735	0.0803	0.0705

Table 3.1: MQM ACC Methods and their Standard Deviations

Short Filters

Previous analysis showed that the performance benefit of longer filters is diminished when the input noise is correlated. Consequently it is pertinent to examine the response of much shorter filters, down to even three points. With much shorter filters it is assumed that the noise will increase appreciably, however using the spliced method may attenuate the noise without compromising the response. The responses of several candidate "short" filters are shown in Figure 3.23a, along with a direct difference from the 3 point classic filter in Figure 3.23b. Table 3.2 shows the standard deviation of the noise transmitted by the filters.
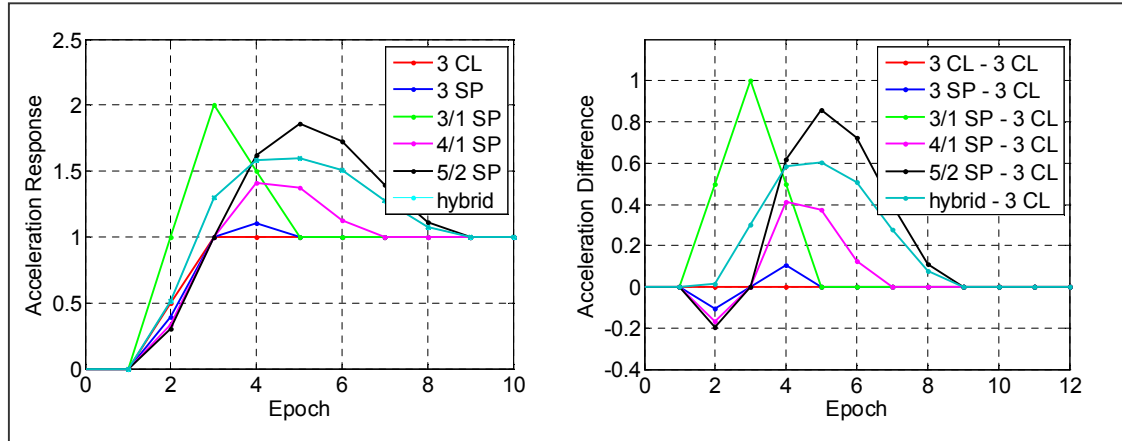
Figure 3.23: a) Response of Acceleration Filters, b) Response Differenced from 3 CL

Method	3 C	3 SP	3/1 OSP	4/1 OSP	5/2 OSP	$3/1_{30}+5/2_{70}$
Std Dev	2.4495	1.6690	5.0990	1.2003	0.9967	1.9278

Table 3.2: MQM ACC Short Filters and their Std. Deviations

Figure 3.24a and Figures 3.24b show each candidate filter's response over the span of correlations. For the 3 point classic filter, there is a monatomic decrease of the filter's noise power with input noise correlation. The response of the long MQM filters (Figure 3.22a) shows why the 9/1 offset filter wasn't used. At higher correlations, the noise standard deviation surpasses the 10 point classic filter. The long hybrid filter is always below the 10 point classic filter noise values, even though it had faster performance. Ultimately, with respect to an acceleration input, it has been demonstrated that the new filter can be more sensitive to a signal and also have a faster response than a classic filter even if it is longer than that classic filter, and this is true despite the potential correlation on the input noise.
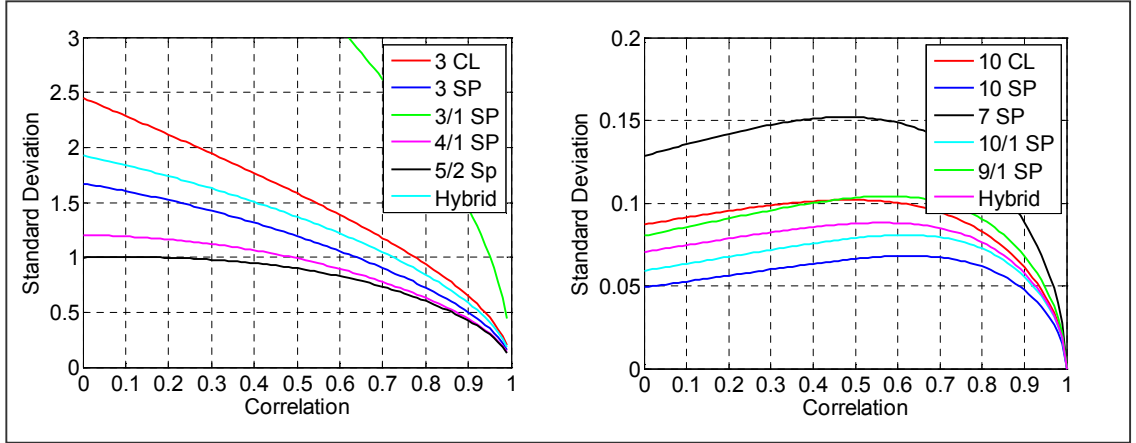
Figure 3.24: Noise Correlation Effect for a) Short, b) Long MQM ACC Filters

When the Gamma values are plotted for these candidate filters (Figure 3.25a and Figure 3.25b), it is clear that the short hybrid filter is superior to the short classic filter. For the longer filters, the spliced method is superior for low noise correlation, but only the hybrid filter's response is strictly greater than the classic filter when considering non-zero correlation.
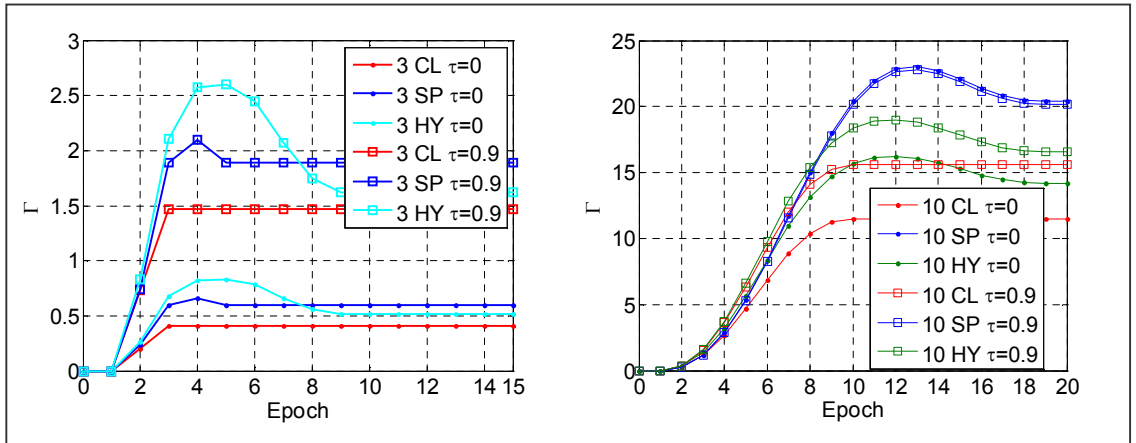
Figure 3.25: Step Responses for a) Short, b) Long MQM ACC Filters

Figure 3.26 shows the direct comparison of a long hybrid filter to a short hybrid filter. For uncorrelated noise, the longer hybrid has a better short term response and clearly a bigger long term response. When the noise is correlated, the shorter filter responds more quickly, but is overshadowed by the long term distinction between the filters.
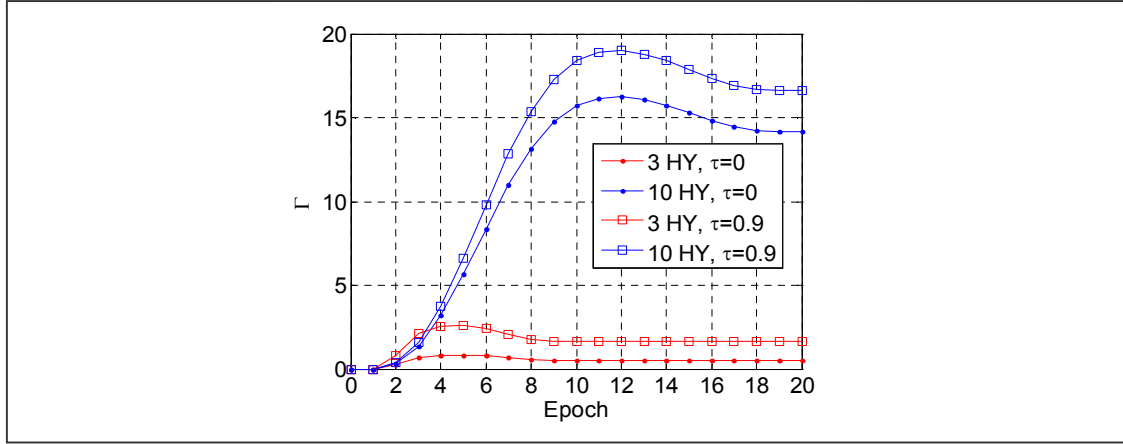
Figure 3.26: Short Hybrid vs. Long Hybrid, under Two Correlations

3.2.8.4 Non-Polynomial Design

At this point, the method of the IMT MQM has been explained and several candidate acceleration filters have been derived and examined using the same fundamental polynomial method the IMT used. There is no requirement to use a polynomial to derive a detection filter/correlator, and that is the focus of the next few sections.

The first intuition for designing the MQM is to use a 2nd order polynomial and take each respective coefficient as an estimate of position, velocity, and acceleration. In the purest form, and when the indices were symmetric, each filter was just a correlator. The velocity filter was a straight line, the acceleration filter was a parabola, although the position filter was also a parabola and not a flat line but this is a consequence of minimizing the sum of squared errors, as intended. The velocity and acceleration filters may be designed separately. The M^t matrix from Equation (3.6) represented all the PVA (position-velocity-acceleration) filters, but a parallel linear algebra form can be created. The values of alpha are the specific coefficients.

$$A = \begin{bmatrix} \alpha_1, & \alpha_2, & \cdots, & \alpha_n \end{bmatrix} \qquad (3.34)$$

70

The first row of the matrix in Equation (3.35) mandates that the response to the mean be zero, the second row mandates the response to velocity be zero, and the third row says the response to an acceleration be one.

3 Point Filter

$$
\begin{pmatrix} 1 & 1 & 1 \\ 3 & 2 & 1 \\ \frac{1}{2}\cdot 3^2 & \frac{1}{2}\cdot 2^2 & \frac{1}{2}\cdot 1^2 \end{pmatrix} \cdot \begin{pmatrix} \alpha_1 \\ \alpha_2 \\ \alpha_3 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix} \quad \rightarrow \quad \begin{pmatrix} \alpha_1 \\ \alpha_2 \\ \alpha_3 \end{pmatrix} = \begin{pmatrix} 1 \\ -2 \\ 1 \end{pmatrix} \qquad (3.35)
$$

4 Point Filter

$$
M_4 \cdot \vec{\alpha} = \begin{pmatrix} 1 & 1 & 1 & 1 \\ 4 & 3 & 2 & 1 \\ 8 & \frac{9}{2} & 2 & \frac{1}{2} \\ ? & ? & ? & ? \end{pmatrix} \cdot \begin{pmatrix} \alpha_1 \\ \alpha_2 \\ \alpha_3 \\ \alpha_4 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 1 \\ ? \end{pmatrix} \qquad (3.36)
$$

Even though the coefficients of the rows reflect different orders of a polynomial, the issue here is that a polynomial is being fit to a finite segment of data. The filter is being designed to have certain response characteristics to particular inputs. Some of those inputs take the form of low order polynomials.

If the last equation/row of Equation (3.36) is open, the first three alpha terms will be solved to minimize the squares error with respect to the model fit. This last equation can be used to "splice" the regression, to prescribe a quick response as well.

$$
\begin{pmatrix} 1 & 1 & 1 & 1 \\ 4 & 3 & 2 & 1 \\ 8 & \frac{9}{2} & 2 & \frac{1}{2} \\ 2 & \frac{1}{2} & 0 & 0 \end{pmatrix} \cdot \begin{pmatrix} \alpha_1 \\ \alpha_2 \\ \alpha_3 \\ \alpha_4 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 1 \\ 1 \end{pmatrix} \qquad (3.37)
$$

This form forces a short/long response, once at epoch two and once at epoch four. That fourth equation can also be used to maximize the transient response. The first several values of the acceleration are shown in Table 3.3.

Epoch	0	1	2	3	4	5	6	7
Acc	0	0.5	2	4.5	8	12.5	18	24.5

Table 3.3: Acceleration Values at each Epoch, Four-Point Non-Polynomial Design

The filter output for the first six epochs of a four epoch acceleration filter in response to an acceleration is given by the Equation (3.38).

$$\begin{pmatrix} A_1 \\ A_2 \\ A_3 \\ A_4 \\ A_5 \\ A_6 \end{pmatrix} = \begin{pmatrix} 0.5 & 0 & 0 & 0 \\ 2.0 & 0.5 & 0 & 0 \\ 4.5 & 2.0 & 0.5 & 0 \\ 8.0 & 4.5 & 2.0 & 0.5 \\ 12.5 & 8.0 & 4.5 & 2.0 \\ 18.0 & 12.5 & 8.0 & 4.5 \end{pmatrix} \cdot \begin{pmatrix} \alpha_1 \\ \alpha_2 \\ \alpha_3 \\ \alpha_4 \end{pmatrix} \tag{3.38}$$

If the desire is to maximize the total output of the first four epochs, ‖A‖, then it must be done while also solving the alpha coefficients to satisfy the underlying linear equations. In reality, the intent is to maximize some ratio Γ^*. Previously, Γ (defined by Equation 3.31) was the response at some epoch N divided by the square root of the filter's noise. Γ^* is generalized as a ratio to express the magnitude of some evaluating function scaled by its noise contribution. Equation (3.39) uses the term "f[]" to describe the evaluating function, which may express the response at a given epoch, the cumulative response up to a certain epoch, or something else.

$$\Gamma^* = \frac{f[A]}{\|\bar{\alpha}\|} \tag{3.39}$$

Equation (3.40) creates a function which maximizes the sum of the first four responses of the input to a pure acceleration correlator. This can be seen in Equation (3.41). Row three of the matrix, {8, 4.5, 2, 0.5}, is equal to one half of the first four integers squared. Moving from right to left, row four is the cumulative sum of row three.

$$f[A] = \sum_{i=1}^{4} A_i = \begin{pmatrix} 15 & 7 & 2.5 & 0.5 \end{pmatrix} \cdot \begin{pmatrix} \alpha_1 \\ \alpha_2 \\ \alpha_3 \\ \alpha_4 \end{pmatrix} \tag{3.40}$$

Equation (3.41) gives an example how an open row in the regression matrix acts as a degree-of-freedom in designing the response of the filter. Using the fourth equation to represent the transient response, Equation (3.41) expresses a formula which seeks to maximize k, and thereby maximizing Γ^*.

$$M_4 \cdot \vec{\alpha} = \begin{pmatrix} 1 & 1 & 1 & 1 \\ 4 & 3 & 2 & 1 \\ 8 & \frac{9}{2} & 2 & \frac{1}{2} \\ 15 & 7 & 2.5 & 0.5 \end{pmatrix} \begin{pmatrix} \alpha_1 \\ \alpha_2 \\ \alpha_3 \\ \alpha_4 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 1 \\ k \cdot \|\vec{\alpha}\| \end{pmatrix} \tag{3.41}$$

This particular problem is not that difficult because of the degrees-of-freedom. Only one truly exists. Since $\vec{\alpha}$ is orthogonal to the first two vectors, M_4 can be reduced. The vectors $\{r_1, \ldots, r_4\}$ in Equation (3.42) represent the rows.

$$M_4^O \cdot \vec{\alpha} = \begin{pmatrix} r_1 \\ r_2 \\ r_3 \\ r_4 \end{pmatrix} \cdot \vec{\alpha} = \begin{pmatrix} 1 & 1 & 1 & 1 \\ 1.5 & 0.5 & -0.5 & -1.5 \\ 0.5 & -0.5 & -0.5 & 0.5 \\ 1.45 & -1.35 & -1.65 & 1.55 \end{pmatrix} \begin{pmatrix} \alpha_1 \\ \alpha_2 \\ \alpha_3 \\ \alpha_4 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 1 \\ k \cdot \|\alpha\| \end{pmatrix} \tag{3.42}$$

This can be solved using the premise from Cauchy-Schwartz (Equation 3.43) that the maximum dot product of one vector with another, normalized, vector is when the first

vector is a scaled multiple of the second. Figure 3.27 provides a useful rendering of the formulation of this problem. In the figure, the term $\{r_1, r_2\}$ is geometrically representing two dimensions in one dimension of the graph. The term *NS* is the Null Space of vectors $\{r_1, r_2, r_3\}$

$$Max\left[\frac{a \cdot b}{\|a\|}\right] \rightarrow (a = k \cdot b) \tag{3.43}$$
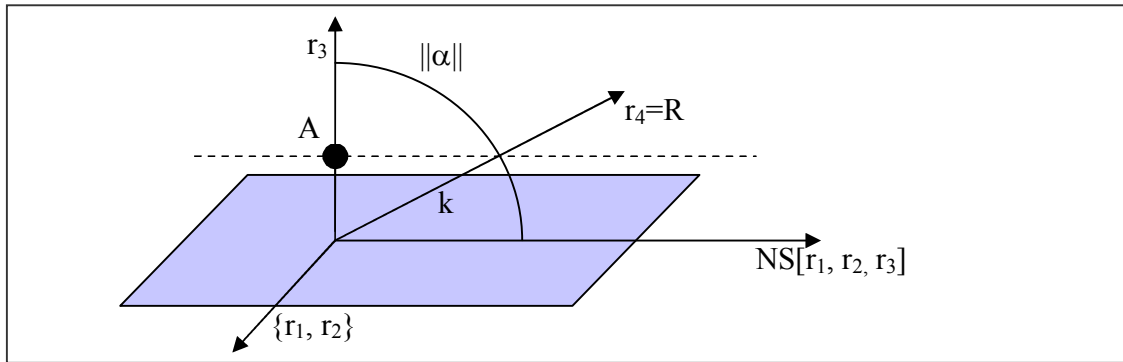
Figure 3.27: Solving for *k* from the Acceleration (*A*), and the Response (*R)*

In Figure 3.27, the vectors r_1 and r_2 are orthogonal to the acceleration estimation, the vector r_3 represents the measured acceleration, and the vector r_4 represents the value to be maximized.

$$[A \cdot r_4 = 3] \rightarrow (k = \tfrac{1}{3}) \rightarrow \alpha = \begin{bmatrix} 0.4833 \\ -0.4500 \\ -0.5500 \\ 0.5167 \end{bmatrix} \tag{3.44}$$

The realized improvement compared to a classic four-point filter is minimal because the degrees-of-freedom were very limited. If more points were allowed, the improvement can become meaningful. If seven points are used, as many as the 4-SP method, the formulation in Equation (3.45) is used. The last equation/row maximizes the response at the fourth epoch, instead of maximizing the total of the first four epochs.

Since the first one or two epochs will logically be insignificant next to the third and fourth and there is no real need to emphasize those. This construction is only slightly different than the short/long response, because here the short response is being maximized, not set to unity.

$$M_7 \cdot \bar{\alpha} = \begin{pmatrix} 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 7 & 6 & 5 & 4 & 3 & 2 & 1 \\ 24.5 & 18 & 12.5 & 8 & 4.5 & 2 & 0.5 \\ 4.5 & 2 & 0.5 & 0 & 0 & 0 & 0 \end{pmatrix} \cdot \begin{pmatrix} \alpha_1 \\ \alpha_2 \\ \alpha_3 \\ \alpha_4 \\ \alpha_5 \\ \alpha_6 \\ \alpha_7 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 1 \\ k \cdot \|\alpha\| \end{pmatrix} \qquad (3.45)$$

$$M_7^O \cdot \bar{\alpha} = \begin{pmatrix} 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 3 & 2 & 1 & 0 & -1 & -2 & -3 \\ 2.5 & 0 & -1.5 & -2 & -1.5 & 0 & 2.5 \\ -0.93 & -0.29 & -0.36 & -1 & -1.14 & -0.29 & 1.57 \end{pmatrix} \cdot \begin{pmatrix} \alpha_1 \\ \alpha_2 \\ \alpha_3 \\ \alpha_4 \\ \alpha_5 \\ \alpha_6 \\ \alpha_7 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 1 \\ k \cdot \|\alpha\| \end{pmatrix} \qquad (3.46)$$

$$[r_3 \cdot r_4 = 5.85] \rightarrow (k = 0.171) \rightarrow \bar{\alpha} = \begin{bmatrix} -0.1590 \\ -0.0496 \\ -0.0615 \\ -0.1709 \\ -0.1949 \\ -0.0496 \\ 0.2684 \end{bmatrix} \qquad (3.47)$$

Squared Condition

Linear combinations are easiest, but a squared version can also be solved, since this is most often associated with power. Going back to the example of a four point acceleration filter, the evaluating function can be made to be the sum of the square of the first four responses, expressed in Equation (3.48).

$$f[A] = \sum_{i=1}^{4} A_i^2$$
$$= \left(0.5\alpha_1\right)^2 \ldots$$
$$+ \left(2\alpha_1 + 0.5\alpha_2\right)^2 \ldots$$
$$+ \left(4.5\alpha_1 + 2\alpha_2 + 0.5\alpha_3\right)^2 \ldots \qquad (3.48)$$
$$+ \left(8\alpha_1 + 4.5\alpha_2 + 2\alpha_3 + 0.5\alpha_4\right)^2$$

The first three equations/rows of the matrix (see Equation 3.39) act as constraints leaving:

$$f[A] = 3.5 - 2 \cdot \alpha_4 + 0.75 \cdot \alpha_4^2 \qquad (3.49)$$

$$\|\bar{\alpha}\| = 6 - 20 \cdot \alpha_4 + 20 \cdot \alpha_4^2 \qquad (3.50)$$

$$\Gamma^* = \frac{f[A]}{\|\bar{\alpha}\|} = \frac{3.5 - 2\alpha_4 + 0.75\alpha_4^2}{6 - 20\alpha_4 + 20\alpha_4^2} \qquad (3.51)$$

$$\Gamma^*_{Max} \rightarrow \left(\alpha_4 = 0.488\right) \rightarrow \bar{\alpha} = \begin{bmatrix} 0.5118 \\ -0.5353 \\ -0.4647 \\ 0.4882 \end{bmatrix} \qquad (3.52)$$

This was rather simple because the three constraints only left one variable in the second order Γ^* function. What if many more variables are present, such as in the seven epoch filter? The equations become cumbersome, but are easily solved using a multivariable Newton-Raphson, executable with a mathematics computer program.

$$\Gamma^* = \frac{f[A]}{\|\bar{\alpha}\|} = \frac{\sum_{i=1}^{7} A_i^2}{\sum_{i=1}^{7} \alpha_i^2} \qquad (3.53)$$

$$f[A] = (0.5\alpha_1)^2 +$$
$$(2.0\alpha_1 + 0.5\alpha_2)^2 +$$
$$(4.5\alpha_1 + 2\alpha_2 + 0.5\alpha_3)^2 +$$
$$(8.0\alpha_1 + 4.5\alpha_2 + 2\alpha_3 + 0.5\alpha_4)^2 + \quad\quad (3.54)$$
$$(12.5\alpha_1 + 8.0\alpha_2 + 4.5\alpha_3 + 2.0\alpha_4 + 0.5\alpha_5)^2 +$$
$$(18.0\alpha_1 + 12.5\alpha_2 + 8\alpha_3 + 4.5\alpha_4 + 2.0\alpha_5 + 0.5\alpha_6)^2 +$$
$$(24.5\alpha_1 + 18.0\alpha_2 + 12.5\alpha_3 + 8\alpha_4 + 4.5\alpha_5 + 2.0\alpha_6 + 0.5\alpha_7)^2$$

$$\bar{\alpha} = \begin{bmatrix} 0.1248 & -0.0049 & -0.0780 & -0.0965 & -0.0652 & 0.0788 & 0.1120 \end{bmatrix}^T \quad (3.55)$$

This method was used to solve the squared f[A] operator using up to 19 variables. This means that a solution can be achieved for acceleration filters of the length discussed in the previous sections. Furthermore, this method is applicable to other filters, such as the Velocity Estimator, but that has not been examined in this thesis. With regard to the acceleration filter, this method only refines the acceleration filter, it doesn't substantially improve the performance, but it does emphasize that the original polynomial method of the IMT is only a conception and not a requirement. Breaking from idea that filter of the MQM must be constructed from polynomials opens up entirely new design methods. One more method for improvement is examined next.

3.2.8.5 Pole Placement

Trackers try to match their outputs to their inputs with complete fidelity. Regulators try to maintain a particular state despite input disturbances. These filters attempt to maximize sensitivity to inputs and minimize sensitivity to noise, essentially a tracker. The acceleration input is nearly a triple impulse integration, and the three point acceleration filter is double differencing, or a double derivative (in discrete time) thus the output resembles a step function. Utilizing poles, an IIR (Infinite Impulse Response) filter can be designed with the intent of creating a resonance which will amplify the

transient response and ideally increase the probability of detection. This idea comes from trying to exploit the overshoot seen in the Spliced Regression method. The general forms of a classic FIR and IIR filters are given below in Equations (3.56) and (3.57)

$$R_{3,FIR} = \frac{z^2 - 2z + 1}{z^2} \tag{3.56}$$

$$R_{3,IIR} = \frac{z^2 - 2z + 1}{z^2 - 2\varsigma\omega z + \omega^2} \tag{3.57}$$

Considering the steady state response, the discrete time final value theorem states,

$$H(z) = (1 - z) \cdot H(z) \tag{3.58}$$
$$_{t \to \infty} \qquad\qquad _{z \to 1}$$

Thus the IIR filter will be.

$$R_{3,IIR} = (z-1)\frac{z^2 - 2z + 1}{z^2 - 2\varsigma\omega z + \omega^2} \cdot \frac{1}{2}\frac{z^2 + z}{(z-1)^3}\bigg|_{z=1} = \frac{1}{1 - 2\varsigma\omega + \omega^2} \tag{3.59}$$
$$_{t \to \infty}$$

Thus the (final) response can grow quite large by making the denominator quite small. This is a long term effect and doesn't affect the rapid response so much. An example of an IIR filter is:

$$R_{3,IIR} = \frac{z^2 - 2z + 1}{z^2 - 1.8z + 0.9} \tag{3.60}$$

Although this filter emphasizes a particular frequency band, it will be referred to by its broader classification of a low pass filter (LPF). The objective of the filter is to detect a sustained acceleration event; however, the classic FIR filters are no more sensitive to an acceleration which has existed for a minute than one which has existed only a second. This seems like an under-utilization of the available data. The numerator

of the LPF shown before will cancel two integrations of the acceleration input, thus leaving a step function representing the acceleration. The denominator will then "pseudo-integrate" this remaining signal. Integrating the acceleration step function would really be deriving velocity (or position) and this would abandon any sense of temporal consistency. Furthermore, pure integration will cause the estimated noise value to increase as well. The LPF can mitigate these concerns. Three sample LPFs are given below; each represents a different aggressiveness of the integration. LPF1 uses imaginary poles. This helps the response because of the overshoot but because of the increased noise transmission, ultimately lowers the Gamma ratio.

$$R_{3,LPF1} = \frac{z^2 - 2z + 1}{z^2 - 1.8z + 0.9} \tag{3.61}$$

$$R_{3,LPF2} = \frac{z^2 - 2z + 1}{z^2 - 1.8z + 0.81} \tag{3.62}$$

$$R_{3,LPF3} = \frac{z^2 - 2z + 1}{z^2 - 1.4z + 0.45} \tag{3.63}$$
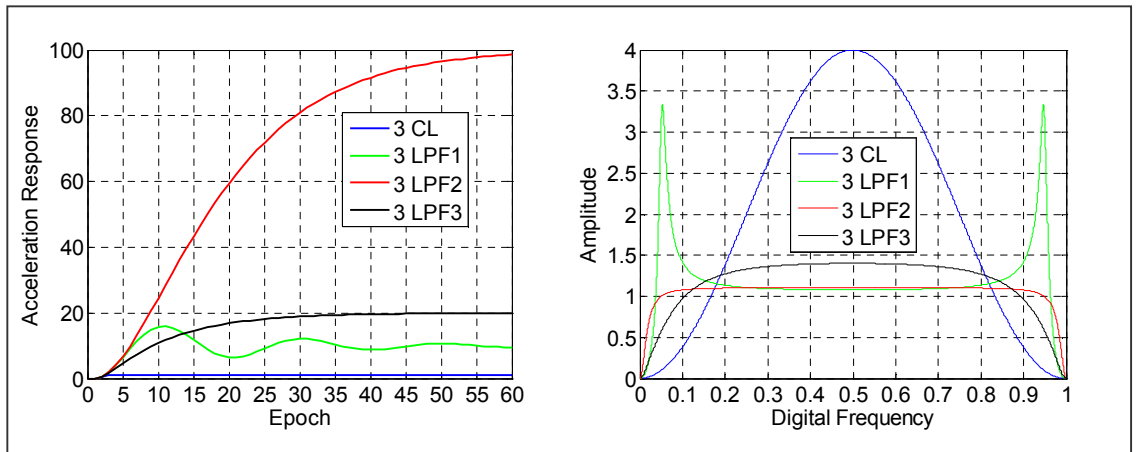
Figure 3.28: Low Pass Filter a) Time, and b) Frequency Response

Figure 3.28b shows that all LPF methods will have less noise under a white noise assumption. If the noise is colored the filters with the most pronounced resonant frequency may be the most effected. The truly interesting result is that the double pole filters are nearly cancelling the double derivative of the acceleration estimate and that

makes the output spectrum very flat. There is certainly dependency upon the input noise correlation, but any concern about the highly correlated output of such a filter can be assuaged.

Figure 3.29 is a plot of noise versus correlation. Obviously the filter with a resonant frequency is affected by colored noise. As long as the correlation is below 0.9, the second LPF has the best performance while always having a lower sigma. Just to provide an LPF with a strictly lower sigma, a fourth LPF was made which backed slightly off the aggressive double 0.9 pole, using instead a double 0.85 pole.

$$R_{3,LPF4} = \frac{z^2 - 2z + 1}{z^2 - 1.7z + 0.7225} \tag{3.64}$$
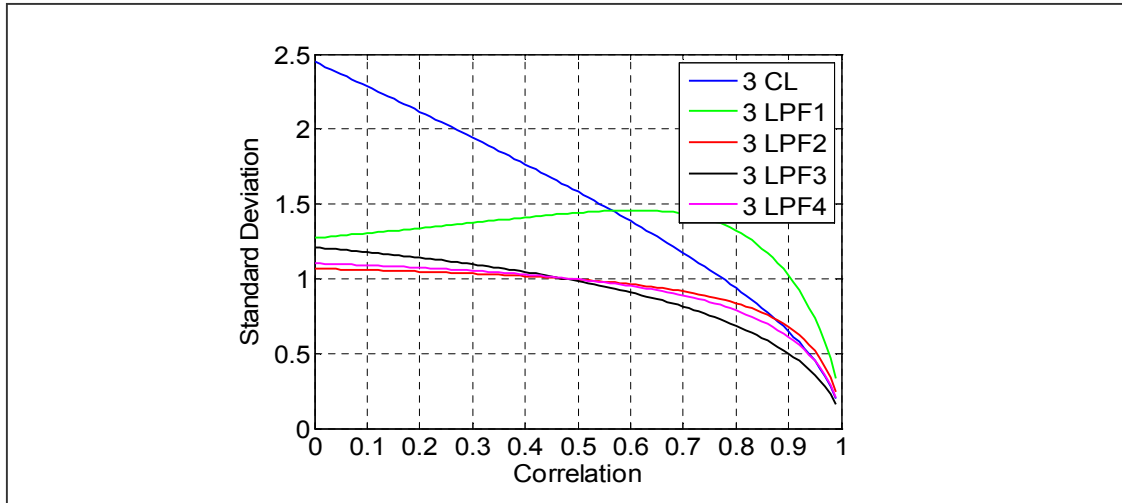
Figure 3.29: LPF Noise Standard Deviation vs. Noise Correlation

Figure 3.30a and Figure 3.30b use LPF4 and show quite clearly that it outperforms the best short and long hybrid filters. The hybrid filters were a combination of all the tricks developed for the FIR filters. They used a linear combination of offset-spliced regression filters.
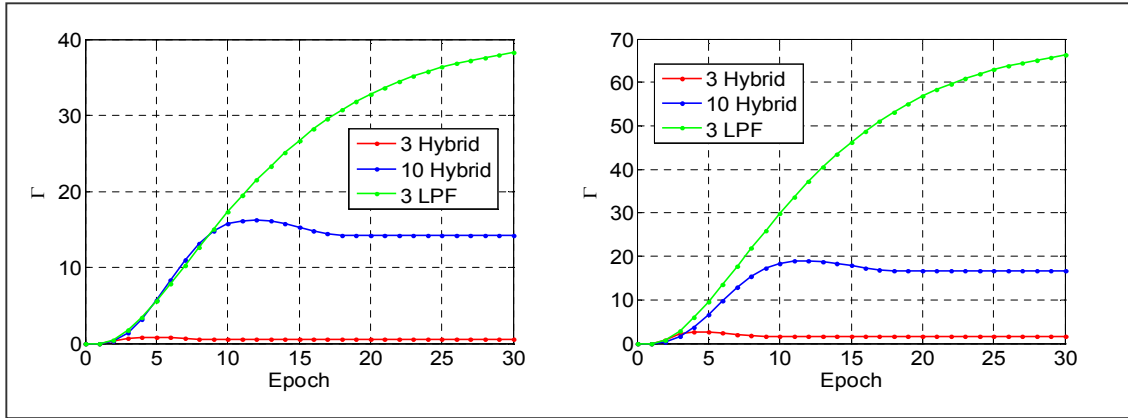
Figure 3.30: Gamma Values for Three Candidate Filters a) τ = 0, b) τ = 0.9

At this point it can be determined that the 3-point, or short filter, has such a substantially lower gamma value than the other methods in contention that it just cannot compete. Figure 3.31a and Figure 3.31b compare the gamma values of the three 10-point contenders (classic, splice, and hybrid) as well as the LPF4 from Equation (3.64).
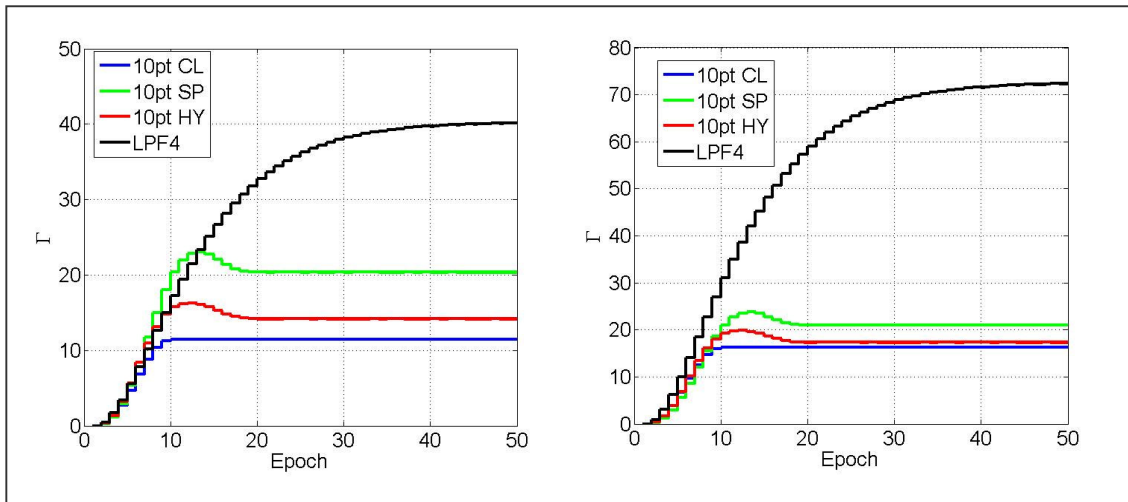
Figure 3.31: Gamma Values for Four Candidate Filters a) τ = 0, b) τ = 0.9

Figure 3.31a above shows that the LPF4 Filter has superior detection in the long term, and that the filters are rather similar in the short term. The filter response for correlated data (Figure 3.31b) shows that the 10 point hybrid filter has the strictly larger response it was designing to have, and the 10 point spliced filter has an initial response below that of the classic filter. Both the spliced and the hybrid filters show the response

hump characteristic of the splicing method common to both. If the LPF4 filter appears to have such a better response, are there any drawbacks of such a filter?

3.2.8.5.1 The Drawback of IIR Low Pass Filters

The potential drawback of using a low-pass filter which semi-integrates a persistent error is the ability to adapt to a reversing input. The following figure shows how the LPF4 Filter will continue to accumulate a constant acceleration and this leads to a potentially delayed detection when the acceleration is reversed. In the case shown in the left panel, the positive and negative accelerations are equal, and the even though the LPF lags the other filters, the argument is that whatever could be detected during the negative acceleration would have already been detected during the positive acceleration event and that channel would already have been flagged thus making the second detection a moot issue. However, in the right panel the positive acceleration is reduced in magnitude to only one quarter that of the negative acceleration. If the detection threshold was set to be be $\Gamma = 10$ (just above the LPF4 limit), there would be no detection during the positive acceleration and the LPF4 filter would be the last one to detect the event during the negative acceleration. Also of note is that the hybrid filter offers better detection time that the 10 point classic filter during the acceleration reversal. The spliced filter does lag the classic filter slightly, but only when the gamma values are small. When the gammas become large enough to initiate a new detection (i.e. larger in response than the positive acceleration), it has already caught the classic filter.
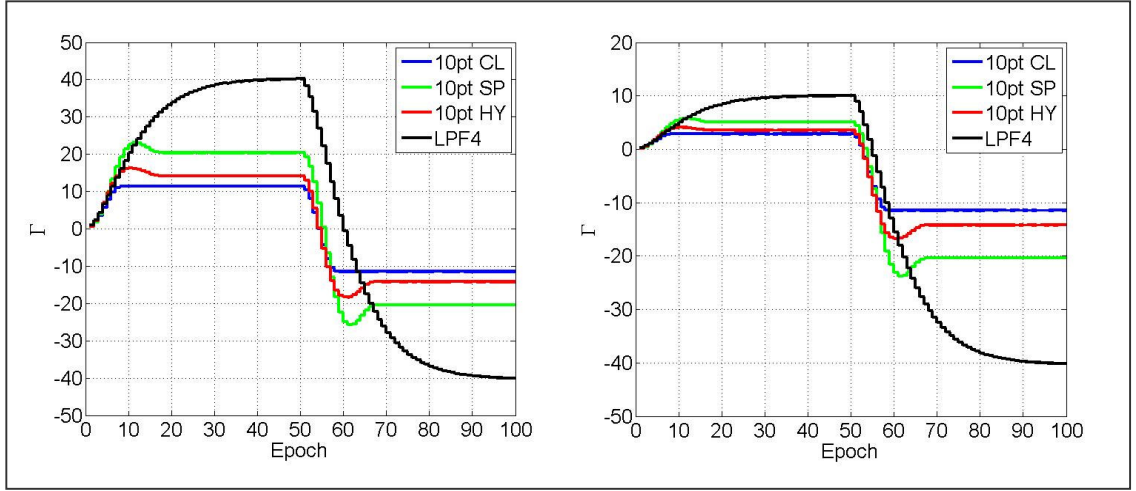
Figure 3.32: Oscillating Acceleration Shows Filter Lag a) τ = 0, b) τ = 0.9

Figure 3.32a and Figure 3.32b above illustrate that although the LPF filters show great promise in detecting small accelerations, it would be risky to rely solely on this method when much of the empirical justification for the IMT MQM has been done using the 10 point classic filter. It is more conservative to adopt the hybrid filter or the spliced filter to increase the performance of the MQM. Those are the two candidate filters which are examined next to the classic filter in the nest section of this chapter, which compares the actual detection rates of the candidate filters under different correlation values of the carrier-phase data.

3.2.8.6 Detection Rates

Here is a direct comparison of the ability of each of the three candidate filters to detect an input acceleration of a given size. That fault size is set to be k times the standard deviation of white noise through the classic filter, and detection occurs when the response reaches seven times the white noise standard deviation of the filter being tested. This makes the input acceleration constant across filters, but allows the thresholds to be scaled according to the particular filter as it would be in its implementation.

In Equation (3.65), n is the epoch number. In Equation (3.66), the threshold, T, is set to be seven time the standard deviation of white noise through that filter, σ_f.

Thresholds are typically set to be six to seven times the noise value and are empirically refined [3].

$$A = k \cdot \sigma_{10CL} \cdot \tfrac{1}{2} \cdot n^2 + \omega_{White} \qquad (3.65)$$

$$T_F = 7 \cdot \sigma_f \qquad (3.66)$$

In this test, each filter was evaluated at k values of {5, 10, 15, and 20}, with white noise, and the test was run for 100 million simulations. Figure 3.33 shows that the spliced and hybrid filters always outperform the classic filter. When $k = 5$, the classic filter cannot reasonably detect the acceleration (recall the threshold is set at $k = 7$). The spliced filter can detect this fault quite well because it is the most sensitive to small accelerations. And finally, the hybrid filter can detect the fault with confidence, but only after a prolonged time; to reach a probability of missed detection takes it slightly over 50 epochs, whereas the spliced filter takes less than 14 epochs. As the acceleration becomes larger the spliced and hybrid filters quickly begin to have very similar responses while the classic filter lags noticeably. At the largest tested input, $k = 20$, both the spliced and hybrid filters will have missed detection rates below 10^{-8} within six epochs. The classic filter takes an additional epoch to reach this confidence level.
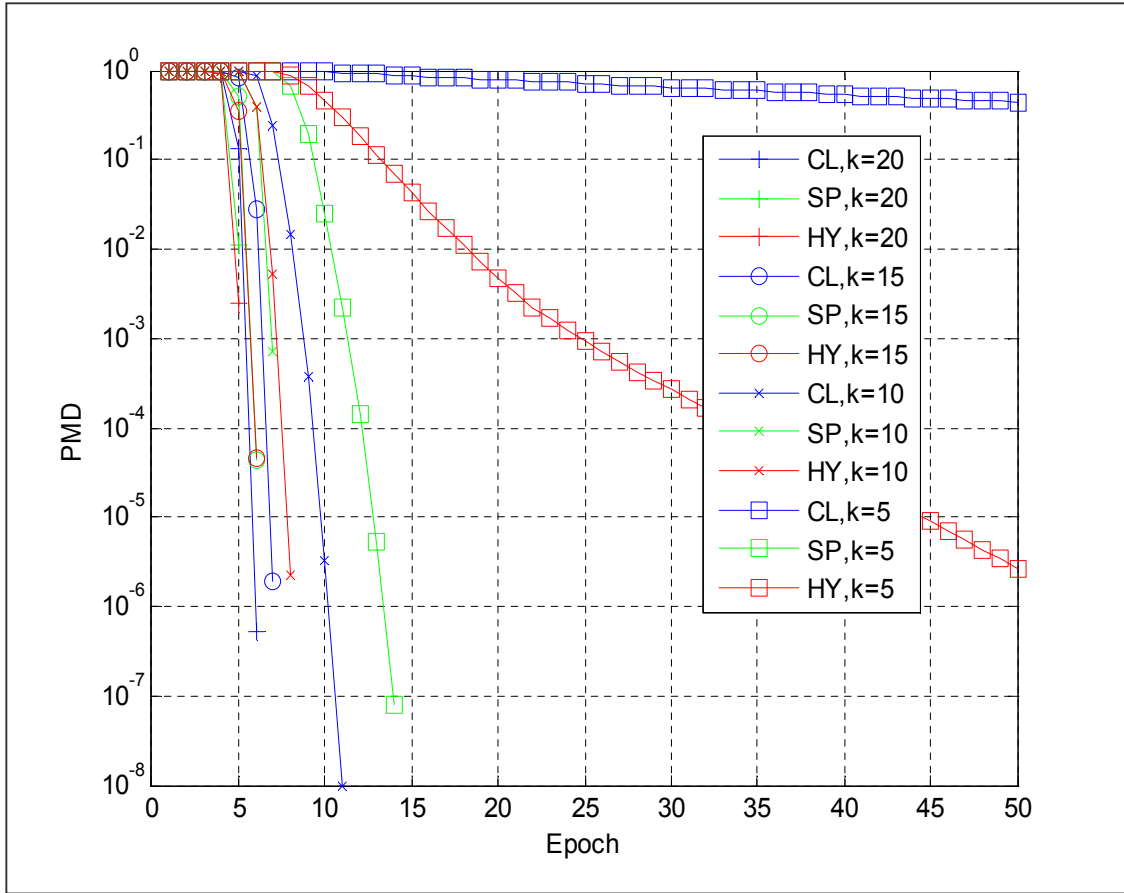
Figure 3.33: Detection Rates of Candidate Filters

3.3   MQM Conclusions

The MQM is one of the monitors which supply the data values for the EXM-I to make its determination of which channels should be used to calculate the corrections which are broadcast to the landing aircraft. There are many different MQM filters which can be used to monitor velocity and acceleration. They can be the result of direct design, or through the simultaneous design of a family of alternative filters.

