

**PERFORMANCE EVALUATION OF A  
SPREAD SPECTRUM LOCAL AREA  
POSITION LOCATION SYSTEM**

by

Martin Jay Feuerstein

Dissertation submitted to the Faculty of the  
Virginia Polytechnic Institute and State University  
in partial fulfillment of the requirements for the degree of

**DOCTOR OF PHILOSOPHY**

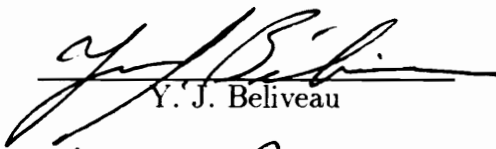
in

Electrical Engineering

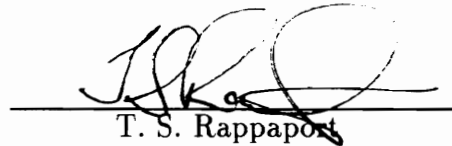
APPROVED:



T. Pratt, Chairman



Y. J. Beliveau



T. S. Rappaport



G. S. Brown



W. L. Stutzman

December, 1990

Blacksburg, Virginia

# **PERFORMANCE EVALUATION OF A SPREAD SPECTRUM LOCAL AREA POSITION LOCATION SYSTEM**

by

Martin Jay Feuerstein

Committee Chairman: Timothy Pratt

Electrical Engineering

(ABSTRACT)

The primary objective of this research was the investigation of position location technologies suitable for use in construction automation applications. Ground based radio frequency spread spectrum techniques were chosen due to their relatively low cost, ease of implementation, and all weather capability. Two types of spectral spreading were examined: frequency hopping and direct sequence. The error performance capabilities of spread spectrum position location systems were studied systematically through theoretical analysis, a series of experimental tests, and by the development of a computer simulation package. A small scale experimental test was conducted as a proof of concept and a test of the prototype hardware. The small scale test was also used to estimate parameters of the computer simulation program. Later, a full scale test was conducted at a site approximately the size of a typical construction project. Computer simulations of the system performance were found to be in good agreement with the experimental results.

The position error performance of a frequency hopping phase measurement type position location system was determined using theoretical, simulation, and experimental means. The primary limitation to position accuracy was found to be propagation channel induced errors. By employing nonlinear filtering methods combined with robust estimation techniques, position error distributions which exceed the performance of classic least squared error methods have been achieved. The relationship between vertical and horizontal errors for practical systems has been determined. Computer simulations have provided estimates of the position accuracy of a direct sequence system which combines time domain techniques for ambiguity resolution determination and phase measurement for fine position. For the first time, phase error distributions of hyperbolic multilateration systems in multipath environments have been studied. The computer simulation tools which have been developed and verified experimentally can be used to analyze the performance of future radio position location systems.

# Acknowledgements

The author would like to convey sincere appreciation to Dr. Timothy Pratt for his patience, guidance, and insight throughout this project. He has contributed immeasurably to my professional and personal growth both as a mentor and a friend. Dr. Yvan Beliveau has been a consistent source of ideas, observations, and encouragement. The members of my graduate committee, by virtue of their differing backgrounds and perspectives, have played an important role in molding this work into its present form.

Dan Purdy, Eric Lundberg, Andrew Dornbusch, and Phil Cataldi deserve thanks for their assistance during the many months of experimental testing. Financial support provided by the National Science Foundation through Grant Number 8717476 is gratefully acknowledged.

This dissertation is dedicated to the memory of Dr. G. Myron Shear, Professor Emeritus, for his commitment to excellence. Cyndi Marshall deserves special thanks for her inspiration and understanding during the seemingly endless hours required to write this dissertation. Finally, without the encouragement and support of my family none of this would have been possible.