Regarding acceleration - for large inputs, it may seem intuitive to use a quick responding short filter, but the increase in noise variance causes deleteriously high thresholds which jeopardize the system's ability to detect smaller errors. Another benefit of a longer filter is the reduction of a heavy tailed distribution into a more Gaussian form

via increased convolution of the data points. The results of this chapter show that the hybridized longer filter provides comprehensively better performance than the currently used 10-point classic filter. The recommendation of this thesis is to use the Hybrid Filter developed in section 3.2.8.3. This filter is described by the notation $(7_{30} + 10/1_{70})$ and the coefficients are shown in the impulse response of Figure 3.3a and compared to the current IMT filter.
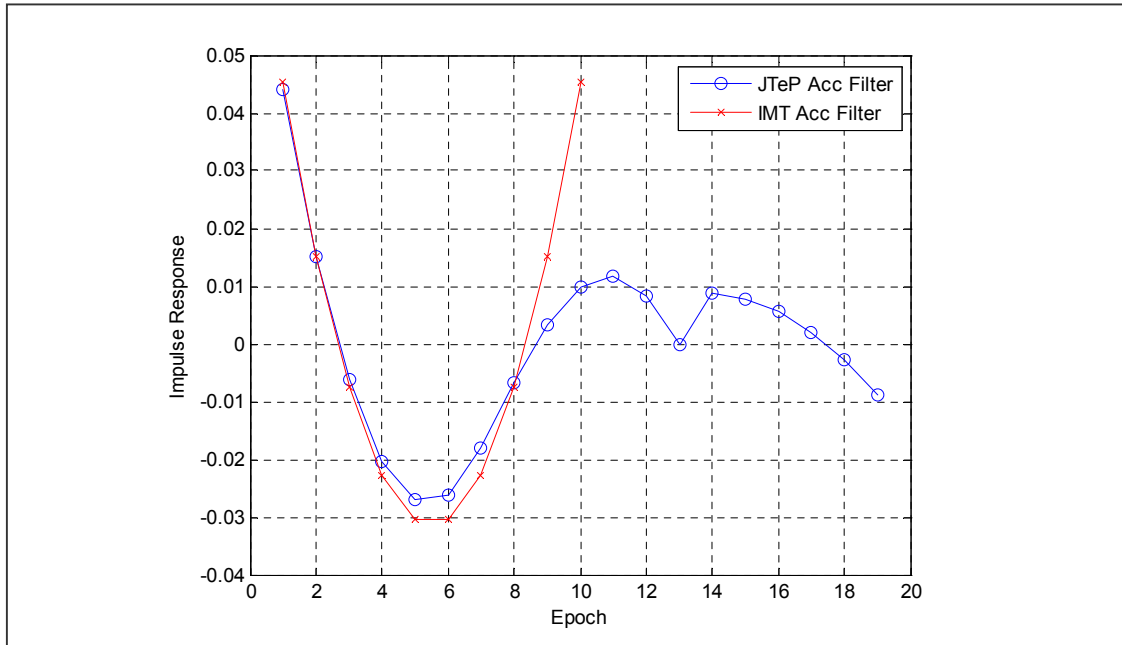
Figure 3.34: Hybrid Acceleration Filter Impulse Response

Lastly, Figure 3.34 shows the recommended JTeP Velocity Filter compared to the IMT Velocity Filter. Currently the IMT Velocity Filter is underperforming and can be significantly improved. The filter should switch from a backwards predicting filter to either a centralized or forward predicting filter to extract a better signal to noise ratio of the filter. Figure 3.35 shows the centralized Velocity Filter of the JTeP and clearly shows how the IMT Velocity Filter is correlated to the acceleration estimate and causes poor performance.
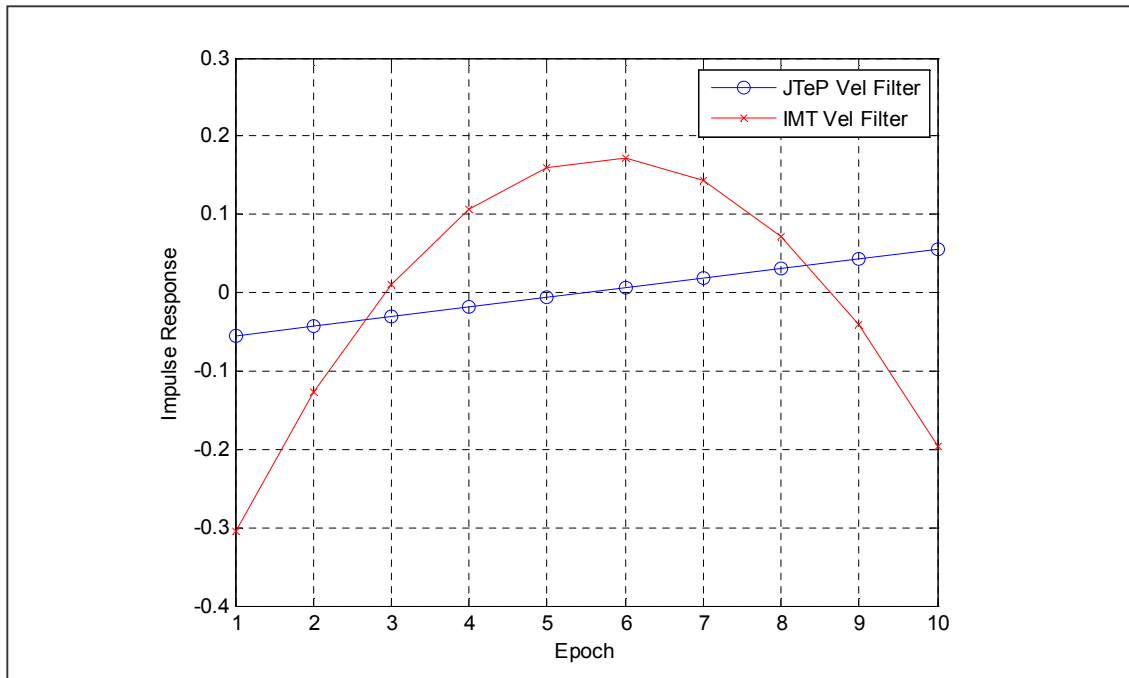
Figure 3.35: Velocity Filter Impulse Response

Chapter 4 - Executive Monitoring Exclusion

4.1   What is EXM Exclusion?

Executive Monitoring (EXM) is a set of algorithms which interprets the pass/fail outputs of various monitors of the larger system. The lower level monitors may test for any number of relatively microscopic problems such as excessive carrier-phase acceleration, or for too large a pseudo-range correction. The Stanford IMT uses a heuristic implementation of single-frequency EXM, but the focus of this chapter is to show how to optimize the EXM decision rules and extend this analysis to dual frequency in Chapter 5. The algorithm improvements in this chapter deal specifically with EXM-I (which interprets the output of the Quality Monitors, see Section 2.2, Figure 2.3), but can be applied to any algorithm which seeks to detect a fault common across multiple channels. As such, the term EXM is used throughout, except where it is appropriate to specifically refer to EXM-I and EXM-II.

We need to create of list of rules to specify when we think a component has failed, be it a satellite or a receiver. The failure of a channel is decided within each monitor, but how do we interpret the cumulative results? Is a satellite faulted if it has channel faults on zero, one, two, or three receivers? Flagging a satellite without any faults would obviously be wasteful (a continuity loss). Flagging a satellite with faults on one or two receivers seems reasonable. But waiting until a satellite is flagged on all three receivers may introduce an integrity concern. These issues reduce to the notion of alpha (false positive) & beta (false negative) errors, or the Type I and Type II errors of Figure 4.1.

		Decision	
		Pass	Fail
Truth	Good	*(Green)*	**Type I**
	Bad	**Type II**	*(Yellow)*

(Green)	Ideal
(Yellow)	Poor Performance
(Purple)	Loss of Continuity
(Red)	Loss of Integrity

Figure 4.1: Basic Premise in Decision vs. Truth

Ideally the system is always in the green zone, meaning that all requirements are met. When in the yellow zone, the system is unable to meet the requirements, but there is no danger, as we are observant and have removed the failed measurements or have taken the system offline. The purple zone is when the system is healthy, but we conclude that a flaw exists and remove one or more measurements, creating inefficiency. The red zone represents the worst scenario; the system is flawed yet we are unaware, and if the system were to be used, there may be a danger to equipment or even to human life.

Our world is not quite as simple as suggested in Figure 4.2 because there is an intermediate stage of observation. Or, more appropriately, since our system invokes redundancy to increase observability, a failure on one receiver or one satellite will most likely be visible on multiple channels. Observing the fault arising on each channel is the responsibility of the quality monitors. The responsibility of interpreting the findings of those monitors belongs to EXM. Figure 4.2 shows how a particular Quality Monitor will either pass or fail a satellite for its given metric. There are N pass/fail observations (where N may represent either the number of satellites, receivers, or frequencies), which EXM compiles into an n-tuple. Based on a decision rule, EXM will then either accept or reject that particular satellite, receiver, or frequency.
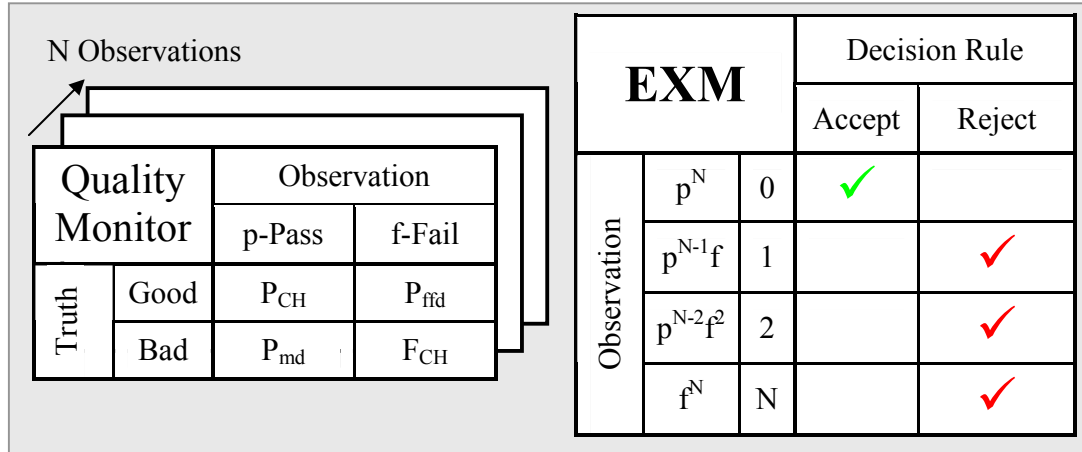
Figure 4.2: Advanced Premise in Decision vs. Truth

Figures 4.2 and 4.3 use the terminology from Section 1.2. P_{ffd} is the probability of a fault-free detection of a channel, whereas P_{FFD} is the probability of a fault-free detection of either a receiver or satellite. EXM will accept or reject a measurement based on the observations of the quality monitors. The workings of the observation monitors are critical, but the P_{MD} and P_{FFD} are strictly input/output relationships. In other words, if in truth the entity is bad and we accept it, there is a Missed Detection. If in truth the entity is good and we fail it, there is a Fault Free Detection. Figure 4.3 shows the propagation from truth to decision.
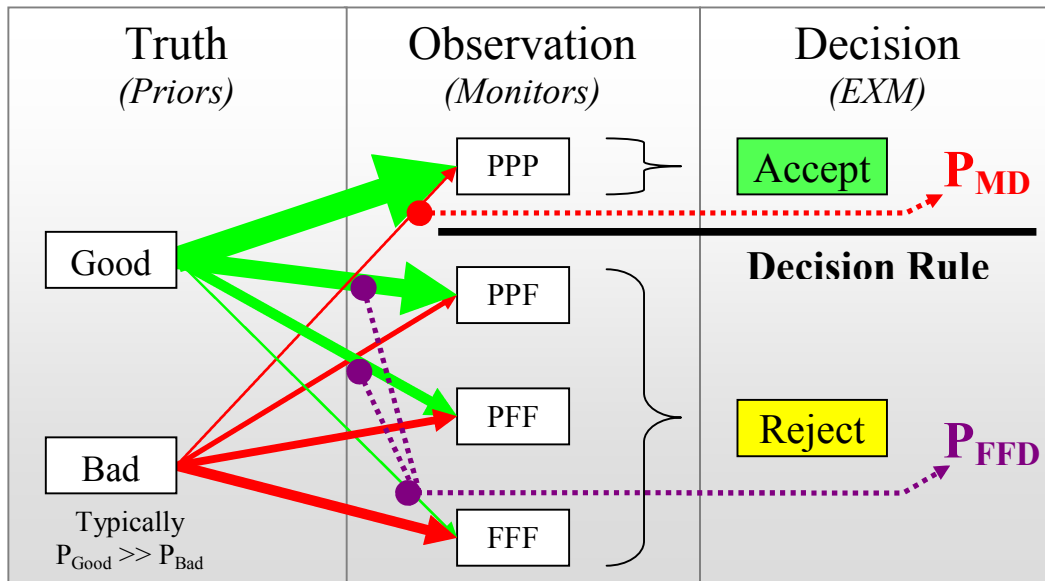
Figure 4.3: Flowchart of Truth to Decision, Three Observation Channels

The GPS integrity monitoring system analyzed in this chapter is the same system initially introduction in Section 2.2; it uses multiple satellites and three receivers. Future versions will likely have four receivers and utilize a second GPS frequency. Figure 4.4 shows a ($10 \times 3 \times 2$) grouping of channels, meaning 10 satellites, three receivers, and two frequencies. This representation is called the VMT Block Form (Vector-Matrix-Tensor, described further in Chapter 5). Channel faults represent one red block, Vector faults are a line, Matrix faults are a plane, and a Tensor fault would mean that every channel is faulted. For Vector faults, the term V_{RX} states that all three receivers for a particular satellite and frequency are faulted, implying that a particular satellite signal is flawed. V_{SV} and V_{FQ} and defined similarly. For Matrix faults, the term M_{RX} states that all the channels for a given receiver are faulted. Again, M_{SV} and M_{FQ} are defined similarly. What is examined in this chapter is how to best detect a V_{RX} fault. This is a fault for a given satellite which is evident across all three receivers. The source of the fault is unspecified; it may be the satellite or the atmosphere affecting the signal propagation.
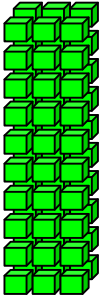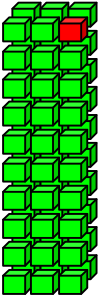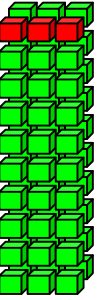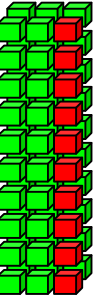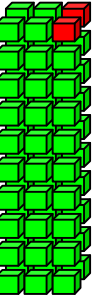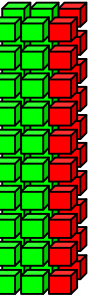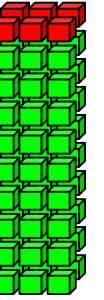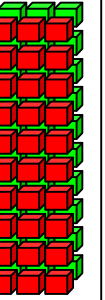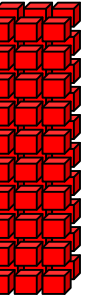
Figure 4.4: Fault Forms, Block Style

4.2   Methodology

The intent of this chapter is to determine if EXM is detecting V_{RX} faults in a way which maximizes system availability and continuity.   However it is difficult to test a system which has required values of $P_{FFD} = 10^{-8}$ and $P_{MD} = 10^{-8}$, for example. This would mean that nominally, the system will only flag 1 of 100 million epochs of "normal" data. It is possible to do functional testing on particular fault modes [6].   This involves injecting an error and running only a few (3 - 5) trials to ensure detection occurs every time. Instead of running the entire system over the collected data, it is possible to isolate a particular monitor in EXM by pre-calculating the input and then applying the algorithm logic and evaluating the output for optimum detection.

The work done here demonstrates that a fundamental change can be made to the EXM of the Stanford IMT in order to improve its detection ability while providing robustness to spurious faults. The extant algorithm used by the Stanford IMT is quite robust to a fault on a single channel. The enhancement made in this chapter to increase the detection ability of that algorithm involves averaging across receivers. Consequently, one channel (i.e. one receiver) has the potential to affect others and care must be taken to ensure that a large single-channel fault for a given satellite on one receiver does not cause the exclusion of that satellite on the other receivers tracking that satellite. Since this work is taken from concept to conclusion, the following method was used to show the efficacy of the modifications.   It starts with proving the concept in theory under simple assumptions, then progressing to showing an improvement with a small set of data, and finally modeling the data to demonstrate an improvement at the large threshold, low fault-free exclusion rates associated with this method.

1. Theory:
   a) Show that an algorithmic improvement can be made using Gaussian CDF models.
   b) Use a Gaussian, white-noise, high-threshold simulation to show improvement.

2. Practice:

    a) Use real data with low thresholds to show improvement under realistic conditions.

3. Projection:

    a) Model the data variance, kurtosis, and correlation and show that a high-threshold simulation for that model shows improvement.

4.3   Theory

This section contains an introduction to the concepts employed to increase the performance of EXM in its balance of fault detection and fault-free operation. One of the core concepts is to use a parameter which can summarize this balance in a single value for a particular exclusion logic scenario. This parameter is called the MDE, or the Minimum Detectable Error.

4.3.1   Why MDE?

There are many factors to consider when designing an optimum detection rule. A rule which minimizes false detections and minimizes false positives is ideal. It is possible to incorporate both of these concepts into one value to enable direct comparisons of methods. The Minimum Detectable Error (MDE) accomplishes this by providing the size of the error that can be detected with the associated P_{FFD} and P_{MD}. T_{FFD} is the monitor detection threshold which determines P_{FFD}, and T_{MD} is the "missed-detection buffer" value beyond T_{FFD} which achieves the P_{MD}. MDE is the sum of T_{FFD} and T_{MD}, given in Equation 4.1 [7].

$$MDE = T_{FFD} + T_{MD} \qquad\qquad (4.1)$$

Figure 4.5 shows a graphical representation of the MDE. The upper (green) Gaussian curve describes a hypothetical probability density function, and the lower (red)

93

Gaussian curve shows how far the faulted distribution would have to be shifted in order for P_{FFD} and P_{MD} to be below their prescribed values.
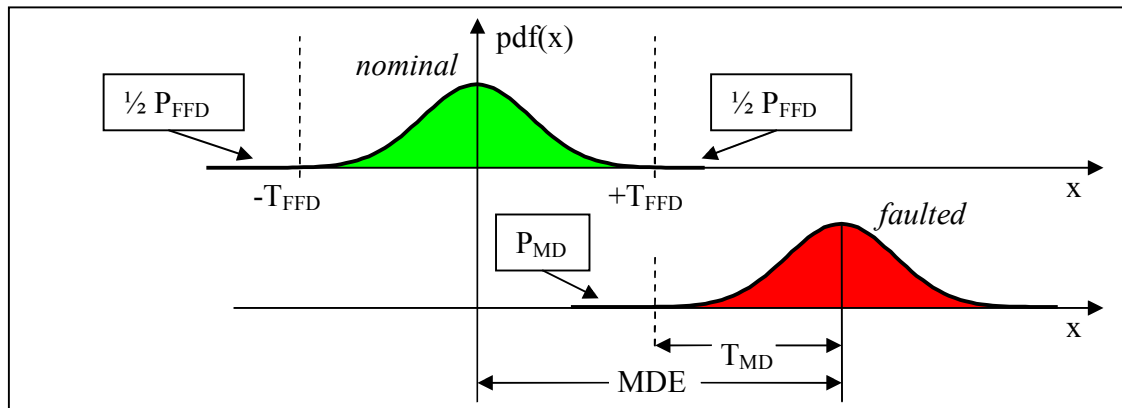
Figure 4.5: How P_{FFD} and P_{MD} affect MDE

Referring to Figure 4.6, P_{MD} is typically calculated by considering everything below the positive detection threshold $(+T_1)$ as passing the monitor test (i.e., not causing a flag). Technically, a datum falling to the left of the negative threshold $(-T_1)$ would also fail the test, but because $Cdf(-T_1)$ is miniscule compared to $Cdf(+T_1)$, ignoring this possibility creates a conservative, although nearly exact, estimate of the actual P_{MD}. Figure 4.6 shows that if the thresholds of $(\pm T_1)$ were narrowed to $(\pm T_2)$, there would be a concern that $Cdf(-T_2)$ was not negligible compared to $Cdf(+T_2)$. This issue is relevant to identify, but such tight thresholds would could an unacceptably large PFFD, so it is a moot circumstance.
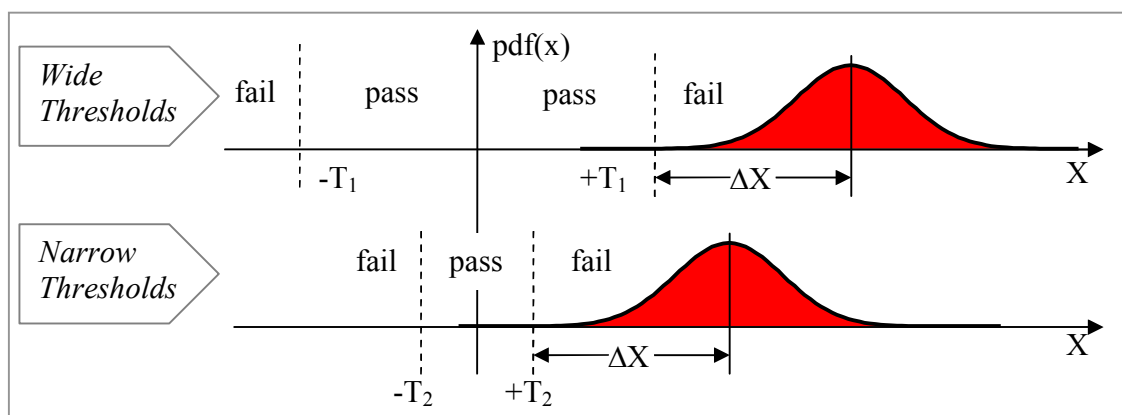
Figure 4.6: $Cdf(-T_2) \approx Cdf(+T_2)$ May Prevent One-Sided P_{MD} Calculation

4.3.1.1 Why Shape Matters

The purpose of EXM-I exclusion is to determine if a fundamental fault has occurred. More specifically, does the information contained in the channel-specific information indicate that perhaps a satellite is performing abnormally? There are many different ways to establish thresholds on the individual channels in a multi-channel system. Figure 4.7 shows a two-dimensional Gaussian distribution described by a radius, with mean (μ) = 0, and standard deviation = σ. The two thresholds, T_{ISO} (Isotropic Radial Bound) and T_{AVG} (Average Bound), are based on maintaining an equal exclusion area, or P_{FFD}. For reference later, the standard Cartesian quadrant notation is included in the figure [32].

When the system is nominal and centered around the origin, the shape of the isotropic Fault Free threshold is a circle that provides equal protection in all directions. This is the Isotropic Radial Bound. If the sole concern is a same-sign vector fault; i.e. a shift of mean, then setting the threshold at T_{AVG} will provide the smallest MDE for a given P_{FFD} and P_{MD}. Or, consequently, the distribution which has shifted to point A will have the lowest P_{MD}, while the distribution at point B will have a very high P_{MD}. The corners are shaded to indicate their respective threat levels. The upper right and lower left corners are red because they strongly indicate a process shift, and the effects of this shift will not be averaged out. The upper left and lower right corners also strongly indicate an aberration, but since the broadcast correction will be averaged, the effect of such an occurrence is smaller.

Equation (4.2) is the probability density function of a zero-mean Gaussian distribution with standard deviation, σ. Equation (4.3) is the probability density function a two-dimensional, zero-mean Gaussian distribution with standard deviation, σ, expressed as a function of the, distance from the origin,. Equation (4.4) is the cumulative density function of the distribution of Equation (4.3), and relates the radial bound to an exclusion probability.

$$pdf_X = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x}{\sigma}\right)^2} \tag{4.2}$$

$$pdf_R = \frac{R}{\sigma^2} \cdot e^{-\frac{1}{2}\left(\frac{R}{\sigma}\right)^2} \tag{4.3}$$

$$cdf_R = 1 - e^{-\frac{R^2}{2}} \tag{4.4}$$
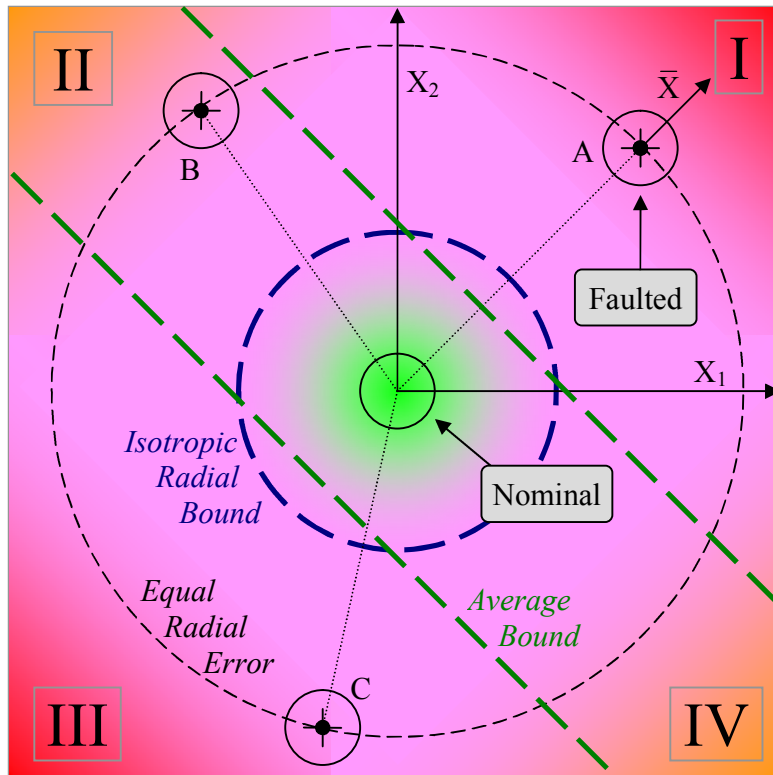
Figure 4.7: Average Method Threshold vs. Isotropic Threshold

Figure 4.7 shows that the Averaging Method's threshold is tighter than the isotropic threshold with respect to a shift in mean, but also that, for a given direction, the optimal curvature for the threshold is a line perpendicular to that vector. However, the IMT's determination of a "vector" fault is a compilation of the estimations of "channel" faults. Figure 4.8 shows how using this method to process the channels separately and then combine their pass/fail outputs will result in a "corner," and forces an increase in the MDE. Reducing the MDE for any given fault mode will increase the integrity

monitoring system's availability because it is better able to both identify true errors and minimizing false positives. There are two other methods described in Section 4.3.2 which will be quickly identified here. For a two-channel system, there can be a rule which declares a fault if either channel's absolute value exceeds a certain threshold. This is called the $(1^+/2)$ rule, meaning one or more flagged measurements out of two channels. In general, this rule is referred to as the (m^+/n) method (or rule) to signify m or more flags out of n channels. Another rule is the $(2/2)$ rule, meaning that two out of two tested channels have values exceeding a certain threshold. Figure 4.8 shows that both the $(1^+/2)$ and $(2/2)$ methods create those "corners," which make both these methods inefficient in detecting a bias existing across both channels. In the figure, Point A is the center of the two faulted distributions.
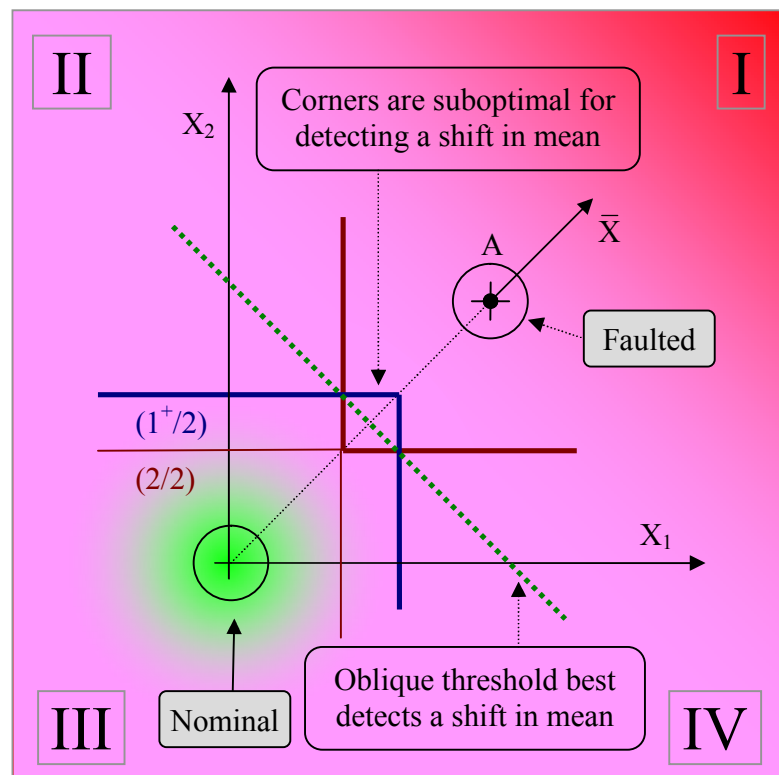
Figure 4.8: "Corners" Increase the MDE

Table 4.1 shows the thresholds for a hypothetical P_{FFD} and P_{MD}. The detailed equations used to compute these thresholds are given in Appendix I. This table suggests that the Averaging Method (MA) provides the best performance for a 2-D detection

system. However, this holds only under the assumption that the noise on each channel is independent.

	T_{FFD} $P_{FFD} = 10^{-8}$	T_{MD} $P_{MD} = 10^{-4}$	MDE
$1^{+}/2$	5.8472	2.3262	8.1734
2/2	3.8906	3.8906	7.7812
Avg.	4.0522	2.6297	6.6819

Table 4.1: Comparison of Two-Channel Exclusion Methods

In Table 4.2, the same calculations are performed for a three-receiver installation using the equations given in Appendix J. Again, the Averaging Method has the lowest MDE under the condition of statistical independence. Section 4.5, which uses real GPS data, and Section 4.6, which run simulations using modeled data, both show the effect of correlated noise across receivers.

	T_{FFD} $P_{FFD} = 10^{-8}$	T_{MD} $P_{MD} = 10^{-4}$	MDE
$1^{+}/3$	5.9143	1.6806	7.5949
$2^{+}/3$	4.0218	2.5250	6.5468
$3^{+}/3$	3.0680	5.7991	8.8671
Avg.	3.3086	2.1472	5.4558

Table 4.2: Comparison of Three-Channel Exclusion Methods

4.3.1.2 What is the Effect of Subsequent (channel) Exclusions?

Referring back to Figure 4.7, quadrants II and III were presumed to be safer than quadrants I and IV. This is because the actual threat to the system is attenuated since the broadcast correction is calculated by averaging across receivers. This may suggest that

Averaging Method thresholds can be relaxed due to a lack of a threat. However, it is possible that in a monitor subsequent to EXM-I, one of those channels is flagged and later excluded by EXM-II (see Section 2.2). If the remaining channels both possess a fault then the averaging will be of no benefit. This is not a concern for the two-receiver case because losing one receiver provides insufficient redundancy to proceed. It is possible to construct an example to demonstrate the potential danger for a three-receiver setup. Using the notation addressed in Section 4.3.1.1, suppose all the channels were faulted with the sign of the faults being $\{+,+,-\}$. Mathematically, this would be expressed as $X_1 > T_1$, $X_2 > T_2$, and $X_3 < T_3$. While the channel-specific values may be large in magnitude, causing them to fail individually, their conflicting signs do not suggest a clear fault mode. If the signs of the faults were $\{+,+,+\}$, this would clearly indicate a shift in mean. Using the $\{+,+,-\}$ fault scenario, what if, in one the monitors which succeed EXM, one of the channels were excluded? Depending on which channel is excluded, the remaining channels could have the reduced structure $\{+,-\}$ or $\{+,+\}$. This implies that the averaging calculation could be more or less effective than with three channels. The $\{+,-\}$ scenario is more likely to average to zero, whereas the $\{+,+\}$ scenario will still have a prominent bias regardless of averaging. There is then a multitude of prior probabilities and correlations amongst monitors. What is the probability of a $\{+,+,-\}$ scenario? What is the probability that another monitor will exclude either the $\{+\}$, or $\{-\}$ channels? Finally, what is the probability that averaging across only two channels may reduce the integrity of the corrections provided by the system? For the analysis in this chapter, identical thresholds are used for the same-signed fault mode and the mixed-signed fault mode.

4.3.2 EXM Methods: Past and Present

Several exclusion methods are discussed in this section, beginning in two dimensions to better render certain figures and then moving to three dimensions once the process is clear. The discussion centers around when to declare a satellite tracked on three receivers to be unhealthy. Although the three-dimensional case is given as the final

result, the monitoring station may well be equipped with four receivers or operate with only two receivers if one of them seems to have failed. The results of this chapter are directly applicable to the case of using more than three receivers; however the extent of the potential improvement using these methods is not explored in detail here.

4.3.2.1 The Boolean Method, M_B

This is the method used by the Stanford IMT (Section 2.2). It requires that there must be m or more apparent failures (i.e. monitor test statistics that exceed their detection thresholds) out of n channels to declare a fault. This will be generically referred to as the (m^+/n) method. Figure 4.9 below shows a graphical representation of the $(1^+/2)$ and $(2/2)$ methods.
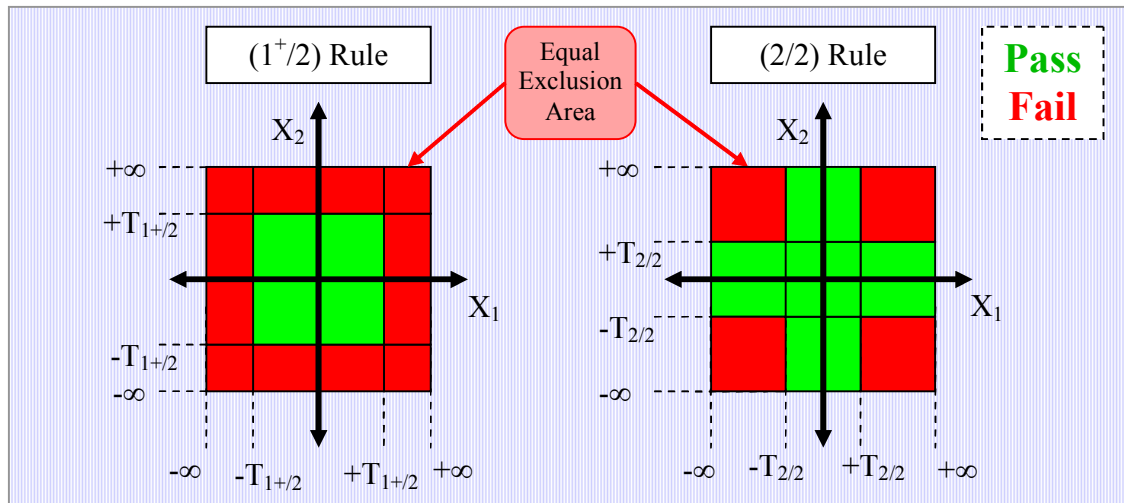
Figure 4.9: Decision Squares for the Two-Channel Case

Section 4.3.1.2 introduced the notion of detection faults with same or opposite signs. The concept of fault detection of both same-sign and opposite-sign faults will continue to be an issue throughout this chapter. Figure 4.9 shows that the $(2/2)$ rule will exclude not only a large x_1 and large x_2 of the same sign but also those of different sign. This means that the detection rule is not detecting a single pattern (such as a shift in mean) but, in general, inexplicably large values. Figure 4.10 below shows the actual

100

faulted distributions under the hypothesis that there has been a shift in mean. The shift can either be positive or negative and is considered a same-sign fault since all affected channels experience the same fault.
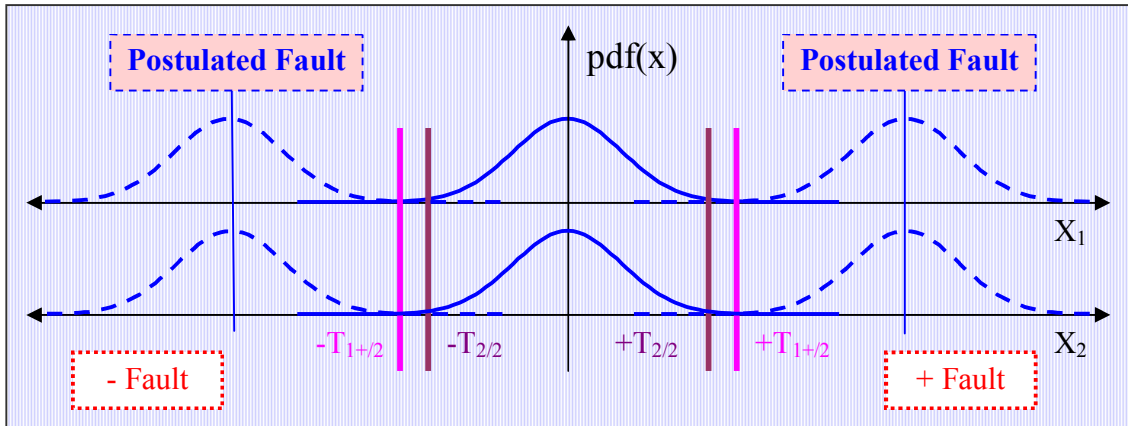
Figure 4.10: A Shift in Mean for Positive and Negative Faults

If Figure 4.10 above, showing the decision squares, was modified only to look for same-sign fault modes, the decision square would be modified as shown in Figure 4.11.
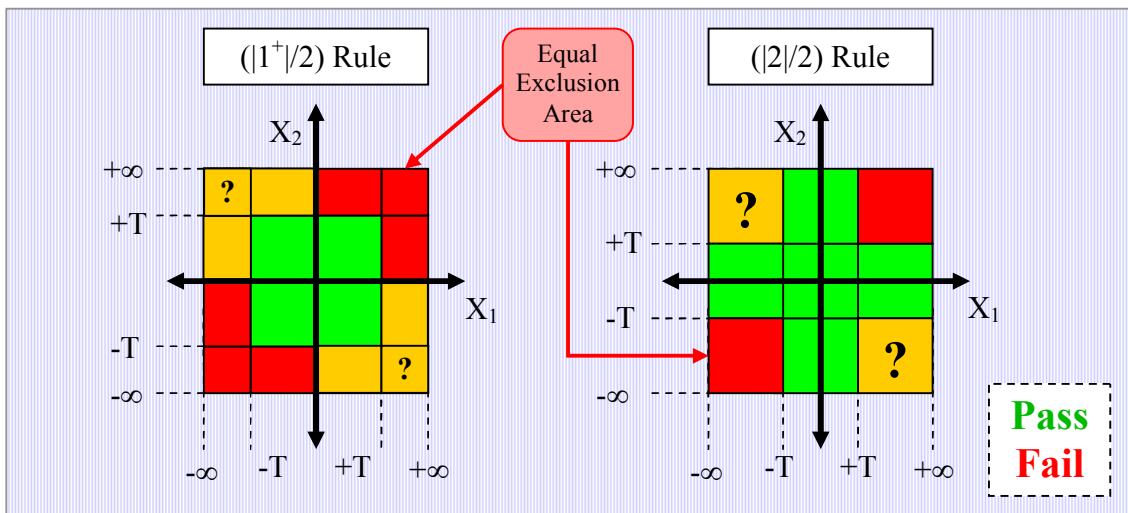
Figure 4.11: Decision Squares for the Two-Channel Case, Same-Sign Faults

There is no significant benefit to modifying the rule to look for same-sign faults because this would make the system oblivious to different-sign faults, which are certainly off-nominal, though not necessarily as threatening as same-sign faults. The correction

broadcast to the user aircraft for each satellite will be the average correction for that satellite across all receivers it is tracked on and declared healthy. Thus, a same-sign error will permeate through, while a different sign-error will be attenuated, but by no means can it be assumed to cancel. Furthermore, eliminating the different-sign exclusion region will double the allotted fault-free detection area, but there will be a minimal benefit to reducing the fault-free detection threshold. The reason is that these are small probabilities and large thresholds, so changing the probability by a factor of two will have a minor impact on the threshold. For example, take a Gaussian Distribution with zero mean and Standard Deviation = σ, then $CDF^{-1} (10^{-7}) = 5.2\sigma$ and $CDF^{-1} (2 \times 10^{-7}) = 5.07\sigma$. In other words, doubling the exclusion area only reduces the threshold by 2.5%. The missed-detection threshold will be negligibly affected, as it is too far away from the postulated same-sign fault to have any impact on the probabilities. The conclusion is that both same-sign and different-sign fault modes should be tested for.

Figure 4.12 below shows which regions are associated with which postulated circumstances: vector (satellite) faults, channel faults, or multiple unrelated channel faults.
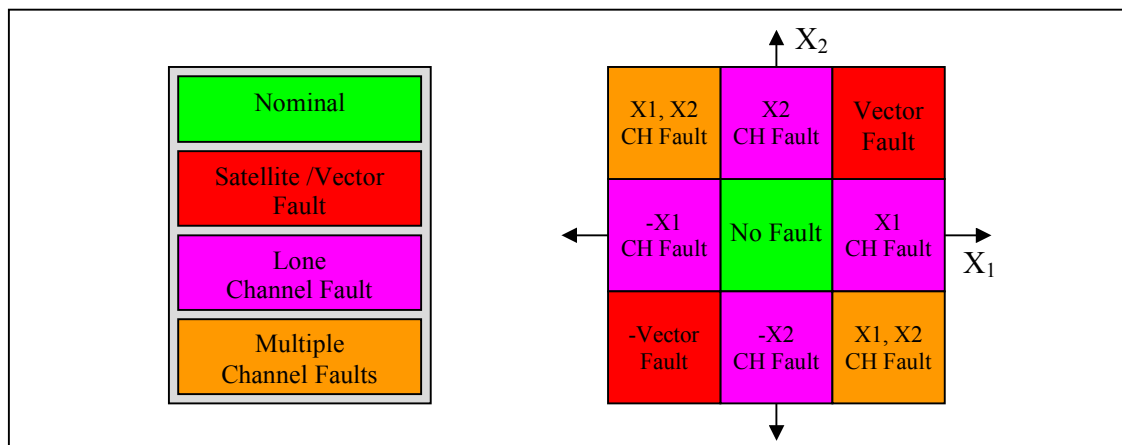
Figure 4.12: The Areas of Nominal, Faulted Channel, and Vector Faults

The attributed explanation for the red zones is a satellite fault, for the magenta zones, a single channel fault, and for the orange zones, a multiple-channel fault (which would likely be the result of an inflated error sigma). Figure 4.13 takes this concept a

step further with hypothesized distributions. Both the channel faults (magenta) and vector faults (red) show a shift in the distribution mean, while the multiple-channel different signed faults (orange) show a distribution which has not shifted its mean but has an increased variance.
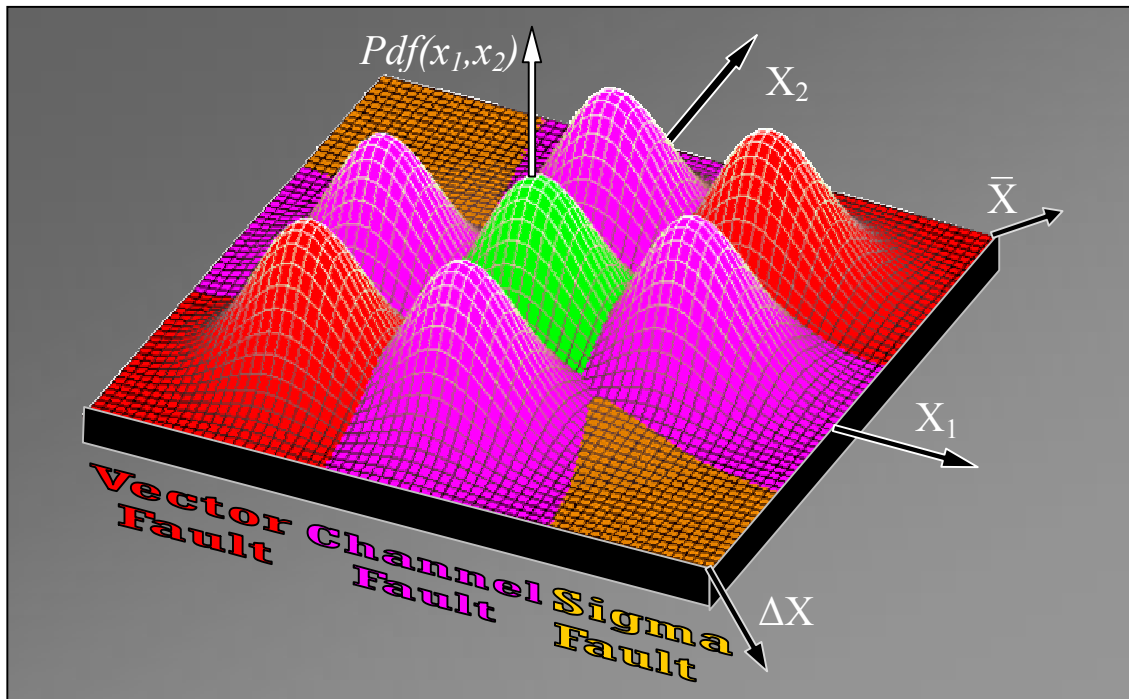
Figure 4.13: Distributions of: Nominal, Faulted Channel, and Vector Faults

4.3.2.2 The Averaging Method, M_A

Applied specifically to MQM (Chapter 3), the Averaging Method simplifies the integrity monitoring process by immediately averaging the corrected carrier-phase values across the three receivers. This is a general method though, and has the potential to be applied to other monitors. The benefit is simplicity and, ideally, noise reduction. A decision square for the averaging rule is shown in Figure 4.14 along with the (|2|/2) rule.
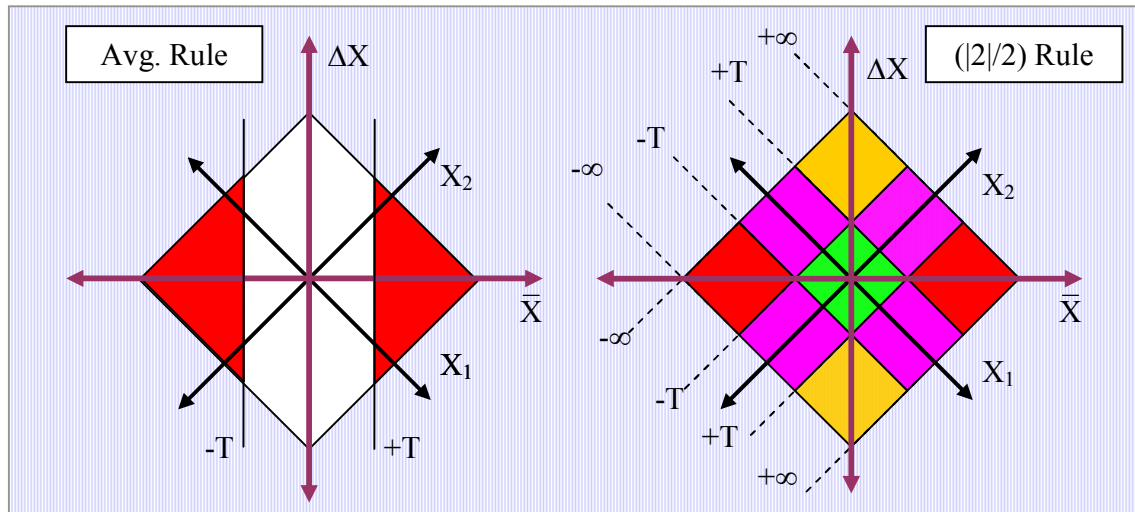
Figure 4.14: Averaging Method Compared to the (|2|/2) Method

The most obvious frailty of the Averaging Method is that there is no protection against ΔX. Even though the broadcast correction will be in the form of $\overline{X}$, it is possible that the three-receiver average is acceptable, whereas a two-receiver average would not be. Therefore, protection against ΔX is also needed. Aside from being oblivious to different-sign faults, the primary drawback is that the averaging method is now more sensitive to single-channel faults. The (m^+/n) method has intrinsic protection against extremely large single-channel errors, but because the Averaging Method does not assess the value of any individual channel, it has no such defense.

It is possible to create a hybrid detection rule that combines the sensitivity of the averaging method with the robustness of the (m^+/n) method. This is done by "drawing in" the thresholds of the averaging rule.
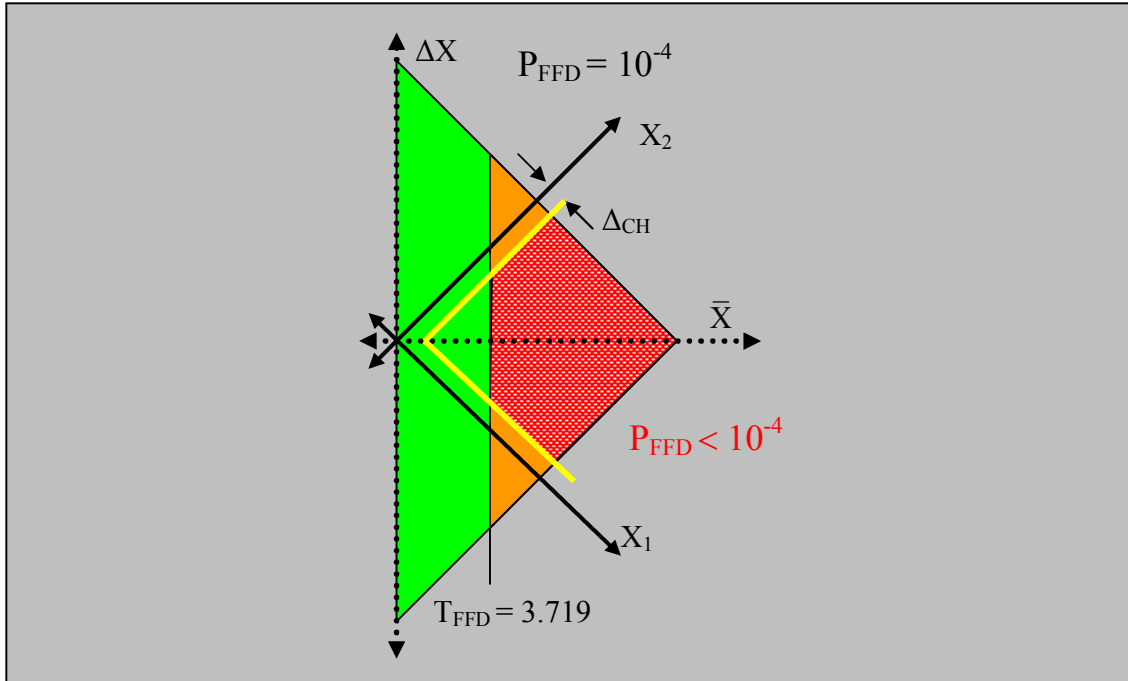
Figure 4.15: Drawing-in the Avg. Threshold for Single-Channel Robustness

Figure 4.15 uses the rotated form to emphasize that the vector fault is the fault of interest as opposed to single-channel faults. The figure shows that the detection threshold for the Averaging Method is $T_{FFD} = 3.719$ to achieve a $P_{FFD} = 10^{-4}$. This P_{FFD} includes both the orange and red areas. The threshold can incorporate a new facet using the term Δ_{CH} to express the distance from the channel axes. When $\Delta_{CH} = -\infty$, the threshold is equivalent to the Averaging Method. Figure 4.15 shows the new threshold when Δ_{CH} is a small positive number, leaving the area in red to be the exclusion area. The P_{FFD} would then be less that 10^{-4}, but more importantly, the probability that a single-channel fault causes an exclusion will be significantly lower. The values of T_{FFD} and Δ_{CH} can be manipulated in such a way as to achieve a steady P_{FFD} with varying levels of robustness to a single-channel fault. The finalized concept is illustrated in Figure 4.16 and is a lead-in to the Oblique Method.
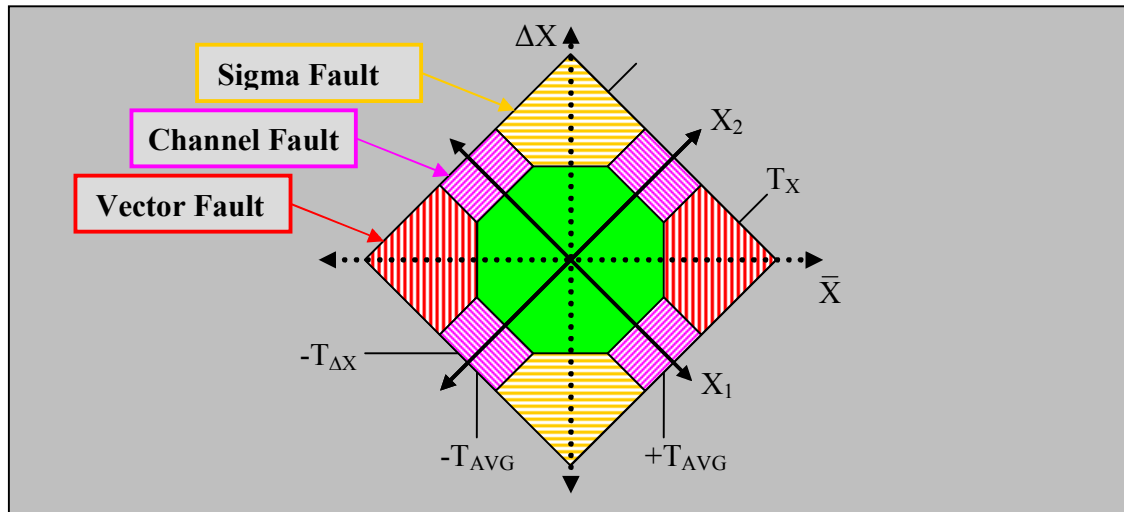
Figure 4.16: Postulated Faults with Modified Thresholds

4.3.2.3 The Oblique Method, M_O

The term "oblique" means to be slanted or indirect. This encapsulates the process of blending the traditional (m^+/n) Boolean Method with the Averaging Method. There are two ways of rounding corners depending on whether the structure is convex or concave. Fillets are concave roundings of interior corners; whereas chamfers are convex roundings of exterior corners. A common fillet is the weld region of two objects, while an example of chamfering is the beveled corners on furniture.

Figure 4.17: Fillets and Chamfers – Two Ways to Round a Corner

Referencing the 2-D decision squares shown earlier, and focusing on one corner, Figure 4.18 shows the continuum that exists when moving from the ($|2|/2$) method through the Averaging Method and on to the ($1^+/2$) method. The value X_C determines the point where the threshold curve intersects the $X_1 = X_2$ vector (dashed line), and X_C is

106

chosen to keep the exclusion area at a constant (specified) value of P_{FFD}. Two plots are shown separately in Figure 4.18 for visual clarity.
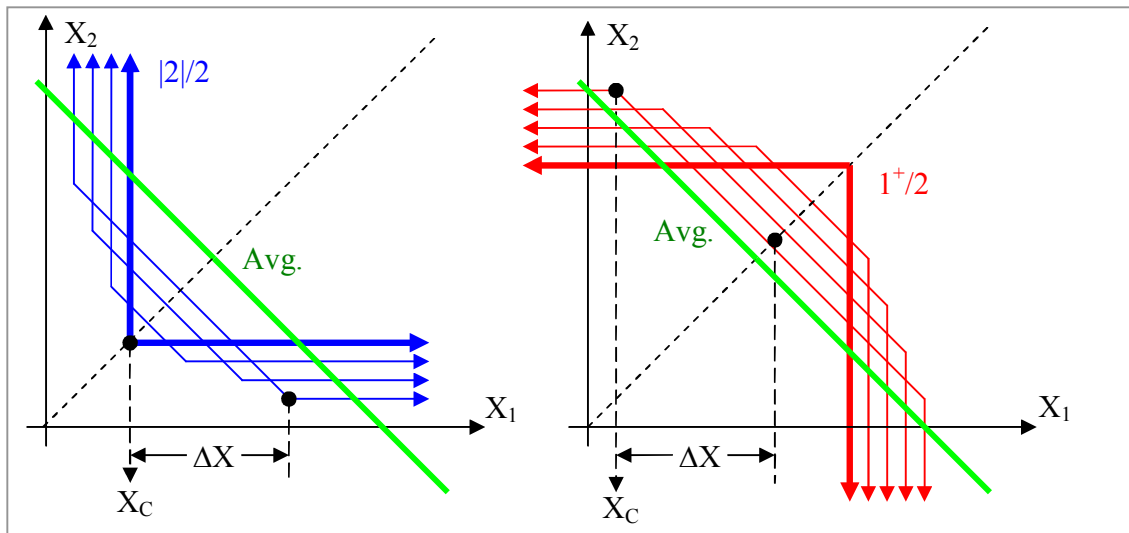
Figure 4.18: Oblique Corners in 2-D

As started, the central point $\{X_C, X_C\}$ in Figure 4.18 can be chosen independently of ΔX, but the goal is to maintain a constant P_{FFD} and to determine the effect of the resulting threshold shape on MDE and $P_{FFD/CH}$. In this scenario $P_{FFD/CH}$ is the probability of a fault-free detection of a satellite given that there is a single channel fault. The actual ΔX vs. X_C relationship is shown in Figure 4.19 below. The point at which the two curves meet is not the value of $X_C = 4.05$ that was calculated as the M_A threshold from Table 4.1 because this new result considers all four fault quadrants, not just the "mean" faults of quadrants I and IV.
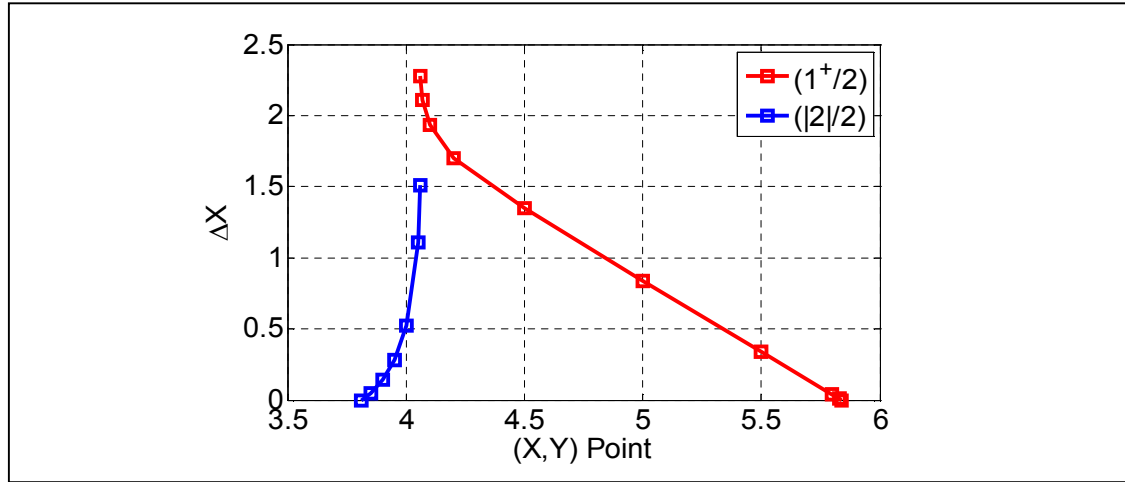
Figure 4.19: The X_C and ΔX Relationship to Maintain $P_{FFD} = 10^{-8}$

2-D Oblique Results

Figure 4.20 shows MDE as a function of X_C as it varies from the extremes of $(|2|/2)$ to $(1^+/2)$. $P_{FFD/CH}$ vs. X_C is also plotted to demonstrate the sensitivity of each method to a channel fault. Here, $P_{FFD} = 10^{-8}$ and $P_{MD} = 10^{-4}$. The results show that the smallest MDE occurs when using the Averaging Method, and this result agrees with the general result derived in Section 4.3.1.1. When also considering the $P_{FFD/CH}$ criterion, $P_{FFD/CH}$ for the $(|2|/2)$ Method can be orders of magnitude less than for the Averaging Method depending on the expected Channel Fault size. For the $(1^+/2)$ Method, P_{FFD} quickly converges to one when the probability of a single-channel fault increases. This is because that lone faulty channel is all that is needed to exclude the entire satellite under the $(1^+/2)$ rule.
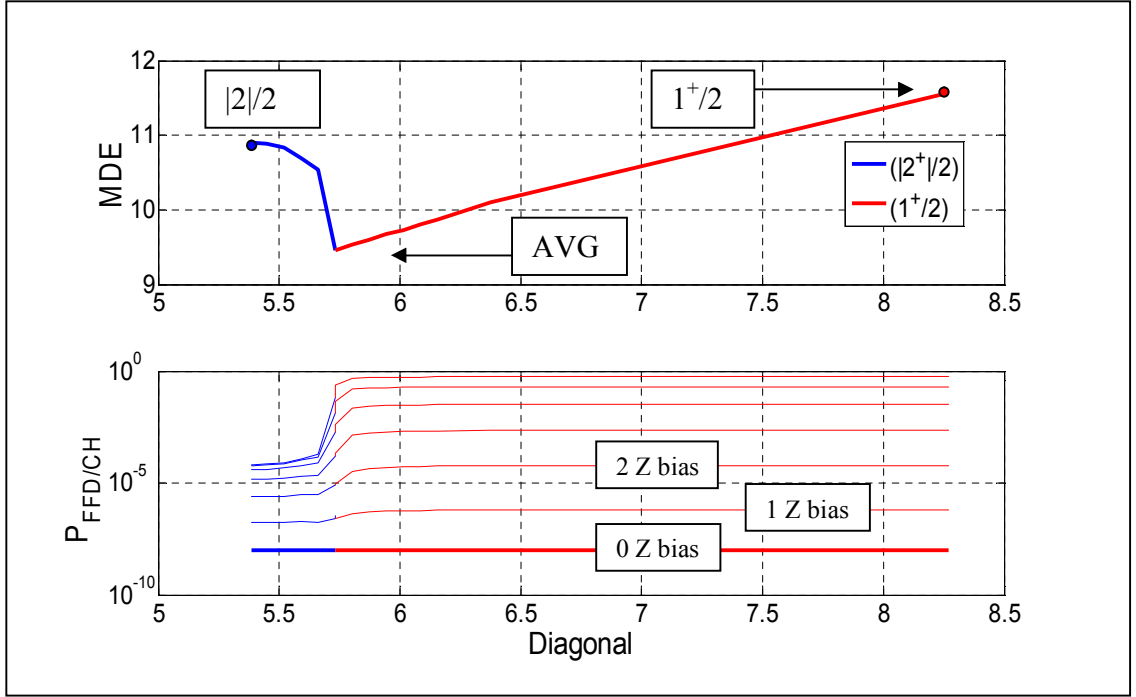
Figure 4.20: Minimum MDE is for Averaging, but P$_{FFD/CH}$ Favors (|2|/2)

Decision Squares

Combining the Averaging and Boolean Methods requires a conjunction operator. It can be the AND ($\cap$) or the OR ($\cup$) operator. Each gives a unique Decision Square such as those shown in Figure 4.21 below. The subscript annotations of F (e.g. (1+/2),$\cap$) in Equations (4.5) through (4.8) express how the Averaging and Boolean Methods are combined.

$$F_{1^+/2,\cap} \equiv \left[\sum_{i=1}^{2}(|x_i| > T) \geq 1\right] \cap \left[(|\bar{x}| > T_{AVG}) \cup (|x_\Delta| > T_\Delta)\right] \tag{4.5}$$

$$F_{1^+/2,\cup} \equiv \left[\sum_{i=1}^{3}(|x_i| > T) \geq 1\right] \cup \left[(|\bar{x}| > T_{AVG}) \cup (|x_\Delta| > T_\Delta)\right] \tag{4.6}$$

$$F_{2/2,\cap} \equiv \left[\sum_{i=1}^{2}(|x_i| > T) = 2\right] \cap \left[(|\bar{x}| > T_{AVG}) \cup (|x_\Delta| > T_\Delta)\right] \tag{4.7}$$

$$F_{2/2,\cup} \equiv \left[\sum_{i=1}^{2}(|x_i| > T) = 2\right] \cup \left[(|\bar{x}| > T_{AVG}) \cup (|x_\Delta| > T_\Delta)\right] \tag{4.8}$$
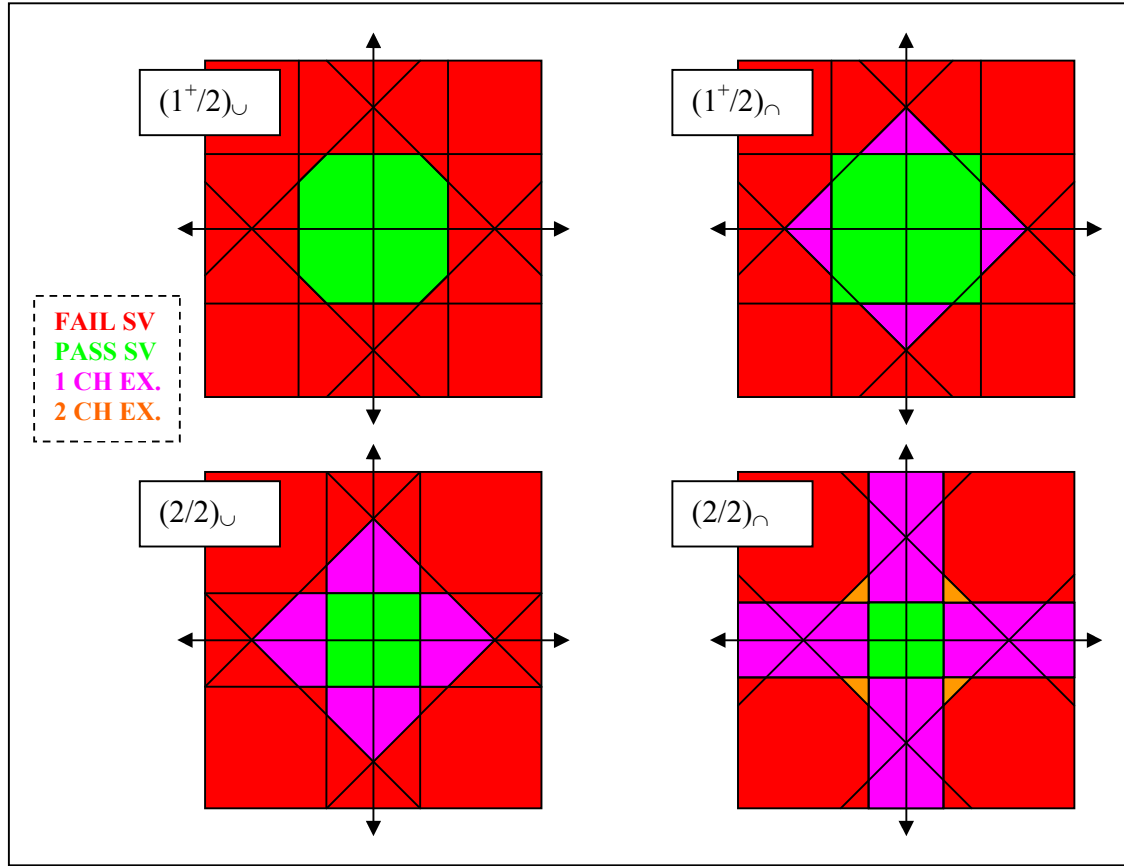
Figure 4.21: Oblique Decision Squares

The (2/2)∩ method suffers from the following contradiction: if either channel exceeds the channel thresholds, they are excluded. However, there is a small triangular section where, if both channels exceed their individual thresholds, then they are both excluded, but the SV is not flagged. The consequence of having no usable channels would of course keep the satellite from being used thereafter, but we have not made a direct action of finding fault with the SV even though all channels have been excluded from use. Figure 4.22 demonstrates how this is resolved. Previously, the threshold to determine if any channel has failed was independent from other channels. After this change, the passing region looks very much like the (2/2)∪ method.
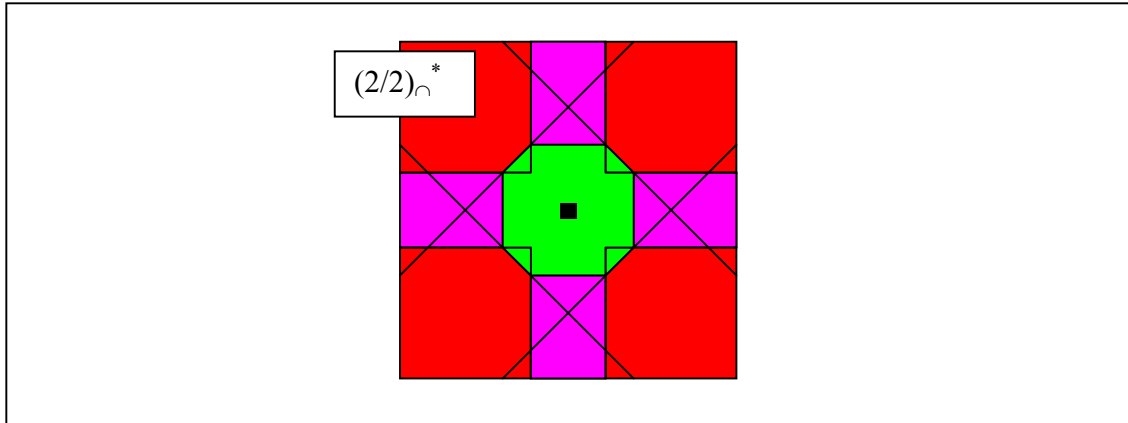
Figure 4.22: Channel Thresholds are now Dependent

4.3.2.4 Methods in 3-D

3-D Decision Cubes

Figure 4.23 extends the Decision Squares into Decision Cubes for the 3-D case to graphically represent the consequences of using Boolean logic.
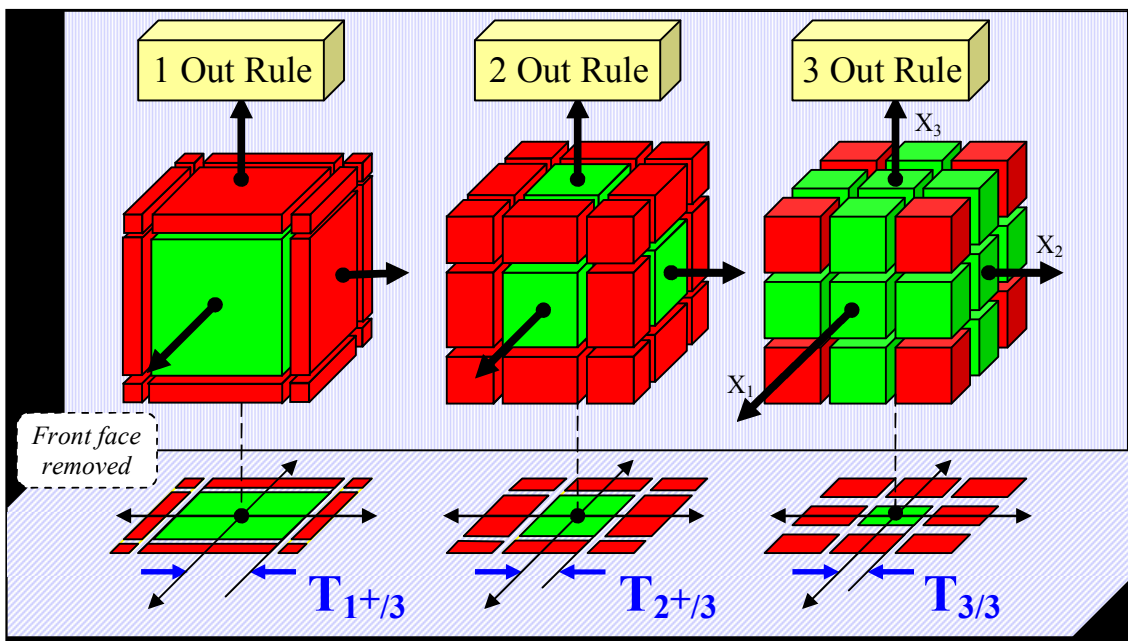
Figure 4.23: Decision Cubes and their Thresholds

The natural question is, "what about a radial bound?" Figure 4.24 shows that either a cube or sphere will exhibit very similar behavior simply because of their roughly comparable shapes. Using a radial bound on n channels is quite similar to the $(1^+/n)$ case and suffers higher $P_{FFD/CH}$; thus the sphere in Figure 4.24 behaves very similarly to a cube $(1^+/3)$ rule.
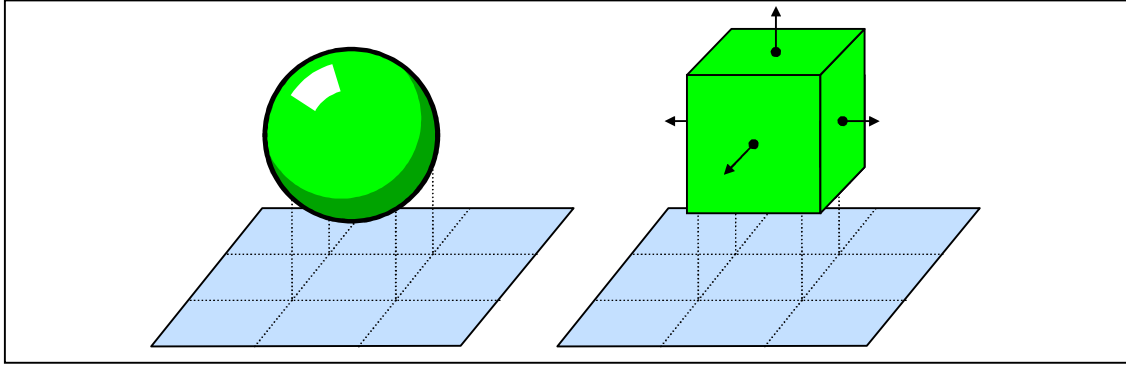
Figure 4.24: What about a Spherical Bound?

Equations (4.9-4.14) express the six possible formulas for combining the Averaging Method with the three-channel Boolean Method. These are analogous to the equations of (4.5-4.8) which expressed the four possible formulas for combining the Averaging Method with the two-channel Boolean Method.

$$F_{1^+/3,\cap} \equiv \left[\sum_{i=1}^{3}(|x_i|>T)\geq 1\right] \cap \left[(\bar{x}>T_{AVG})\cup(\bar{x}>T_\Delta)\right] \tag{4.9}$$

$$F_{1^+/3,\cup} \equiv \left[\sum_{i=1}^{3}(|x_i|>T)\geq 1\right] \cup \left[(\bar{x}>T_{AVG})\cup(\bar{x}>T_\Delta)\right] \tag{4.10}$$

$$F_{2^+/3,\cap} \equiv \left[\sum_{i=1}^{3}(|x_i|>T)\geq 2\right] \cap \left[(\bar{x}>T_{AVG})\cup(\bar{x}>T_\Delta)\right] \tag{4.11}$$

$$F_{2^+/3,\cup} \equiv \left[\sum_{i=1}^{3}(|x_i|>T)\geq 2\right] \cup \left[(\bar{x}>T_{AVG})\cup(\bar{x}>T_\Delta)\right] \tag{4.12}$$

$$F_{3/3,\cap} \equiv \left[\sum_{i=1}^{3}(|x_i|>T)= 3\right] \cap \left[(\bar{x}>T_{AVG})\cup(\bar{x}>T_\Delta)\right] \tag{4.13}$$

$$F_{3/3,\cup} \equiv \left[\sum_{i=1}^{3} \left(|x_i| > T \right) = 3 \right] \cup \left[\left(\bar{x} > T_{AVG} \right) \cup \left(\bar{x} > T_{\Delta} \right) \right] \qquad (4.14)$$

The concept of the Averaging Method (octahedral) is merged with the Boolean Method (cube) in Figure 4.25, which shows the various resulting geometric shapes. Only the passing volumes are shown; the remaining volumes are the failing regions. Equations (4.9-4.14) are shown graphically in parts (C-H) of Figure 4.25.
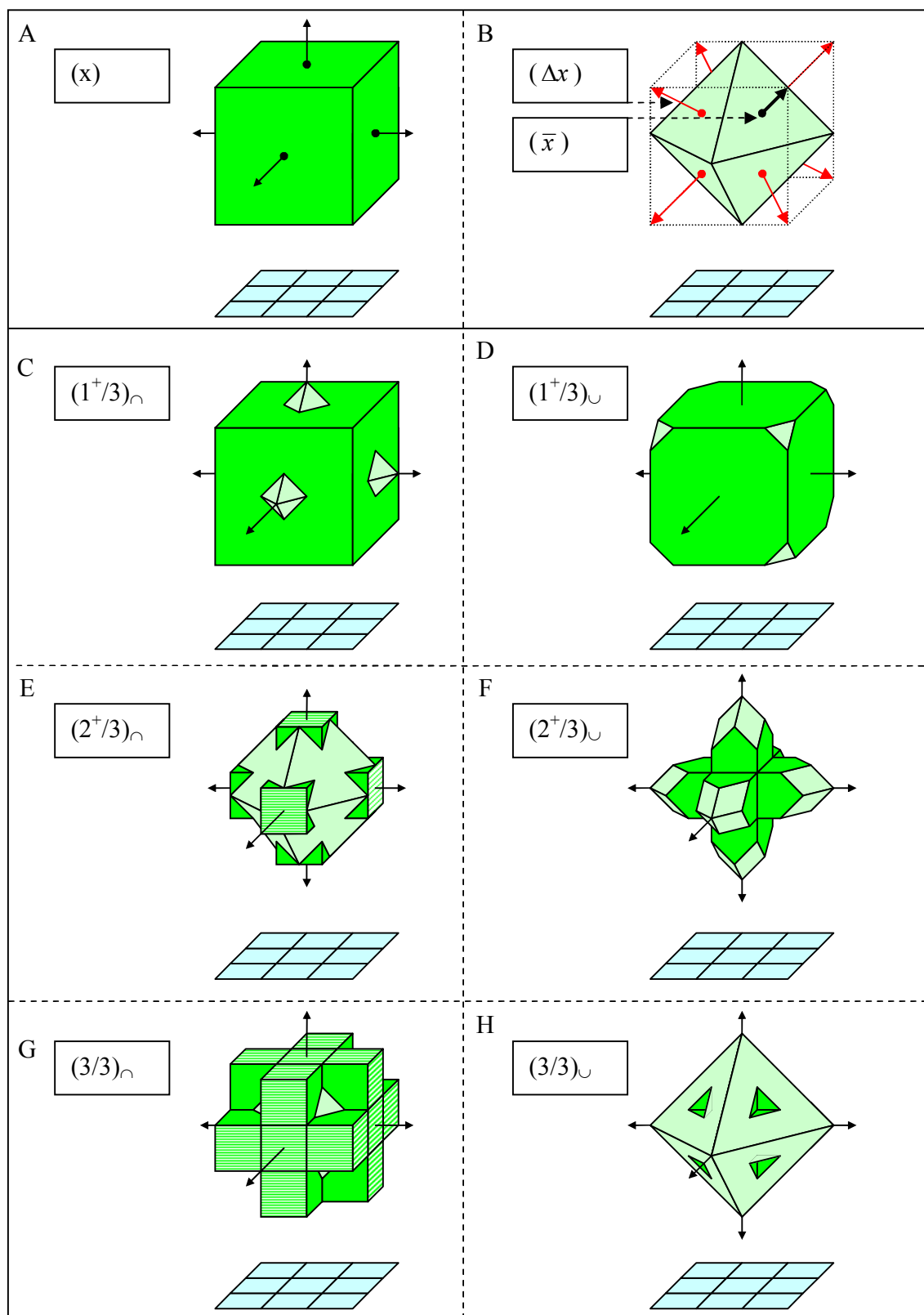
Figure 4.25: Oblique Polyhedra Inclusion Areas

114

3-D Oblique Results

The first objective is to identify which of the methods shown in Figure 4.25 are suitable for our objectives. Methods D, E, & G, are the ones with the oblique face on the diagonal., and it is this facet which promotes the smallest MDE. Again having a small MDE enables better fault detection. Figure 4.26 is analogous to Figure 4.20 but for three receivers and for ($P_{FFD} = 10^{-8}$, $P_{MD} = 10^{-4}$). The "steps" are the coarser quantized sample sizes compared to Figure 4.20 because of the need to numerically integrate the extra dimension with finite computational resources. Here, the smallest MDE is again for the Averaging Method at (diagonal) $X_C \cong 3.46$. These results define the trade space between MDE and $P_{FFD/CH}$ and allow a trade-off between achieving lower MDE by sacrificing $P_{FFD/CH}$. In Figure 4.26, the new MDE (at 6.1) is set to be about 10% higher than the MDE corresponding to the Averaging Method (at 5.5), which is still about 10% better than the ($2^+/3$) Method. Increasing the MDE slightly allows $P_{FFD/CH}$ to be lowered from 2% to 0.02%, compared to 0.01% for the ($2^+/3$) Method.
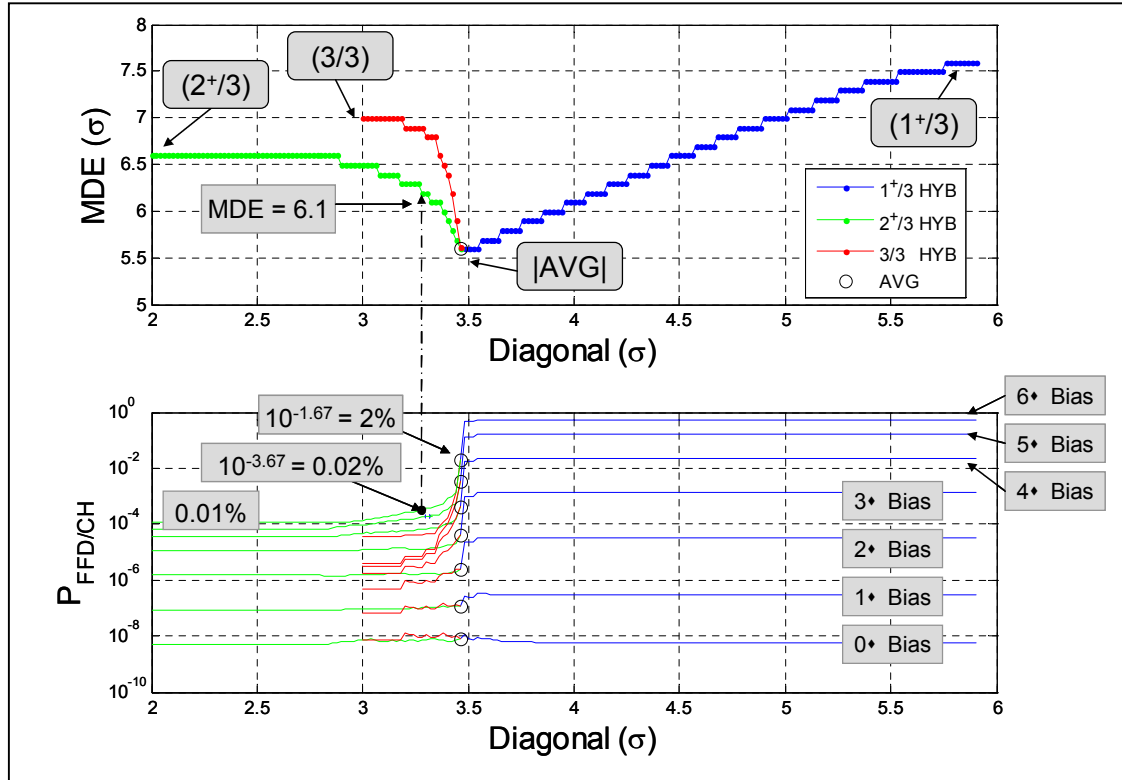
Figure 4.26: 3-D Oblique Results

4.5 Practice

Now that the theory has been examined, the resulting concepts can be applied to real data. Much tighter thresholds are used compared to an operational LAAS so that the performance can be compared without having to collect 10^8 or more samples. One hour of L1 GPS data is used from a set that was collected over 24 hours on March 11, 2003, starting at 22:00 PST, using the Stanford IMT with three NovAtel Pinwheel antennas and OEM4 Receivers. The statistic being tested is the acceleration estimate (ACC) from MQM.

4.5.1 Evaluating the Oblique Detection Method

In Section 4.3.2.4 the Oblique Method was shown to decrease MDE when compared to the Boolean Method used in the original IMT. Given a constant $P_{FFD} = 10^{-3}$,

the intent here is to show that the Oblique Method will be able to detect a vector fault of a given size with a lower P_{MD} than the Boolean Method but with a higher P_{MD} than the Averaging Method (as anticipated from theory). Immediately following that result is a comparison of the performance of the methods when there is a single-channel bias of a given size. Table 4.3 shows the empirically determined thresholds to achieve the specified P_{FFD} of 10^{-3}. The fact that real data was used is the reason that the Oblique Method has the same T_{AVG} threshold as the Averaging Method while also having $T_{BOOL} = 2.87$. The values in Table 4.3 describe the smallest thresholds such that the $P_{FFD} \leq 10^{-3}$. Holding $T_{AVG} = 3.00$, the T_{BOOL} threshold increases from zero until the point at which enough additional data are excluded to cause the P_{FFD} to exceed 10^{-3}.

	$(1^+/3)$	$(2^+/3)$	$(3/3)$	AVG	OBL
T_{BOOL}	4.40	3.20	2.40	-	2.87
T_{AVG}	-	-	-	3.00	3.00
P_{FFD}	10^{-3}	10^{-3}	10^{-3}	10^{-3}	10^{-3}

Table 4.3: Empirical Thresholds for Various Methods for $P_{FFD} \leq 10^{-3}$

SV Fault Effect

In this set of experiments, all MQM acceleration estimates have been normalized by dividing by their elevation-dependent threshold. At each epoch and for each satellite tracked over all three receivers, an error was injected into the data for each of the receivers. This error was the same across all three receivers and was increased in size from zero to six (in normalized units). Figure 4.27 shows the results of this experiment. As expected, the Averaging Method had the least number of missed detections, as this is the type of fault the Averaging Method is optimized to detect. Also as expected, the $(2^+/3)$ Boolean Method did better than either the $(1^+/3)$ or $(3/3)$ Methods. The Oblique Method outperformed the (2+/3) Method by only having roughly half the missed detections.
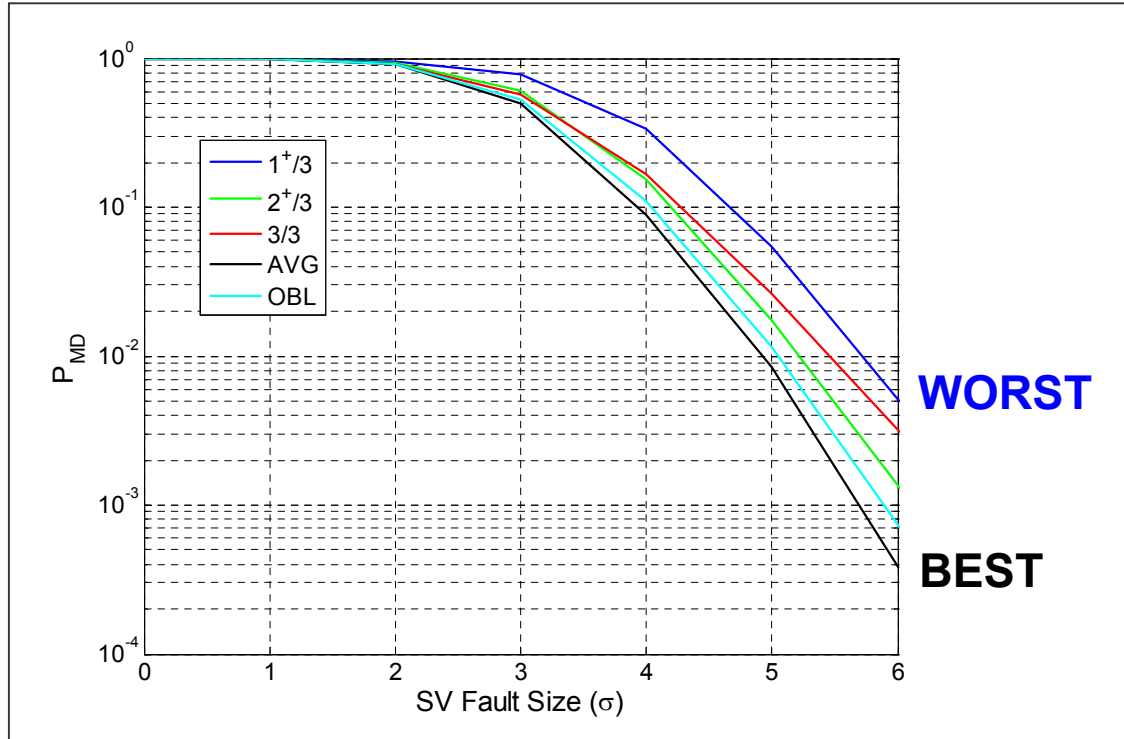
Figure 4.27: Oblique Method Outperforms the $(2^+/3)$ Boolean Method in P_{MD}

CH Bias Effect

The next experiment uses the same empirically-derived thresholds and tests each method's fault-free detection (of the SV) against a single-channel fault. The terminology presented in Section 1.2 stated that the term $P_{FFD/CH}$ represents the probability of a fault-free detection of a satellite (or receiver) given that one channel for that satellite (or receiver) is faulted. This means that there is no implicit fault on the satellite or receiver (making it fault-free) but that on one particular channel a fault was declared. It is comparable to the argument for fault-free detection of channel. That fact that the channel was flagged is due to there being an overly large estimate of some parameter for that channel, but it is the fact that there is no implicit, or true fault with the channel that makes it a fault-free detection.

Here, the channel fault is applied to a single randomly-chosen receiver. In this scenario, the (3/3) Boolean Method would appear to be the winner; however, any (*n*/*n*)

Boolean Method is unrealizable because it would suggest that ($n-1$) channels can be faulted without the system also flagging the SV. Essentially, if the satellite in question is only deemed to be healthy on one channel, there is insufficient redundancy for that satellite to allow it to be used in a safety-critical system. Therefore, although such a rule cannot be implemented, it serves to complete the picture when comparing the set of possible rules.

Figure 4.28 shows the single-channel robustness performance comparison. With the (3/3) Method disqualified, the ($2^{+}/3$) Boolean Method just edges out the Oblique Method with six-sigma-channel-fault $P_{FFD/CH}$ values of 0.011 and 0.015, respectively. These results are consistent with the previous theoretical findings of Section 4.3.2.
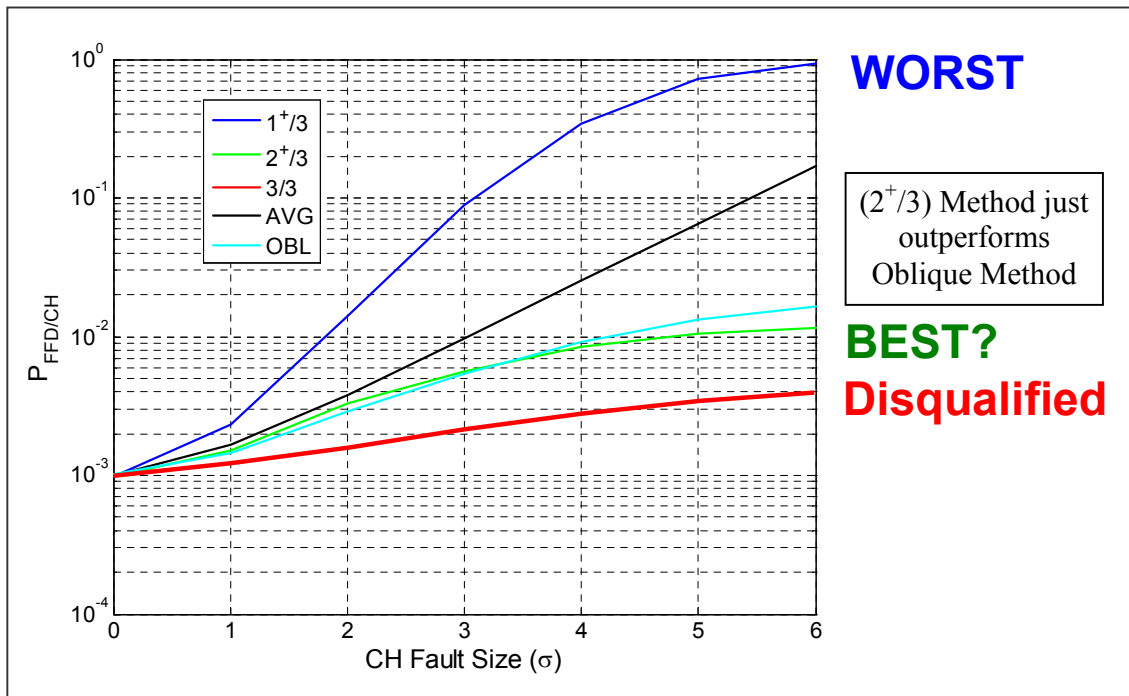
Figure 4.28: ($2^{+}/3$) Boolean Method Outperforms Oblique Method in $P_{FFD/CH}$

4.5.2   The Oblique Radial Method, M_{OR}

It was shown that Oblique Method could reduce the MDE by increasing $P_{FFD/CH}$, but it would be a stronger argument in favor of the Oblique Method if this newly derived

method could always at least equal and mostly outperform the old method. Although real data typically has different statistical behavior across different receivers, such as correlation and higher-order moments, in this simulation the probability space is taken to be uniform for each direction/channel, i.e. a sphere. It is possible to manipulate the shape of the threshold limits in a way that does not alter the overall P_{FFD} but does reduce $P_{FFD/CH}$. By attacking right where the Boolean Threshold meets the Averaging Threshold, the higher-density probability region in this 3-D probability space is sacrificed to collect more volume in return. As Figure 4.29 states, this is an equal probability swap in 3-D which results in a net probability gain (for passing the system) in 2-D under the assumption of a channel fault. This means that the P_{FFD} remains the same while the $P_{FFD/CH}$ actually decreases. This alteration is what defines the Oblique Radial Method.

Figure 4.29 shows this trade graphically. It shows all three axis/channels, with channel X_3 coming out of the page. The line of equal probability (the "Probability Isocline") shows that what is lost will have a higher probability density because of its proximity to the origin. Consequently, the "Gained" (blue) area is larger than the "Lost" (yellow) area even though they are equal in total probability in the 3-D space. When there is a large channel fault on X_1, what is left is a net probability gain in the remaining two dimensions. The effect is that there is less likely to be a fault-free detection on either of those two channels; thus there will be an overall reduction in $P_{FFD/CH}$.

120
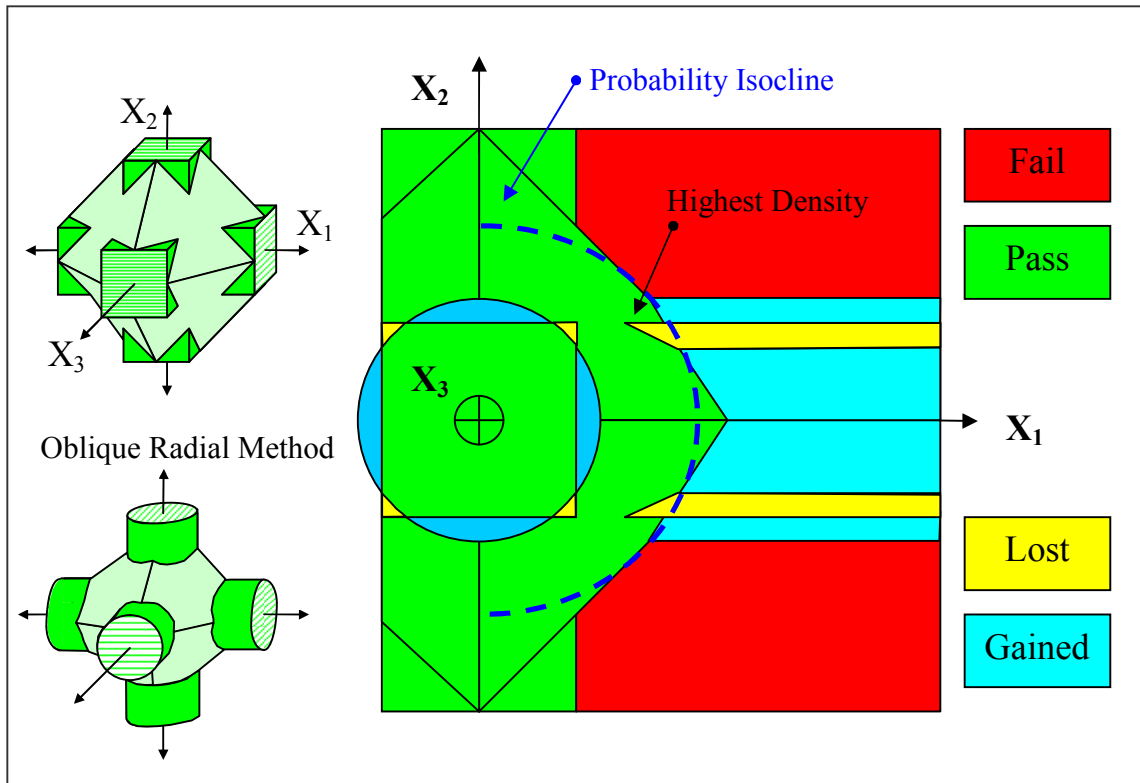
Figure 4.29: Equal Probability Swap in 3-D Space is a Net Gain in 2-D Space

To review, the analysis methodology used to reach this point has progressed through several steps. Initially, the intent was to make a comparison of the Boolean Method with an Averaging Method and determine if any form of compromise could be made to improve overall performance. This resulted in the design of the Oblique Method (M_O). After determining that there was another characteristic of interest, $P_{FFD/CH}$, the results of the new method, M_O, was examined using this criterion. The last directive was to modify M_O if possible to sustain the improvements in MDE but to also provide the best $P_{FFD/CH}$ performance. This resulted in the Oblique Radial Method (M_{OR}). Figure 4.30 briefly sums up this progression.