# Contents

<b>1</b>	<b>Introduction</b>	<b>1</b>
1.1	Background . . . . .	2
1.2	Technology Options . . . . .	4
1.3	Historical Perspective . . . . .	6
1.4	Research Objectives . . . . .	8
1.5	Organization of Dissertation . . . . .	10
<b>2</b>	<b>Review of Literature</b>	<b>12</b>
2.1	Inertial Methods . . . . .	12
2.2	Acoustical Methods . . . . .	14
2.3	Optical Methods . . . . .	16
2.4	Radio Frequency Methods . . . . .	17
2.4.1	Tracking Radar . . . . .	21
2.4.2	Direction Finding . . . . .	23
2.4.3	Multilateration . . . . .	23
<b>3</b>	<b>Position Location by Radio</b>	<b>43</b>
3.1	Tracking Radar . . . . .	43
3.2	Direction Finding . . . . .	50
3.3	Multilateration . . . . .	56
3.3.1	Ranging Method . . . . .	56
3.3.2	Hyperbolic Method . . . . .	63
3.3.3	Solving Nonlinear Systems of Equations . . . . .	76
3.3.4	Time Delay Measurement . . . . .	82
3.3.5	Phase Delay Measurement . . . . .	84
3.3.6	Frequency Shift Measurement . . . . .	90

<b>4</b>	<b>Phase System Theory and Simulation</b>	<b>92</b>
4.1	System Description . . . . .	93
4.2	Computer Simulation . . . . .	103
4.2.1	Multipath Channel Modeling . . . . .	114
4.3	Simulation Results . . . . .	118
<b>5</b>	<b>Phase Measurement Experimental Tests</b>	<b>125</b>
5.1	Hardware Description . . . . .	126
5.2	Small Scale Site . . . . .	132
5.2.1	Experimental Procedure . . . . .	133
5.2.2	Experimental Results . . . . .	135
5.3	Full Scale Site . . . . .	137
5.3.1	Experimental Procedure . . . . .	139
5.3.2	Experimental Results . . . . .	141
5.3.3	Comparison of Simulation and Experiment . . . . .	143
5.3.4	Vertical Coordinate Accuracy . . . . .	145
5.4	System Improvements . . . . .	150
<b>6</b>	<b>Time System Theory and Simulation</b>	<b>154</b>
6.1	System Description . . . . .	158
6.2	Computer Simulation . . . . .	161
6.3	Simulation Results . . . . .	162
6.4	System Implementation . . . . .	166
<b>7</b>	<b>Conclusions</b>	<b>170</b>
<b>A</b>	<b>Median Filtering Theory</b>	<b>186</b>
<b>B</b>	<b>Maximum Likelihood Estimation</b>	<b>190</b>
<b>C</b>	<b>Solving Nonlinear Systems</b>	<b>195</b>

# List of Figures

- 2.1 Low earth orbit satellite geometry for geometric dilution of precision calculation . . . . . 36
- 2.2 Geometric dilution of precision versus satellite viewing time . 37
- 3.1 Monopulse antenna patterns and error signal . . . . . 45
- 3.2 Block diagram of amplitude-comparison monopulse tracking radar . . . . . 46
- 3.3 Low-angle tracking with surface reflection present . . . . . 48
- 3.4 Single baseline interferometer geometry . . . . . 52
- 3.5 Unambiguous binary baseline interferometer . . . . . 55
- 3.6 Ranging method of multilateration . . . . . 57
- 3.7 Hyperbolas of constant differential distance for a pair of base stations . . . . . 65
- 3.8 Example of position location using a pair of hyperbolas of constant differential distance for two pairs of base stations . . 66
- 3.9 Hyperbolic pattern for two base stations showing expansion function . . . . . 71
- 3.10 Magnification of hyperbolic intersection and error analysis geometry . . . . . 72
- 3.11 Hyperbolic lines of constant phase difference for one base station pair . . . . . 86
- 3.12 Determination of ambiguity digits in two dimensions using multiple frequencies . . . . . 89
- 3.13 Ambiguity resolution using slope of the phase versus frequency curve . . . . . 91
- 4.1 Reflection coefficients as a function of grazing angle for vertical polarization and horizontal polarization . . . . . 101

4.2	Physical layout of measurement area and position location system . . . . .	104
4.3	Block diagram of simulation program main routines . . . . .	107
4.4	Block diagram of simulation program ambiguity resolution routines . . . . .	110
4.5	Phase versus frequency data before and after median filtering .	112
4.6	Simulated and model phase error probability distribution for small test site . . . . .	120
4.7	Simulated position error cumulative probability distribution for 4 receivers, 8 receivers, and majority vote . . . . .	122
5.1	Electrical block diagram of prototype phase multilateration system . . . . .	127
5.2	Simplified electrical diagram of the phase measurement system	128
5.3	Small scale site base station geometry . . . . .	134
5.4	Small scale site position error cumulative probability distributions for dry and wet conditions . . . . .	136
5.5	Small scale site position error cumulative probability distributions for dry conditions and computer simulation . . . . .	138
5.6	Full scale site measurement locations and base station positions	140
5.7	Full scale test position error cumulative distribution . . . . .	142
5.8	Full scale site experimental position error with median filtering and robust estimation compared to least mean squared error analysis . . . . .	144
5.9	Full scale site simulated versus experimental position error distribution . . . . .	146
5.10	Full scale site simulated versus experimental phase error distribution . . . . .	147
5.11	Divergence of phase lanes for two different baseline lengths . .	149
6.1	Example of PSK pulse and corresponding correlator output .	157
6.2	Block diagram of time domain spread spectrum system . . . .	160
6.3	Block diagram of direct sequence spread spectrum simulation .	163
6.4	Output from the time of arrival correlator with no multipaths and 10 multipaths . . . . .	165
6.5	Time delay system position error distribution . . . . .	167