Step 1) Focus on MDE
  Compare Boolean, AVG methods

$(2^+/3)$ AVG

Step 2) Hybridize methods to reduce MDE
  Identify $P_{FFD/CH}$ criteria, observe effect *(passive)*

M_O

Step 3) Design method to minimize $P_{FFD/CH}$ given MDE
  Use a radial 2-CH bound
  Pass $\equiv (A_i^2 + A_j^2 \leq R^2) \cap (\bar{A} \leq \bar{A}_{LIM})$

M_{OR}
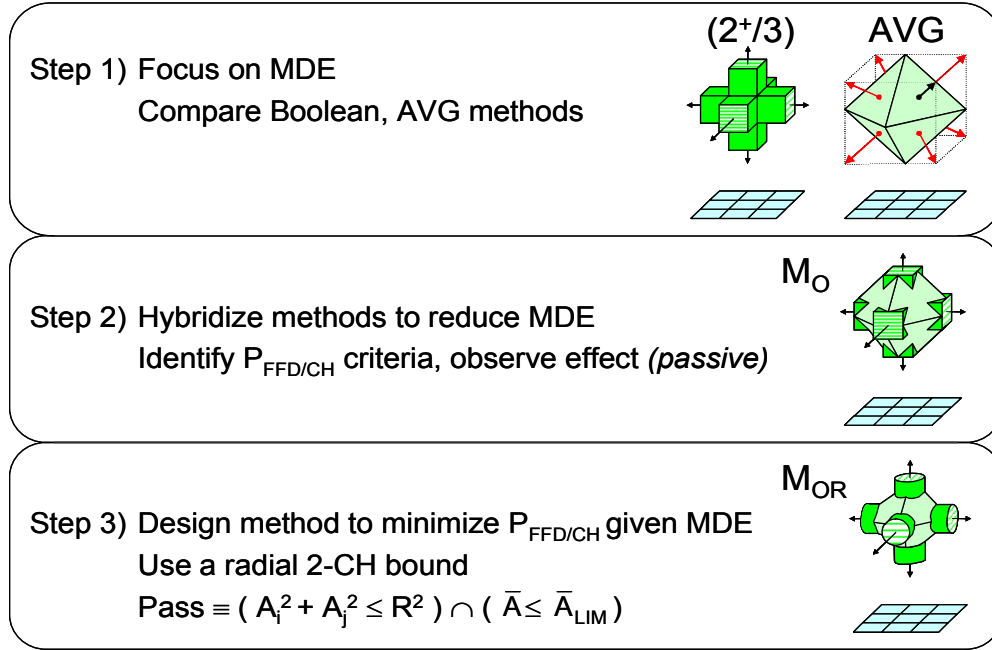
Figure 4.30: Engineering Methodology & Progression of Design Complexity

SV Fault Effect

Previously, the theoretically derived Oblique Method was applied to the data and showed equal fault detection performance compared to the $(2^+/3)$ Method but worse performance for a single-channel fault scenario. Figure 4.31 shows the results of applying the data, which simulates an SV fault, to the newly devised Oblique Radial Method. The Oblique Radial Method has the same vector fault performance as the Oblique Method, as expected.
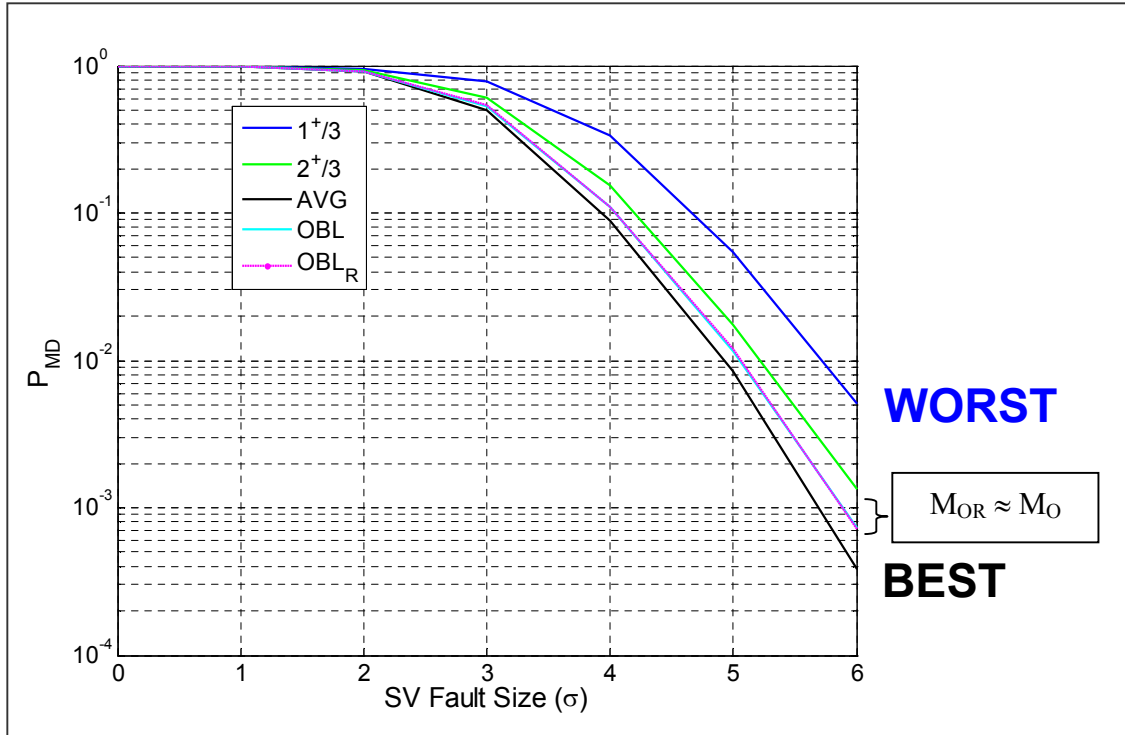
Figure 4.31: M_{OR} Performance Comparison

Figure 4.31 shows that this modification to the Oblique Method to create the Oblique Radial Method did not affect the MDE with respect to the Oblique Method. This was deliberate and desired, and due to the fact that only the thresholds along the axes were changed. The threshold for the mean vector, i.e. where $X_1 = X_2 = X_3$ was not changed and this is the primary factor affecting MDE. Figure 4.32 shows that the objective of reducing $P_{FFD/CH}$ was also achieved. This means that the Oblique Radial Method outperforms the previous Boolean Method in both the primary objective (smaller MDE) and the newly identified secondary objective (smaller $P_{FFD/CH}$).
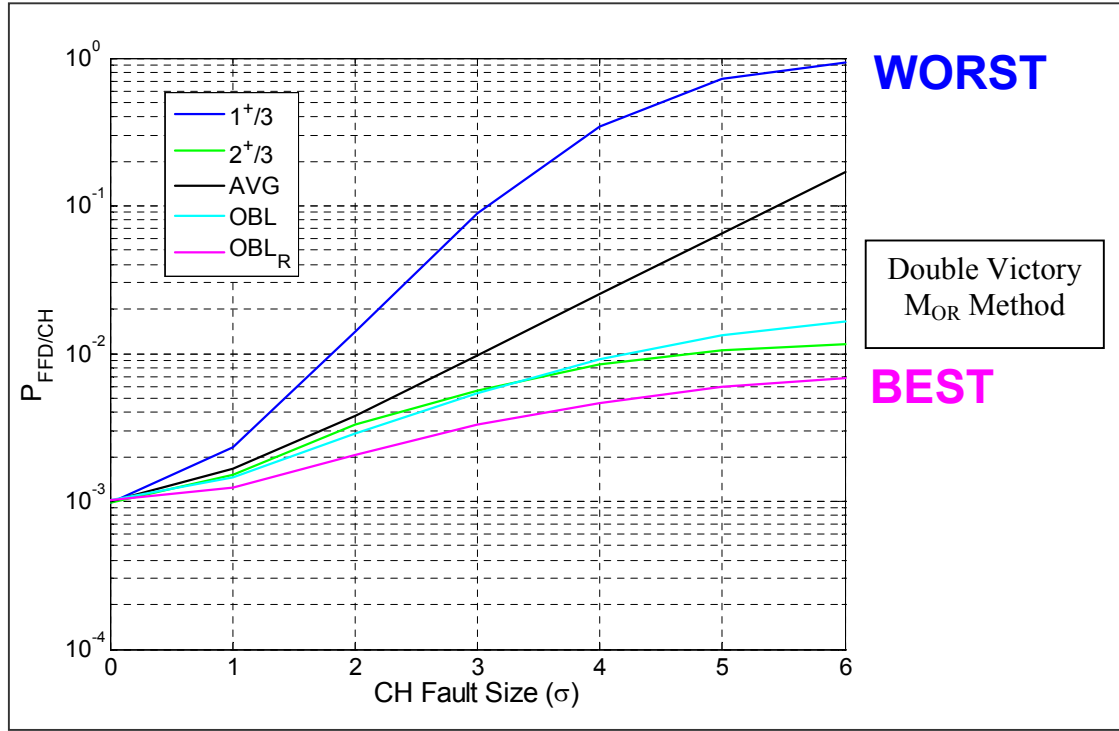
Figure 4.32: $P_{FFD/CH}$ Performance Advantage of M_{OR} Method

4.5.3 Summary of Data Results

The previous section showed the successful results of the newly designed method in terms of being more sensitive to vector faults while being more robust to single-channel faults. For this low-threshold comparison test, the primary objective was to lower the MDE while the secondary criterion was to lower the $P_{FFD/CH}$. Table 4.4 summarizes a comparison of the performance of the newly devised Oblique Radial Method against the $(2^+/3)$ Method. Keeping the P_{FFD} and P_{MD} constant for both methods, the Oblique Radial Method (M_{OR}) exceeds the performance of the $(2^+/3)$ for the criteria identified thus far.

Criterion	$(2^+/3)$	M_{OR}
P_{FFD}	=	=
P_{MD}	=	=
MDE	**X**	✓
$P_{FFD/CH}$	**X**	✓

Table 4.4: Overall Comparison of $(2^+/3)$ M_B and M_{OR} Methods

4.6 Projection / Data Characterization

The data analysis covered in the following sections is applicable to other GPS monitoring systems, however, the carrier-phase acceleration estimates were generated using the Stanford IMT MQM. The Stanford IMT thresholds have been refined over the examination of many 24-hour periods of data. Using a different form of EXM logic while still meeting the continuity and integrity criteria requires a re-examination of the GPS data. The monitor used here for the assessment of EXM is the MQM Acceleration estimator. This data needs to be characterized in order to determine what the thresholds should be for each EXM method. The three concepts to consider for the MQM acceleration estimates are: 1) the correlation across receivers; 2) the standard deviation and 3) the kurtosis, to determine if the distribution is heavy-tailed.

It is important to determine the association between these three issues. For example, MQM Acceleration estimates are affected by satellite elevation. At low elevations, the signals from GPS satellites traverse more of the ionosphere. The ionosphere is a region of charged particles roughly 50 km to 1000 km above the Earth [9]. This error source is thus correlated among the three receivers at the ground station. Even through the thresholds are already a function of elevation, should this potential correlation be considered to further modify the thresholds? Two issues to examine with respect to the three concepts mentioned above are elevation dependency and the time-of-day effect.

4.6.1 Elevation Dependency

Referring to Figure 4.33a, MQM Acceleration test statistic values are plotted as a function of elevation. These figures show the full 24 hours of the dataset. The increase in variance for lower elevations is obvious. The MQM Acceleration values are scaled by an elevation-dependent normalizing function defined within the IMT. Figure 4.33b which shows the normalized acceleration estimates. There appears to be a drop-off in the range of Acceleration estimated values, but this is in fact the results of a change in the sampling size; i.e., there are fewer satellites above 70° elevation in this data set.
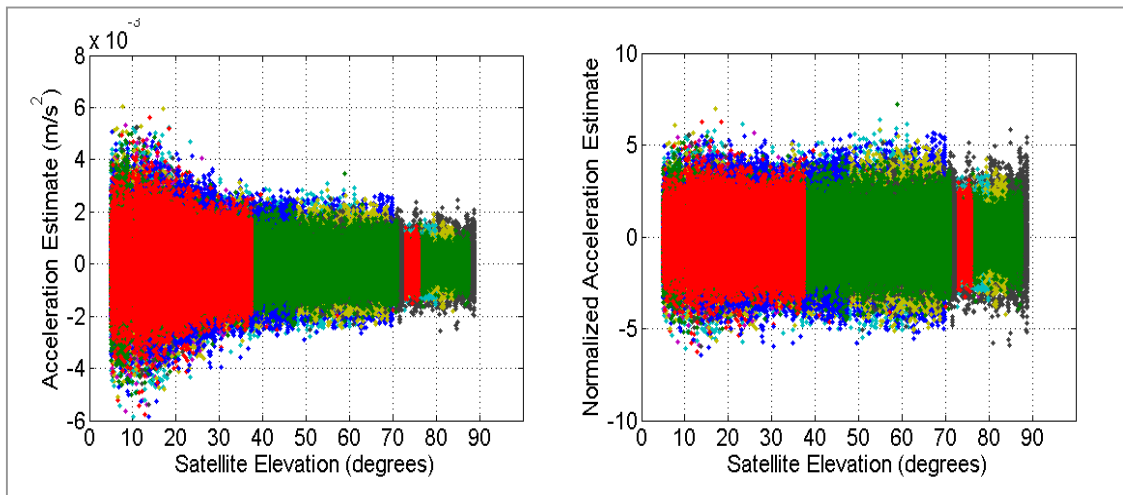
Figure 4.33: a) Raw Acc. and, b) Normalized Acc. Estimate vs. Elevation

Figure 4.34a shows the elevation vs. time for all the tracked satellites. There are actually more than 31 peaks because some satellites make two shallow (low-elevation) passes per day. Figure 4.34b shows a histogram of those elevations. Around 70°, there is a peak with an immediate drop off. This is due to the fact that several satellites reached their zenith within this elevation bin. Again, what appears to be a decrease in the variance is really just a decrease in the data population. There is a sharp drop off for low-elevation satellites because of the difficulty in tracking below 10° with the IMT antennas, which are sited in a rooftop location susceptible to multipath. However, the $(0^o - 5^o)$ bin is not empty, because this histogram reflects the tracked data, not the processed measurements, which have an elevation mask angle of 7.5^o.

126
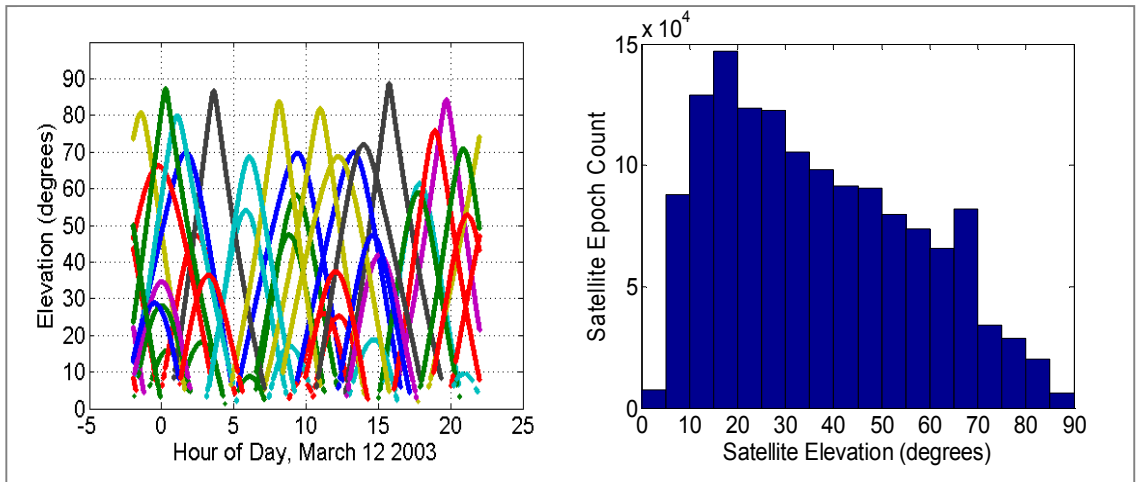
Figure 4.34: a) Time Plot and, b) Histogram of Satellite Elevations

4.6.2 Time-of-Day Effect

What is also neglected in the current IMT is that thresholds are only dependent upon satellite elevation (and receiver), but there are other measurable factors which can directly effect the amount of noise on the measurements. Figure 4.35 (a & b) shows the Acc data of two satellites, PRN 5 and PRN 8. Clearly all signals are not created equal.
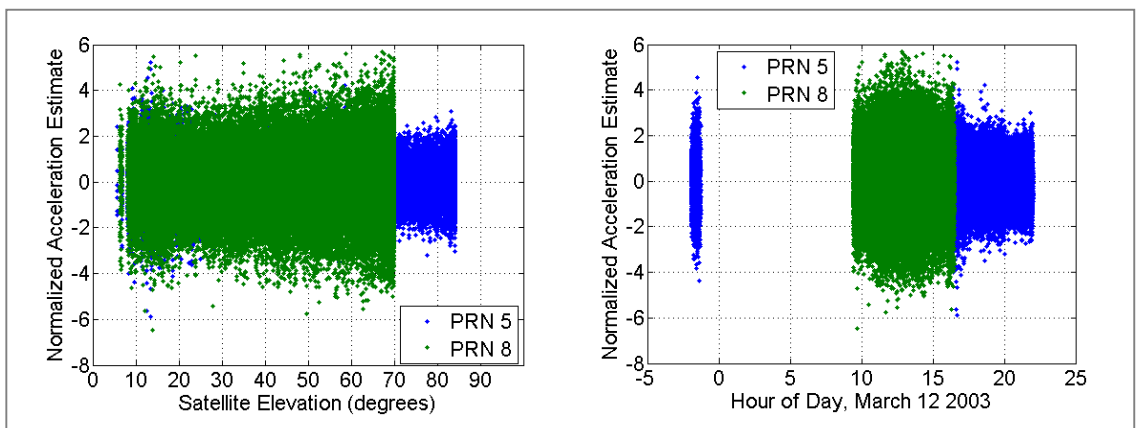
Figure 4.35: Normalized Acceleration vs. a) Satellite Elevation, b) Time-of-Day

Plotting the satellite data with as much description as possible can expose any distinguishing factors which are relevant in describing the data. For example, when the

data is subdivided by satellite, sometimes a particular satellite will have made two shallow passes through the sky instead of one somewhat-overhead pass. It is important to identify this because satellites with shallow passes have more of a tangential trajectory with respect to a ground-based receiver, and their expected rate of change of the carrier phase will be different than for a satellite at equal (low) elevation but on a path directly towards and overhead of the receiver. Figure 4.36 uses PRN 1 and PRN 11 with different colors and terms to refer to the issue of how many passes the satellite makes in one day. PRN 1 makes one pass and will be denoted as (1/1). PRN 11 makes two visible passes, and these two passes will be described as (1/2) and (2/2), respectively.
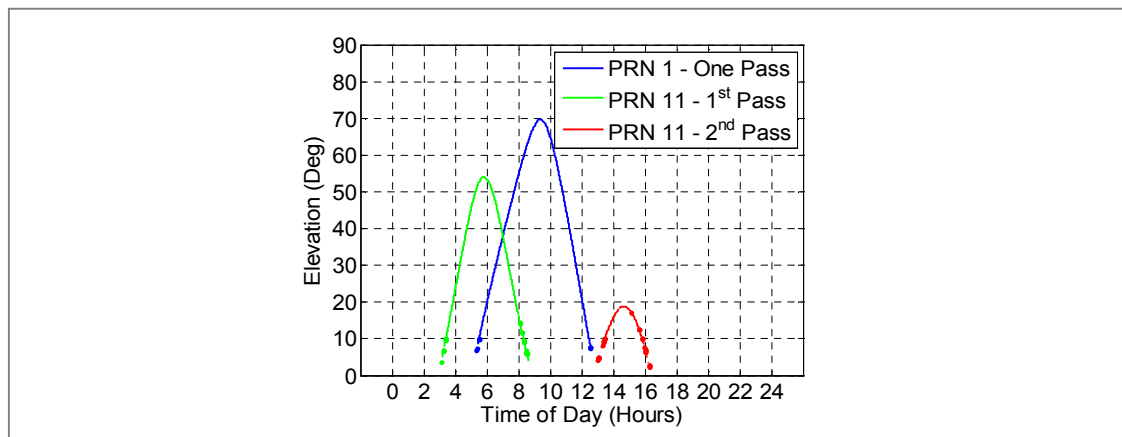
Figure 4.36: One-Pass and Two-Pass Satellites

Figure 4.37 (a & b) use the conventions described above and combine the estimated correlation of each channel with the calculated standard deviation of that data set. The peak elevation of each satellite is included (face color – the color of the interior of the shape) in addition to the receiver number (edge color – the color of the border of the shape). The shape of the icon declares what type of pass the satellite made: (1/1), (1/2), or (2/2). Figure 4.37 (a & b) shows two plots of the normalized standard deviation for a particular satellite and for a particular time of day. The time of day is given as the time of zenith for each satellite.

128
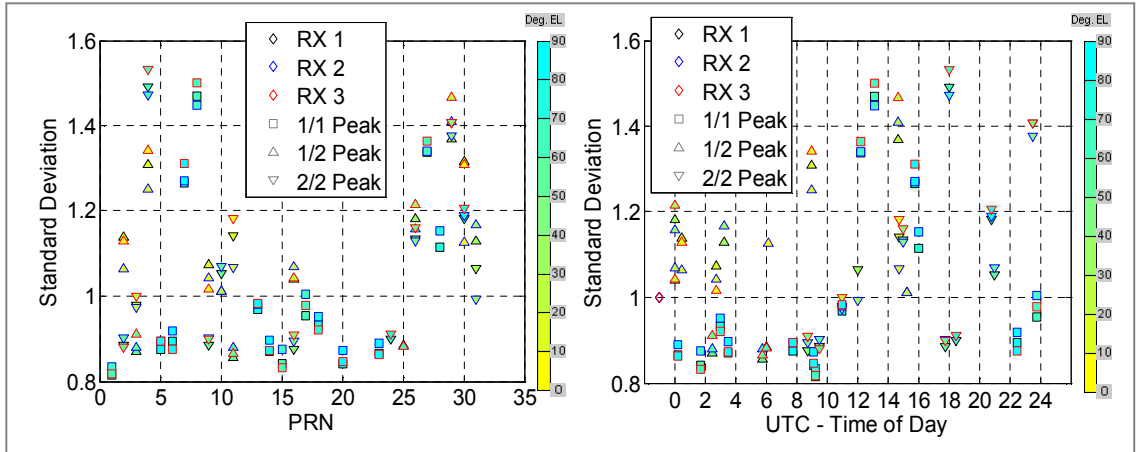
Figure 4.37: Standard Deviation vs. a) PRN, and b) Time of Day

Figure 4.37b suggests there may be some sigma inflation around the hours of 2 - 4 PM. The Klobuchar model of the ionosphere expresses the maximum delay at 2 pm local time [1]. The change in delay associated with the ionosphere is not itself a concern when estimating acceleration of the carrier-phase. The rate of change of this delay is too slow to affect the acceleration estimates, but it is possible that along with the increase in delay there is also an increase in the variation of the delay.

Figure 4.38 shows the kurtosis estimate as a function of the time of day. Kurtosis is the statistic describing the fourth-order moment of a distribution and is used to asses whether a distribution is heavy-tailed [8]. For quick reference, a Gaussian distribution has a kurtosis of three, and a distribution with a kurtosis above three is considered to be heavy-tailed, meaning there is more probability at higher values given a mean equal to zero and a variance of unity. Kurtosis is further addressed in Section 4.6.5.

129
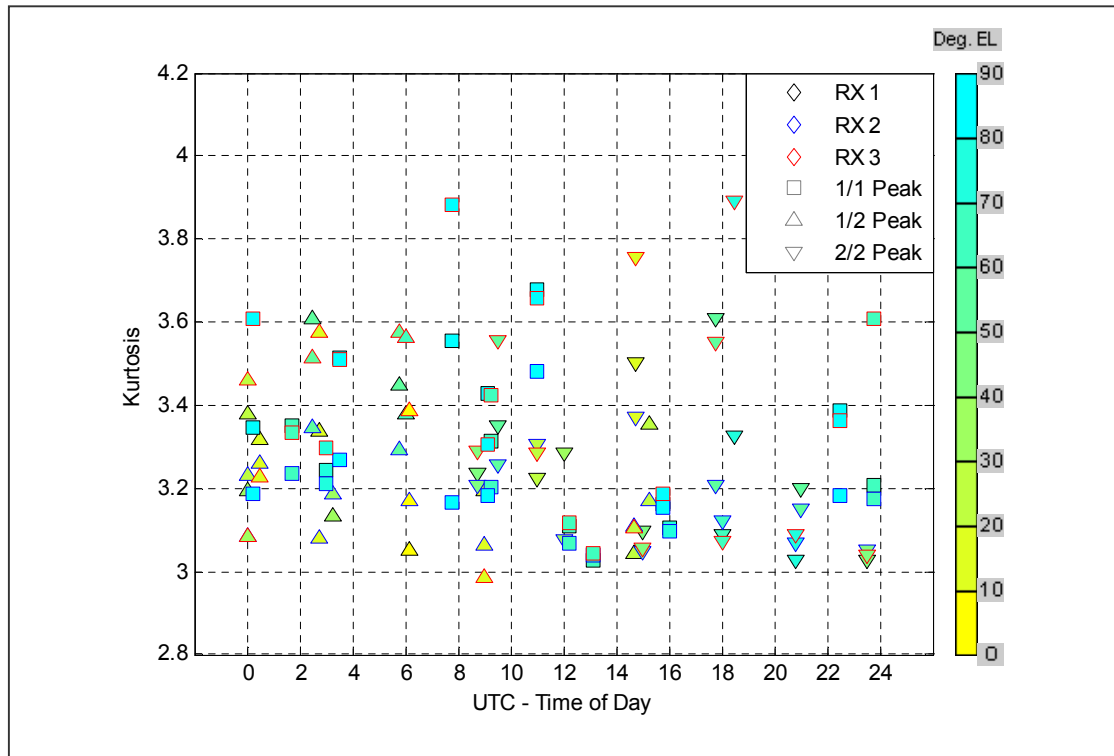
Figure 4.38: Kurtosis vs. Time of Day

In this simplified analysis of the effect of the ionosphere, the time of day is associated with the time of satellite zenith. It is appropriate to also consider the local time of the Ionosphere Pierce Point (IPP), which is the location where the GPS signal penetrates the ionosphere based on the 2-D ionosphere "shell model," which is typically represented by a thin shell roughly 350 km above the Earth's surface [10]. With respect to associating the effect of the ionosphere with the variance of the estimated parameter, Figure 4.39 illustrates that the most relevant time of day is not calculated at the ground but at the IPP.
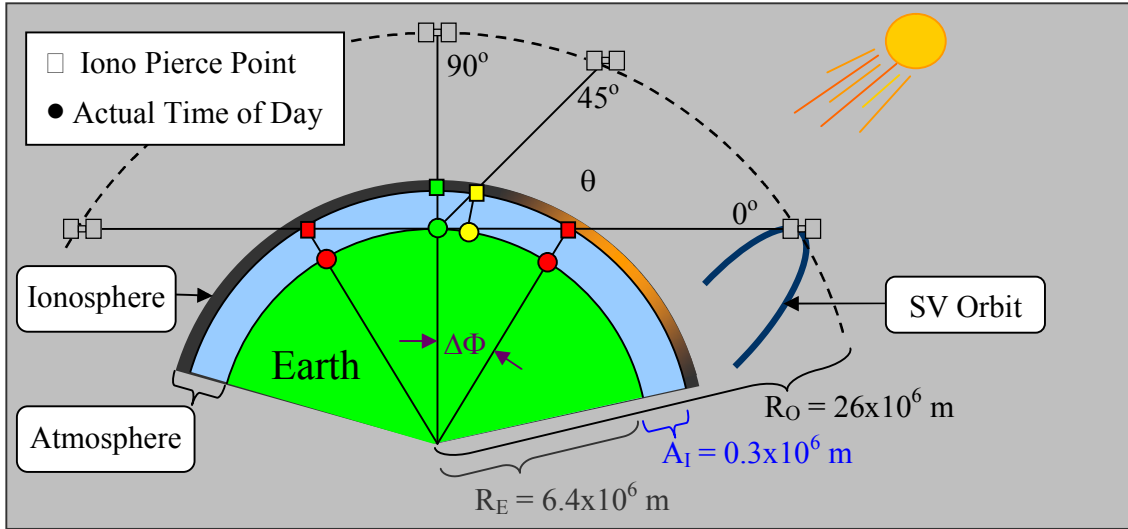
Figure 4.39: Ionosphere Pierce Point and Time of Day

In Figure 4.39, the relative thickness of the atmosphere is exaggerated with respect to the radius of the Earth, but it shows that the two satellites at opposite horizons may experience vastly different ionospheric conditions even though they are assigned the same Time-of-Day values in the previous analysis (because they reach zenith simultaneously). The signal from the satellite to the right passes through an ionosphere being affected by the Sun much more than the signal from the satellite to the left. Further complicating the issue is the fact that the angle $\Delta\Phi$ may express a difference in longitude or latitude. A change in latitude will have the same IPP Time-of-Day, but the effects of a GPS signal traversing the polar ionosphere vs. the equatorial ionosphere are quite different [33]. In general, we will analyze $\Delta\Phi$ as though it were a difference in longitude. Equation (4.15) gives the relationship between the satellite elevation, θ, and the change in longitude, $\Delta\Phi$, given also the radius of the Earth, R_E (~6,400 km), and the approximate altitude (or 2-D shell height) of the ionosphere, A_I (350 km), is:

$$\tan(\theta) = \frac{(R_E + A_I) \cdot \cos(\Delta\phi) - R_E}{(R_E + A_I) \cdot \sin(\Delta\phi)} \tag{4.15}$$

Figure 4.40 shows how much the longitude of the IPP would change at the equator given the elevation of the satellite.
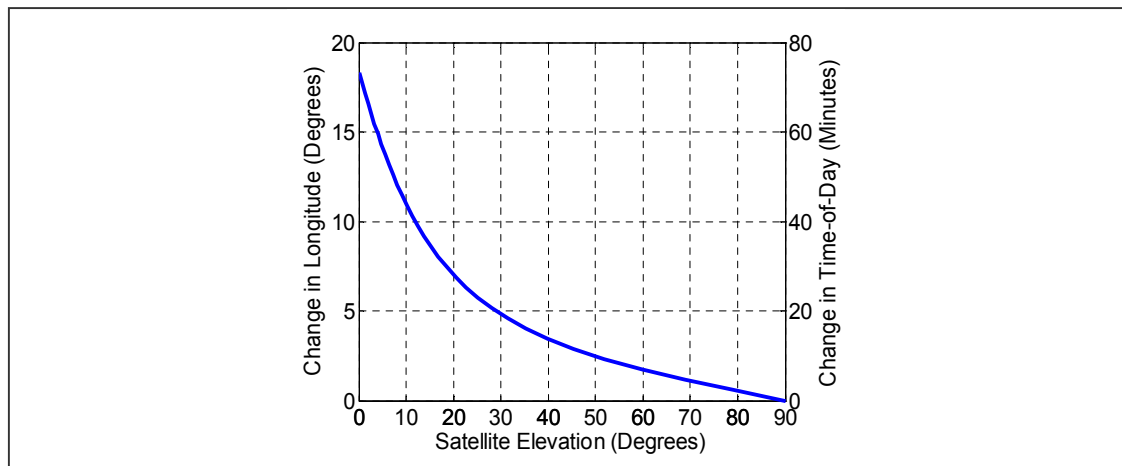
Figure 4.40: Longitude/Time of Day Effect vs. Satellite Elevation

Figure 4.40 shows that any satellite which reaches zenith at the horizon (i.e., a 0° elevation angle) will have a 18° shift. At 30° elevation or more, there is less than a 5° shift of longitude. At 15°/hr (the longitudinal rate of change caused by Earth's rotation), this results in only a 20-minute effect on the actual time of day. The significance of this is that the maximum error of any datum would only be 20 minutes, and any effect between time of day and Ionosphere on the measurements would still be observable. This figure can also apply to a latitudinal shift. The effect of the ionosphere on MQM measurement noise is not a consideration of the current IMT. This section has shown that an ionosphere dependent testing threshold may be appropriate for future refinement of the MQM. In order to evaluate the performance of the EXM, the following section of this thesis will use an estimate of correlation, standard deviation, and kurtosis for use in the large-threshold-model simulation and leave it to future work to determine if the normalizing thresholds should consider more parameters than just elevation.

4.6.3   Correlation

Three-receiver data is used in the IMT, but to allow an easier visual comparison, two-receiver data is shown in Figure 4.41. The horizontal axis shows the normalized Acc estimate (or Z-Score) for Receiver 1, while the vertical axis shows the same estimate

calculated on Receiver 2. Receiver 3 data is not shown in this plot. The data for two PRNs are shown in different colors. The correlation across receivers is unmistakable for PRN 8 (green) because of its elliptical scatter plot. PRN 5 (blue) has an insignificant amount of correlation, evident by its circular scatter plot. The reason for correlation across receivers lies in what is common across receivers, which is the satellite, the transmission medium, and the path of travel. It is also possible for the measurement processing to induce a correlation in the estimated correlation across receivers, but the effect is not as pronounced as from the list just mentioned. Since the receiver clock bias is removed at each epoch, if one satellite was experiencing a large acceleration the other satellites would see a small negative image of that acceleration on each receiver. Accordingly, this would cause a small correlation across receivers for each satellite. However, if the actual acceleration of the faulted satellite was large, it would be detected and removed, thus creating no image on the other satellites. What is addressed here is the impact the correlation has on the exclusion thresholds. Is there any association between correlation and high sigma/low kurtosis data or the low sigma/high kurtosis data? Instead of plotting all correlation estimates separately, they are combined with the standard deviation plots in the next section.
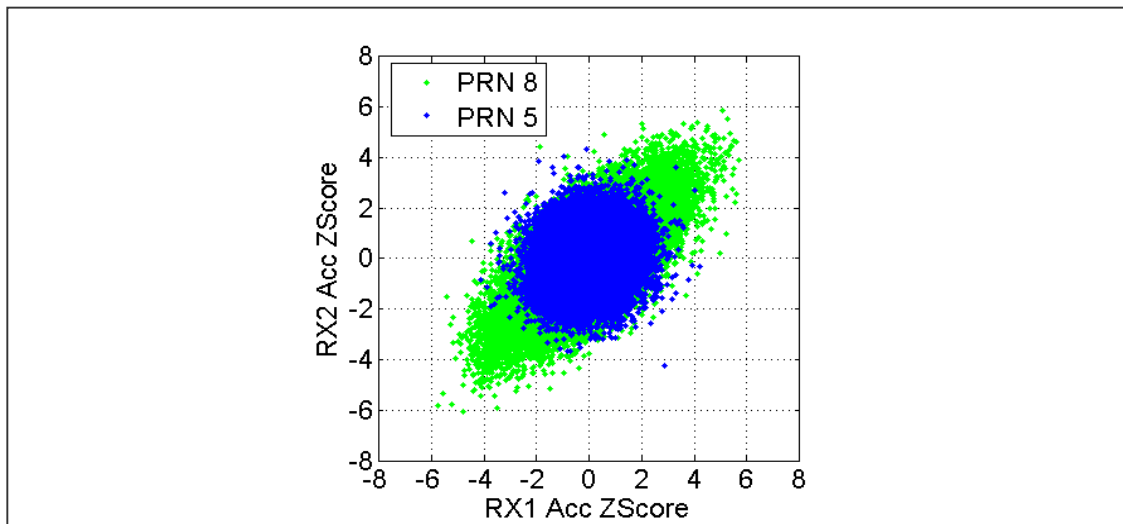
Figure 4.41: Evidence of Prominent Correlation

4.6.4 Standard Deviation

It is cumbersome to represent the cross correlation of a set of three elements with only one parameter. Therefore, in the plot below, when the couplet of correlation and standard deviation is given for Receiver 1, it denotes the correlation between Receivers 1 and 2 and the standard deviation of Receiver 1. When the couplet is given for Receiver 2, it denotes the correlation between Receivers 2 and 3 and the standard deviation of Receiver 2. Finally, when the couplet is given for Receiver 3, it denotes the correlation between Receivers 3 and 1 and the standard deviation of Receiver 3. These parameter definitions are formalized in the following equations:

$$RX_1 \equiv \left\{ \sigma_{RX_1}, \rho_{RX_1 RX_2} \right\} \tag{4.16}$$

$$RX_2 \equiv \left\{ \sigma_{RX_2}, \rho_{RX_2 RX_3} \right\} \tag{4.17}$$

$$RX_3 \equiv \left\{ \sigma_{RX_3}, \rho_{RX_3 RX_1} \right\} \tag{4.18}$$

Figure 4.42 shows a clear connection between correlation and standard deviation. Remember that the data shown in this figure is for the carrier-phase acceleration estimates. The higher-elevation passes strongly support this relation, whereas the lower-elevation passes show the most deviation from this trend. Figure 4.42 shows that PRN 5 reaches 85° elevation and PRN 8 reaches 75° elevation. Both have high elevation passes and have kurtosis and cross-receiver correlations consistent with the trend line in Figure 4.42. Within the IMT, the aggregate standard deviation is normalized; thus the average standard deviation of the data shown in the plot should equal unity. With a clear view from the receivers to the satellite and little multipath, the dominant source of correlation across the receivers for high-elevation satellites will be the ionosphere. At lower elevations, multipath exerts much more of an influence. The locations of the IMT antennas were constrained by the limited size of the HEPL roof and are known to be significantly impacted by multipath. Previous work has shown there to be correlated multipath on these receivers, though this figure clearly shows the correlation to be strongest for higher-elevation satellites.
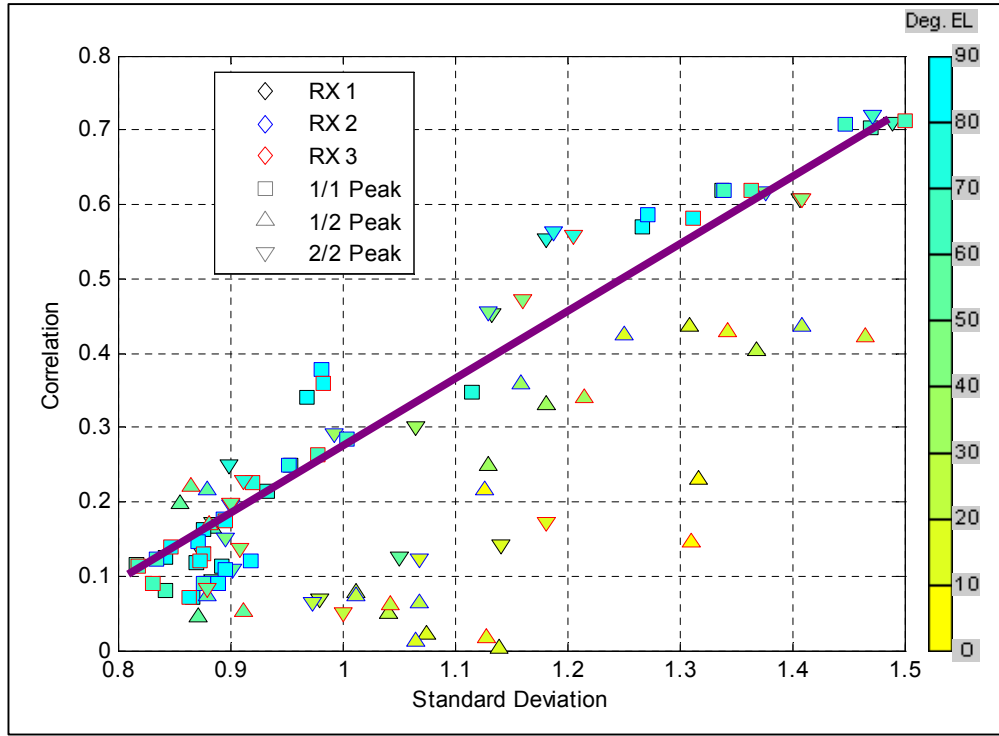
Figure 4.42: Correlation vs. Standard Deviation

4.6.5 Kurtosis

Since much GBAS design and testing is intended to detect faults which might otherwise exist at the outer tails of an un-faulted distribution, there must be as accurate a representation of those tails as possible to be able to statistically separate faulted from rare but un-faulted behavior. Those tails can be approximately described by the kurtosis estimate, which is the ratio of the fourth moment to the square of the second moment of the relevant probability distribution (assuming it is known).

$$k = \frac{M_4}{M_2^2} = \frac{\int\limits_{-\infty}^{\infty} (p(x) - \mu)^4}{\left(\int\limits_{-\infty}^{\infty} (p(x) - \mu)^2\right)^2} = \frac{\int\limits_{-\infty}^{\infty} (p(x) - \mu)^4}{(\sigma^2)^2} \qquad (4.19)$$

135

The multivariate Student t-distribution is used to capture the significance of the increased kurtosis [8]. Figure 4.43 (a & b) shows a comparison of the probability density functions of a Gaussian distribution and a scaled Student t-distribution. The Student t-distribution is described solely by its degree-of-freedom parameter (*n*), and both distributions have been scaled to have a mean of zero and a variance of unity. The log-scale plot of Figure 4.43b most clearly shows the distinction between the distributions. Despite having a variance equal to the Gaussian distribution, the Student t-distribution has more probability concentrated in the tails.
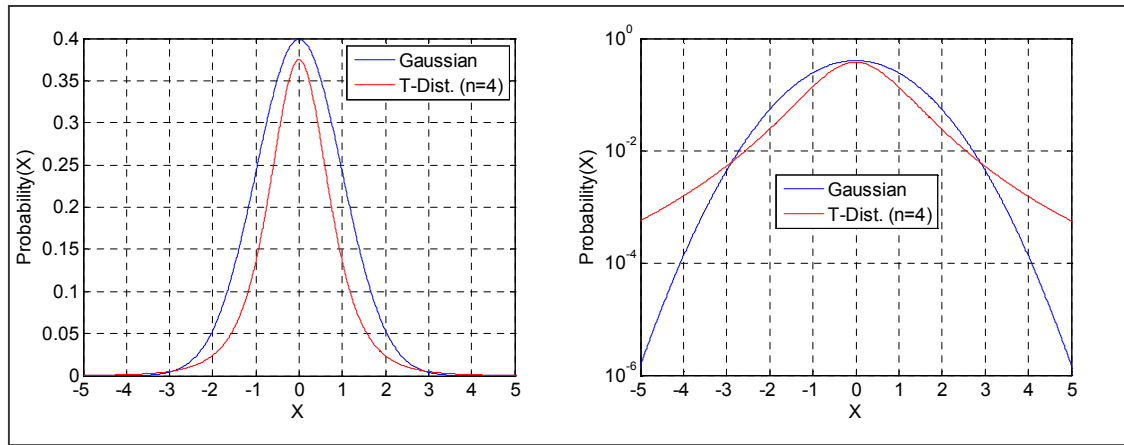
Figure 4.43: Gaussian and Student t-Distributions, a) Linear, b) Log

Since the distribution tails are the primary concern, the distribution can be overbounded with a Gaussian curve with even larger sigma, although this tends to be very conservative [11]. Typically, a Gaussian curve is used for overbounding but that is not obligatory. A distribution which is more representative of the data allows for more accurate simulations and results. Because the IMT data examined here appears to be heavy tailed, the well-known Student t-distribution is an appropriate choice. Table 4.5 compares the first four parameters of a Gaussian distribution to a Student t-distribution. The fact that the Student t-distribution is heavy-tailed in comparison to the Gaussian distribution is evident in the fact that the kurtosis of the Student t-distribution can be greater than that of the Gaussian distribution.

136

Distribution	Gaussian	Student t
Parameters	{ μ - mean, σ^2 - variance}	n - Degrees of Freedom
Mean	μ	0
Variance	σ^2	$\dfrac{n}{n-2}$
Skew	0	0
Kurtosis	3	$\dfrac{6}{n-4}+3$

Table 4.5: Comparing Gaussian and Student t-distribution

With respect to the data, Figure 4.44 shows the difference between two satellites, PRN 5 and PRN 8. There is certainly a difference in both the standard deviation and the shape of these two distributions. PRN 8 has a relatively large sigma and appears more Gaussian in the tails. PRN 5 has a noticeably smaller sigma but is decidedly non-Gaussian.
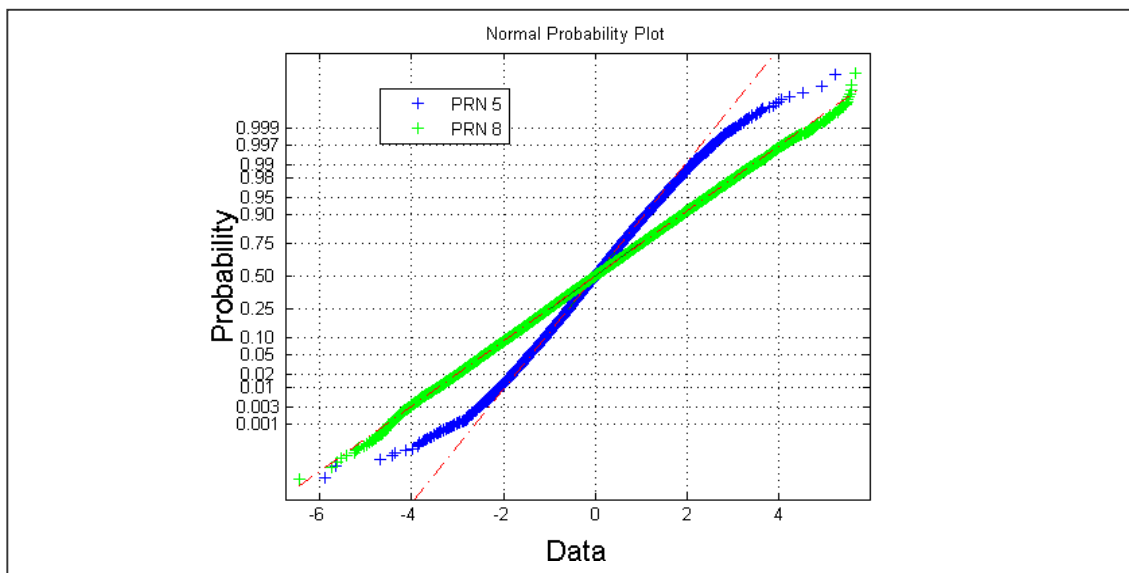
Figure 4.44: Normal Probability Plots PRN 5 & PRN 8

Figure 4.45 shows that when kurtosis is plotted against standard deviation, there is a mild but clear inverse relationship, particularly among higher-elevation satellites. The

specific IMT receiver also plays a role. It appears as though Receiver 2 has the lowest consistent kurtosis, whereas Receiver 3 has the highest. Referring back to Figure 2.2, Receiver 3 is the one on the Lower Roof of the Stanford HEPL Building and suffers from the highest multipath error. The data for lower standard deviations is varied, but as the standard deviation increases, the kurtosis becomes consistently lower until it is nearly that of a Gaussian distribution ($k = 3$).
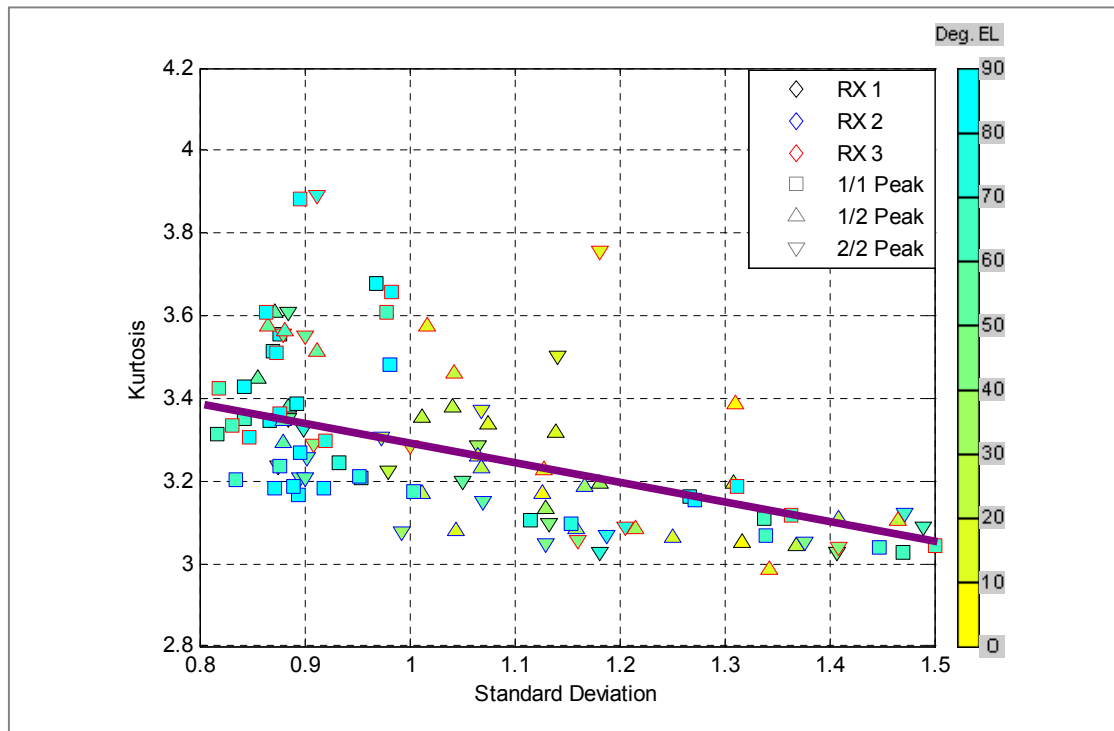
Figure 4.45: Kurtosis vs. Standard Deviation

The data is consistent with an inverse relationship between kurtosis and standard deviation. Particular outliers in the data can be identified and examined to see if anything peculiar is driving that data point or if these values of kurtosis and standard deviation truly characterize that data set. As shown in Figure 4.46a, PRN 13 on Receiver 1 ($k = 3.6792$, $\sigma = 1.185$) is inconsistent with other values and is driven by a section of data when PRN 13 was at about 20° elevation. Taking only the data after 8:30 AM UT (i.e., 8.5 Hrs) reduces the kurtosis to $k = 3.3981$, which fits perfectly with the other data. Also, the two non-contiguous segments of Acc data on the zoomed-in plot are when the

satellite was rising and the receiver lost lock at low elevation. It is possible that the particular receiver being examined was experiencing unusual multipath during this time. Examining data from all three of the receivers give kurtoses of { 3.6792, 3.4808, 3.6564 } for the entire PRN 13 data set, and kurtoses of { 3.3981, 3.3668, 3.4518 } for the reduced PRN 13 data set.

The most relevant question now is whether to include such aberrant data in setting thresholds. For this work, the answer is no. The intent here is to determine the relationship between correlation, standard deviation, and kurtosis for representative data that can be used to test new EXM algorithms. One of the core issues in determining how to set thresholds is whether or not to exclude what seems to be spurious data. Ultimately, the true intent of such a detection algorithm is to detect rare events with large values that may pose a threat to the operation of the system. The reason those points may be excluded here is because the IMT prototype system that collected the data is known to be non-ideal because of its rooftop siting constraints. This is evident by the effect of multipath discussed in Section 4.6.4. An operational LAAS would be implemented at a site which has greater freedom in separate the antennas and minimize the potential for correlated multipath. However, if a more comprehensive data set with ideally sited antennas still contains numerous or sustained aberrations, then those events are no longer aberrations but valid characteristics of the data.
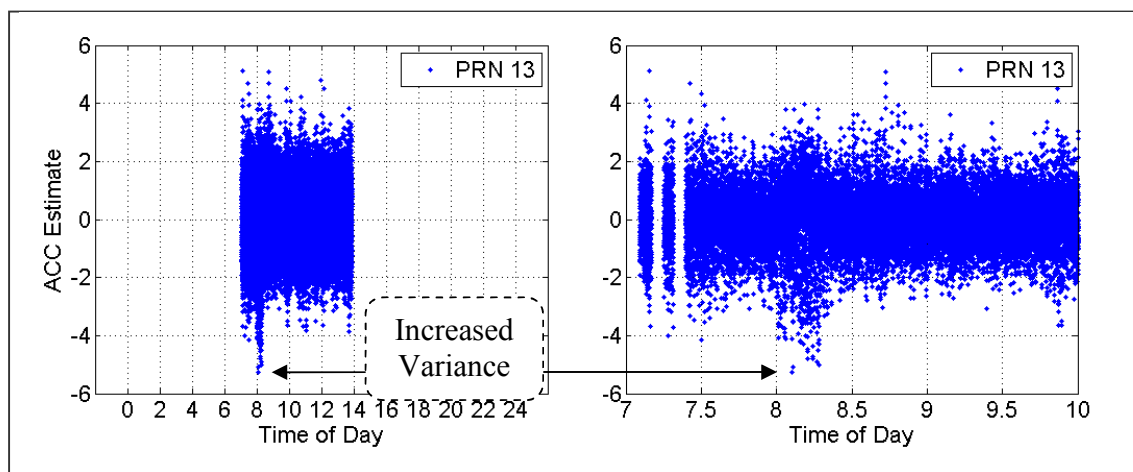
Figure 4.46: a) Unusually High Kurtosis for PRN 13, b) Zoomed-In View

Summarizing the results of Sections 4.6.3-4.6.5, the model used to represent the data for testing of large thresholds has a standard deviation of $\sigma = 1$, a kurtosis of $k = 3.3$ (from Figure 4.45), and a correlation of $\rho = 0.3$ (from Figure 4.42).

4.6.6   Full Simulation Results

Using the specifications of $\mu = 0$, $\sigma = 1$, $k = 3.3$, $\rho = 0.3$, and the equation relating kurtosis to number of degrees-of-freedom from Table 4.5, this leads to $n = 24$ degrees-of-freedom. Once the degrees-of-freedom value is set, the distribution needs to be scaled to match the required variance. If a random variable is scaled by α, then the variance, σ^2, becomes $(\alpha \cdot \sigma)^2$, but the kurtosis remains unchanged because it is actually scaled by the square of the variance. This property is clear from Equation (4.19). The derivation of the degrees-of-freedom, n, is given in Equation (4.20). The solution for the scaling factor, α, is given in Equation (4.21) using the requirement that $(\alpha \cdot \sigma)^2 = 1$.

$$
\begin{aligned}
n &= \frac{6}{k-3} + 4 \\
&= \frac{6}{3.3-3} + 4 \\
&= 24
\end{aligned}
\tag{4.20}
$$

$$
\begin{aligned}
\alpha &= \sqrt{\frac{n-2}{n}} \\
&= \sqrt{\frac{24-2}{24}} \\
&= 0.957
\end{aligned}
\tag{4.21}
$$

Equation (4.22) specifies that the random variable x is sampled from a Student t-distribution with 24 degrees-of-freedom and is scaled by the factor 0.957 from equation (4.21). This equation describes the distribution for a particular axis, but the multivariate Student t-distribution is required to describe the system as a vector.

$$x \equiv 0.957 \cdot \text{TDist}[\ n = 24\ ] \tag{4.22}$$

The Multivariate Student t-distribution

When the input vector is simply one element, the probability density function (pdf) of the Student t-distribution is defined by Equation (4.23) [8].

$$pdf(x) = \frac{1}{\sqrt{n \cdot \pi}} \cdot \frac{\Gamma\left(\frac{1}{2}(n+1)\right)}{\Gamma\left(\frac{1}{2}n\right)} \cdot \left(1 + \frac{x^2}{n}\right)^{-\frac{1}{2}(n+1)} \tag{4.23}$$

Equation (4.25) expresses the multivariate form of the probability density function of the Student t-distribution, where p is the number of elements in the vector x, n is the degrees-of-freedom, and ρ is the correlation coefficient in the covariance matrix of Equation (4.24) [12].

$$\Sigma = \begin{vmatrix} 1 & \rho & \rho \\ \rho & 1 & \rho \\ \rho & \rho & 1 \end{vmatrix} \tag{4.24}$$

$$
\begin{aligned}
pdf(x) &= \frac{1}{(n \cdot \pi)^{p/2}} \cdot \frac{\Gamma\left(\frac{1}{2}(n+p)\right)}{\Gamma\left(\frac{1}{2}n\right)} \cdot \left(1 + \frac{1}{n} \cdot (x-\mu)^T \cdot \Sigma^{-1} \cdot (x-\mu)\right)^{\frac{1}{2}(n+p)} \\
&= \frac{1}{(24 \cdot \pi)^{3/2}} \cdot \frac{\Gamma\left(\frac{1}{2}(24+3)\right)}{\Gamma\left(\frac{1}{2}24\right)} \cdot \left(1 + \frac{1}{24} \cdot x^T \cdot \Sigma^{-1} \cdot x\right)^{\frac{1}{2}(24+3)}
\end{aligned}
\tag{4.25}
$$

Equation (4.25) is used to create a correlated three-axis probability distribution when x is a vector of three uncorrelated elements sampled from a Gaussian distribution with zero mean and unity standard deviation. Detection thresholds were derived for each of the detection methods previously examined in order to establish a nominal fault detection rate of $P_{FFD} = 10^{-8}$. With regard to the Oblique Radial Method, there is design flexibility with respect to the multiple thresholds needed in order to alter the relative

performance results of MDE, $P_{MD/SV}$, and $P_{FFD/CH}$. The thresholds chosen for this method were meant to emphasize an overall balance in the improvement of each of these three parameters. The results from the simulation are shown in Figure 4.47 and Figure 4.48. The injected bias and channel faults were simulated in increments of 1σ from 0 to 16 times the standard deviation of the estimate noise in order to confidently compare the MDE values for each method when P_{MD} and P_{FFD} were as small as 10^{-8}. Figure 4.47 and Figure 4.48 show that the Oblique Radial Method (M_{OR}) has the best overall performance of the methods examined here.
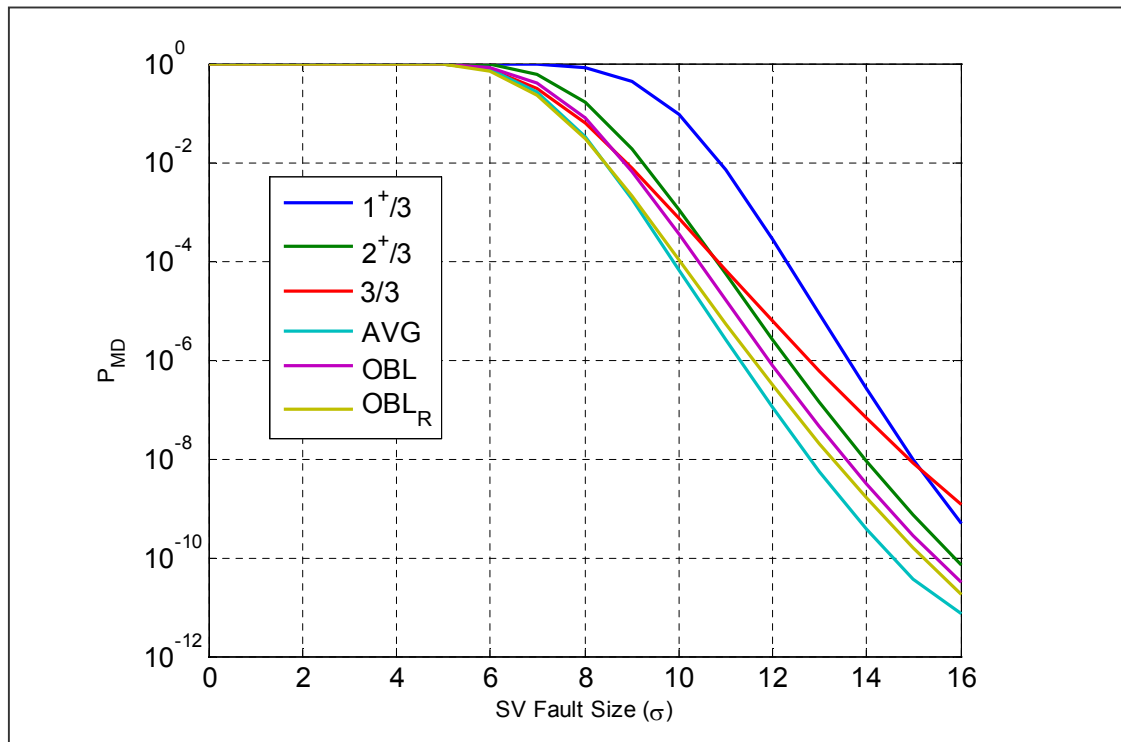
Figure 4.47: Simulation Results for P_{MD}
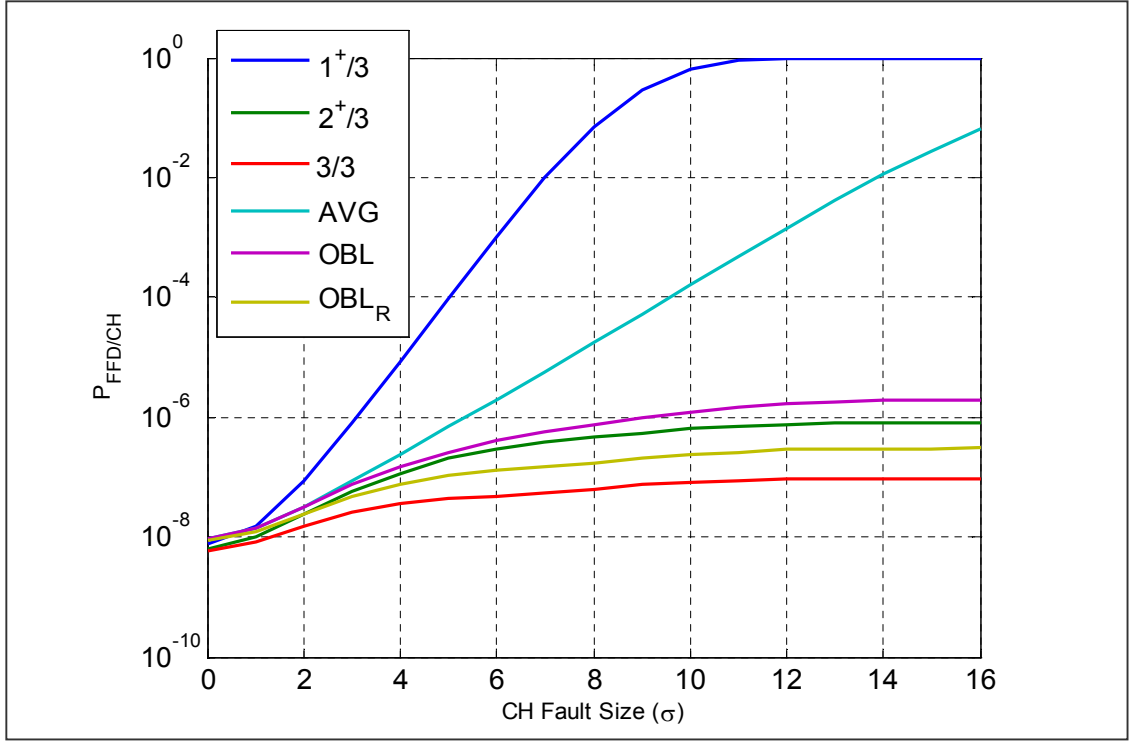
Figure 4.48: Simulation Results for $P_{FFD/CH}$

These results of Figure 4.47 and Figure 4.48 are also summarized in Table 4.6. The conclusion is that, by using the M_{OR} method instead of the old ($2^+/3$) method, we can reduce $P_{FFD/CH}$ by a factor of 2.6 and P_{MD} by a factor of 3.9 (at a fault magnitude of 16σ) while simultaneously lowering MDE, albeit by a modest 4%. These results are consistent with the small-threshold simulation done earlier, which was able to improve $P_{FFD/CH}$ and P_{MD} by roughly a factor of two out to the maximum fault size tested.

	$(1^+/3)$	$(2^+/3)$	$(3/3)$	AVG	M_O	M_{OR}
T_{BOOL}	9.51	6.06	5.73	-	6.79	-
T_{AVG}	-	-	-	6.53	6.06	6.06
T_{Radial}	-	-	-	-	-	8.57
P_{FFD}	10^{-8}	10^{-8}	10^{-8}	10^{-8}	10^{-8}	10^{-8}
MDE ($P_{FFD} = 10^{-8}$) ($P_{MD} = 10^{-8}$)	15.1	13.9	14.9	12.8	13.5	13.3
$P_{MD/SV} = 16\sigma$	4.99×10^{-10}	7.03×10^{-11}	1.17×10^{-9}	7.36×10^{-12}	3.13×10^{-11}	1.80×10^{-11}
$P_{FFD/CH} = 16\sigma$	1	7.88×10^{-7}	9.49×10^{-8}	6.44×10^{-2}	1.95×10^{-6}	3.02×10^{-7}

Table 4.6: Methods, Thresholds, and Simulation Results

4.7   EXM Exclusion Conclusions

This chapter demonstrates that there is a better way to determine if a satellite is worthy of being flagged than the previous method. The method of the previous Stanford IMT removes a satellite if an error was detected on two out of three channels; i.e., receivers tracking that satellite. This method uses a Boolean form of logic and disregards the fact that a faulty satellite is likely to have an error which is observable across all receivers. If the detection method only utilized averaging across receivers, there is a possibility that an egregiously bad measurement on just one of the receivers could cause the decision logic to exclude that satellite. The method devised in this chapter shows a way to use a modified version of the Boolean Method with the method of averaging to create a method called the Oblique Radial Method (M_{OR}). The results of theoretical analysis, testing of real data under small thresholds, and a simulation capturing the main characteristics of the data distribution all show that the newly devised method will modestly, but comprehensively, outperform the previous method.

Chapter 5 - The Vector-Matrix-Tensor Method

The "Vector-Matrix-Tensor" or "VMT" method is an algorithm for determining the output probabilities of the EXM decision logic by simulating the span of pass/fail input channels with a prescribed probability for each channel failure. Specifically, it determines how the candidate EXM logic responds to the vector, matrix, or tensor faults described earlier in Figure 4.4. This chapter demonstrates how detection algorithms meant to target one fault may or may not affect the detection rates of other types of faults.

5.1   Why do it this way?

In order to calculate the exclusion probabilities in the simulation from Chapter 5, significant computation was needed to simulate the three-element vector probability density function of the t-distribution. That was just for three channels. Here, there are three dimensions and 60 ($10 \times 3 \times 2$) channels being simulated. Calculating the conditional probabilities of the ($RX \times SV \times FQ$) tensor is impractical; thus a systematic method to determine the EXM output for a prescribed set of inputs is required. The formulation of the conditional probability equations is ungainly, but actually interpreting the multiple layers of EXM logic is the true inhibitor. For example, it might not be too onerous to derive the probability of detecting a SV Vector Fault when the "truth" was an RX Vector Fault ($P_{SV/RX}$), but calculating the probability of that happening without FQ or SV Matrix Detection also occurring includes too many varieties to operate upon numerically (without resorting to Monte Carlo simulation). Furthemore, how does EXM respond when there is a fourth reference receiver? What about the use of signals on a second GPS frequency? What if the receivers also make use of Galileo (the future European satellite navigation system) and/or GLONASS (the Russian satellite navigation system)? An additional advantage of this approach is that each satellite can be given its

own detection profile. This allows each channel to be given its own statistical characteristics of variance, correlation, and kurtosis, as described in detail in Section 4.6.

5.2   Untracked Channels?

Since the EXM exclusion logic (Section 4.1) only apply to flagged channels, the possibility of untracked channels does not affect this method. VMT has the ability to register de-facto exclusion, but the untracked case must be handled differently. This is because a satellite going untracked (when it is expected to be tracked based on the knowledge of its orbit location) does not by itself support a fault diagnosis. In other words, not being included is fundamentally different from being excluded. If the satellite were flagged as being faulted, it must undergo a recovery period where it is scrutinized before being allowed to be included in any future solution. That aspect of the EXM is not addressed here; it is covered thoroughly in [15].

5.3   Channel Correlation

Correlation across multiple receivers tracking a satellite was addressed in Chapter 4.6.3 in the development of MQM test thresholds. In that discussion, a model with variance, kurtosis, and correlation across receivers was used to describe the entire population of estimated carrier-phase acceleration data in order to determine the frequency of fault detections. That method used a sampling of the 3-D probability density function of the homoskedastic (i.e., having equal and finite variance), multivariate t-distribution and accumulated the sample probabilities of all points deemed as faulted by a given detection rule. This method required a large sample size and forced a common t-statistic degree of freedom (which dictates the distribution variance) across channels. This method is not helpful for deriving the probability distribution when the channels have unequal variance and thus require a heteroskedastic model. Furthermore, it becomes infeasible when considering additional channels because of the exponential sample-size growth due to the extra dimensions of sampling and integration.

In this chapter, EXM logic is restricted to operating on a simple pass/fail Boolean vector structure. For example, for a three receiver installation, each element of the Boolean vector describes whether that receiver detected some fault with that satellite. Such a structure for a satellite tracked on three receivers might be {Pass, Pass, Fail}. That does not imply that the first two receives have passed the satellite (they may not be tracking it), only that they have no reason to declare a fault for that satellite. Using pre-defined probabilities for fault-free detections and missed detections, those structures could be derived from the Binomial distribution, which requires only two parameters, n (the number of elements) and p (the probability of an element being True/One). However, the standard Binomial distribution does not incorporate a correlation term, and this will prove to be a clear impediment towards accurately modeling the failure probabilities.

To examine this further, in a three-receiver system, the probability of a fault on a given channel can be set to p (by manipulating the detection threshold), but using the Binomial distribution shown in Equation (5.1) means that this will also immediately determine the set of probabilities shown in Equations (5.2-5.5), which assume statistical independence and identical distributions for faults on these three receivers. The reason for this is that, when the number of channels, n, is specified, the only real degree on freedom in this probability equation is p, as k is just an index for the number of occurrences.

$$P_B\left(k, n, p\right) = \binom{n}{k} \cdot p^k \cdot \left(1 - p\right)^{n-k} \tag{5.1}$$

Equations (5.2-5.5) evaluate Equation (5.1) where k is the number of detected faults and p is replaced by p_{FFD} (the probability of a fault-free detection on a channel).

$$P\left[k = 0\right] = \left(1 - p_{ffd}\right)^3 \tag{5.2}$$

$$P[k=1] = 3 \cdot p_{ffd} \cdot \left(1 - p_{ffd}\right)^2 \tag{5.3}$$

$$P[k=2] = 3 \cdot p_{ffd}^2 \cdot \left(1 - p_{ffd}\right) \tag{5.4}$$

$$P[k=3] = p_{ffd}^3 \tag{5.5}$$

Using a correlated Binomial model gives the freedom to assign different probabilities to different satellites to reflect realistic differences among them. Models like this have been derived in the financial world, including Moody's Correlated Binomial distribution, but this model is limited to correlation among identically-distributed random variables, which is not helpful in this case because the distributions of the parameters being estimated cannot be assumed to be identical across receivers [34].

An empirical approach to this problem is to simply measure the detection rates for a set of data of a given satellite, but then there is the limitation in the amount of sampled data, as mentioned above. Figure 5.2 shows the detections rates for the two satellites PRN 5 and PRN 8 which were analyzed in Chapter 4. However, there were only about 40,000 samples, and to model the detection response down to the order of 10^{-8}, the data has to be extrapolated.
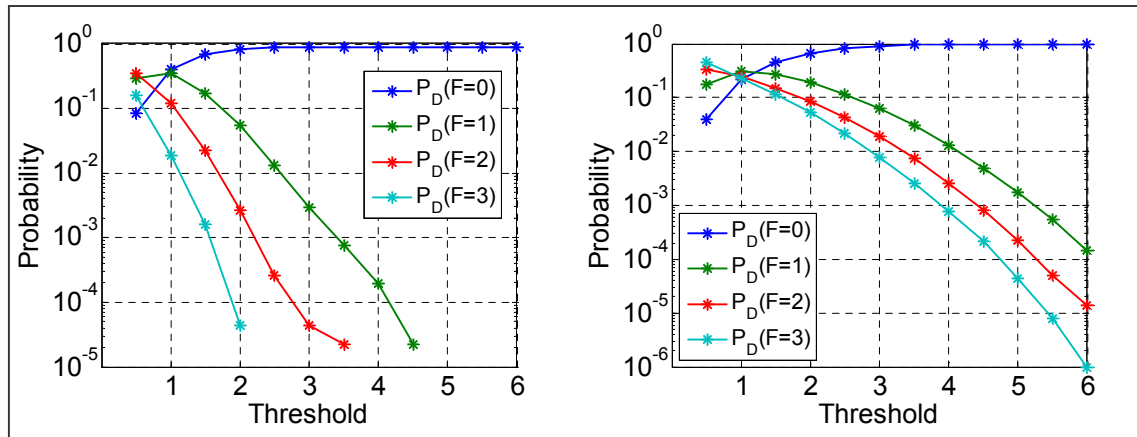
Figure 5.1: Detection Response for a) PRN 5, b) PRN 8

148

5.3.1 Correlated Multivariate Student t-distribution

The approach proposed in this section is to derive a model for the measured data that can be used to generate a stream of random data which mirrors the fundamental properties of the measured data. That data can then be tested to determine its detection profile down to the required specification and achieve the necessary extrapolation.

There are many forms which can satisfy a finite set of requirements, but we are looking for a computationally direct method which can easily be expanded to additional dimensions without requiring multivariate solutions. The following lower triangular matrix representation allows for the solution of the full covariance in addition to the kurtosis of each channel. The parameters are easy to solve for, and the simulated results agree well with the data, as is described in detail below.

Equation (5.6) describes the model where $\{x_1, x_2, x_3\}$ simulates the three-element vector representing data of one satellite tracked on three receivers. The correlation across channels is created by the matrix multiplication of the randomly-generated t-distribution values $\{t_1, t_2, t_3\}$.

$$\begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = \begin{pmatrix} a_1 & 0 & 0 \\ b_1 & b_2 & 0 \\ c_1 & c_2 & c_3 \end{pmatrix} \cdot \begin{pmatrix} t(v_1) \\ t(v_2) \\ t(v_3) \end{pmatrix} \tag{5.6}$$

PRN 5 is used as an example to calculate the values of the lower triangular representation. This method requires relationships for the second and fourth moments for a t-distributed random variable as well as their cross-correlations. The values for PRN 5 are given in Equations (5.7-5.9), while the symbolic forms follow in Equations (5.10-5.12) [35].

$$\Sigma_{PRN5} = \begin{vmatrix} 0.7149 & 0.1220 & 0.1347 \\ 0.1220 & 0.7938 & 0.1342 \\ 0.1347 & 0.1342 & 0.7460 \end{vmatrix} \tag{5.7}$$

$$E\left[PRN5^4\right] = \begin{vmatrix} 1.6601 & 1.9788 & 1.8200 \end{vmatrix} \tag{5.8}$$

$$k_{PRN5} = \begin{vmatrix} 3.2478 & 3.1407 & 3.2709 \end{vmatrix} \tag{5.9}$$

$$E\left[t_k^2\right] = \frac{v_k}{v_k - 2} \tag{5.10}$$

$$E\left[t_k^4\right] = \frac{3 \cdot v_k^2}{(v_k - 2) \cdot (v_k - 4)} \tag{5.11}$$

$$k = \frac{E\left[t_k^4\right]}{E^2\left[t_k^2\right]}$$
$$= \frac{6}{v_k - 4} + 3 \tag{5.12}$$

Equation (5.12) immediately gives the value v_1 from the kurtosis of x_1.

$$v_1 = \frac{6}{k_1 - 3} + 4 \tag{5.13}$$

$$v_{1,PRN5} = \frac{6}{3.2478 - 3} + 4$$
$$= 28.2131 \tag{5.14}$$

With v_1 and the equation of the second moment (5.10), a_1 can be solved for.

$$E\left[x_1^2\right] = a_1^2 \cdot E\left[t_1^2\right]$$
$$= a_1^2 \cdot \left(\frac{v_1}{v_1 - 2}\right) \tag{5.15}$$

$$a_1 = \sqrt{E\left[x_1^2\right] \cdot \left(\frac{v_1 - 2}{v_1}\right)} \tag{5.16}$$

150

$$a_{1,\text{PRN 5}} = \sqrt{0.7149 \cdot \left(\frac{28.2131 - 2}{28.2131} \right)} \qquad (5.17)$$

$$= 0.8150$$

Using the first correlation Equation (5.18) the value of b_1 value can be solved for.

$$E\left[x_1 \cdot x_2 \right] = a_1 \cdot b_1 \cdot E\left[t_1^2 \right]$$
$$= a_1 \cdot b_1 \cdot \left(\frac{v_1}{v_1 - 2} \right) \qquad (5.18)$$

$$b_1 = \frac{E\left[x_1 \cdot x_2 \right]}{a_1} \cdot \left(\frac{v_1 - 2}{v_1} \right) \qquad (5.19)$$

$$b_{1,\text{PRN 5}} = \frac{0.1220}{0.8150} \cdot \left(\frac{28.2131 - 2}{28.2131} \right) \qquad (5.20)$$
$$= 0.1391$$

The value b_2 is solved simultaneously with the number of degrees of freedom, v_2.

$$E\left[x_2^2 \right] = b_1^2 \cdot \left(\frac{v_1}{v_1 - 2} \right) + b_2^2 \cdot \left(\frac{v_2}{v_2 - 2} \right) \qquad (5.21)$$

$$E\left[x_2^2 \right]_{\text{PRN 5}} = 2.0820 \times 10^{-2} + b_2^2 \cdot \left(\frac{v_2}{v_2 - 2} \right) \qquad (5.22)$$

$$E\left[x_2^4 \right] = \frac{3 \cdot b_1^4 \cdot v_1^2}{\left(v_1 - 2 \right) \cdot \left(v_1 - 4 \right)} + 6 \cdot b_1^2 \cdot b_2^2 \cdot \left(\frac{v_1}{v_1 - 2} \right) \cdot \left(\frac{v_2}{v_2 - 2} \right) + \frac{3 \cdot b_2^4 \cdot v_2^2}{\left(v_2 - 2 \right) \cdot \left(v_2 - 4 \right)} \qquad (5.23)$$

$$E\left[x_2^4 \right]_{\text{PRN 5}} = 1.4217 \times 10^{-3} + 0.1249 \cdot b_1^2 \cdot \left(\frac{v_2}{v_2 - 2} \right) + \frac{3 \cdot b_2^4 \cdot v_2^2}{\left(v_2 - 2 \right) \cdot \left(v_2 - 4 \right)} \qquad (5.24)$$

$$\{ b_2, v_2 \} = \{ 0.8593, 44.583 \} \qquad (5.25)$$

The process is repeated for the values c_1 to c_k.

$$E[x_1 \cdot x_3] = a_1 \cdot c_1 \cdot E[t_1^2]$$

$$= a_1 \cdot c_1 \cdot \frac{v_1}{v_1 - 2} \tag{5.26}$$

$$c_1 = \frac{E[x_1 \cdot x_3]}{a_1} \cdot \frac{v_1 - 2}{v_1} \tag{5.27}$$

$$c_{1,\text{PRN 5}} = \frac{0.1347}{0.8150} \cdot \frac{28.2131 - 2}{28.2131}$$

$$= 0.1536 \tag{5.28}$$

$$E[x_2 \cdot x_3] = b_1 \cdot c_1 \cdot E[t_1^2] + b_2 \cdot c_2 \cdot E[t_2^2]$$

$$= b_1 \cdot c_1 \cdot \left(\frac{v_1}{v_1 - 2}\right) + b_2 \cdot c_2 \cdot \left(\frac{v_2}{v_2 - 2}\right) \tag{5.29}$$

$$c_2 = \left(E[x_2 \cdot x_3] - b_1 \cdot c_1 \cdot \left(\frac{v_1}{v_1 - 2}\right)\right) \cdot \left(\frac{v_2 - 2}{v_2}\right) \cdot \frac{1}{b_2} \tag{5.30}$$

$$c_{2,\text{PRN 5}} = \left(0.1342 - 0.1394 \cdot 0.1538 \cdot \left(\frac{28.2131}{28.2131 - 2}\right)\right) \cdot \left(\frac{44.8968 - 2}{44.8968}\right) \cdot \frac{1}{0.8593} \tag{5.31}$$

$$= 0.1236$$

The values of c_3 and v_3 are solved simultaneously using Equations (5.33) and (5.35).

$$E[x_3^2] = c_1^2 \cdot \left(\frac{v_1}{v_1 - 2}\right) + c_2^2 \cdot \left(\frac{v_2}{v_2 - 2}\right) + c_3^2 \cdot \left(\frac{v_3}{v_3 - 2}\right) \tag{5.32}$$

$$E[x_3^2]_{\text{PRN 5}} = 4.1381 \times 10^{-2} + c_3^2 \cdot \left(\frac{v_3}{v_3 - 2}\right) \tag{5.33}$$

$$E[x_2^4] = \frac{3 \cdot c_1^4 \cdot v_1^2}{(v_1 - 2) \cdot (v_1 - 4)} + \frac{3 \cdot c_2^4 \cdot v_2^2}{(v_2 - 2) \cdot (v_2 - 4)} + \frac{3 \cdot c_3^4 \cdot v_3^2}{(v_3 - 2) \cdot (v_3 - 4)} +$$

$$6 \cdot c_1^2 \cdot c_2^2 \cdot \left(\frac{v_1}{v_1 - 2}\right) \cdot \left(\frac{v_2}{v_2 - 2}\right) + 6 \cdot c_1^2 \cdot c_3^2 \cdot \left(\frac{v_1}{v_1 - 2}\right) \cdot \left(\frac{v_3}{v_3 - 2}\right) + \tag{5.34}$$

$$+ 6 \cdot c_2^2 \cdot c_3^2 \cdot \left(\frac{v_2}{v_2 - 2}\right) \cdot \left(\frac{v_3}{v_3 - 2}\right)$$

152

$$E\left[x_2^4\right]_{\text{PRN 5}} = 5.3346 \times 10^{-3} + 0.2483 \cdot c_3^2 \cdot \left(\frac{v_3}{v_3 - 2}\right) + \frac{3 \cdot c_3^4 \cdot v_3^2}{(v_3 - 2) \cdot (v_3 - 4)} \qquad (5.35)$$

$$\{c_3, v_3\} = \{0.8034, 23.8259\} \qquad (5.36)$$

The numerical results for the model given in Equation (5.6) are summarized in Equation (5.37). Even though the degrees of freedom, v, of the Student t-distribution are typically integer values, the t-distribution is a specific case of the Gamma distribution, and the Gamma distribution allows the use of non-integer values for this parameter [35].

$$\begin{vmatrix} x_1 \\ x_2 \\ x_3 \end{vmatrix} = \begin{vmatrix} 0.8150 & 0 & 0 \\ 0.1391 & 0.8592 & 0 \\ 0.1536 & 0.1236 & 0.8034 \end{vmatrix} \cdot \begin{vmatrix} t_{28.2131} \\ t_{44.5830} \\ t_{23.8259} \end{vmatrix} \qquad (5.37)$$

Equations (5.38-5.40) show the results of this process. Using the initial conditions for the data (PRN5) from Equations (5.7-5.9), a model (*M*) has been created which matches (approximately) the data for the descriptive statistics of variance and fourth-order moment.

$$\Sigma_{PRN5} = \begin{vmatrix} 0.7149 & 0.1220 & 0.1347 \\ 0.1220 & 0.7938 & 0.1342 \\ 0.1347 & 0.1342 & 0.7460 \end{vmatrix}, \qquad \Sigma_M = \begin{vmatrix} 0.7178 & 0.1232 & 0.1347 \\ 0.1232 & 0.7944 & 0.1333 \\ 0.1348 & 0.1333 & 0.7461 \end{vmatrix} \qquad (5.38)$$

$$E\left[(\text{PRN 5})^4\right] = \begin{vmatrix} 1.6601 & 1.9788 & 1.8200 \end{vmatrix}, \quad E\left[M^4\right] = \begin{vmatrix} 1.6702 & 1.9810 & 1.8177 \end{vmatrix} \qquad (5.39)$$

$$k_{(\text{PRN 5})} = \begin{vmatrix} 3.2478 & 3.1407 & 3.2709 \end{vmatrix}, \qquad k_M = \begin{vmatrix} 3.2418 & 3.1393 & 3.2709 \end{vmatrix} \qquad (5.40)$$

This simple algorithm only requires basic calculations using the data itself and simple algebraic solutions for the parameters. The method of lower triangular representation is summarized in Table 5.1.

Step	Measurement	Equation Variables	Variables Known	Solved
1a	k_1	$f(v_1)$	$\{\}$	v_1
1b	$E[x_1^2]$	$f(v_1, a_1)$	$\{v_1\}$	a_1
2a	$E[x_1 \cdot x_2]$	$f(a_1, v_1, b_1)$	$\{a_1, v_1\}$	b_1
2b	$E[x_2^2]$	$f(b_1, v_1, b_2, v_2)$	$\{a_1, v_1, b_1\}$	$\{b_2, v_2\}$
	$E[x_2^4]\}$	$f(b_1, v_1, b_2, v_2)$		
3a	$E[x_1 \cdot x_3]$	$f(a_1, v_1, c_1)$	$\{a_1, v_1, b_1, b_2, v_2\}$	c_1
3b	$E[x_2 \cdot x_3]$	$f(b_1, c_1, v_1, b_2, c_2, v_2)$	$\{a_1, v_1, b_1, b_2, v_2, c_1\}$	c_2
3c	$E[x_3^2]$	$f(c_1, v_1, c_2, v_2, c_3, v_3)$	$\{a_1, v_1, b_1, b_2, v_2, c_1, c_2\}$	$\{c_3, v_3\}$
	$E[x_3^4]$	$f(c_1, v_1, c_2, v_2, c_3, v_3)$		

Table 5.1: Algorithm for Correlated Multivariate t-distribution Model

This formulation can be extended to a larger number of variables by extending Equation (5.6). There is an exact match between the number of degrees of freedom in the formulation and the number of descriptive parameters being modeled. Starting with one variable, x_1, all that is needed is a way to express the variance and kurtosis. These are effectively determined by the terms a_1 and v_1, respectively. Adding a second variable, x_2, requires three additional degrees of freedom, the correlation to the previous variable, the variance, and the kurtosis. Again, these are effectively determined by the terms b_1, b_2, and v_2, respectively. The process is repeated to incorporate additional variables such as x_3 and beyond.

More parameters can be used to express higher-order relationships across receivers, but this leads to calculations such as $E[(x_1 \cdot x_2 \cdot x_3)^2]$, or potentially even higher order terms if more channels are included. This criterion will give weight to the outliers and, given the limited number of samples, can add appreciable error into the model. Lastly, additional terms in the matrix model would destroy the simplicity of the algebraic solutions. Equations (5.41-5.42) show the general relationship between model and data terms and how this model can be extended to a larger number of channels. Note that it is

154

adding additional parameters and not additional channels which is the basis for the increases in the order of the calculations (referred to above). By using the model derived in this section, adding channels does not increase the order, or exponent, applied to the data, and the highest order terms would still only be $E[\ (x_n \cdot x_m)^2\ ]$ and $E[\ (x_n)^4\ ]$.

$$
\begin{vmatrix} a_1 & 0 & 0 \\ b_1 & b_2 & 0 \\ c_1 & c_2 & c_3 \end{vmatrix} \cdot \begin{vmatrix} t_1 \\ t_2 \\ t_3 \end{vmatrix} \approx \begin{vmatrix} \sigma_1^2 & 0 & 0 \\ \rho_{2,1} & \sigma_2^2 & 0 \\ \rho_{3,1} & \rho_{3,2} & \sigma_3^2 \end{vmatrix} \cdot \begin{vmatrix} k_1 \\ k_2 \\ k_3 \end{vmatrix} \tag{5.41}
$$

$$
\begin{vmatrix} a_1 & 0 & 0 & 0 \\ b_1 & b_2 & 0 & 0 \\ \vdots & \ddots & \ddots & 0 \\ n_1 & \cdots & n_{n-1} & n_n \end{vmatrix} \cdot \begin{vmatrix} t_1 \\ t_2 \\ \vdots \\ t_n \end{vmatrix} \approx \begin{vmatrix} \sigma_1^2 & 0 & 0 & 0 \\ \rho_{2,1} & \sigma_2^2 & 0 & 0 \\ \vdots & \ddots & \ddots & 0 \\ \rho_{n,1} & \cdots & \rho_{n,n-1} & \sigma_n^2 \end{vmatrix} \cdot \begin{vmatrix} k_1 \\ k_2 \\ \vdots \\ k_n \end{vmatrix} \tag{5.42}
$$

Both a Monte Carlo simulation and a numerical integration scheme can be used to assess the viability of the method derived which uses the lower triangular representation for the correlated multivariate Student t-distribution. In the Monte Carlo approach, random uncorrelated data from the appropriate Student t-distribution can be generated. That data is then correlated using the appropriate lower triangular matrix. That data can then be compared against a specified threshold, and the number of failures can be counted. The distribution of the $\{0, 1,\ldots, n\}$ number of failures across n channels from the Monte Carl simulation can be compared against the results from the real data. The numerical integration scheme involves integrating a three-dimensional PDF of the Student t-distribution utilizing the transformation matrix from Equation (5.41) to determine the integration limits of each variable.

The Monte Carlo approach is the simplest to implement, and Figure 5.2 shows a comparison between the PRN 5 data and the corresponding model. The fit is excellent down to the expected accuracy of the data. Given that there are about 40,000 data points for PRN 5, we would expect the data sample to reflect the underlying population down to the inverse of about $1/10^{th}$ the sample size, meaning $1/4000 = 2.5 \times 10^{-4}$. This Monte Carlo technique easily generates 10^{10} data points and achieves consistent behavior down

to the 10^{-9} level. The added benefit is that the model can be appended at any time with more data runs under the condition that the underlying random generator is not repeating data from a previous run. Figure 5.2 shows that, if the threshold on the magnitude of each data value to determine whether that channel is passed or failed was set to $T = 3$, then with n being the number of faults on this SV, the probabilities would be:

$$P[n = 0] = 0.996 \tag{5.43}$$

$$P[n = 1] = 3.646 \times 10^{-3} \tag{5.44}$$

$$P[n = 2] = 1.079 \times 10^{-5} \tag{5.45}$$

$$P[n = 3] = 3.646 \times 10^{-8} \tag{5.46}$$

The individual channel fault rates were all at or slightly larger than 10^{-3}. If this system conformed to an uncorrelated binomial distribution, it would be expected that $P[n = 2] \approx 3 \times 10^{-6}$, and $P[n = 3] \approx 1 \times 10^{-9}$. Since the actual results are quite different, it is pertinent to consider incorporating a correlation model within the VMT simulation.
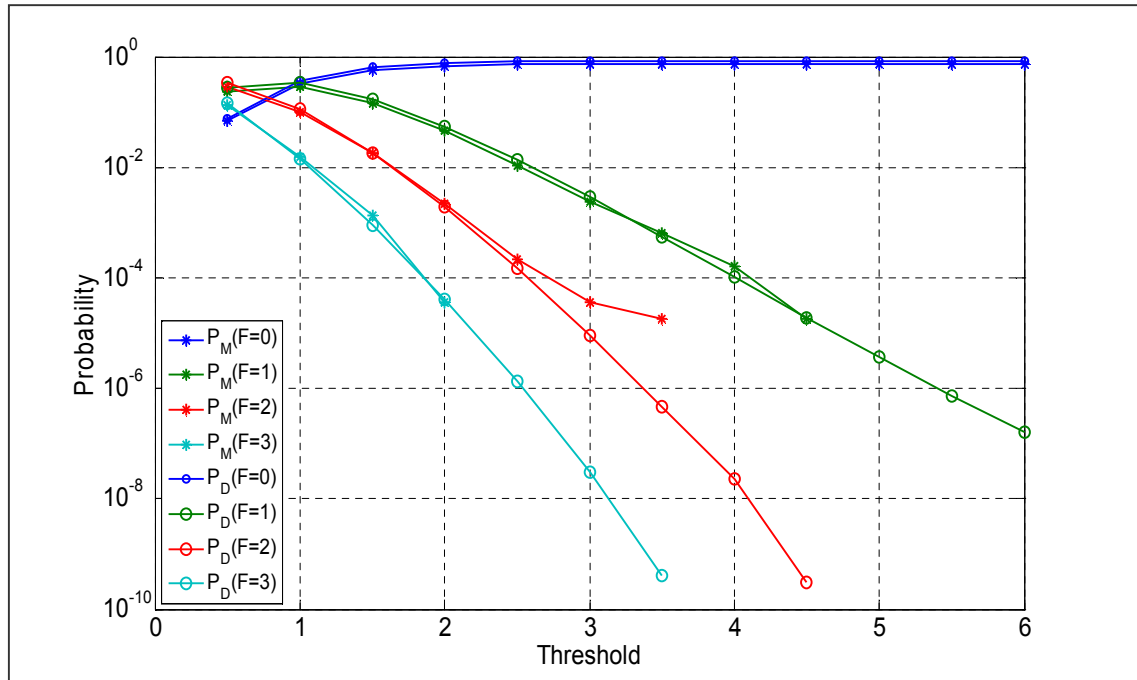
Figure 5.2: LT Model (Monte Carlo) Data vs. PRN 5 Data Fault Rates

Running 10 billion samples within Monte Carlo simulation is not instantaneous, and the obvious drawback is the $\sqrt{n}$ roll-off for convergence, meaning significantly more points would need to be simulated to increase the accuracy of the results. The alternate approach is to perform numerical integration. The process developed in this chapter takes in a vector of sampled values from a Student t-distribution and uses a linear transform to output variables which meet specifications for variance, kurtosis, and correlation. Those sampled values from the Student t-distribution will be generally referred to as *t* variables, and the output values will generally be referred to as *x* variables. Direct integration of those *x* variables is difficult because there is no well-defined probability density function to express a correlated multivariate heteroskcedastic distribution, as was discussed before. It is possible, though, to integrate the *t* variables because they are drawn from independent distributions, and then modify the integration limits based on the parameters of the transformation matrix.