6.6 Time delay and phase measurement system position error distribution (computer simulation for identical multipath and noise conditions) . . . . . 168

# List of Tables

2.1	Comparison of position location techniques . . . . .	13
2.2	Position location and navigation systems with classifications .	22
2.3	Error budget for Transit system with single pass and station- ary user . . . . .	26
2.4	Position errors for several GPS configurations using carrier phase	29
2.5	Satellite outage probabilities for various environments . . . . .	30
4.1	Soil dielectric constant and conductivity versus frequency and moisture content . . . . .	124

If you cannot—in the long run—tell everyone what you have been doing,  
your doing has been worthless.

—ERWIN SCHRÖDINGER

Science . . . means unresting endeavor and continually progressing development  
toward an aim which the poetic intuition may apprehend,  
but which the intellect can never fully grasp.

—MAX PLANCK

Never concerned that the answer may prove disappointing,  
with pleasure and confidence we turn over each new stone  
to find unimagined strangeness leading on  
to more wonderful questions and mysteries—certainly a grand adventure!

—RICHARD P. FEYNMAN

Time marches on,  
and brings with it new and more complicated problems.  
Nature may be fundamentally simple,  
but its simplicity is concealed under a very intricate web.

—RONALD W. P. KING

Instead of concentrating just on finding good answers to questions,  
it's more important to learn how to find good questions!

—DONALD E. KNUTH

The life so short, the craft so long to learn.

—HIPPOCRATES

# Chapter 1

## Introduction

In this dissertation the performance capabilities of a radio frequency spread spectrum position location system are evaluated. The first system implementation studied is a frequency hopping scheme where the measured phase delay is used to calculate estimated position. The second implementation is a direct sequence scheme where a correlator is used to determine time delay and hence estimated position. The error performance of the systems is influenced by many factors but the characteristics of the radio propagation channel often predominate. Closed form analytic evaluation of the error performance of complex systems such as this is difficult and inflexible due to the random nature of the propagation channel. For this reason, a mixture of analytic evaluation, computer simulation, and experimental test methods are employed to determine error performance capabilities.

## 1.1 Background

In the latter part of the twentieth century, the construction industry, fueled by increased pressure from international competition, has been undergoing a shift towards modernization and automation. Every aspect of the construction operation, from the actual fabrication to cost and scheduling estimation, has been affected.

Position location technology plays an important part in the construction process by acting as an interface between the actual fabrication operations and the project computer-aided-design (CAD) database. When particular elements of the project are installed at the construction site, the as-built position, relative to the design position, can be automatically recorded into the project database [1]. A positioning system could also be used to locate specific materials in the project lay-down area [2]. Earth moving projects normally require complex planes for cut and fill operations. With a position location system, a constant real-time update of the vehicle location can be maintained, greatly improving the efficiency of the operation.

The trend toward automation, in construction as well as other industries, may someday reach as far as operatorless robotic equipment [3, 4, 5, 6]. Robotic equipment represents a particularly attractive option for use in hostile environments, such as military zones or toxic waste areas, where a significant risk to human operators exists [7]. It has been predicted that autonomous guided vehicles will account for an important part of the increased productivity of future factories [8, 9]. In order for these robotic systems to be safe and effective, accurate and reliable navigation systems must be

developed.