The convenience of the lower triangular representation of the Correlated Multivariate method rests in the ease of specifying integration limits. The data must be categorized in terms of passing or failing each channel. The domains are designated Ω_i and along with the integration equation are given in Equations (5.47-5.50). In these equations, the terms F and P represent the failing and passing regions, which are determined by the threshold T and all the parameters derived to establish the correlation across variables.

$$P[x_1, x_2, x_3 < T] = \int_{\Omega_1} pdf_1(t_1, v_1) \cdot \left(\int_{\Omega_2} pdf_2(t_2, v_2) \cdot \left(\int_{\Omega_3} pdf_3(t_3, v_3) \cdot dt_3 \right) \cdot dt_2 \right) \cdot dt_1 \qquad (5.47)$$

$$\Omega_1 \equiv -\infty \leq F_{t_1} \leq -\frac{T}{a_1} \leq P_{t_1} \leq \frac{T}{a_1} \leq F_{t_1} \leq \infty \qquad (5.48)$$

$$\Omega_2 \equiv -\infty \leq F_{t_2} \leq \frac{-T - b_1 \cdot t_1}{b_2} \leq P_{t_2} \leq \frac{T - b_1 \cdot t_1}{b_2} \leq F_{t_2} \leq \infty \qquad (5.49)$$

$$\Omega_3 \equiv -\infty \leq F_{t_3} \leq \frac{-T - c_1 \cdot t_1 - c_2 \cdot t_2}{c_3} \leq P_{t_3} \leq \frac{T - c_1 \cdot t_1 - c_2 \cdot t_2}{c_3} \leq F_{t_3} \leq \infty \qquad (5.50)$$

This integration scheme is relatively straightforward and enables a solution for thresholds and probabilities that were unachievable in a practical sense using the sampling method. Figure 5.3 demonstrates this, where M_1 (Method 1) is the Sampling Method, and M_2 (Method 2) is the Numerical Integration Method. The extrapolation to $P_{M2}(F = 3) = 10^{-33}$ may overexert the model, but the point is that a method exists to extrapolate the model out to the realm of very high thresholds and very low probabilities.

The intention of developing a model to be able to simulate the low fault probabilities is centered on incorporating the two major issues that affected the results of the prior chapters, which were the correlation and kurtosis. The model serves as a way to evaluate the effectiveness of the exclusion rules being developed, particularly with respect to the correlation, as exclusions result from multiple channel faults. The probability of those channel faults is a function of detection thresholds which are a degree-of-freedom in the model. This enables the exclusion rules to be meaningfully evaluated over a range of channel fault probabilities. Furthermore, the distributions are ultimately reduced to n-tuples of passing or failing channels, reducing the impact of numerical errors at tails of the distributions.
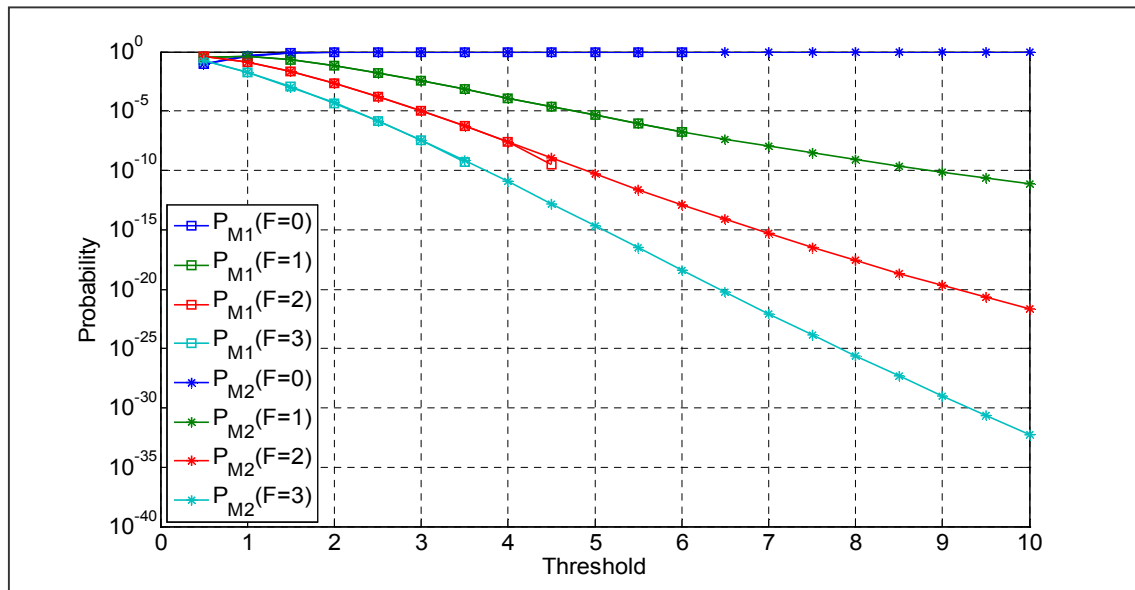
Figure 5.3: Comparison of Monte Carlo and Numerical Integration Techniques

158

So far, a method to calculate a transformation matrix to model satellites simulated in the VMT has been demonstrated. Further, there are two methods to produce the actual fault detection rates for that model. And even though the integration method is superior for the 3-D case, the final consideration is what happens if the system uses a fourth receiver. Let,

k = Number of Receivers (3 – 4 or more)

m = Number of Numerical Integration Subdivisions ($\sim 10^4$)

n = Number of Samples Simulated ($\sim 10^9$)

Generally speaking, $k << m << n$. The number of iterations for the Numerical Integration Method is m^{k-1}. Iterations refer to the number of computational cycles (one of the limiting factors in practical computations), and samples refer to the total number of sample points from the multidimensional probability density function that are ultimately simulated. The reason for the $(k-1)$ exponent is because the known CDF for the Student t-distribution can be used to save one dimension of integration. This means that there are roughly $(10^4)^2 = 10^8$ iterations. The Sampling Method uses $(k \times n)$ samples, which amounts to 3×10^9 iterations. When there are four receivers, this ratio shifts to ($10^{12} : 4 \times 10^9$), meaning that the Sampling Method still has some virtue if the system is expanded because its computational requirement grows linearly, not exponentially. With this utility in hand, the VMT system and its various terms can be examined.

5.4   Defining Matrix Probabilities: P_{UD}, P_D, P_{OD}, P_{CD}

So far, the concepts of P_{FFD} and P_{MD} have been discussed. Next, the following new terms are introduced: P_{UD} (Probability of an Under-Detection), P_D (probability of a Detection), P_{OD} (Probability of an Over-Detection), and P_{CD} (Probability of a Cross-Detection). An Under-Detection implies that the known fault is not fully captured but that what is detected is a subset of that fault. A Detection is a detected fault without

removing more channels than needed (i.e., the optimal result). An Over-Detection means that the fault has been detected but that more channels were excluded than necessary. This could be a natural superset of channels, or it could be that the actual fault and another type of fault were detected – basically a simultaneous Detection and a Cross-Detection. A Cross-Detection is when neither the fault nor the detection channels are subsets of the other. Technically, a fault-free-detection is a form of Over-Detection when no fault exists, and a missed detection is a form of Under-Detection. A Cross-Detection is considered to be an Under-Detection, since the primary concern is whether or not the actual fault is detected and excluded. Figure 5.4 shows these concepts graphically.
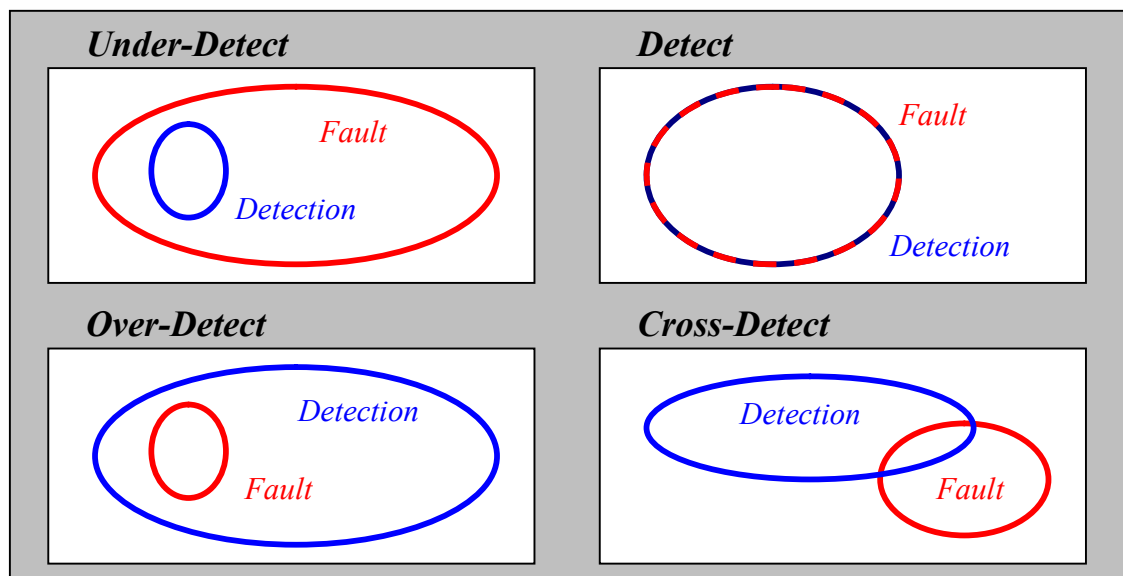
Figure 5.4: Venn Diagrams of Fault/Detection Relationships

The Fault Detection Matrix in Table 5.2 below shows how these labels apply to the categorical faults described earlier. The color coding is meant to imply that, if there is a fault, it is optimal to detect (and exclude) exactly that fault, nothing more or less. An Over-Detection would imply inefficiency, but an Over-Detection is preferable to an Under-Detection because an undetected fault poses a safety threat to the system and user. A Cross-Detection may arguably be the most treacherous because not only has the actual fault been missed, but the system has chosen to exclude an entirely different set of channels. However, it is possible that the integrity monitoring logic may become more conservative in checking for errors if it has detected any type of fault. For example, a

"Check Engine" light on the console of an automobile would suggest that something is wrong with the car's engine. A "Low (Oil) Pressure" light might not directly indicate a problem with the engine, but the driver would likely become more cautious in driving and check the major engine components at the next available time.

		Detection								
		None	CH	V_{RX}	V_{SV}	V_{FQ}	M_{RX}	M_{SV}	M_{FQ}	T
Fault	None	N/A	OD	OD	OD	OD	OD	OD	OD	OD
	CH	UD	D	OD	OD	OD	OD	OD	OD	OD
	V_{RX}	UD	UD	D	CD	CD	CD	OD	OD	OD
	V_{SV}	UD	UD	CD	D	CD	OD	CD	OD	OD
	V_{FQ}	UD	UD	CD	CD	D	OD	OD	CD	OD
	M_{RX}	UD	UD	CD	UD	UD	D	CD	CD	OD
	M_{SV}	UD	UD	UD	CD	UD	CD	D	CD	OD
	M_{FQ}	UD	UD	UD	UD	CD	CD	CD	D	OD
	T	UD	UD	UD	UD	UD	UD	UD	UD	D

Table 5.2: Fault Detection Matrix

The objective of this chapter is to show that rules designed to detect a given fault do not trip another detector with significant frequency and that this allows for effectively independent detection rule construction. This significantly simplifies the design of an integrity/continuity allocated system by removing the need for iterations between monitors in real time.

5.4.1 When is a Fault a Fault?

As explained above, any removal of a set of channels from use in providing corrections should ideally be surgical in precision. If only one channel is faulted, it is wasteful to exclude either the entire receiver or satellite. Similarly, if a receiver or satellite is faulted, it is wasteful to remove an entire frequency which is experiencing

some clustering of faults. Figure 5.5 shows different scenarios where five total faults are observed. What is the appropriate action for each scenario?
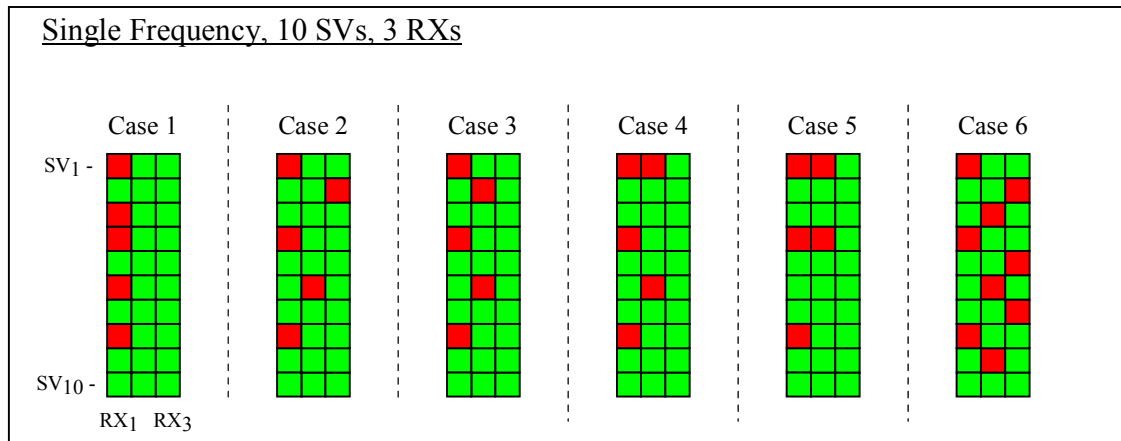
Figure 5.5: Which of these is an SV Fault, RX Fault, or an FQ Fault?

The objective in any system which discriminates faulted channels from healthy ones is to use a clear set of rules. Although the guiding principle may be to use the smallest fault declaration that covers any set of channel flags, this can be ambiguous, as is shown by examining the six cases above. By developing the VMT method, different sets of rules can be applied to all possible channel scenarios to determine which set provides the best protection against faults while minimizing the number of channels unnecessary excluded. The cases examined here use the detection rules from the IMT EXM-I given in Section 2.2.3.1.

Case 1:   This is clearly an RX_1 fault. This receiver would be excluded, and a (10 SV) $\times$ (2 RX) system of channels would remain.

Case 2:   This is likely an RX_1 fault, and this receiver would be excluded. There are other faults, but they do not seem to be correlated and would be considered isolated channel faults.

Case 3:   This is likely an RX_1 and possibly an RX_2 fault. Both would be detected under a $(2^+/10)$ rule used to detect Receiver Faults described in Chapter 5

162

and would result in system unavailability, as only one receiver would remain and provide no comparison (or averaging) for the broadcast corrections.

Case 4:   Likely an SV_1, RX_1, and possibly an RX_2 fault. If SV_1 and RX_1 are excluded, RX_2 is salvageable and the system could remain operational.

Case 5:   Is this two bad receivers or two bad satellites and an unrelated faulty channel? The former seems more likely and would result in system unavailability.

Case 6:   There is no obvious pattern of faults, though there are sufficient faults per receiver to exclude each of them. The sheer number of supposedly-rare faults indicates that the entire frequency may be jammed (or the entire system is faulty); thus exclusion of all measurements is required.

5.5   What are Isomorphs?

Using the $(10 \times 3 \times 2)$ representation of information described earlier, it can be determined when the different rules combine, overlap, or interplay. It would be difficult to directly construct all of the conditional probabilities to do this. For example, what would be the probability of detecting an SV Fault given two RX Faults, one with two channel faults and one with three channel faults, all without detecting an FQ Fault? This would require a massive set of combinatorics. The alternate approach would be to simulate every possible fault profile and see how they respond to a given set of rules. The daunting issue is the sheer number of fault profiles that can result from approximately $10 \times 3 \times 2$ or 60 channels. The fact is that many of these fault profiles will behave identically because, with respect to the rules created for detection and exclusion, they are isomorphs.

Isomorphs are entities which have identical form, even if it is not immediately evident. Mathematically, an isomorphism is a one-to-one correspondence between the elements of two sets such that the result of an operation on elements of one set corresponds to the result of the analogous operation on their images in the other set. For example, with respect to calculating matrix determinants, the following two matrices are isomorphs.

$$m_1 = \begin{pmatrix} 1 & 0 \\ 2 & 1 \end{pmatrix} \tag{5.54}$$

$$m_2 = \begin{pmatrix} 1 & 2 \\ 0 & 1 \end{pmatrix} \tag{5.55}$$

$$\left| m_1 \right| = \left| m_2 \right| \tag{5.56}$$

The method of reducing all, potential fault profiles into a manageable set of unique profiles (in order to enable comprehensive fault simulation) is described in the following sections.

5.5.1 A simplified version

To begin, consider a system having only three satellites and two receivers.

	RX$_1$	RX$_2$	1 = Pass
SV$_1$	1	0	0 = Fail
SV$_2$	1	1	
SV$_3$	1	1	
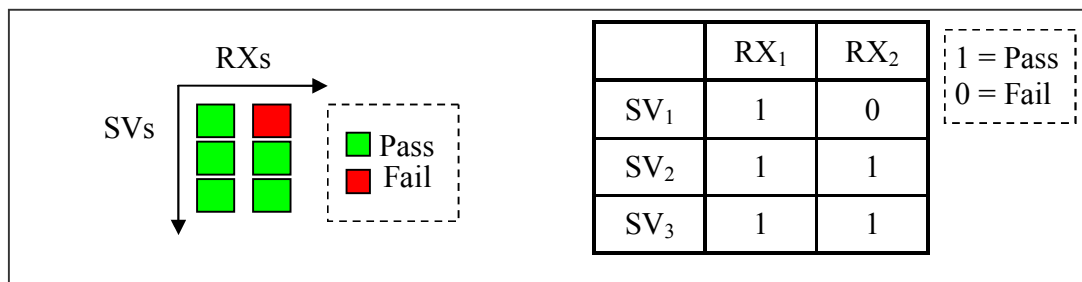
Figure 5.6: a) Block Form of EXM, b) Table Form of EXM

Using the (3 × 2) matrix from Figure 5.6a, there are $2^{(2\times3)} = 64$ potential forms of this matrix. Since each channel is either "pass" or "fail", we can treat this as Boolean

data and assign pass = 1 and fail = 0. Thus, each matrix is really just a unique binary structure as depicted in Figure 5.6b [4]. This structure represents the EXM channel flags described in Section 2.2.3.

Root Form		Iso$_\#$	Faults	List	Freq	Sum
0		η_1	6	{ 0 }	1	1
1		η_2	5	{ 1, 2, 4, 8, 16, 32 }	6	6
2		η_3	4	{ 3, 12, 48 }	3	
3		η_4	4	{ 5, 10, 17, 20, 34, 40 }	6	15
4		η_5	4	{ 6, 9, 18, 24, 33, 36 }	6	
5		η_6	3	{ 21, 42 }	2	
6		η_7	3	{ 22, 25, 26, 37, 38, 41 }	6	20
7		η_8	3	{ 7, 11, 19, 22, 25, … }	12	
8		η_9	2	{ 27, 30, 39, 45, 54, 57 }	6	
9		η_{10}	2	{ 23, 29, 43, 46, 53, 58 }	6	15
10		η_{11}	2	{ 15, 51, 60 }	3	
…	Pass = 1	η_{12}	1	{ 31, 47, 55, 59, 61, 62 }	6	6
63	Fail = 0	η_{13}	0	{ 63 }	1	1
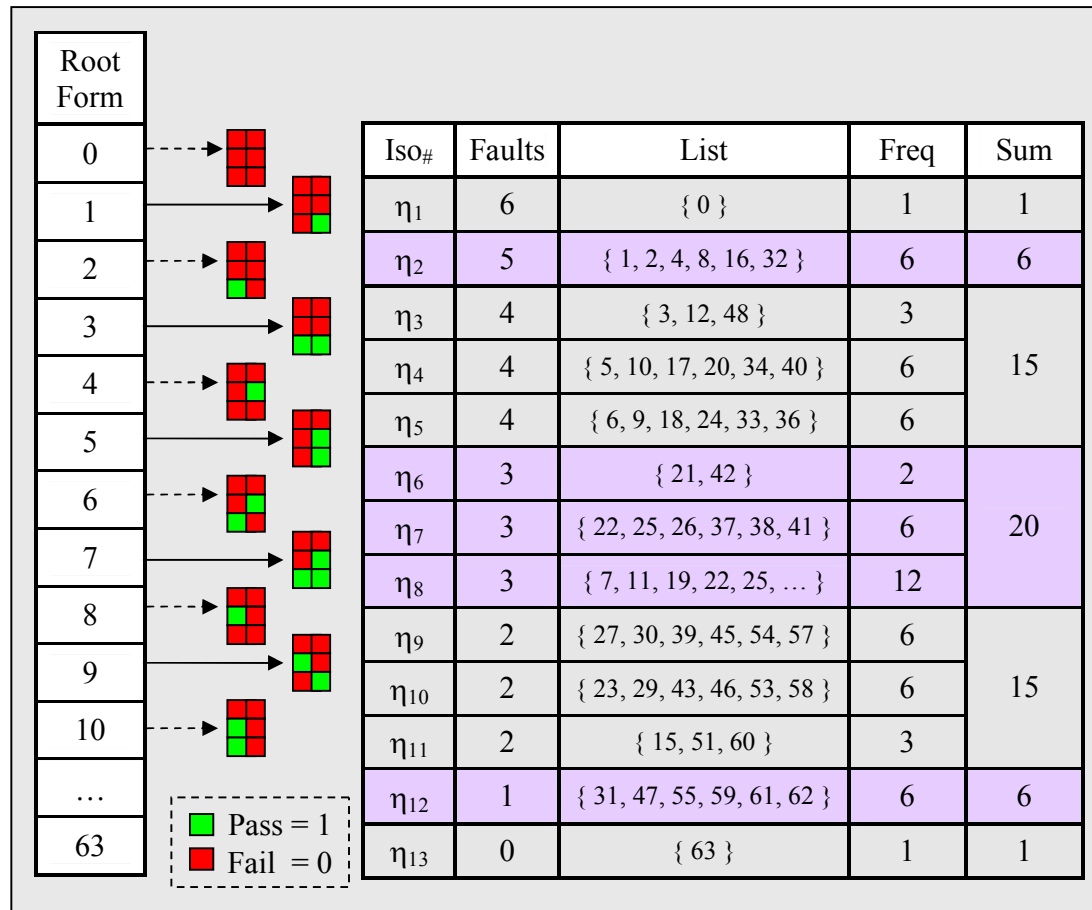
Figure 5.7: Description of Isomorphs for the (3 × 2) Channel Case

Figure 5.7 shows that, although there are 64 forms of the (3 × 2) channel case, there are actually only 13 unique cases. Each isomorph is shown as a row in the table which includes the list of the Form numbers. Root Form numbers are the decimal interpretations of the Boolean Root Form structures (the red and green matrices). The bits count just like reading a page, starting from the top and reading from left to right and then wrapping to the line below. The most significant bit is in the upper-left corner and the least significant bit is in the lower right corner. If the box/bit is red, that corresponds to a zero. If the box/bit is green, that corresponds to a one. For example, Root Form 0

results in isomorph η_1. Root Forms 1 and 2 are isomorphic with respect to each other but are distinct from Root Form 0, so the result is isomorph η_2. There are other Root Forms which are isomorphs of η_2, and they are listed in the table of Figure 5.7. The first column of the table indicates the isomorph number. The second column indicates how many faults, or red squares, there are in that isomorph. The third column is the list of all the Root Forms which are equivalent to that isomorph. All the decimal numbers of zero through 63 are accounted for in the table List, although for η_8 the list is abbreviated. The fourth column indicates the frequency, or number of elements, in the List. The last column demonstrates how this derivation adheres to the Binomial distribution. Identifying each isomorph and its frequency, or number of repetitions, allows a significant reduction in computation. Here, the number of isomorphs is only 20% of the initial number of Root Forms. For larger systems, the computational load will be reduced substantially more.

5.5.2   Finding Isomorphs

For a system which is (SV × RX × FQ), there are $2^{(SV \times RX \times FQ)}$ Root Forms. A (10 × 3 × 2) system has 1.153×10^{18} Root Forms and $10! \times 3! \times 2! = 87$ million modes of symmetry; i.e., the number of permutations of rows, columns, and pages. Consequently, it is prohibitive to simply iterate through all of the Root Forms, check their modes of symmetry, and reduce them to the set of isomorphs. There are two stages to determining the amount of computation required. The first stage involves finding out how much complexity exists in creating the starting set, and the second involves how efficiently each candidate form can be processed to determine if it is unique or is instead a repetition of a previous form. There are essentially four methods of attack to resolve the number of isomorphs for a given number of satellites, receivers, and frequencies:

1) Simulate all Root Forms and resolve by their modes of symmetry.
2) Simulate a smaller set which still contains all isomorphs and resolve by symmetry.

3) Recursively solve this set of isomorphs from a smaller system of isomorphs

4) Recursively solve for a reduced set of isomorphs which have the highest probability of occurrence based on having the smallest number of faults.

From the (3 SV × 2 RX) case in Figure 5.6a, it was shown how each Root Form can be ascribed identification numbers 0 to (2^n-1) and also what the final number of isomorphs is, but this was done by examining each of the 64 possible forms. This represents Method 1 above. The total computational requirement for each method is a function of the number of iterations and the number of calculations done per iteration. For Method 1 the number of iterations is the size of the $2^{(SV \times RX \times FQ)}$ Root Forms. The number of calculations per iteration is low because each number is quickly converted to the Boolean matrix structure shown in Figure 5.6b. Ultimately, Method 1 is impractical because of the sheer number of iterations. Leveraging the symmetry of the system can reduce the initial search size and consequently reduce the number of iterations.

Table 5.3 expresses the computational requirements of the creation of each candidate set. Methods 2, 3, and 4 use progressively fewer iterations to generate the initial set but at the expense of more computational complexity per iteration. Lastly, Method 4 has the lowest total computational requirement but results in a reduced set of isomorphs. When a failure is considered to be a rare event, then it is much more likely to have a system with few failures than many. Consequently, isomorphs with large numbers of failures are highly improbable and can be ignored. Each method and their computation requirements are examined this chapter.

	# of Iterations	# of Computations	Result
1	Massive	Small	Complete set of Isomorphs
2	Large	Medium	Complete set of Isomorphs
3	Medium	Medium	Complete set of Isomorphs
4	Small	Medium	Reduced set of Isomorphs

Table 5.3: Computational Considerations for Isomorph Resolution

5.5.2.1 The Row Unique Form

Since the satellites, or rows (in matrix form), have an irrelevant order with respect to which exact satellite may be passing or failing, we can try to remove the replication of individual satellites from the system. Furthermore, it is wise to attack along the largest dimension (the number of satellites) because it has the most symmetry and will reduce the initial search space the most. The (3 SV × 2 RX) case in Figure 5.6 is used again here. There are two receivers, and each one can be represented by either a zero or a one, forming a row. Each row is given a number equivalent to its binary value, and there are four possible row values for this example. These values are given in Equations (5.57-5.60)

$$
\begin{aligned}
\{0,0\} &= 0 \\
\{0,1\} &= 1 \\
\{1,0\} &= 2 \\
\{1,1\} &= 3
\end{aligned}
\qquad (5.57\text{-}5.60)
$$

Now that each row has been reduced to one number, these numbers (one for each row/satellite) can also be reduced into a number. This means that every form can be written as a three-digit, base-four number. It is base four because there are $2^2 = 4$ unique types of rows and three digits because that is the number of satellites. Here are two examples for the (3 SV × 2 RX) case.

$$
\begin{pmatrix} 1 & 0 \\ 1 & 1 \\ 0 & 1 \end{pmatrix} = \begin{pmatrix} 2 \\ 3 \\ 1 \end{pmatrix} = 231
$$

$$
\begin{pmatrix} 1 & 1 \\ 0 & 1 \\ 1 & 0 \end{pmatrix} = \begin{pmatrix} 3 \\ 1 \\ 2 \end{pmatrix} = 312
$$

$$(5.61\text{-}5.62)$$

168

The next transform alters this number from a three-digit base-four number into a form that expresses the total count of each of the unique row types. In other words, it tallies up the number of "1" digits, "2" digits, etc. This is the "Frequency Form", and because it only expresses the frequency of each row, the actual order of the rows is irrelevant. It is like a histogram in that it only expresses how often a value appears in a data set, not the order of its appearance in that set. Continuing this example in Equations (5.63) and (5.64), the right-most number "0111", shows that there are zero of Row 0, one of Row 1, one of Row 2, and one of Row 3. This implies that the numbers 123, 132, 213, 231, 312, and 321 are equal in the Frequency Form.

$$
\begin{pmatrix} 1 & 0 \\ 1 & 1 \\ 0 & 1 \end{pmatrix} = \begin{pmatrix} 2 \\ 3 \\ 1 \end{pmatrix} = 231 = 0111 \tag{5.63}
$$

$$
\begin{pmatrix} 1 & 1 \\ 0 & 1 \\ 1 & 0 \end{pmatrix} = \begin{pmatrix} 3 \\ 1 \\ 2 \end{pmatrix} = 312 = 0111 \tag{5.64}
$$

Table 5.4 shows the forms of the (3 SV × 2 RX) system reduced to the Unique Frequency Form, which are just the unique elements of the set of Frequency Forms. The number of repetitions of each Unique Frequency Form is simple to calculate: it is the multinomial calculation using the digits of the Unique Frequency Form number. For the Unique Frequency Form "0111" from Equation (6.63), the multinomial calculation and number of repetitions is given in Equation (5.65).

$$
\#\text{Reps}\left("0111"\right) = \left(\frac{3}{\{1,1,1\}} \right) = \frac{3!}{1! \cdot 1! \cdot 1!} = 6 \tag{5.65}
$$

.   In Table 5.4, Root Form #4 has no Unique Frequency Form because it has already been determined to be identical to Root Form #1. In other words, they have identical Frequency Forms.

Root Form	Row Form	Frequency Form	Unique Freq. Form	#Repetitions
0	{000}	{3000}	{3000}	1
1	{001}	{2100}	{2100}	3
2	{002}	{2010}	{2010}	3
3	{003}	{2001}	{2001}	3
4	{010}	{2100}	-	
5	{011}	{1200}	{1200}	3
6	{012}	{1110}	{1110}	6
7	{013}	{1101}	{1101}	6
8	{020}	{2010}	-	
9	{021}	{1110}	-	
10	{022}	{1020}	{1020}	3
…	…	…	…	…
63	{333}	{0003}	{0003}	1
Count	64	64	20	64

Table 5.4: From Root Forms to Unique Frequency Forms

Table 5.4 shows that all 64 Root Forms for the (3 SV × 2 RX) system do not need to be simulated. Technically, one could start with only the Unique Frequency Form. Creating the set of Unique Frequency Forms is highly recursive and can be done with a minimal amount of computation. All of the Unique Frequency Forms for the (3 SV × 2 RX) system are shown in Table 5.5.

{3000}	{2100}	{1200}	{0300}
	{2010}	{1110}	{0210}
	{2001}	{1101}	{0201}
		{1020}	{0121}
		{1011}	{0112}
		{1002}	{0103}
			{0030}
			{0021}
			{0012}
			{0003}

Table 5.5: Unique Frequency Form for Three Satellites and Two Receivers

Now, starting with the Unique Frequency Form from Table 5.4, column symmetry can be extracted. This is done by determining the mapping from one Row Form to another based upon a pre-defined process of column swapping. Given that there are only two columns, there are only two operations, "swap" and "no swap". This is where the RX! × FQ! degrees of symmetry come from.

Old Row Form	Old Root Form		New Root Form	New Row Form
0	{00}	⇔	{00}	0
1	{01}	⇔	{10}	2
2	{10}	⇔	{01}	1
3	{11}	⇔	{11}	3

Table 5.6: Extracting Column/Receiver Symmetry

Table 5.6 shows that, under a column swap, the original Root Form of {01} becomes {10}. In Row Form, this means that a (1) becomes a (2), and vice versa. Using Root Form #1 from Table 5.4, which has a Row Form of {001}, each digit of that number is converted using Table 5.6. This means that Row Form {001} would become {002}. For Frequency Forms, the indices, not the digits, are changed. The Frequency Form would change from {2100} to {2010}. The convention defined here is that first element of an array is considered to be in the zero-place. The zero-place element and the three-place element remain unchanged, while the one and two-place elements are swapped.

The term "Equivalent Form" is introduced here and is critical in understanding the process of determining if each new form resulting from a column swap is in fact a new isomorph. Figure 5.8 depicts the concept of Equivalent Forms graphically. Each ellipse in the figure is one of the Unique Frequency Forms. When two Unique Frequency Forms are shown to be equivalent due to a column operation, they form an Equivalence Set. Graphically this is shown as two ellipses connected by a line. The smallest "number" in the set represents that Equivalence Set and is shown as a colored ellipse. That "number" is the decimal interpretation of the Unique Frequency Form. The number of repetitions for that Equivalence Set is the sum of repetitions of its members. Note that, when

performing the column swaps on a Unique Frequency Form, and other Unique Frequency Forms are shown to be in its Equivalence Set, those other Unique Frequency Forms do not have to undergo the columns swaps, as they are already known not to be new isomorphs. This reduces computation time.
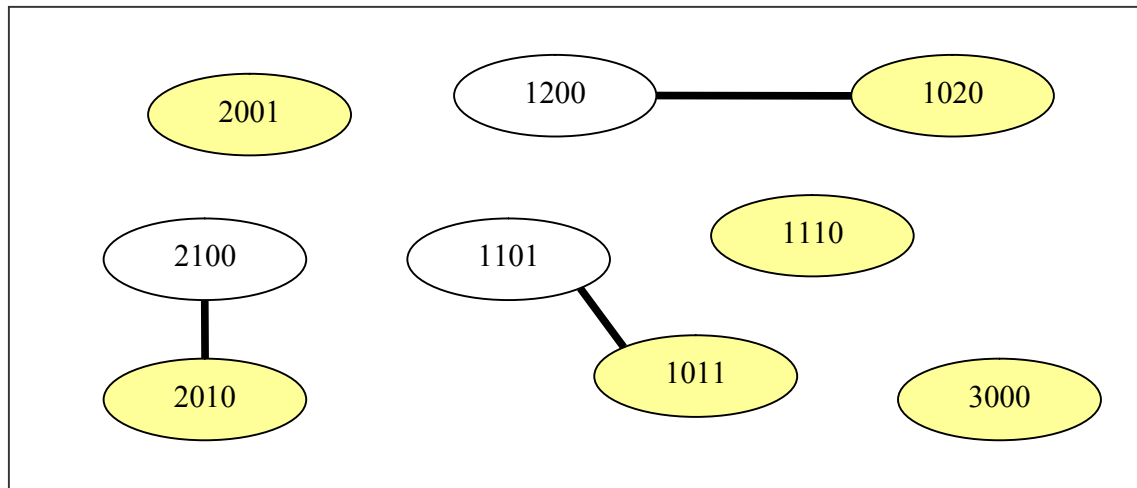
Figure 5.8: Each Equivalence Set Represents One Isomorph

Table 5.7 shows the results of applying the column-swap operation to the Unique Frequency Forms of Table 5.4 to determine which of the Unique Frequency Forms results in new isomorphs. What is mechanically happening in Table 5.7 is most easily explained by example. Starting with the first Unique Frequency Form at the top of the table {3000}, performing a column swap causes the new Frequency Form to be {3000}, which is identical (and therefore shown in gray). The next Row Unique Form is {2100}. A column swap yields the number {2010}, which is a new Equivalent Form. The next number is {2010}, but it was shown to be an Equivalent Form of {2010}, and so it will not result in a new isomorph. By the end of this process, it turns out that there are only 13 isomorphs (shown in green). Those isomorphs are converted back to their matrix (or tensor) representations, and their respective numbers of repetitions are tallied from their Equivalence Set and stored.

Unique Frequency Form	Pre-Swap Comments	Column Swap	Post-Swap Comments
3000	New	3000	No Change
2100	New	2010	New Equivalent Form
2010	Already Found	2100	No need to Swap
2001	New	2001	No Change
1200	New	1020	New Equivalent Form
1110	New	1110	No Change
1101	New	1011	New Equivalent Form
1020	Already Found	1200	No need to Swap
1011	Already Found	1101	No need to Swap
1002	New	1002	No Change
0300	New	0030	New Equivalent Form
0210	New	0120	New Equivalent Form
0201	New	0021	New Equivalent Form
0120	Already Found	0210	No need to Swap
0111	New	0111	No Change
0102	New	0012	New Equivalent Form
0030	Already Found	0300	No need to Swap
0021	Already Found	0201	No need to Swap
0012	Already Found	0102	No need to Swap
0003	New	0003	No Change

Table 5.7: From Unique Frequency Form to Isomorphs

The cost savings for the (3 SV × 2 RX) system solution is small. Instead of comparing 64 Root Forms, there are only 20 Row Unique Forms to start with, requiring 13 simple operations (20 − 7) to reach the end result of 13 isomorphs shown in Table 5.5. However, in a more-practical (10 SV × 2 RX) system, the number of Root Forms is $2^{30} =$ 1,073,741,824, but the number of Row Unique Forms is only 19,448, reducing the initial search space to less than $1/100,000^{th}$ of its original size.

Table 5.8 shows the relationship between the number of Root Forms and the number of Unique Frequency Forms. In general, the number of Unique Frequency Forms is given by Equation (5.65), where R is the number of receivers, S is the number of satellites, and F is the number of frequencies. The first line of Equation (5.66) uses the standard notion for the binomial coefficient [38]. Each cell of Table 5.8 represents a certain number of satellites, receivers, and frequencies. For each cell, the top number is the number of Root Forms and the lower number is the number of Unique Frequency

Forms. For example, for a 10 SV × 3 RX × 1 FQ system, the reduction is from roughly 1×10^9 Root Forms down to 2×10^4 Unique Frequency Forms. This can be seen in the highlighted cells of Table 5.8.

$$N_{RF} = \binom{2^{R \cdot F} + S - 1}{S}$$
$$= \frac{\left(2^{R \cdot F} + S - 1\right)!}{\left(2^{R \cdot F} - 1\right)! \cdot S!}$$

(5.66)

# SV	1 RX		2 RX		3 RX		4 RX	
	1 FQ	2 FQ	1 FQ	2 FQ	1 FQ	2 FQ	1 FQ	2 FQ
1	2	4	4	16	8	64	16	256
	2	4	4	16	8	64	16	256
2	4	16	16	256	64	4×10^3	256	7×10^4
	3	10	10	136	36	2×10^3	136	3×10^4
3	8	64	64	4×10^3	512	3×10^5	4×10^3	2×10^7
	4	20	20	816	120	5×10^4	816	3×10^6
4	16	256	256	7×10^5	4×10^3	2×10^7	7×10^4	4×10^9
	5	35	35	4×10^3	330	8×10^5	4×10^3	2×10^8
5	32	1×10^3	1×10^3	1×10^6	3×10^4	1×10^9	1×10^6	1×10^{12}
	6	56	56	2×10^4	792	1×10^7	2×10^4	1×10^{10}
6	64	4×10^3	4×10^3	2×10^7	3×10^5	7×10^{10}	2×10^7	3×10^{14}
	7	84	84	5×10^4	2×10^3	1×10^7	5×10^4	4×10^{11}
7	128	2×10^4	2×10^4	3×10^8	2×10^6	4×10^{12}	3×10^8	7×10^{16}
	8	120	120	2×10^5	3×10^3	1×10^9	2×10^5	2×10^{13}
8	256	7×10^4	7×10^4	4×10^8	2×10^7	3×10^{14}	4×10^8	2×10^{19}
	9	165	165	5×10^5	6×10^3	1×10^{10}	5×10^5	5×10^{14}
9	512	3×10^5	3×10^5	7×10^{10}	1×10^8	2×10^{16}	7×10^{10}	5×10^{21}
	10	220	220	1×10^6	1×10^4	9×10^{10}	1×10^6	1×10^{16}
10	1×10^3	1×10^6	1×10^6	1×10^{12}	1×10^9	1×10^{18}	1×10^{12}	1×10^{24}
	11	286	286	4×10^6	2×10^4	6×10^{11}	4×10^6	4×10^{17}
11	2×10^4	4×10^6	4×10^6	2×10^{12}	9×10^9	7×10^7	2×10^{13}	3×10^{26}
	12	364	364	8×10^6	3×10^4	4×10^{12}	8×10^6	1×10^{19}
12	4×10^4	2×10^7	2×10^7	3×10^{14}	7×10^{10}	5×10^{21}	3×10^{14}	8×10^{28}
	13	455	455	2×10^7	5×10^4	3×10^{13}	2×10^7	2×10^{20}

Table 5.8: Comparing the Numbers of Root Forms to Unique Frequency Forms

5.5.3 Finding Isomorphs of Larger Systems

Table 5.8 shows that resolving all isomorphs for a $(10 \times 3 \times 2)$ system would require creating a list of roughly 621 billion Unique Frequency Forms and then processing them to extract their remaining symmetry. Despite the fact that this is a tremendous reduction from the 1.2 quintillion Root Forms, this is still an encumbrance. Fortunately, the isomorphs can be "built up" from the previous solution of Isomorphs for a smaller system size rather than "stripping down" the Unique Frequency Forms of this system. This method is straightforward and simply involves adding one satellite to the solution of a smaller system. In other words, to resolve the $(10 \times 3 \times 2)$ system, the existing $(9 \times 3 \times 2)$ solution is used (building up these solutions one at a time from smaller cases) and run through all permutations of one additional satellite. Now the initial search space is only $2^{(\#RX \times \#FQ)}$ times the number of isomorphs of the preceding scenario; in this case, $2^{(3 \cdot 2)} = 64$ times larger. The results of this method are shown in Figure 5.10 in terms of the number of isomorphs as a function of the number of satellites.
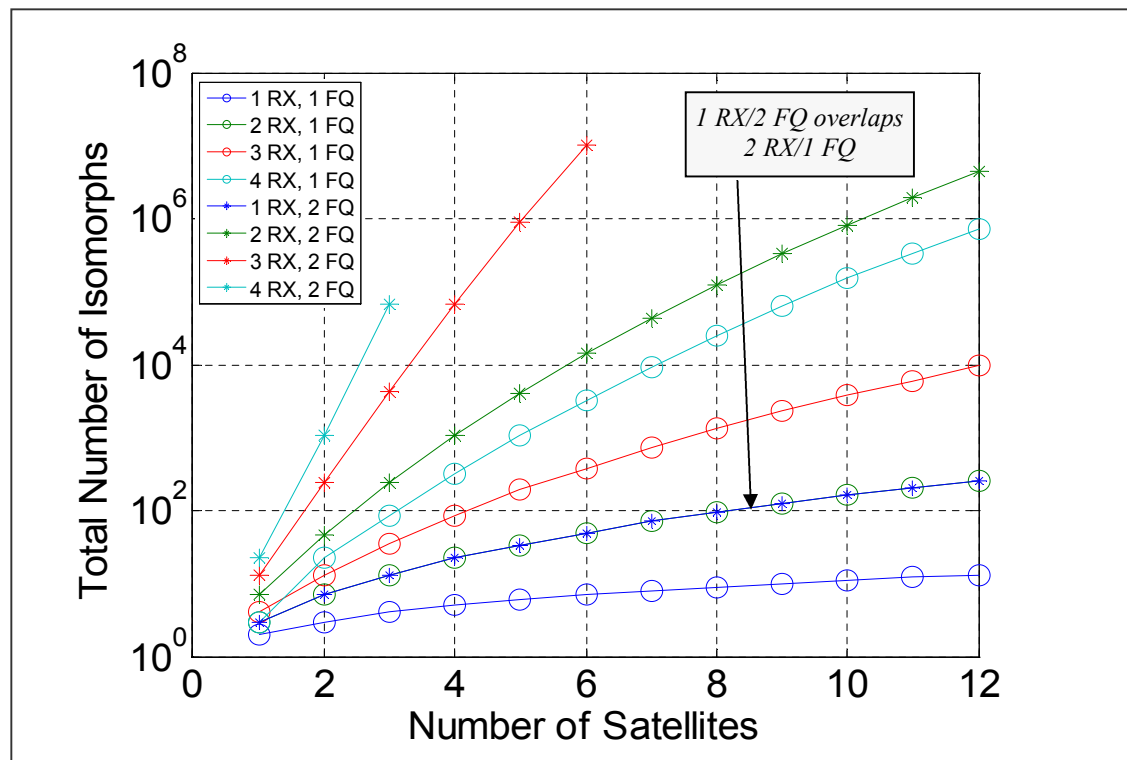
Figure 5.9: Number of Isomorphs given #SV, #RX, & #FQ

Figure 5.9 shows the number of isomorphs for up to twelve satellites with as many as four receivers and two frequencies. Data for the larger number of satellites for the three and four-receiver two-frequency cases could not be calculated due to the computational requirements of those calculations. This figure demonstrates the significant increase in the number of isomorphs when moving from one frequency to two, especially when using three (or even four) receivers. Note that the {1 RX, 2 FQ} curve lies atop the curve for the {2 RX, 1 FQ} case because of the natural symmetry between these two scenarios.

5.5.3.1 Modeling the Number of Isomorphs vs. Channel Faults

Figure 5.10 shows that, even when using only seven satellites for the three-receiver, two-frequency case, there are still about 10^7 isomorphs.
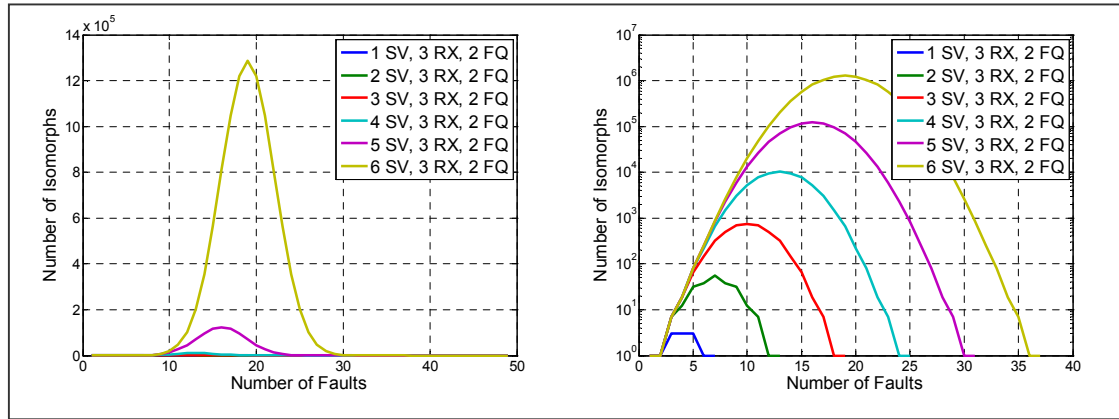
Figure 5.10: Isomorphs per Fault (3 RX, 2 FQ), a) Linear-scale, b) Log-scale

There is considerable regularity in the logarithmic plot of Figure 5.10b. This can be used to model curves for larger systems. Those models can then be used to estimate the total number of isomorphs for larger systems. There are a few identifying features that can be used to model the curvature. First, the curves are nearly parabolic in (base-10) logarithmic space. The distribution of the number of faults in the ($n_{SV} \times n_{RX} \times n_{FQ}$) system is described by the standard Binomial distribution (Equation 5.67). The number

of total channels is *n*, the number of faults is *x*, and the probability of a single fault is *p*. When the number of degrees of freedom becomes large while the probability of the fault remains constant, the distribution of Equation (5.68) will approach the Normal distribution given by Equation (5.69) [36].

$$pdf_B(x,n,p) = \binom{n}{x} p^x (1-p)^{n-x} \tag{5.67}$$

$$pdf\left(\frac{x - n \cdot p}{\sqrt{n \cdot p \cdot (1-p)}}\right)_{n \to \infty} = pdf_N(x,0,1) \tag{5.68}$$

$$pdf_N(x,\mu,\sigma) = \frac{1}{\sigma\sqrt{2\pi}} \cdot e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} \tag{5.69}$$

Taking the (natural) Log of the Normal distribution reduces it to Equation (5.70) and reveals that, in log space, the distribution is a quadratic function of *x*.

$$Log\left[pdf_N(x,\mu,\sigma)\right] = -\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2 + Log[\frac{1}{\sigma\sqrt{2\pi}}] \tag{5.70}$$

The next useful feature is that the number of isomorphs at both the endpoints and midpoint of each curve in Figure 5.10b are known, as they have already been computed, and this is sufficient data to construct a model of a parabola. For larger systems, however, it is possible to calculate the endpoints but not the midpoints. It is those midpoints that must be estimated by extrapolating the midpoints of the smaller systems. The reason that the endpoints of larger systems can always be calculated is because the number of faults (at the endpoints) is still small despite the fact that the total number of channels has increased. When the number of faults is less than or equal to the smallest dimension of the structure, additional channels only increase the amount of symmetry and not the number of isomorphs. The $(10 \times 3 \times 2)$ system can be used as an example here. The smallest dimension is two (the number of frequencies), and as long as the

number of faults is two or less, then adding a satellite, receiver, or frequency will not increase the number of isomorphs for the specific number of faults (i.e. 0, 1, or 2).

Using those models for the smaller systems, the peak values (midpoints) of the modeled parabolas can be modeled and extrapolated to estimate the number of faults at the midpoint of larger systems. Having the number of isomorphs for the endpoints and midpoint of the larger systems allows a new parabola to be estimated to describe the number of isomorphs as a function of the number of faults in a larger system. A third-order polynomial works sufficiently well to estimate the peak number of isomorphs per satellite from Figure 5.10b, and the resulting model is given in Equation (5.70), where N is the number of isomorphs at the midpoint of a system with n_{SV} satellites, three receivers and two frequencies.

$$\log\left[\widehat{N}\right] = \left(9.32 \times 10^{-4}\right) \cdot n_{SV}^3 - \left(3.61 \times 10^{-2}\right) \cdot n_{SV}^2 + \left(1.3382\right) \cdot n_{SV} - \left(0.82\right) \quad (5.71)$$

Equation (5.71) is plotted in Figure 5.11 and is compared against the known results for cases with fewer numbers of satellites, demonstrating excellent agreement.
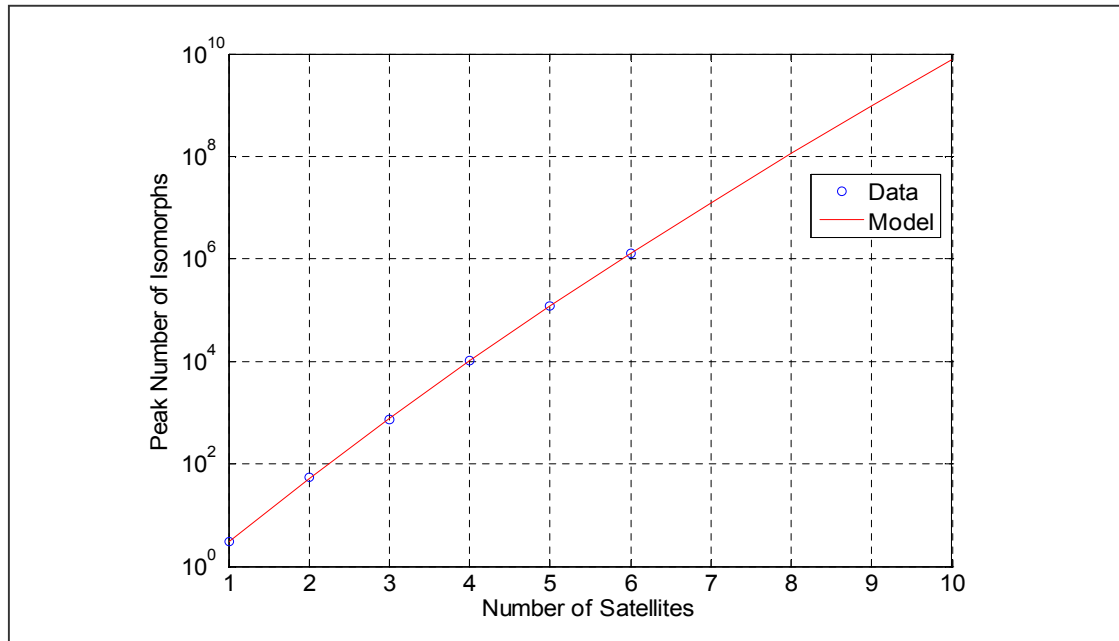
Figure 5.11: Peak Number of Isomorphs per Satellite

The model in Equation (5.71) was created by estimating the peak midpoints of each of the six parabolas in Figure 5.10b. That model was then used to create individual models for each of those six parabolas. These new models are given by Equations (5.72-5.77). The model in Equation (5.71) was also used to extrapolate out for the number of satellites and yields the models given in Equations (5.78-5.83) The first and last data points were neglected because of the deviation from the model. For any sized system, there is only one isomorph where there is either zero faults or one fault. In the models of Equations (5.72-5.83) this is at the outermost part of the parabola where the slope of the line should be steepest, but including the outermost data point (the data point for zero faults) would make the slope zero. This can be seen in Figure 5.12. Consequently, the data point for zero faults is discarded, as is the data points for a system with all faults (i.e., completely faulted) by an argument of symmetry.

$$p_{1,3,2} = 10^{0.483 \cdot \left(1 - \left(\frac{x-3}{2}\right)^2\right)} \tag{5.72}$$

$$p_{2,3,2} = 10^{1.720 \cdot \left(1 - \left(\frac{x-6}{5}\right)^2\right)} \tag{5.73}$$

$$p_{3,3,2} = 10^{2.895 \cdot \left(1 - \left(\frac{x-9}{8}\right)^2\right)} \tag{5.74}$$

$$p_{4,3,2} = 10^{4.015 \cdot \left(1 - \left(\frac{x-12}{11}\right)^2\right)} \tag{5.75}$$

$$p_{5,3,2} = 10^{5.086 \cdot \left(1 - \left(\frac{x-15}{14}\right)^2\right)} \tag{5.76}$$

$$p_{6,3,2} = 10^{6.112 \cdot \left(1 - \left(\frac{x-18}{17}\right)^2\right)} \tag{5.77}$$

Using Equation (5.71), we can derive similar approximate equations for the three-receiver, two-frequency models for scenarios with seven to twelve satellites.

$$p_{7,3,2} = 10^{7.100 \cdot \left(1 - \left(\frac{x-21}{20}\right)^2\right)} \tag{5.78}$$

$$p_{8,3,2} = 10^{8.055 \cdot \left(1 - \left(\frac{x-24}{23}\right)^2\right)} \tag{5.79}$$

$$p_{9,3,2} = 10^{8.982 \cdot \left(1 - \left(\frac{x-27}{26}\right)^2\right)} \tag{5.80}$$

$$p_{10,3,2} = 10^{9.888 \cdot \left(1 - \left(\frac{x-30}{29}\right)^2\right)} \tag{5.81}$$

$$p_{11,3,2} = 10^{10.779 \cdot \left(1 - \left(\frac{x-33}{32}\right)^2\right)} \tag{5.82}$$

$$p_{12,3,2} = 10^{11.658 \cdot \left(1 - \left(\frac{x-36}{35}\right)^2\right)} \tag{5.83}$$

All of these equations are plotted in Figure 5.12. The agreement between model and data is sufficient to estimate what the distributions of the higher-order systems will look like, which is all that is needed here.
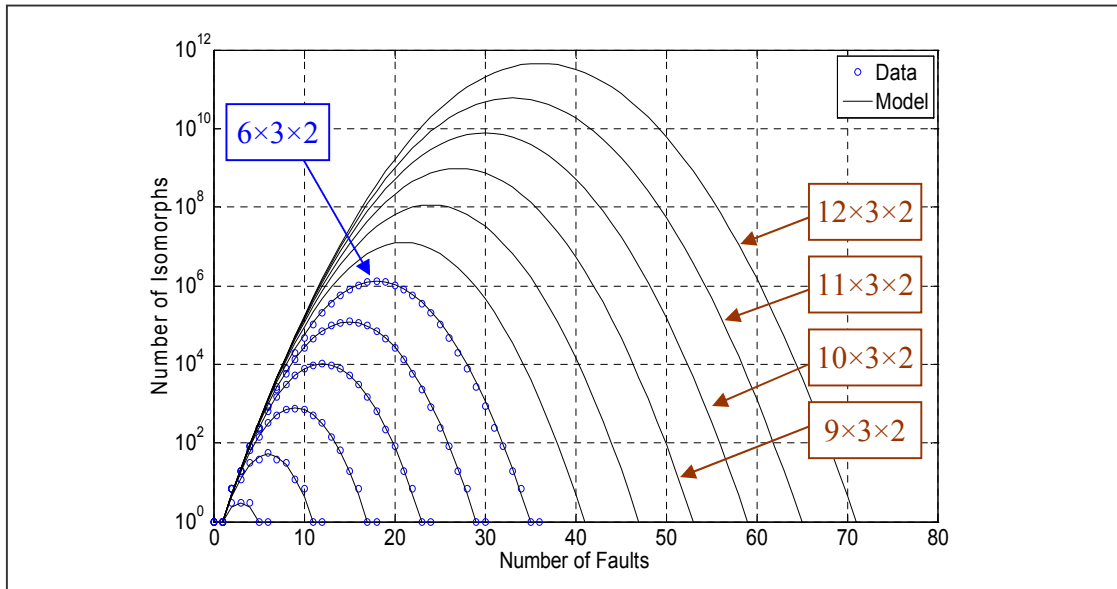
Figure 5.12: Extrapolating Isomorphs for Larger Numbers of Satellites

Figure 5.9 showed the total number of isomorphs as a function of the number of satellites, with each curve representing a given number of receivers and signal frequencies. Although it is possible to extrapolate those curves for the three and four-receiver, two-frequency case, it is more reasonable to simply integrate the equations plotted in Figure 5.12, since these curves represent the number of isomorphs as a function of the number of faults for a given system. The ($n_{SV} \times 3$ RX $\times 2$ FQ) curve of Figure 5.9 describes the total number of isomorphs for a certain-sized system. The parabolas in Figure 5.12 describe the number of isomorphs as a function of the number of faults of that sized system. Therefore, each point of the ($n_{SV} \times 3$ RX $\times 2$ FQ) curve of Figure 5.9 is equivalent to the integrated sum of each parabola in Figure 5.12. In other words, the total number of isomorphs for a given sized system is equal to the integral of the number of isomorphs per fault over the number of faults for a same-sized system. The objective of this section was to understand the curves of Figure 5.12 in order to extrapolate these curves and ultimately extrapolate the points of Figure 5.9. Again, integrating the curves describing the number of isomorphs per fault will estimate the total number of isomorphs for a given system size. The results are given in Equations (5.84-5.89) and are the sum of the parabola models (Equations 5.78-5.83) evaluated at each number of faults.

$$N_{7,3,2} = \sum_{n=0}^{7 \cdot 3 \cdot 2} 10^{7.100 \cdot \left(1 - \left(\frac{x-21}{20}\right)^2\right)} = 1.10 \times 10^8 \tag{5.84}$$

$$N_{8,3,2} = \sum_{n=0}^{8 \cdot 3 \cdot 2} 10^{8.055 \cdot \left(1 - \left(\frac{x-24}{23}\right)^2\right)} = 1.07 \times 10^9 \tag{5.85}$$

$$N_{9,3,2} = \sum_{n=0}^{9 \cdot 3 \cdot 2} 10^{8.982 \cdot \left(1 - \left(\frac{x-27}{26}\right)^2\right)} = 9.73 \times 10^9 \tag{5.86}$$

$$N_{10,3,2} = \sum_{n=0}^{10 \cdot 3 \cdot 2} 10^{9.888 \cdot \left(1 - \left(\frac{x-30}{29}\right)^2\right)} = 8.33 \times 10^{10} \tag{5.87}$$

$$N_{11,3,2} = \sum_{n=0}^{9 \cdot 3 \cdot 2} 10^{10.779 \cdot \left(1 - \left(\frac{x-33}{32}\right)^2\right)} = 6.84 \times 10^{11} \tag{5.88}$$

$$N_{12,3,2} = \sum_{n=0}^{10\cdot3\cdot2} 10^{11.658\cdot\left(1-\left(\frac{x-36}{35}\right)^2\right)} = 5.45\times10^{12} \tag{5.89}$$

The modeling and extrapolation results for the scenarios addressed here are summarized in Table 5.9 below. With only a few data points for the four-receiver, two-frequency cases, it is difficult to extrapolate the number of isomorphs. For the four-satellite, four-receiver, two-frequency case, the number of isomorphs will be on the order of 10 million.

# SV	1 RX		2 RX		3 RX		4 RX	
	1 FQ	2 FQ	1 FQ	2 FQ	1 FQ	2 FQ	1 FQ	2 FQ
1	2	3	3	7	4	13	5	22
2	3	7	7	46	13	237	22	1.1×10^3
3	4	13	13	237	36	4.2×10^3	87	6.6×10^4
4	5	22	22	1.1×10^3	87	6.6×10^4	317	$*1.0 \times 10^7$
5	6	34	34	4.1×10^3	190	8.8×10^5	1.1×10^3	?
6	7	50	50	1.4×10^4	386	1.0×10^7	3.3×10^3	?
7	8	70	70	4.3×10^4	734	$*1.1 \times 10^8$	9.3×10^3	?
8	9	95	95	1.2×10^5	1.3×10^3	$*1.1 \times 10^9$	2.5×10^4	?
9	10	125	125	3.3×10^5	2.3×10^3	$*9.7 \times 10^9$	6.4×10^4	?
10	11	161	161	8.2×10^5	3.8×10^3	$*8.3 \times 10^{10}$	1.6×10^5	?
11	12	203	203	1.9×10^6	6.1×10^3	$*6.8 \times 10^{11}$	3.6×10^5	?
12	13	252	252	4.4×10^6	9.5×10^3	$*5.5 \times 10^{12}$	7.9×10^5	?

Table 5.9: Number of Total Isomorphs (* *estimated*)

Table 5.9 shows that, although the number of isomorphs for the $(10 \times 3 \times 2)$ case is well below $(2^{60} = 1.2 \times 10^{18})$ Root Forms, or even 631 billion Unique Frequency Forms, the estimated 83 billion remaining isomorphs are still a challenge to any real application of this method. The objective of this process was to reduce the number of cases so that not only could this method be used to understand the effect that different EXM logic has on exclusion rates but also that it could be done rapidly enough that the EXM logic could be iterated repeatedly for design purposes.

Importantly, not all isomorphs are created equal. When the probability of a fault is low, the probability of many faults happening is extremely low. This means that only those isomorphs with a small number of faults need to be simulated in order to evaluate the performance of the EXM logic. This argument applies to the system when it is operating nominally and a fault is assumed to be rare, but this argument also applies to the system when a fault is assumed to exist. Because the isomorph is an abstract representation, meaning that it describes a relationship between two symbols, passing and failing, those symbols can simply be reversed. What was once labeled as a low probability fault can now be labeled as a low-probability passing channel, and the prior argument holds. The number of rare events (either failing or passing) that must be simulated in order to accurately calculate the probabilities for the different EXM logic rules requires an equation such as (5.84-5.89) (for the $10 \times 3 \times 2$ case) to describe the number of isomorphs as a function of the number of channel faults.

5.5.4   How Many Faults must be simulated?

The objective of the VMT is to be able to simulate the detection-rate probabilities of various EXM detection rules. In order to do so, the VMT must be able to simulate the system down to an acceptably low residual probability. In this case, the residual probability is the sum of the probabilities of all fault-channel scenarios which are not simulated. Given a system with $n = (n_{SV} \times n_{RX} \times n_{FQ})$ channels and a probability p of there being a fault on any given channel, integrating the binomial pdf of Equation (5.67) will determine the residual probability.

Figure 5.13 shows the residual probability (the amount of probability left unevaluated) of the simulation as a function of the number of faults in the entire system. Each curve represents a different probability of a fault on each channel. For this calculation, all channels (SV, RX, and FQ) are assumed to be independent in order to use the cumulative density function for the binomial distribution. The curves are derived from the "binocdf" function in MatLab (Equation 5.67) and illustrate that the VMT simulation is limited to representing a number down to 10^{-16}, as that is the limit of

machine precision for a system running MatLab v14. When the required integrity and continuity probabilities are on the order of 10^{-8} to 10^{-10}, the issue of machine precision is not a concern because even with a fault probability as high as 10^{-2}, only up to ten faults would need to be simulated.
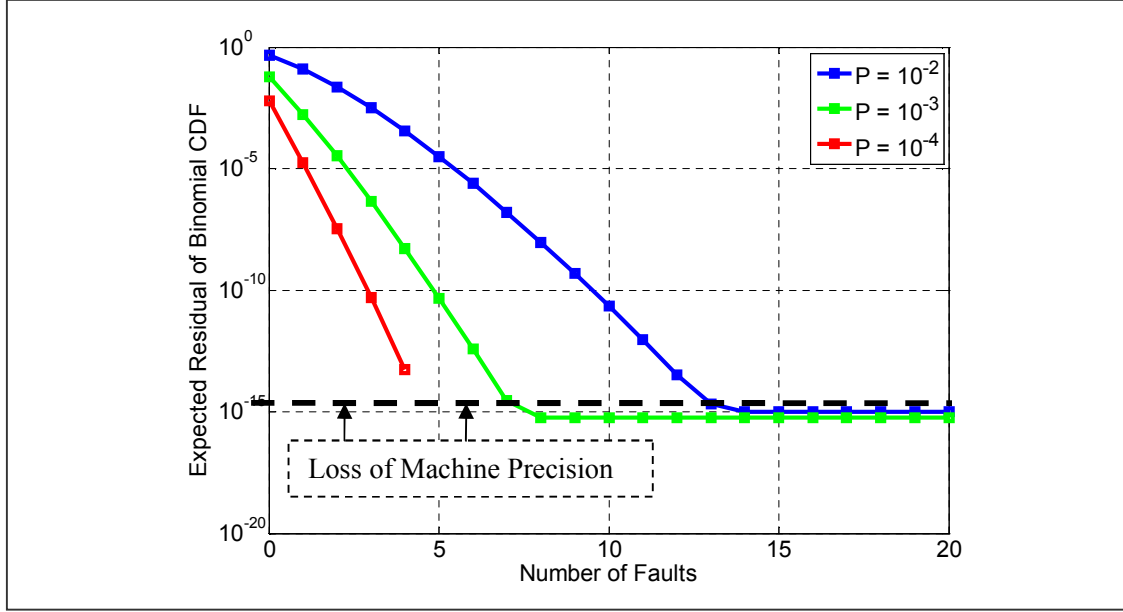
Figure 5.13: Residual Probability Based on Fault Probability and Number of Faults

The methods described in Section 5.5.3 for resolving isomorphs described solving for the full set of isomorphs. The two ways this was accomplished were by either determining all the Unique Frequency Forms for the entire system and then determining which of those were actually isomorphs, or by building up the system size from the set of isomorphs of a smaller system. Recall that the full set of isomorphs for the $(n_{SV} \times 3 \times 2)$ system can be found from the full set of isomorphs from the $((n_{SV} - 1) \times 3 \times 2)$ system. In the same manner, the set of isomorphs with P channel faults of the $(10 \times 3 \times 2)$ system can be found from the set of isomorphs with $(P-1)$ faults for the same system. Given that each new fault can only have a total of $((n_{SV} \times n_{RX} \times n_{FQ}) - P)$ locations, or $60-P$ locations for the $(10 \times 3 \times 2)$ case, this method is efficient for this purpose.

Referring back to the parabolas of Figure 5.12, there are roughly 50,000 isomorphs for the ($10 \times 3 \times 2$) system with 10 faulted channels. The cumulative sum of isomorphs of this system from zero to ten faulted channels is approximately 73,000. This implies a search space of not more than $60 \times 73,000 = 4.4$ million, which is far less than the number of Root Forms (1.8 quintillion), Unique Frequency Forms (631 billion), or even total isomorphs for that system (83 billion) found earlier.

5.5.4.1 Building the Isomorph Set one Fault at a Time

The main benefit of the method described in the previous section is that is uses a smaller search space and requires less memory. However, it is not as efficient in manipulating the previously-used representation forms because it has difficulty utilizing the "row-independent" concept. The reason is that, in this method, the rows must be sorted to remove the significance of satellite order, and sorting is a significant computational burden.

Despite this limitation, the numbers of isomorphs for up to 13 faulted channels of the ($10 \times 3 \times 2$) scenario were found using this method. Figure 5.14 shows that, for 13 faulted channels, there were 1.3 million isomorphs, while the estimation model predicted 3 million. This conservative prediction is consistent with the method of modeling whereby the parabolic fit is more exact toward the median number of faults and is essentially stretched or widened to meet the boundary criteria.

185
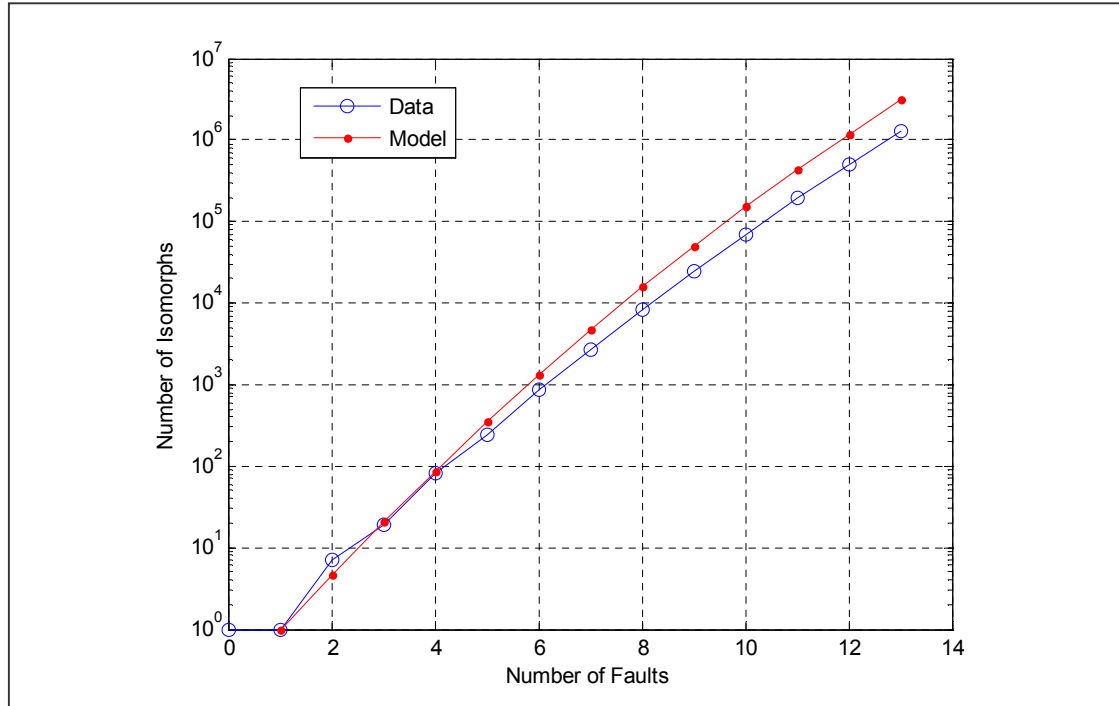
Figure 5.14: Isomorphs vs. Number of Faults for the $(10 \times 3 \times 2)$ Scenario

5.5.5 Mixed Tensors

The set of isomorphs for the $(10 \times 3 \times 2)$ case has been determined, but in order to check the actual ability of EXM to detect a particular event, isomorphs are needed which recognize both the faulted and un-faulted segments of channels. Figure 5.15 uses the representation first shown in Figure 4.4 but adds the notion that both faulted and un-faulted segments can have passing and failing channels and that these two entities are not isomorphic because they are not representing equivalent objects. In Figure 5.15, the yellow blocks represent faulted channels which have been flagged (i.e., have failed some monitor threshold test), while the blue blocks represent faulted channels which are not flagged (i.e., have passed all monitors).
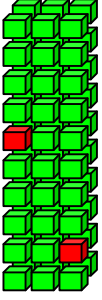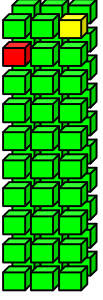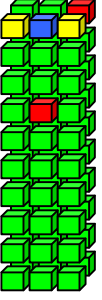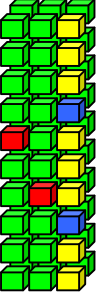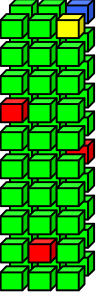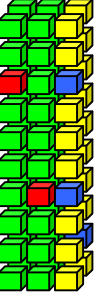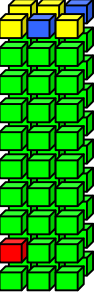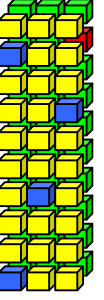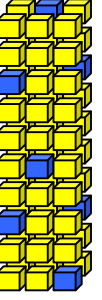
Figure 5.15: Tensor Representation for Designated Fault Modes

Using the familiar simplified example of three satellites and two receivers, it was previously determined that there are 64 Root Forms and 11 isomorphs. This was under the assumption that all channels were equal. Consider the circumstance when one satellite is declared to be faulty. An example of a mixed tensor is given in Figure 5.16. In this figure, each one of the (3×2) matrices is an isomorph when the top row represents a faulted satellite. The value of n_F down the left side of the figure gives the number of faulted channels in the faulted satellite. The value of n_{UF} across the top of the figure gives the number of faulted channels for the rest of the system.
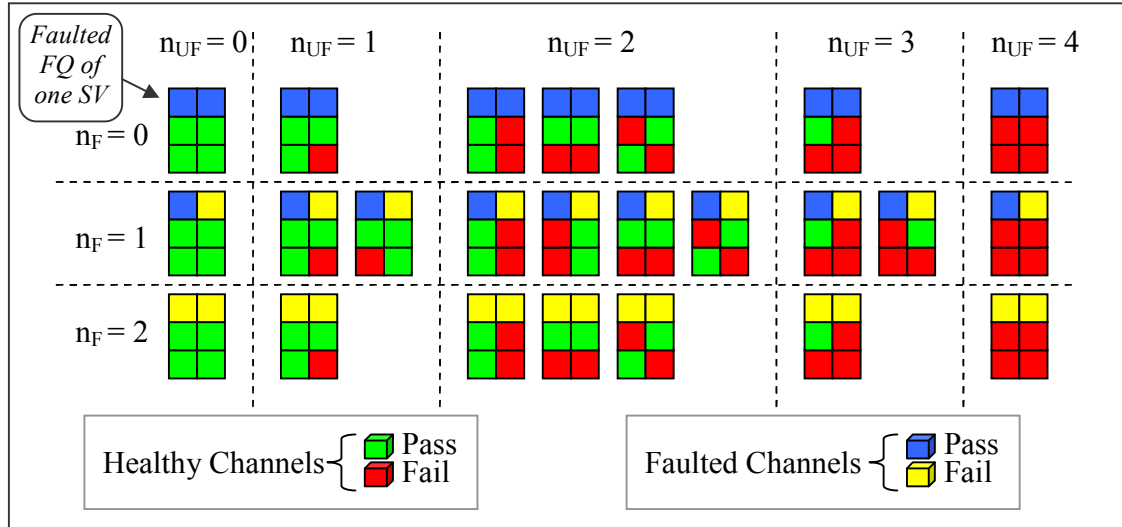
Figure 5.16: Simplified Example – Distinguishing Faulted from Un-faulted Channels

The methods described before for determining the un-faulted isomorphs can also determine the isomorphs for the Vector, Matrix, and Tensor faults shown in Figure 5.16. The number of isomorphs increases from 11 to 24 because a new potential state has essentially been added (i.e. whether the channel is *actually* healthy or not), but because those isomorphs which are highly improbable can be discarded, there is still a manageable set with which to work.

5.6 A Litany of Rules

This section defines a set of rules that govern what form of detection and exclusion will occur given a particular scenario of passing and failing channels. The faults described in bullets (i) through (vi) are referred to as "System Faults". Many sets of rules are possible to define these System Faults. One example set of rules is the following:

i) SV Vector Fault: 2 or more channel faults

ii) RX Vector Fault: 2 or more channel faults

iii) FQ Vector Fault: 2 channel faults (assuming only 2 signal frequencies)

iv) SV Matrix Fault: 3 or more channel faults

v)   RX Matrix Fault: 5 or more channel faults

vi)   FQ Matrix Fault: 7 or more channel faults

vii) Tensor Fault: 9 or more channel faults

viii) When two System Faults are declared, and if System Fault 1 precludes System Fault 2 from occurring, then System Fault 1 is the one chosen, but only if System Fault 2 requires a larger number of channel faults to trip. Otherwise, both System Faults are chosen. (This is relevant when it takes only two channel faults to trip either a V_{RX} or V_{SV} Fault with one or more common channels.)

ix)   If fewer than two receivers remain for a given frequency and satellite, that satellite may be "excluded" on that frequency due to lack of redundancy. This is similar to the single-frequency requirement that a satellite must be unflagged on at least two satellites to avoid being excluded [3].

Rules (i-vii) are self evident and use the nomenclature that has been used consistently throughout this thesis. Rules (viii) and (ix) are explained using the V_{SV} fault in Figure 5.17. All satellites on Receiver 3, Frequency 2 are declared to be faulty and are assumed to result in seven flagged channels (yellow) and three missed-detection channels (blue). Furthermore, Satellite 1 has "fault-free" detections on two more channels. The most obvious action here is to remove the SV-Vector on {RX 3, FQ 2}. This will be referred to as $V_{RX3,FQ2}$. The other questions are:

Q1)   What to do about Frequency 2?

A1)   Based on Rule (vi) all of FQ 2 would be excluded, but Rule (viii) intervenes because a V_{SV} Fault keeps a M_{FQ} Fault from being chosen, and the M_{FQ} Fault requires more channel faults than the V_{SV} Fault.

Q2)   What to do about Receiver 3?

A2)   Based on Rule (v), all of RX 3 would be excluded, but Rule (viii) intervenes.

Q3)   What to do about SV 1?

A3)   Based on Rule (iv), SV 1 would be excluded, but Rule (viii) intervenes.

Q4)   What to do about $V_{SV1,FQ2}$?

A4)   Based on Rule (ii), $V_{SV1,FQ2}$ would be excluded. With $V_{RX3,FQ2}$ being excluded, it turns out that $V_{SV1,FQ2}$ will also be excluded by rule (ix). $V_{RX3,FQ1}$ would be allowed to survive and contribute useful information to the system.

This set of rules becomes more complicated if there are four receivers, and it would be even more complex if the reference receivers tracked signals from other navigation satellite systems. For example, if one receiver had multiple faults in tracking Galileo satellites, should that influence the detection logic for the channels tracking GPS satellites? Among other things, the right answer would depend on what processes or hardware are common inside the receiver, such as whether there is a common clock/frequency standard or if the tracking loops are assisting each other.
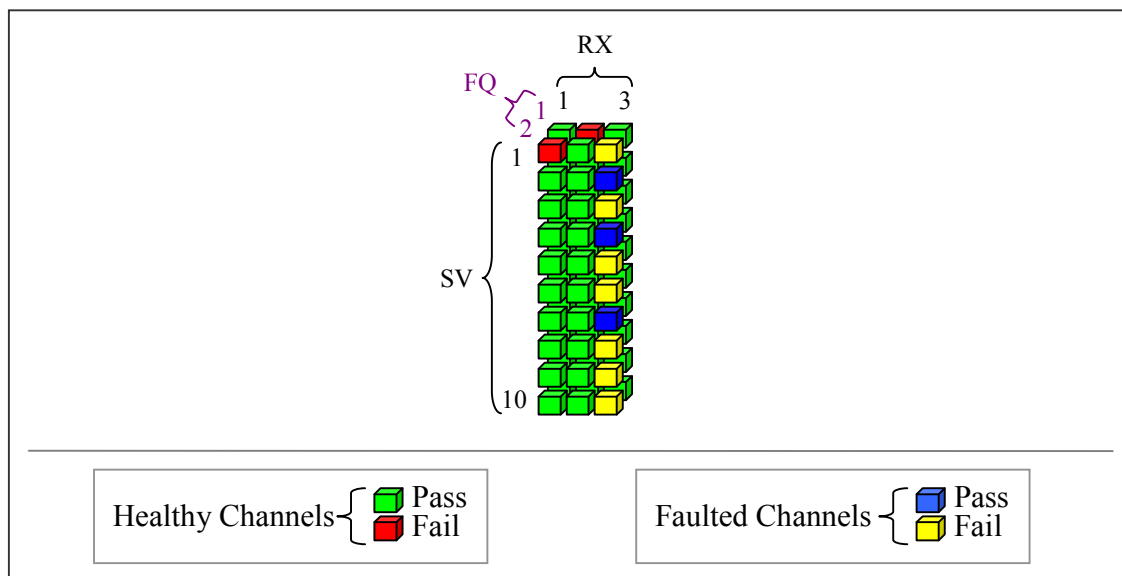
Figure 5.17: Explaining the Rules of Detection

5.7   Results and Conclusions

Using the rules given above, coupled with specific values for P_{md} and P_{ffd}, each of the channel scenarios represented by an isomorph (together with the knowledge of the number of repetitions of each isomorph) can be simulated to determine how frequently each of the fault modes shown in Figure 5.15 occurs. Table 5.11 and Table 5.12 show the VMT simulation results using the rules and probabilities given in Table 5.10. Each entry in Table 5.11 and Table 5.12 shows the probability of each type of System Fault being declared. In this table, the column headings show the input fault modes, and the row headings describe the response probabilities (the color coding just aids in distinguishing the rows). The first value (P_{CDF}) shows the probability of the event happening according to the binomial CDF given the selected P_{FFD}, P_{MD}, detection threshold, and number of channels of the ($n_{SV} \times n_{RX} \times n_{FQ}$) system (see Table 5.10 for these values). What follows is a comprehensive list of the terms used in the VMT output of Table 5.11 and Table 5.12:

P_{CDF}:   Probability of the event happening according to the binomial CDF

$1-P_{CDF}$: Complement of the P_{CDF}

P_{UD}:   Probability of an under-detect; what is detected is a subset of the true fault

P_D: Probability of accurately detecting the true fault

P_{OD}:   Probability of an over-detect; what is detected is a superset of the true fault

P_{RES}:   The residual probability in the simulation (described in Section 5.5.4).

P_{None}:   Probability that there is no detection at all

P_{SVF}:   Probability that there is an SV Vector detection of the declared fault

P_{SVG}:   Probability that there is an SV Vector detection other than the declared fault

P_{RVF}:   Probability that there is an RX Vector detection of the declared fault

P_{RVG}:   Probability that there is an RX Vector detection other than the declared fault

P_{FVF}:   Probability that there is an FQ Vector detection of the declared fault

P_{FVG}:   Probability that there is an FQ Vector detection other than the declared fault

P_{SMF}:   Probability that there is an SV Vector detection of the declared fault

P_{SMG}:   Probability that there is an SV Matrix detection other than the declared fault

P_{RMF}:   Probability that there is an RX Matrix detection of the declared fault

P_{RMG}:   Probability that there is an RX Matrix detection other than the declared fault

P$_{FMF}$:   Probability that there is an FQ Matrix detection of the declared fault

P$_{FMG}$:   Probability that there is an FQ Matrix detection other than the declared fault

P$_{TEN}$:   Probability that there is a Tensor detection (the entire system is declared faulted)

Table 5.10 gives the required number of faults to cause an exclusion of each mode. These are the parameters used in the simulation that generates the results in Table 5.11 and Table 5.12.

P$_{ffd}$	P$_{md}$	#V$_{SV}$	#V$_{RX}$	#V$_{FQ}$	#M$_{SV}$	#M$_{RX}$	#M$_{FQ}$	#T
10^{-8}	10^{-4}	2	2	2	3	5	5	9

Table 5.10: Simulation Parameters– Required Number of Faults for Exclusion

		Declared Fault Mode			
		FF	**V$_{SV}$**	**V$_{RX}$**	**V$_{FQ}$**
Detection Result	P$_{CDF}$	N/A	1	$1 - (3.00 \times 10^{-8})$	$1 - (2.00 \times 10^{-4})$
	$(1\text{-}P_{CDF})$	N/A	$\leq 1 \times 10^{-15}$	3.00×10^{-8}	2.00×10^{-4}
	P$_{UD}$	0	0	3.00×10^{-8}	2.00×10^{-4}
	P$_D$	$1 - (3.92 \times 10^{-14})$	$1 - (3.11 \times 10^{-14})$	$1 - (3.30 \times 10^{-7})$	$1 - (2.00 \times 10^{-4})$
	P$_{OD}$	3.6×10^{-14}	2.85×10^{-14}	3.00×10^{-7}	2.20×10^{-7}
	P$_{RES}$	3.22×10^{-15}	2.66×10^{-15}	3.11×10^{-15}	3.11×10^{-15}
	P$_{None}$	$1 - (3.92 \times 10^{-14})$	0	3.00×10^{-8}	2.00×10^{-4}
	P$_{SVF}$	$\leq 1 \times 10^{-15}$	$1 - (2.66 \times 10^{-15})$	2.70×10^{-7}	1.80×10^{-7}
	P$_{SVG}$	2.70×10^{-14}	2.25×10^{-14}	0	0
	P$_{RVF}$	0	2.00×10^{-7}	$1 - (3.00 \times 10^{-8})$	4.00×10^{-8}
	P$_{RVG}$	6.00×10^{-15}	0	2.43×10^{-14}	0
	P$_{FVF}$	0	1.00×10^{-8}	3.00×10^{-8}	$1 - (2.00 \times 10^{-4})$
	P$_{FVG}$	3.00×10^{-15}	0	0	2×10^{-15}
	P$_{SMF}$	0	$\leq 1 \times 10^{-15}$	3.00×10^{-8}	$\leq 1 \times 10^{-15}$
	P$_{SMG}$	$\leq 1 \times 10^{-15}$	0	0	0
	P$_{RMF}$	$\leq 1 \times 10^{-15}$	$1 - (2.66 \times 10^{-15})$	$\leq 1 \times 10^{-15}$	$\leq 1 \times 10^{-15}$
	P$_{RMG}$	0	0	0	0
	P$_{FMF}$	$\leq 1 \times 10^{-15}$	$1 - (2.66 \times 10^{-15})$	3.5×10^{-14}	$\leq 1 \times 10^{-15}$
	P$_{FMG}$	0	0	0	0
	P$_T$	0	$1 - (4.50 \times 10^{-7})$	0	0

Table 5.11: VMT Simulation Results (Part 1)

		Declared Fault Mode			
		M_{SV}	M_{RX}	M_{FQ}	T
Detection Result	P_{CDF}	$1 - (2.00 \times 10^{-11})$	1	1	1
	$(1-P_{CDF})$	2.00×10^{-11}	$\leq 1 \times 10^{-15}$	$\leq 1 \times 10^{-15}$	$\leq 1 \times 10^{-15}$
	P_{UD}	2.00×10^{-11}	0	0	0
	P_{D}	$1 - (2.00 \times 10^{-11})$	$1 - (2.21 \times 10^{-14})$	$1 - (1.97 \times 10^{-14})$	$1 - (3.22 \times 10^{-15})$
	P_{OD}	2.97×10^{-14}	2.00×10^{-14}	1.65×10^{-14}	0
	P_{RES}	2.89×10^{-15}	2.11×10^{-15}	3.22×10^{-15}	3.22×10^{-15}
	P_{None}	$\leq 1 \times 10^{-15}$	0	0	0
	P_{SVF}	5.40×10^{-7}	$1 - (2.11 \times 10^{-15})$	$1 - (3.22 \times 10^{-15})$	$1 - (3.22 \times 10^{-15})$
	P_{SVG}	0	1.80×10^{-14}	1.35×10^{-14}	0
	P_{RVF}	$1 - (3.77 \times 10^{-15})$	3.60×10^{-10}	$1 - (3.22 \times 10^{-15})$	$1 - (3.22 \times 10^{-15})$
	P_{RVG}	5.40×10^{-15}	0	3.00×10^{-15}	0
	P_{FVF}	$1 - (8.00 \times 10^{-12})$	$1 - (2.11 \times 10^{-15})$	3.00×10^{-7}	$1 - (3.22 \times 10^{-15})$
	P_{FVG}	2.70×10^{-15}	2.00×10^{-15}	0	0
	P_{SMF}	2.00×10^{-11}	6.00×10^{-15}	3.00×10^{-7}	$1 - (3.22 \times 10^{-15})$
	P_{SMG}	$\leq 1 \times 10^{-15}$	0	0	0
	P_{RMF}	$\leq 1 \times 10^{-15}$	$1 - (2.11 \times 10^{-15})$	$1 - (3.22 \times 10^{-15})$	$1 - (3.22 \times 10^{-15})$
	P_{RMG}	0	0	0	0
	P_{FMF}	7.02×10^{-14}	$1 - (2.11 \times 10^{-15})$	$1 - (3.22 \times 10^{-15})$	$1 - (3.22 \times 10^{-15})$
	P_{FMG}	0	0	0	0
	P_{T}	$\leq 1 \times 10^{-15}$	$1 - (2.11 \times 10^{-15})$	$1 - (3.22 \times 10^{-15})$	$1 - (3.22 \times 10^{-15})$

Table 5.12: VMT Simulation Results (Part 2)

The results in Table 5.11 and Table 5.12 will vary with the prior probabilities and the particular rules defining the EXM exclusion procedures. For each input fault mode, there is a very strong connection to the detected fault, as expected. When there is no input fault (i.e., FF, or fault-free), the system only has a probability of $(1 - 3.92 \times 10^{-14})$ of declaring a vector, matrix, or tensor fault. It does not include the probability of an

individual channel fault. To avoid confusion with regard to the fault-free case, the P_D from Table 5.10 does not only imply a detection; it also means that the accurate action was taken in not faulting any assortment of channels. If anything else had been faulted, that would constitute and over-detection and show in the value of P_{OD}.

Examining the simplest fault type, V_{FQ}, there are only two actual fault possibilities, and each has a specified P_{md} of 10^{-4}. A missed detection of this fault would require either of the two channels to be un-faulted. Equation (5.92) shows that P_{UD} (the probability of an Under-Detection) in this case is approximately equal to 2×10^{-4}.

$$Q_{md} = 1 - P_{md} \tag{5.90}$$

$$1 = P_{md}^2 + 2 \cdot P_{md} \cdot Q_d + Q_{md}^2 \tag{5.91}$$

$$\begin{aligned} P_{UD} &= 2 \cdot P_{md} \cdot Q_{md} + P_{md}^2 \\ &= 2 \cdot 10^{-4} \cdot (1 - P_{md}) + P_{md}^2 \\ &\approx 2 \cdot 10^{-4} \end{aligned} \tag{5.92}$$

An Over-Detection means that the actual fault was detected, but another fault was also declared when it shouldn't have been. This concept was discussed in Section 5.4. For an Over-Detection to occur, one of two things has happened:

1) A fault has been declared which contained the actual fault that exists. An example is that the entire satellite is declared to be faulted (M_{SV}) when only one receiver on one satellite is faulted (V_{FQ}).

2) A fault has been declared which matches the actual fault ($V_{FQ,}$) and another faulted has been declared which is neither a superset nor subset of the actual fault. An example is when both a V_{FQ} and V_{RX} fault are declared when only the V_{FQ} fault actually exists.