Industrial safety is another important situation where position location can reduce the risks of an accident. Safety is a particular problem when large tower cranes, lifting equipment, or excavating equipment are in use. Often the equipment operators cannot visually see the position of their hook or bucket relative to other equipment and personnel. The equipment operator usually has poor information about the location of hidden components; such as underground pipes, wiring, and other facilities. Position location systems, integrated with computerized databases of the work areas, could provide the operator with the precise location of his equipment. Safety could be greatly improved by giving warnings to the operator before an accident occurs. Ground controllers at airports could also make use of positioning systems to track aircraft in the vicinity of the terminal, reducing the risk of collision.

In the years since World War II, farmers have seen large increases in productivity combined with a narrowing of profit margins. One of the most expensive aspects of farming is field application of chemicals, including fertilizers, pesticides, and herbicides. Using today's relatively crude field navigation techniques, typical overlap in the chemical spraying of a farm field is about 10%. Accurate position location systems may be able to reduce the required overlap to a fraction of this value, eliminating waste and providing a potential savings to farmers [10, 11]. This same economic benefit may also be possible in other businesses; for instance, open pit mining, dredging, and ship docking.

The digital cellular telephone network, which is currently being imple-

mented, could be upgraded to provide the user with position location information. Studies have been performed which indicate that using position location information as a cell hand-off criteria, rather than received signal strength, minimizes the number of hand-overs required [12]. Position location information could be used to drive rolling map displays for vehicle route guidance or emergencies. Automotive companies have realized for years that there is a huge untapped market for these types of services, if the cost can be reduced sufficiently [13] .

The unprecedented success of both cellular telephones and portable navigation systems shows clearly the huge commercial market which exists for portable communications and navigation services. The United States, Europe, and Japan are now in a race to develop intelligent vehicle highway systems which will provide a host of new capabilities for the general public [13, 14]. Europe and Japan have invested over two billion dollars into this effort. Intelligent vehicle systems, updated with position location and traffic flow information, could be used to reduce the ever increasing problem of traffic congestion in urban areas. An important part of the intelligent vehicle services is the requirement for low cost navigation capabilities.

## 1.2 Technology Options

Several different technologies present themselves as candidates for use in a construction automation type position location system. Each of these methods has specific advantages and disadvantages; no single technique meets the requirements for all system applications.

Acoustic techniques can be used for position location over extremely localized areas. The main drawback of acoustic methods is the dependence of the speed of sound in air upon pressure, humidity, and temperature. These environmental factors may vary considerably over a large construction site causing calibration problems and introducing errors into the system.

Laser systems offer the potential of extremely high accuracy [1, 15, 8]. The chief disadvantage of laser systems is the necessity that line-of-sight be maintained at all times. This is particularly a problem on sites with large altitude variations or where dust and fog can be present. In a cluttered setting, for example on a construction site, the requirement for line-of-site to a number of points can be restrictive.

High pixel count cameras, using charge-coupled device technology, can be employed in conjunction with lasers to do precise imaging. The detailed images returned from the cameras may be analyzed using pattern recognition algorithms to determine the position of objects with excellent accuracy. As with laser systems, the requirement for line of sight is the main limitation.

The Global Positioning System is a satellite-based navigation system which will be fully operational by 1993. Operation of the system, the advantages, and limitations are discussed in detail in several excellent references [16, 17, 18]. Important questions have been raised about the system's adequacy to meet the high accuracy near real-time requirement of the construction application at a reasonable cost. Satellite blockages become a problem in urban areas and in hilly or mountainous regions [19]. The system is operated by the United States Department of Defense for military purposes; therefore, the civilian user is subject to intentional accuracy degradation for



national security reasons [20].

Inertial navigation systems making use of gyroscopes and accelerometers, like those used extensively in the aircraft industry, are costly and complex. Dead reckoning systems based on less expensive sensors, for example, magnetometers and differential odometers, are a more cost effective alternative [21]. Unfortunately, these approaches require frequent calibration and are subject to large errors due to wheel slip and magnetic influences [22], factors which are hard to control on a construction site.

Radio frequency systems, which are discussed in detail in Sections 4 and 6, offer the advantages of low cost and potentially high accuracy. Radio systems may operate, to a limited extent, in cases where line of sight is not present. This is an advantage in the potentially foggy and dusty conditions present on a typical construction site. The chief disadvantage of radio systems is their susceptibility to multipath interference and the constraints imposed due to limited spectrum availability.

### **1.3 Historical Perspective**

Electronic navigation systems have been in existence, albeit in a crude form, since the 1930's. During World War I a method was invented for locating enemy guns by measuring the differences between the times at which their firings were heard at distant listening posts [23]. While a system using sound waves such as this was never used for navigation per se, it is similar in principle to the radio navigation systems developed later.

During World War II progress in electronic navigation developed at a

feverish pace. Of the systems developed during the war effort, Loran is probably the most famous. A modern version of this system, Loran-C, is still in use today for maritime and aircraft navigation [16, 17]. Loran-C is capable of absolute accuracies from 185 to 463 meters with a repeatability from 15 to 90 meters [17, page 114].

During the late 1960's and early 1970's, a number of systems were developed for metropolitan vehicle location using phase or time-delay measurement techniques. An excellent synopsis of these developments is presented in the tutorial articles by Roth [24] and Riter and McCoy [25]. Most of these systems made use of a radio frequency carrier modulated by a single audio tone. At a receiver, the audio phase was used as a measure of the propagation delay; accuracies of several hundred meters were achieved [26].

The advent of satellite navigation began with the Transit system, which was introduced for military use in 1964 and commercial use in 1977. A Transit receiver makes use of Doppler information from the satellite to estimate position with typical accuracies of several hundred meters in two dimensions for stationary users. The successor to Transit is the Global Positioning System, a constellation of 18 satellites, which uses time delay to estimate range. Accuracies for typical receivers are about 20 meters, although more sophisticated techniques may be used to improve the accuracy. The Global Positioning System is scheduled to be fully operational in 1993.

As the modern needs for more precise position location increased, manufacturers and researchers responded with a wide variety of alternatives. A comprehensive survey of the commercially available ground-based positioning systems, as of 1985, is presented by Sampson [27]. Each system offers

a different combination of range, accuracy, dynamics, operating modes, and inherent limitations. Horizontal accuracies for multiuser systems are typically in the range of  $\pm 1$  to  $\pm 5$  meters for both phase measurement and pulse timing techniques.

## 1.4 Research Objectives

The wide variety of possible applications for position location systems makes it apparent that no single system will meet the requirements of all users. A number of different factors enter into the decision as to which type of system is best for a particular application. The operating environment, accuracy, cost, ruggedness, size, and update rate are but a few of the variables which must be considered.

The objectives of this research are the investigation of position location technologies, which could be interfaced to computer project databases, to meet the needs of the construction automation effort. Therefore, it is pertinent to discuss briefly the position location requirements of the construction industry, as they relate to the objectives of this research.

A typical construction site is a local area of less than a square kilometers in size. For many stages of the construction operation the site will be in an extremely rugged state. Large altitude variations can be expected across the site and line-of-sight cannot be guaranteed between points on the site. Many large obstructions may exist on the site; including storage areas, buildings, cranes, and piles of earth. Construction operations continue in environmental conditions which may be far from optimum; such as fog, rain,

dust, temperature and humidity extremes.

Many different vehicles can be in operation on the site at one time along with personnel on foot; all of which may require position location information simultaneously. Equipment moving on the site may need position updates extremely rapidly; ten updates per second is a common requirement. A position location system for a construction site must be able to perform acceptably under all of these adverse conditions.

The different construction operations place a variety of accuracy requirements on a position location system [1]. Lay-down yard control and equipment optimization are two applications which require limited accuracy of  $\pm 3$  meters in the horizontal plane. Earth moving operations such as cross-sectional analysis of the site and quality control require a medium accuracy of  $\pm 0.25$  meters in horizontal as well as vertical dimensions. Operations such as layout and robotic control require high accuracy position location systems capable of  $\pm 5$  millimeter accuracy in three dimensions.

The position location system is to interface to a 3-D CAD database of the construction project. Any information which is presented to the equipment operators and other site personnel will first be processed through the project database. In general, the equipment operator needs to have a direction indication and does not necessarily require his or her exact location. The information display at the user's location can be a simple terminal which gives indications such as up, down, left, and right. This arrangement favors a position location system where the majority of the processing and analysis is done at a central location rather than at the user's position. In this manner, the equipment which is located within the construction area can be kept

simple, inexpensive, and rugged. The more complex analysis equipment is located at a central point probably on the periphery of the site.

## **1.5 Organization of Dissertation**

A survey of the literature on position location and navigation is presented in Chapter 2. The various technology options which may be utilized are compared and contrasted. Radio frequency position location techniques are covered from their infancy through futuristic systems.

The concepts of radio frequency position location are presented in Chapter 3. The fundamental differences between systems which make absolute measurements and those which make relative measurements are discussed. The differences between measuring the phase delay or time delay of a signal are examined. Methods of solving for the estimated position and geometric limitations on accuracy are presented.