195

The V_{FQ} Fault is the one shown in Figure 5.16, which examines mixed isomorphs. It is shown again in Figure 5.18 to illustrate the two types of Over-Detection mentioned above.
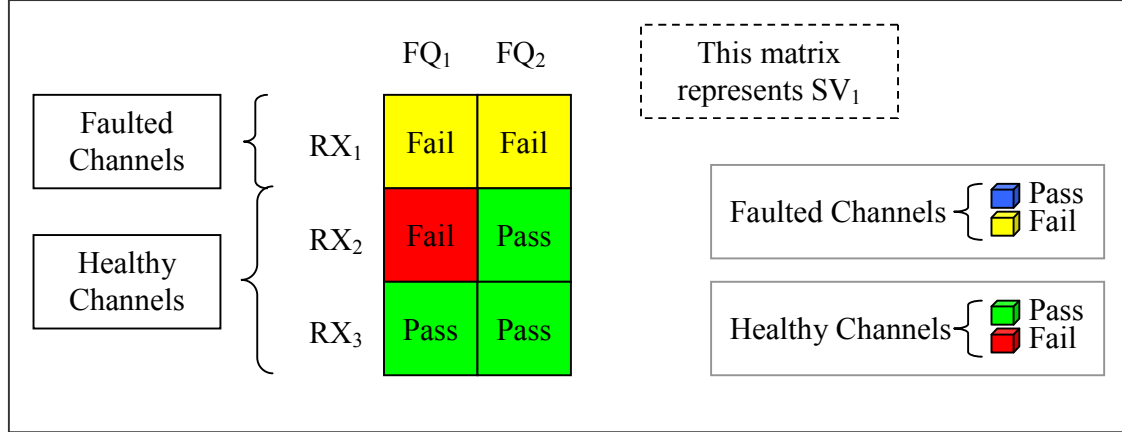

Figure 5.18: V_{FQ} Over-Detections

Figure 5.18 shows that for a given satellite, RX_1 is declared *a priori* to be unhealthy and is flagged on both frequencies. If that happens, the potential fault on $\{RX_1, FQ_1\}$ will cause some form of an Over-Detection. For a V_{FQ} fault, the most likely superset would be the M_{SV} fault. This can be restated as follows. For a given satellite and receiver, if the channels for both frequencies are declared *a priori* to be faulted, then the most likely System Fault to be declared which is larger than (and contains) the actual fault, would be if the entire satellite were to be declared faulted. This is the M_{SV} fault. Based on Table 5.10, a M_{SV} fault requires three failing channels out of the six total channels on a satellite to declare a M_{SV} fault. The two channels of the V_{FQ} fault have a $(1-P_{md})^2$ probability of both failing. When P_{md} is small, it is very likely that these two channels will both be failed. There needs to be only one failing channel among the other four nominal channels of the satellite to cause a MSV fault. This requires one fault-free detection of probability P_{ffd}. The term Q_{ffd} is the complement of P_{ffd} and is nearly unity. Equation (5.93) derives the probability of this type of "superset" Over-Detection.

$$P_{OD,SS} = 4 \cdot P_{ffd} \cdot Q_{ffd}^3 \cdot \left(1 - P_{md}\right)^2$$
$$\approx 4 \cdot 10^{-8}$$
$$(5.93)$$

196

When there is an *a priori* V_{FQ} fault, there is also the likelihood of V_{RX} detection. In the case of Figure 5.18, this means that FQ_1 across all receivers would be declared faulted. However, the detection rules stated at the beginning of Section 5.6 imply that both an M_{SV} and V_{RX} fault are possible. Rule (viii) does not apply because declaring a V_{RX} fault doesn't preclude an M_{SV} fault from happening. In this case, both System Faults would be declared, and this type of fault is represented by the "superset" Over-Detection probability shown in Equation (5.93)

The most likely Cross-Detection is the V_{SV} System Fault. This means that all the satellites for given frequency and receiver pairing are declared faulted. Referring to Figure 5.18, assuming the two channels {RX_1, FQ_1} and {RX_2, FQ_2} are failed, then if there was a channel failure on another satellite for these receiver-frequency pairings, that would result in a V_{SV} System Fault. When there are nine other satellites for the two frequencies on RX_1, then there are eighteen channel failures which would result in a V_{SV} System Fault. Equation (5.94) shows the calculation of the probability of this outcome.

$$P_{OD,CD} = 2 \cdot 9 \cdot P_{ffd} \cdot Q_{ffd}^8 \cdot \left(1 - P_{md}\right)$$
$$\approx 18 \cdot 10^{-8}$$

(5.94)

This covers the two cases of Over-Detection for the VFQ fault mode. Therefore the probability of an Over-Detection, P_{OD} = 2.2 × 10^{-7}, is the sum of the probabilities derived in Equations (5.93) and (5.94). This value is given in Table 5.11.

Table 5.11 and Table 5.12 provide a way to look at the ultimate effect of any particular monitor used within an Integrity Monitoring system. Each monitor that calculates a quality metric and compares it to a threshold has the potential to make fault free detections or to miss the detection of an underlying fault. The method developed in this chapter demonstrates how to apply the probabilities of fault-free detection (P_{ffd}) and missed-detection (P_{md}), along with a codified set of rules using the VMT, to establish what the actual effect would be on the exclusion of channels which may result in a

dropped receiver, satellite, or signal frequency. The decision logic used in the IMT (the predecessor to the JTeP) dealt with only a single frequency. The addition of a second frequency increases the complexity of the decision logic, and the VMT provides a method to determine if decision rules which are geared towards detecting certain faults are actually causing a significant amount of Cross-Detections. That is, based on a candidate set of decision rules, does a fault with a receiver commonly get misinterpreted as a fault with a satellite?

The development of a correlated multivariate Student t-distribution using the lower triangular representation provides a way to vary prior probabilities to simulate a mixture of dissimilar satellites instead of a set of purely homogenous ones. Chapter Five showed that the data being tested in MQM is both non-Gaussian and correlated across receivers. Simulating data from a correlated multivariate Student t-distribution enhances the ability of the VMT to predict the system's response to more realistic data. Given P_{ffd}'s and P_{md}'s that are sufficiently low ($\leq 10^{-4}$), the simulation results shown in this section demonstrate that EXM will not contribute any substantial risk of Cross-Detection. This conclusion suggests that each monitor algorithm that contributes to EXM can be developed independently provided that each individual algorithm satisfies its own sub-allocated P_{ffd} and P_{md}.

Chapter 6 - Conclusion

GPS has become ingrained in the everyday lives of more people than ever envisioned when the system was originally conceived. Our society will come to depend on GPS devices for critical applications which directly affect the health and safety of many. One of those applications is the landing of aircraft, both civilian and military. The contributions of this thesis enhance the ability of a landing system to provide accurate guidance with very high confidence in order to safeguard the lives of those who come to depend on such systems.

The Stanford GPS Lab has developed a software system called the Integrity Monitor Testbed (IMT) in order to meet the requirements of the civilian Local Area Augmentation System (LAAS). The IMT operates as a reference station in that it calculates and broadcasts range corrections for each GPS satellite it tracks and deems healthy. The IMT has a comprehensive system of monitoring algorithms which determine which satellites are operating nominally and which satellites should be excluded because they could result in hazardously misleading information (HMI) at the aircraft.

The IMT, developed under the LAAS program, was the starting point of the system developed in this thesis called the JPALS Testbed Platform. JPALS stands for the Joint Precision and Approach Landing System. It is military program, comparable to LAAS, which seeks to use GPS for landing aircraft. LAAS reference stations are installed at commercial airports and benefit from having well-sited antennas and a low occurrence of either deliberate or accidental electronic interference. JPALS, on the other hand, would operate on military airstrips or aircraft carriers which may well be rapidly deployed to less hospitable environments. For a ground based system, this means that there might be more uncertainty associated with the location of the antennas if there is insufficient time to survey the antenna locations. For a system deployed on a large aircraft carrier experiencing structural flexure, there is an error associated with

connecting the exact location of the GPS antennas to the touchdown point. Also, in both cases, there is less freedom to choose the antenna location, so multi-path may be correlated across antennas. Lastly, there may be other antennas nearby creating unintentional interference in addition to the possibility that unfriendly forces may try to deliberately interfere with the GPS signal.

All of these circumstances serve to inhibit the ability of an integrity monitoring system to detect any anomalous behavior in GPS measurements. This is the motivation to increase the Measurement Quality Monitor detection performance in the current IMT. It also is motivation to refine the Executive Monitor decision logic in the IMT, which decides whether or not a failure on one or more channels in an indication that either a satellite or GPS receiver is faulty. If this decision logic is inefficient, it may unnecessarily exclude useable ranging information because it is operating over-cautiously. Both of these improvements were tested using the JPALS Testbed. Also, the IMT currently operates on only one GPS frequency, L1. Future integrity monitoring systems will operate on multiple frequencies, L1, L2, and L5. This thesis also examined the effect of an additional frequency on the decision algorithms.

6.1   Summary of Results

This thesis is comprised of three contributions. The first contribution is a thorough examination of the Measurement Quality Monitor algorithm, which demonstrates that the filter used to estimate the acceleration of the GPS carrier phase can be improved. The second contribution looks at the first Executive Monitoring Logic segment of the IMT and shows that more effective decision logic can be used to detect a fault which is common across all receivers for a given satellite. The third contribution of this thesis derives a method to calculate the probability of Cross-Detection for the channel exclusion rules of the Executive Monitoring Logic for a two frequency system.

6.1.1 Measurement Quality Monitoring (MQM)

The IMT subsystem known as Measurement Quality Monitoring (MQM) is responsible for detecting anomalies in carrier phase measurements, including abnormally large accelerations. This thesis has shown in Chapter Three that using a filter which is a combination of two spliced polynomials can improve detection performance for carrier-phase acceleration when compared against the filter used by the IMT. The newly derived filter is simultaneously more robust to noise while also being more responsive to an acceleration on the carrier phase. Overall this increases the filter's sensitivity and performance. Additionally, this chapter demonstrates the weakness of the velocity filter currently under consideration for use in the IMT. By applying a different set of regression indices, a new velocity filter is derived which markedly outperforms the IMT velocity filter.

6.1.2 Executive Monitoring (EXM) Logic

This thesis has shown that EXM can leverage measurement processing across multiple receivers to further increase its ability to detect faults common across receivers. The obvious difficulty in doing this is that a set of bad measurements on one receiver could jeopardize measurements on other receivers. The methods developed in Chapter Four show that a unique form of multi-receiver thresholds can both increase the ability to detect faults while also keeping the system robust to single-channel faults.

6.1.3 The Vector-Matrix-Tensor (VMT) Method

Lastly, Chapter Five of this thesis demonstrates how to determine the detection rates when the system incorporates a second frequency. The result of this chapter shows how detection rules for receivers and satellites can be independent, thus enabling the design of these detection rules without affecting each other. This result is a function of the exclusion rules defining the EXM and the channel fault-free probability (P_{ffd}) and missed detection probability (P_{md}).

Taken together, these contributions improve the performance of the Stanford IMT by increasing the ability of the MQM to detect certain faults and by optimizing the decision logic of the first phase of EXM in order to avoid unnecessarily exclusion of un-faulted channels, thereby increasing the availability of the system. These improvements apply equally to JPALS as well.

6.2   Future Work

The improvements made to the acceleration and velocity filters of the MQM require testing and refinement of the exact thresholds before implementation. In Chapter Three, the new acceleration filter was designed to outperform the IMT acceleration filter under the full range of first-order correlation values of the noise. It is possible that some higher-order correlation exists on the carrier phase that would affect the results achieved in Chapter Three. Similar consideration should be made towards the improvements to EXM derived in Chapter Four. The ideas of kurtosis and correlation across receivers were examined extensively, both by using real data and by modeling data so that simulations could be run out to high thresholds. The results by theory and simulation each showed improvement, but it is prudent to test the recommendations of this chapter against more collected data. The results of Chapter Five are applicable to a two-frequency system, but are currently computationally limited from applying to a three-frequency system. Further work is needed to develop a method that can support an additional frequency.

6.3   Closing

The best way to describe the role of GPS in aviation is taken from the International Civil Aviation Organization *U.S. Satellite Navigation Program Status Report* which reads:

GPS is here to stay for use in civil aviation navigation worldwide. The Global Positioning System has proven its value, first in providing new military capabilities, and more recently for such diverse interests as hiking, farming, emergency service, and virtually all modes of transportation. There are many pilots today who rely more on unaugmented GPS than on conventional navigation and landing systems. Golfers use it, truckers use it, and, of course, sailors use it, but the system in place today falls short of the high standards required for all phases of civil flight [1].

This is the impetus for development of the LAAS and WAAS programs. It is also the impetus for developing JPALS, a military version for supporting GPS based aircraft landings. Much work is needed to reach the broad objective of providing landing assistance in a timely manner with the highest levels of accuracy and integrity. This thesis has made progress towards that end.
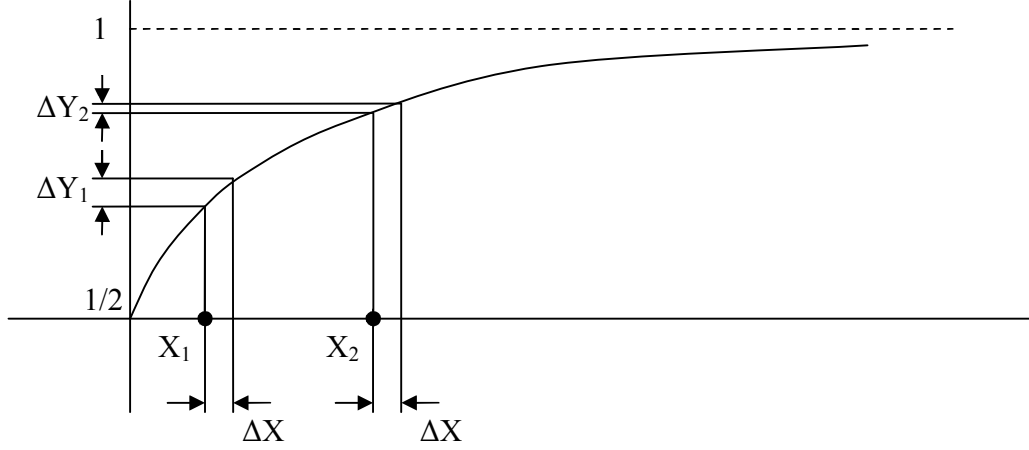
Appendix A: CDF Proof

Figure A1: Gaussian CDF

This curve is the CDF of x (with zero mean and unity variance) and satisfies:

$$pdf(x) = \frac{1}{\sqrt{2\pi}} e^{\frac{-x^2}{2}}$$
Eq. A1-1

$$y(x) = cdf(x) = \int_{-\infty}^{x} pdf(\tau)d\tau$$
Eq. A1-2

$$y(x) \le y(x+a),\ a \ge 0$$
Eq. A1-3

$$y_x(x) \ge y_x(x+a),\ [a \ge 0, when\ x \ge 0]$$
Eq. A1-4

Eq. A1-4 quickly implies that $\Delta y_1 \ge \Delta y_2$, but the certainty follows:

$$\Delta y_1 = \int_{x_1}^{x_1+\Delta x} f_x(x)dx$$
Eq. A1-5

$$\Delta y_2 = \int_{x_2}^{x_2+\Delta x} f_x(x)dx = \int_{x_1}^{x_1+\Delta x} f_x(x+(x_2-x_1))dx$$
Eq. A1-6

$$\Delta y_1 - \Delta y_2 = \int\limits_{x_1}^{x_1+\Delta x} f_x(x)dx - \int\limits_{x_1}^{x_1+\Delta x} f_x(x+(x_2-x_1))dx$$

$$= \int\limits_{x_1}^{x_1+\Delta x} \left(f_x(x) - f_x(x+(x_2-x_1)) \right) dx$$

$$= \int\limits_{x_1}^{x_1+\Delta x} \left(f_x(x) - f_x(x+a) \right) dx, \quad [a \geq 0]$$

$$\geq 0$$

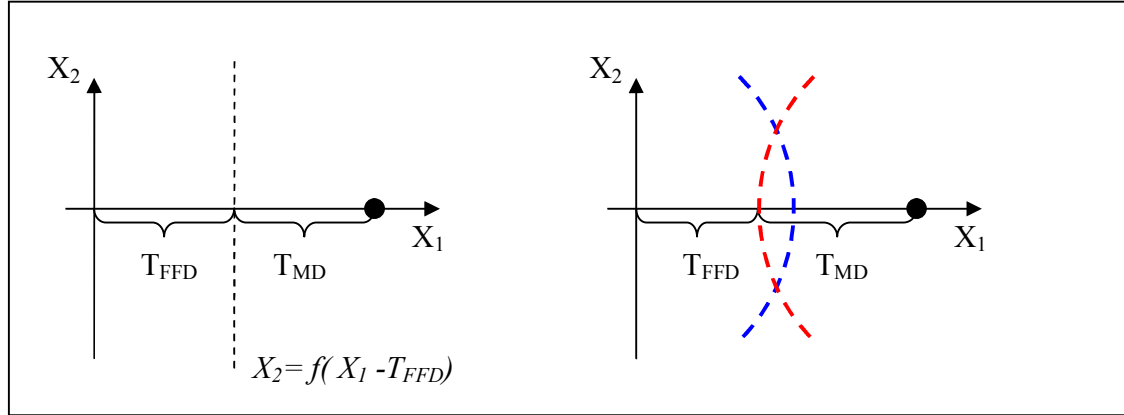Eq. A1-7

Appendix B: Optimal Threshold Proof

Figure B1: 2-D Optimal Detection Threshold Curves

The point is to show that the equation $x_2 = f(x_1)$ that minimizes the distance $T_{FFD} + T_{MD}$, is a straight line.

Independence

Since x1 is independent of x_2, the process can be determined solely along the x_1 axis, and point T_{FFD} determined. This point is then projected straight up and down. If considering x_2 could optimize detection of an x_1 bias, that implies it is better to deliberately introduce a random variable into the system. Induction reasoning would suggest adding x_3, x_4, etc. would be even better. But is counterintuitive, and the solution lies within analyzing x_1 alone.

Symmetry

If the curve of $x_2 = f(x_1 - T_{FFD})$ were anything other than a straight line, then it would be an equal solution to flip it such that $x_2 = f(T_{MD} - x_1)$. Why is the red curve correct and not the blue? Equal exclusion probabilities assumed.
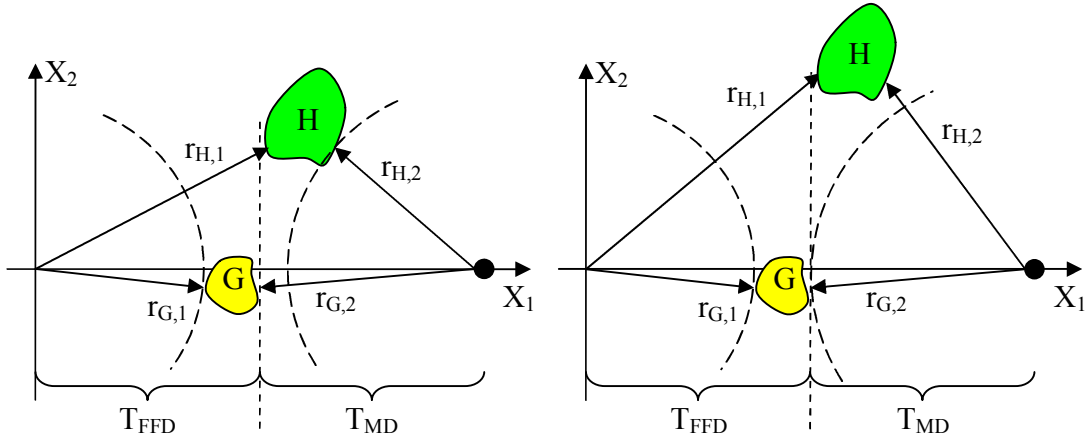
Derivative

Figure B2: Symmetry of Probability

In this rendering the dividing line of T_{FFD} has been specified, and we are examining the effects of perturbing the line. The areas G and H are also assumed to be imperceptibly small. If area G is removed from the FFD probability and area H included, there are certain declarations. Because $r_{G,1} < r_{H,1}$, the area of G is less than the area of H, $A_G < A_H$ so that P(G) = P(H). Now, to the P_{MD} side, if $r_{G,2} > r_{H,2}$ then clearly P(H) > P(G), and T_{MD} must be increased to compensate.

When $r_{G,2} < r_{H,2}$, then it is true that $r_{G,2} > r_{G,1}$, and $r_{H,2} < r_{H,1}$. Using the conclusion of A-1 that $\Delta y_1 - \Delta y_2 > 0$, implies that $\Delta x_2 - \Delta x_1 > 0$, when $\Delta y_1 = \Delta y_2$. This translates to say that A_H is larger than A_G when $P(H_1) = P(G_1)$. Since $r_{G,2} > r_{G,1}$ (with A_G constant), $P(G_1) > P(G_2)$ and similarly, $P(H_2) > P(H_1)$. The chain is $P(H_2) > P(H_1) = P(G_1) > P(G_2)$, or $P(H_2) > P(G_1)$. Consequently there is too much included probability on the MD side and T_{MD} must be increased to reduce the probability densities of G and H.

Appendix C: System of Logic

Five methods of experimental reasoning distinguished by John Stuart Mill (1843)

Used in determining what factors cause a specific effect, E, under a specific set of circumstances

1. **The Method of Agreement:**
    Look for factors present on all occasions when E occurs

2. **The Method of Difference:**
    Look for some factor present on some occasion when E occurs and absent on an otherwise similar occasion when it does not.

3. **The Joint Method of Agreement and Difference:**
    Combine the previous two methods.

4. **The Method of Residues:**
    When part of E is explicable by reference to known factors, the "residue" is attributed to the remaining circumstances under which E has occurred.

5. **The Method of Concomitant Variation:**
    When E can be present in various degrees; if we identify a factor F whose variations are positively or negatively correlated with variations in E then we can infer that F is causally connected with E.

Appendix D: MQM Covariance

$$M = \begin{pmatrix} 1 & \dfrac{1-n}{2} & \frac{1}{2}\left(\dfrac{1-n}{2}\right)^2 \\ \vdots & \vdots & \vdots \\ 1 & \dfrac{n-1}{2} & \frac{1}{2}\left(\dfrac{n-1}{2}\right)^2 \end{pmatrix}$$

$$M^{\dagger} = \left(M^T M\right)^{-1} M^T$$

$$\begin{aligned} \Sigma_{M^{\dagger}} &= M^{\dagger} \cdot \left(M^{\dagger}\right)^T \\ &= \left(M^T M\right)^{-1} M^T \cdot \left(\left(M^T M\right)^{-1} M^T\right)^T \\ &= \left(M^T M\right)^{-1} M^T \cdot M \left(M^T M\right)^{-1} \\ &= \left(M^T M\right)^{-1} \end{aligned}$$

$$\left(M^T M\right) = \begin{pmatrix} n & 0 & \frac{1}{2}\left(\dfrac{n^3 - n}{12}\right) \\ 0 & \left(\dfrac{n^3 - n}{12}\right) & 0 \\ \frac{1}{2}\left(\dfrac{n^3 - n}{12}\right) & 0 & \frac{1}{4}\left(\dfrac{3n^5 - 10n^3 + 7n}{240}\right) \end{pmatrix}$$

$$\Sigma_{M^{\dagger}} = \left(M^T M\right)^{-1} = \begin{pmatrix} \dfrac{21 - 9n^2}{16n - 4n^3} & 0 & \dfrac{30}{4n - n^3} \\ 0 & \dfrac{12}{n^3 - n} & 0 \\ \dfrac{30}{4n - n^3} & 0 & \dfrac{720}{n^5 - 5n^3 + 4n} \end{pmatrix}$$

Appendix E: Noise Normalization

$$y_n = \tau \cdot y_{n-1} + \rho \cdot x_n$$

When the process is zero mean, the variance and the expected square are equal

$$
\begin{aligned}
\sigma^2_{y(n)} &= E[y_n^2] - E^2[y_n] \\
&= E[y_n^2] \\
&= E[(\tau \cdot y_{n-1} + \rho \cdot x_n)^2] \\
&= E[\tau^2 y_{n-1}^2 + 2\tau\rho y_{n-1} x_n + \rho^2 x_n^2] \\
&= \tau^2 E[y_{n-1}^2] + \rho^2 E[x_n^2] \\
&= \tau^2 \sigma^2_{y(n-1)} + \rho^2 \sigma^2_x
\end{aligned}
$$

Assuming the variance is stationary

$$\sigma^2_{y(n)} = \tau^2 \sigma^2_{y(n)} + \rho^2 \sigma^2_x$$

$$\sigma^2_y (1 - \tau^2) = \rho^2 \sigma^2_x$$

Solving for rho in order for sigma y to equal sigma x.

$$\rho = \sqrt{(1 - \tau^2)}$$

Appendix F: Reciprocal Solution

If the $(1^+/3)$ rule is used, then one or more channels faults results in a vector fault. This is tacitly a fault-free analysis. If we assume the vector is faulted, then again one or more faults causes a flag, but this is desired. It is actually zero or less (trivial here) that causes the vector to pass (a missed detection).
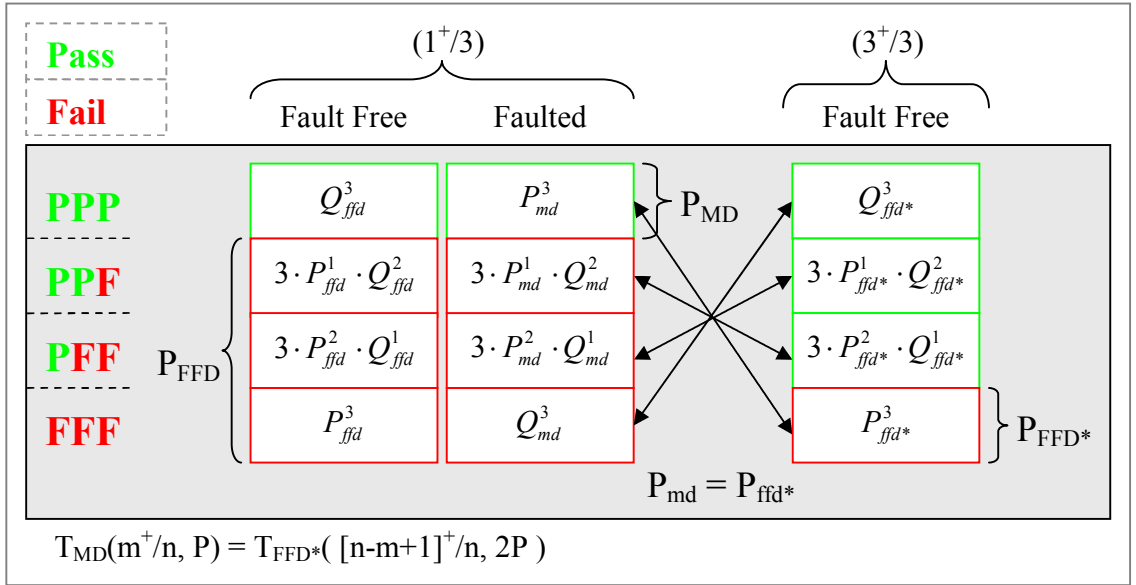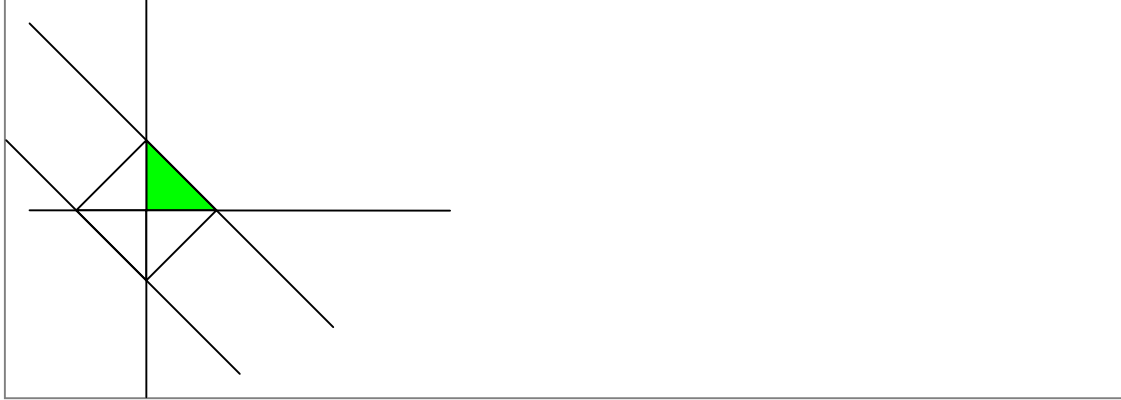

Figure F1: Reciprocal Solution

Appendix G: Gaussian Triangles 2-D

$$A = \int\limits_{0}^{A} \int\limits_{0}^{1-x} pdf(x)\, pdf(y)\cdot dy dx$$

$$= \frac{1}{2\pi} \int\limits_{0}^{A} \int\limits_{0}^{1-x} e^{-\frac{1}{2}(x^2+y^2)}\cdot dy dx$$

$$= \frac{1}{2\pi} \int\limits_{0}^{A} e^{-\frac{1}{2}x^2} \cdot Erf\left[\frac{A-x}{\sqrt{2}}\right]\cdot dx$$

At this point, a solution is not solvable in this form. We can alter the problem, realizing a change of variable to xbar and deltax.

$$P_{\bar{x}} = \int\limits_{-A/\sqrt{2}}^{A/\sqrt{2}} pdf[x]\cdot dx$$

$$= Erf\left[\tfrac{A}{2}\right]$$

$$A = \frac{1}{4}\cdot P_{\bar{x}}^2$$

$$= \frac{1}{4}\cdot Erf\left[\tfrac{A}{2}\right]^2$$

Square Integral

Included area

$$A = \frac{1}{4}\cdot Erf\left[\tfrac{A}{\sqrt{2}}\right]^2$$

Excluded area

$$A = \frac{1}{4}\cdot \left(Erf\left[\tfrac{A}{\sqrt{2}}\right]-1\right)^2$$

Appendix H: Gaussian Triangles 3-D

What was once so easy is two dimensions is purely enviable in three. Symbolic integrations programs couldn't even achieve the results for two dimensions and thus we used a symmetry existing to the Boolean forms. This is lost in three dimensions because the shape of the abs(AVG) is an octahedral. We can use a numerical integration through, with a polynomial fit.

We integrate the 2-D footprint, with the value T being the apex, or the dimension of the corner.
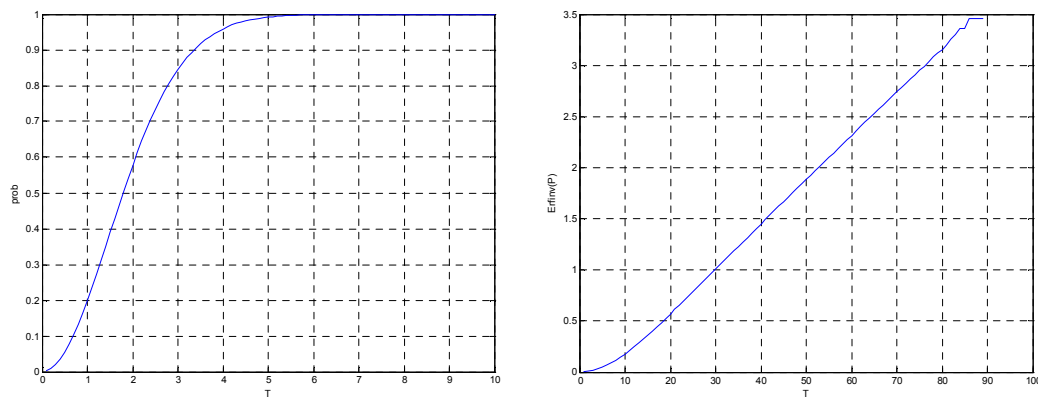
Taking the log of the probability shows an exponential trend.
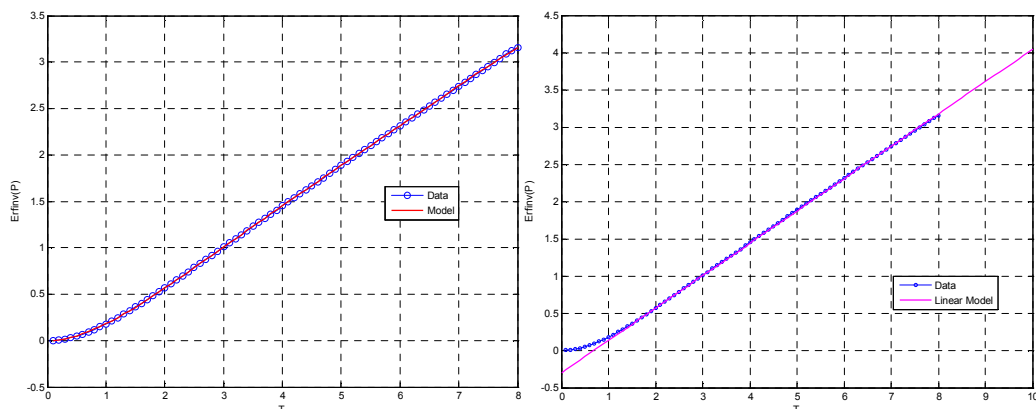
This figure shows the model fit, which is quite good at larger values, through for really small values the model is marginally off. The resulting equation uses a fifth order polynomial.

$$C = \begin{vmatrix} -9.7179e-3 \\ 3.1421e-2 \\ 2.0163e-1 \\ -4.5929e-2 \\ 4.8959e-3 \\ -1.9922e-4 \end{vmatrix}$$

$$P = Erf\left[\sum_{i=0}^{5} C_i \cdot (3 \cdot T)^i \right]$$

But this causes problems at large values. We use a linear form, first then an approximating linear form.

At T = 2, P* = 0.5677012

At T = 12, P* = 4.83

$$P = Erf\left[0.5677012 + \frac{4.83 - 0.5677012}{12 - 2}(T-2) \right]$$

$$= Erf\left[0.5677 + 0.42623 \cdot T \right]$$

Appendix I: Boolean CDFs 2-D

$$P_{FFD,1+/2} = 1 - \left(cdf(t) - cdf(-t) \right)^2$$
$$= 1 - Erf\left(\tfrac{t}{\sqrt{2}} \right)^2$$
$$P_{MD,1+/2} = cdf(-t)^2$$
$$= \tfrac{1}{4}\left(1 - Erf\left(\tfrac{t}{\sqrt{2}} \right) \right)^2$$
$$P_{FFD,2/2} = 4 \cdot cdf(-t)^2$$
$$= \left(1 - Erf\left(\tfrac{t}{\sqrt{2}} \right) \right)^2$$
$$P_{MD,2/2} = 1 - cdf(t)^2$$
$$= 1 - \tfrac{1}{4}\left(1 + Erf\left(\tfrac{t}{\sqrt{2}} \right) \right)^2$$
$$A^* = 1 - Erf(t)^2$$

$$T_{FFD,1+/2} = \sqrt{2} \cdot Erf^{-1}\left[\sqrt{1 - P_{FFD,1/2}} \right]$$
$$T_{MD,1+/2} = \sqrt{2} \cdot Erf^{-1}\left[1 - 2\sqrt{P_{MD,1/2}} \right]$$
$$P_{FFD,2/2} = \sqrt{2} \cdot Erf^{-1}\left[1 - \sqrt{P_{FFD,2/2}} \right]$$
$$P_{MD,2/2} = \sqrt{2} \cdot Erf^{-1}\left[\sqrt{4\left(1 - P_{MD,2/2} \right) + 1} \right]$$

Appendix J: Boolean CDFs 3-D

$$P_{FFD} = cdf[t] - cdf[-t]$$
$$= Erf\left[\tfrac{t}{\sqrt{2}}\right]$$
$$Q_{FFD} = 1 - P_{FFD}$$
$$= 1 - Erf\left[\tfrac{t}{\sqrt{2}}\right]$$
$$Q_{MD} = 1 - cdf[-t]$$
$$= \tfrac{1}{2}\left(1 + Erf\left[\tfrac{t}{\sqrt{2}}\right]\right)$$
$$P_{MD} = 1 - Q_{MD}$$
$$= \tfrac{1}{2}\left(1 - Erf\left[\tfrac{t}{\sqrt{2}}\right]\right)$$

$$P_{FFD,m/n} = \sum_{i=m}^{n} \binom{n}{i} \cdot Q_{FFD}^{n-i} \cdot P_{FFD}^{i}$$

$$P_{MD,m/n} = \sum_{i=0}^{m-1} \binom{n}{i} \cdot P_{MD}^{n-i} \cdot Q_{MD}^{i}$$

$$P_{FFD,1+/3} = 1 - Erf\left[\tfrac{t}{\sqrt{2}}\right]^{3}$$
$$P_{MD,1+/3} = \tfrac{1}{8} Erfc\left[\tfrac{t}{\sqrt{2}}\right]^{3}$$
$$P_{FFD,2+/3} = Erfc\left[\tfrac{t}{\sqrt{2}}\right]^{2} \cdot \left(1 + 2 \cdot Erf\left[\tfrac{t}{\sqrt{2}}\right]\right)$$
$$P_{MD,2+/3} = \tfrac{1}{4}\left(2 - 3 \cdot Erf\left[\tfrac{t}{\sqrt{2}}\right] + Erf\left[\tfrac{t}{\sqrt{2}}\right]^{3}\right)$$
$$P_{FFD,3/3} = Erfc\left[\tfrac{t}{\sqrt{2}}\right]^{3}$$
$$P_{MD,3/3} = 1 - \tfrac{1}{8}\left(1 + Erf\left[\tfrac{t}{\sqrt{2}}\right]\right)^{3}$$

$$T_{FFD,1+/3} = \sqrt{2} \cdot Erf^{-1}\left[\sqrt[3]{\left(1 - P_{FFD,1+/3}\right)} \right]$$

$$T_{MD,1+/3} = \sqrt{2} \cdot Erf^{-1}\left[\sqrt[3]{\left(1 - 8 \cdot P_{FFD,1+/3}\right)} \right]$$

$$T_{FFD,2+/3} = \sqrt{2} \cdot Erf^{-1}\left[\frac{1}{\left(2\sqrt{P(P-1)} + 2P - 1\right)^{1/3}} + \left(2\sqrt{P(P-1)} + 2P - 1\right)^{1/3} \right]$$

$$P_{MD,2+/3} = \sqrt{2} \cdot Erf^{-1}\left[\frac{1}{2}\left(1 - \frac{1}{\left(2\sqrt{P(P-1)} - 2P + 1\right)^{1/3}} - \left(2\sqrt{P(P-1)} - 2P + 1\right)^{1/3} \right) \right]$$

$$P_{FFD,3/3} = \sqrt{2} \cdot Erf^{-1}\left[\sqrt[3]{1 - P_{FFD,3/3}} \right]$$

$$P_{MD,3/3} = \sqrt{2} \cdot Erf^{-1}\left[\sqrt[3]{1 - 8 \cdot \left(1 - P_{MD,3/3}\right)} \right]$$

Appendix K: Boolean CDFs 4-D

$$T_{FFD,1+/4} = 1 - Erf\left[\tfrac{t}{\sqrt{2}}\right]^4$$

$$T_{MD,1+/4} = \tfrac{1}{16} Erfc\left[\tfrac{t}{\sqrt{2}}\right]^4$$

$$P_{FFD,2+/4} = 1 + Erf\left[\tfrac{t}{\sqrt{2}}\right]^3 \left(3 \cdot Erf\left[\tfrac{t}{\sqrt{2}}\right] - 4\right)$$

$$P_{MD,2+/4} = \tfrac{1}{16} Erfc\left[\tfrac{t}{\sqrt{2}}\right]^3 \left(5 + 3 \cdot Erf\left[\tfrac{t}{\sqrt{2}}\right]\right)$$

$$P_{FFD,3+/4} = Erfc\left[\tfrac{t}{\sqrt{2}}\right]^3 \left(4 - 3 \cdot Erfc\left[\tfrac{t}{\sqrt{2}}\right]\right)$$

$$P_{MD,3+/4} = \tfrac{1}{16} Erfc\left[\tfrac{t}{\sqrt{2}}\right]^2 \left(11 + 10 \cdot Erf\left[\tfrac{t}{\sqrt{2}}\right] + 3 \cdot Erf\left[\tfrac{t}{\sqrt{2}}\right]^2\right)$$

$$P_{FFD,4/4} = Erfc\left[\tfrac{t}{\sqrt{2}}\right]^4$$

$$P_{MD,4/4} = 1 - \tfrac{1}{16}\left(1 + Erf\left[\tfrac{t}{\sqrt{2}}\right]\right)^4$$

References

1. Misra, Pratap and Per Enge. <u>Global Positing System: Signals Measurements, and Performance</u>, Lincoln, MA: Ganga-Jamuna Press, 2001.

2. Pervan, Boris. "Navigation Integrity for Aircraft Precision Landing Using the Global Positioning System." Stanford Dissertation, 1996.

3. Luo, Ming, et al. "Development and Testing of the Stanford LAAS Ground Facility Prototype." *Proceedings of the 2000 National Technical Meeting of the Institute of Navigation*, January 26 - 28, 2000, pp. 210-219.

4. Xie, Gang, et. al. "Integrity Design and Updated Test Results for the Stanford LAAS Integrity Monitor Testbed." *Proceedings of the 57th Annual Meeting of the Institute of Navigation*, June 11 - 13, 2001, pp. 681 – 693.

5. Rife, Jason. "Vertical Protection Levels for a Local Airport Monitor for WAAS." *ION 2005, Proceedings of the 61st Annual Meeting*, June 27 - 29, 2005, pp. 745 – 758.

6. Pullen, Sam and Ming Luo et. al. "GBAS Validation Methodology and Test Results from the Stanford LAAS Integrity Monitor Testbed." *13th International Technical Meeting of the Satellite Division of the Institute of Navigation*, September 18-22, 2000, pp 1191-1201.

7. Rife, Jason and Eric Phelts. "Formulation of a Time Varying Maximum Allowable Error for Ground Based Augmentation Systems." *Proceedings of the 2006 National Technical Meeting of the Institute of Navigation*, January 18 - 20, 2006, pp. 441 – 453.

8.   Walpole, Ronald E. and Raymond H. Meters. Probability and Statistics for Engineers and Scientists 5th Edition.

9.   Luo, Ming, and Sam Pullen, et. al. "LAAS Ionosphere Spatial Gradient Threat Model and Impact of LGF and Airborne Monitoring." *Proceedings of the 16th International Technical Meeting of the Satellite Division of the Institute of Navigation ION GPS/GNSS 2003*, pp. 2255 – 2274.

10.   Rajagopal, Sriram and Todd Walter, et. al. "Correlation Structure of the Equatorial Ionosphere." *Proceedings of the 2004 National Technical Meeting of the Institute of Navigation*, San Diego, CA. pp. 542 – 550.

11.   Pullen, Sam and Ming Luo et. al. "LAAS Ground Facility Design Improvements to Meet Proposed Requirements for Category II/III Operations." *Proceedings of the 15th International Technical Meeting of the Satellite Division of the Institute of Navigation ION GPS 2002*, Portland, OR, pp. 1934 – 1947.

12.   Kotz, Samuel; Nadarajah, Saralees. Multivariate t Distributions and Their Applications. Cambridge University Press, 2004.

13.   Walter, Todd and Andrew Hansen et. al. "Validation of the WAAS MOPS Integrity Equation." *Meeting Proceedings of the 55th Annual National Technical Meeting,* June 28-30, 1999, pp. 217-226.

14.   Mittelman, Alexander.   "Signal Quality Monitoring for GPS Augmentation Systems." Doctoral Dissertation, Stanford University, 2004.

15.   Xie, Gang.   "Optimal On-Airport Monitoring of the Integrity of GPS-Based Landing Systems." Doctoral Dissertation, Stanford University, 2004.

16   Fontana, Richard, Wai Cheung, and Tom Stansell. "The Modernized L2 Civil Signal: Leaping Forward in the 21st Century." *GPS World* (Sept 2001), pp. 28-34.

17   Phelts, Robert Eric. "Multicorrelator Techniques for Robust Mitigation of Threats to GPS Signal Quality." Doctoral Dissertation, Stanford University, 2001.

18   Kim, Ung Suok, et. al. "Precise Phase Calibration of a Controlled Reception Pattern GPS Antenna for JPALS." Presented at the IEEE Position, Location and Navigation Symposium, Monterey, CA, April 27-29, 2004.

19   J. Lee, S. Pullen, et. al. "LAAS Sigma Monitor Analysis and Faulture-Test Verification." *ION 57th Annual Meeting, Albuquerque*, NM, June 11-13, 2001, pp. 694 – 704.

20   "Specification: Category I Local Area Augmentation System Non-Federal Ground Facility". U.S. Federal Aviation Administration, Washington, D.C., FAA/AND710-2937, May 31, 2001.

21   S. Pullen, et. al. "The Use of CUSUMs to Validate Protectional Level Overbounds for Ground-Based and Space Based Augmentation System," *Proceedings of ION GPS 2000*. Salt Lake City, UT., Sept. 19-22, 2000, pp. 1234-1244.

22   J. Spilker and A.J. Van Dierendonck. "Proposed New L5 Civil GPS Codes," *Journal of the Institute of Navigation*, vol. 48, no. 3, pp. 135-144.

23   Walter, Todd, et. al. "The Effects of Large Ionospheric Gradients on Single Frequency Airborne Smoothing Filters for WAAS and LAAS". *Proceedings of the 2004 National Technical Meeting of the Institute of Navigation,* San Diego, California, January 26 - 28, 2004, pp. 103 – 109.

24   Xie, Gang, et. al. "Detecting Ionospheric Gradients with the Cumulative Sum (CUSUM) Method," *Proceedings of 21st International Communication Satellite Systems Conference and Exhibit*, Yokohama, Japan, AIAA 2003-2415, April 15-19, 2003.

25   Luo, Ming, et. al. "Ionosphere Spatial Gradient Threat for LAAS: Mitigation and Tolerable Threat Space." *Proceedings of the 2004 National Technical Meeting of the Institute of Navigation*, 2004, pp. 490 – 501.

26   Luo, Ming, et. al. "LAAS Ionosphere Spatial Gradient Threat Model and Impact of LGF and Airborne Monitoring." *Proceedings of the 16th International Technical Meeting of the Satellite Division of the Institute of Navigation ION GPS/GNSS 2003*, Portland, OR. pp. 2255 – 2274.

27   Datta-Barua, Seebany, et. al. "Modeling the 20 November 2003 Ionosphere Storm with GRACE". *Proceedings of* ION *GNSS 2007*, Fort Worth, TX. pp. 2840-2848.

28   Park, Young Shin, et. al. "Data-Replay Analysis of LAAS Safety during Ionosphere Storms", *Proceedings of the 20th International Technical Meeting of the Satellite Division of the Institute of Navigation ION GNSS 2007*, Fort Worth, TX. pp. 404 – 414.

29   Ahlberg, Nilson, and Walsh. The Theory of Splines and Their Applications, New York, New York: Academic Press, 1967.

30   Kaplan, Elliot D., Hegarty, Christopher J., Understanding GPS Principles and Applications, 2nd edition. Boston, MA: Artech House Publishing, 2006.

31 Kayton, Myron, and Fried, Walter R., Avionics Navigation Systems 2nd Edition, John Wiley and Sons, New York, 1997.

32 http://en.wikipedia.org/wiki/Cartesian_coordinate_system

33 Parkinson, Bardford W., and James J. Spilker. Global Positioning System: Theory & Applications (Volume One). Washington D.C., American Institute of Aeronautics and Astronautics Inc., 1996.

34 Mori, Shintaro, et. al. "Correlation Structure of Correlated Binomial Models and Implied Default Distributions", J Physics Soc. Japan Volume 77 (2008)

35 http://en.wikipedia.org/wiki/T_distribution (November, 2009)

36 http://en.wikipedia.org/wiki/Binomial_distribution (November, 2009)

37 http://dictionary.reference.com (November, 2009)

38 http://en.wikipedia.org/wiki/Binomial_coefficient (November, 2009)

39 International Civil Aviation Organization. U.S. Satellite Navigation Program Status. *Tenth Meeting of the Asia/Pacific Air Navigation Planning and Implementation Regional Group* (APANPIRG/10) Bangkok, Thailand, 30 August to 3 September 1999.

40 Normark., et. al. "The Next Generation Integrity Monitor Testbed (IMT) for Ground System Development and Validation Testing," *ION GPS 2001 14th International Technical Meeting of the Satellite Division of the Institute of Navigation*, pp. 1200-1208.

41   Pullen, Sam, et. al. "Ephemeris Protection Level Equations and Monitor Algorithms for GBAS," *ION GPS 2001 14th International Technical Meeting of the Satellite Division of the Institute of Navigation*, September 11-14, 2001. pp. 1738-1749.

42   http://en.wikipedia.org/wiki/Joint_Precision_Approach_and_Landing_System. (November, 2009)