In Chapter 4 the concepts behind the frequency hopping phase measurement system are presented. The theory is reviewed and the details of a computer simulation program discussed. The results of computer simulation trials are presented. Experimental results for the frequency hopping system are discussed in Chapter 5 and a comparison is made between simulation and experiment. Possible improvements to the system are evaluated.

The details of a direct sequence system which measures time delay using an analog correlator are presented in Chapter 6. The structure of a computer simulation is presented along with simulated error performance. A comparison is made between the frequency hopping/phase delay system and

the direct sequence/time delay system. The potential accuracy of a hybrid system which combines time delay and phase information is examined.

Chapter 7 summarizes the results of the performance analysis of spread spectrum position location systems. Some unanswered questions which could serve as the basis for future research are discussed.

# Chapter 2

## Review of Literature

Position location techniques may be broken down into the broad categories of inertial, acoustical, optical, and radio methods. The relevance of each of these techniques toward solving the problem at hand will be discussed separately. Table 2.1 gives a brief comparison of the various methods with their advantages and limitations.

### 2.1 Inertial Methods

Inertial navigation systems making use of gyroscopes and accelerometers, like those used extensively in the aircraft industry, are costly and complex. Dead reckoning systems based on less expensive sensors, for example, magnetometers and differential odometers, are a more cost effective alternative [21]. Gyrocompass-odometer systems typically have errors of about 2% of the cumulative distance traveled on paved roads while the more expensive inertial

Table 2.1: Comparison of position location techniques

<i>TECHNIQUE</i>	<i>ACCURACY</i>	<i>LIMITATIONS</i>
Inertial	Fair	Frequent calibration
Acoustical	Fair	Short range only
Optical	Excellent	Not all weather
Radio	Good	Subject to multipath



land navigators achieve errors of 0.25% of distance traveled at a cost in 1989 of \$300,000 [28].

Inertial navigation systems must be calibrated initially and thereafter attitude and velocity errors increase in proportion to the square root of time [29]. Periodic zero velocity updates, typically for periods of 1 to 5 minutes for every 5 to 20 minutes of operation, are required to maintain acceptable accuracy [28]. Extreme difficulties are encountered when using magnetic sensors on steel vehicles [30] because the vehicle's own magnetic field changes with time, resulting in heading accuracies rarely better than  $2^\circ$  and usually much worse [28]. Differential odometer readings can be used instead of magnetometers to measure heading but wheel slip can be a large source of heading errors [22], particularly in conditions of snow, ice, water, or mud. These heading errors cause position errors in the dead reckoning system which steadily increase with time. In cities, map matching techniques can be used to improve the accuracy of dead reckoning systems, but this approach is not feasible in construction applications where no maps or predetermined routes are generally available.

## 2.2 Acoustical Methods

Acoustic techniques have been used successfully for position location over extremely localized areas. The main accuracy limitation of acoustic methods is the dependence of the speed of sound in air upon pressure, humidity, and temperature. Roberts and Miller [31] give the following relationship for the

velocity of sound,  $v$ , in a real gas

$$v = \sqrt{\gamma \frac{RT}{M} \left[ 1 + \frac{9}{64} \frac{p}{p_c} \frac{T_c}{T} \left( 1 - 6 \frac{T_c^2}{T^2} \right) \right]} \quad (2.1)$$

where  $T_c$  is the critical temperature of the gas,  $p_c$  is the critical pressure,  $p$  is the pressure,  $T$  is the absolute temperature,  $R$  is the gas constant,  $\gamma$  is the ratio of the specific heats, and  $M$  is the molecular weight. This equation reveals that the velocity is a strong function of the temperature and pressure of the air. These environmental factors may vary considerably over a large site causing calibration problems and introducing errors.

Generally, ultrasonic frequencies must be used to avoid disturbance to humans and interference from extraneous noise sources [32]. The use of ultrasonics severely limits the maximum range of acoustic systems because of attenuation in the air due to viscosity, heat conduction, and absorption [33]. The absorption mechanisms are due to the excitation of modes of vibration in certain molecules, most notably water vapor [34]. Sivian gives a plot showing the overall attenuation versus the frequency for typical temperature, pressure, and humidity conditions based on lab measurements [35]. The curve indicates that for frequencies above 50 kHz attenuations of greater than 2.5 dB/m are observed, limiting the scale of most practical systems to about a hundred meters.

Acoustic waves suffer from multipath reflection problems but the slow speed of propagation, relative to the speed of light, allows the time difference between the direct signal and any reflections to be determined. This can be done by using a swept frequency transmitter and a narrowband tracking filter on the receiver to separate the direct path from the delayed multipaths.

## 2.3 Optical Methods

Laser systems offer the potential of extremely high accuracy [1, 15, 8]. The chief disadvantage of laser systems is the necessity that line-of-sight be maintained at all times, which is particularly a problem on sites with large altitude variations or when atmospheric visibility is poor. The atmospheric medium may exhibit a number of effects on the laser propagation: absorption and scattering due to aerosols and hydrometeors (rain, snow, dust, smoke); amplitude and phase fluctuations due to turbulence (scintillation); and extraneous light sources may be present [36].

The limitations on atmospheric optical systems imposed by low visibility weather conditions like haze, rain, fog, and snow has been well documented [37, 38]. At most visible and many infrared wavelengths the extinction induced by these weather conditions is due primarily to scattering rather than absorption [39, 40]. Shapiro has performed experiments indicating that laser attenuation through dust and smoke clouds is typically on the order of 5 to 40 dB [36] and may last for periods of a minute.

Several optical position location systems have been developed using either visible or infrared tracking techniques [41, 42, 43]. These systems require sophisticated servo-mechanical tracking of the target vehicle by a telescope or motorized theodolite. For example, in one system a single target may be tracked at any given time with quoted accuracies of 5 cm at a distance of 914 m [41].

High pixel count cameras, using charge-coupled device technology, can be employed in conjunction with lasers to do precise imaging. The detailed

images returned from the cameras may be analyzed using pattern recognition algorithms to determine the position of objects with excellent accuracy [44]. Bayer is attempting to produce a system with an accuracy of 1 cm/km for tracking slow moving machines on a construction site [45], actual accuracies are not currently available.

Modern survey techniques make use of laser Electronic Distance Measuring (EDM) instruments to measure the distance between a control point and a target [46, 47]. Most of these devices transmit an infrared laser beam modulated at several frequencies typically less than 15 MHz. The transmitted signal hits the retroreflector target and is received back at the EDM. Phase comparison between the transmitted and received modulation signal yields the distance between the EDM and target. Several different modulation frequencies are used to resolve the ambiguities resulting from the unknown integer number of wavelengths. If atmospheric temperature, pressure, and humidity are measured and appropriate compensations made then accuracies of one part in 25,000 are achievable [46]. The EDM's laser must be pointed directly at the retroreflector target and only a single target position may be calculated at a time.

## **2.4 Radio Frequency Methods**

A number of different methods exist for estimating position using radio frequency techniques. Radar tracking, direction finding, and multilateration systems all are capable of providing estimates of position but with vastly differing complexities and accuracies.

Position location systems can be broken down into several broad classifications with the mathematical details of each of these classes presented in detail in Chapter 3. Before discussing the literature on radio position location it is necessary to at least briefly introduce the system classifications and some terminology.

Multilateration systems are those which make use of the range or difference in range between the user's position and the location of a set of base stations at known locations. In a common implementation of multilateration systems, ranges are measured between the user and three base stations. From the three ranges the user's unknown position can be estimated using trilateration.

One class of multilateration systems, termed *ranging type systems*, are those which measure the absolute distance between the user's unknown location and a number of base stations. In a two dimensional system, each range measurement between a base station and the user defines the unknown position to be on a circle of constant radius centered at the base station. For this reason *ranging type systems* are also called *circular type systems* in the literature. To determine the user's location uniquely in two dimensions the distances between the user and three base stations must be measured, with the intersection of the three circles defining the user position. In three dimensions the range between the user and a base station locates the user on a sphere centered at the base station and consequently four ranges are required to uniquely specify the user's location. Most practical *ranging type systems* are not able to measure the range between the user and a base station directly but instead measure the range plus a bias term. For this reason, systems of

this type are often called *pseudo-ranging type systems*. The bias term can be calculated using a range measurement to an additional base station, resulting in a system of 4 equations and 4 unknowns. In this dissertation all systems of this class will be referred to as *ranging systems*.

Another class of system, called *differential range type systems*, are those which measure the difference in distance between the user's location and a pair of base stations. The difference in distance between the user and a pair of base stations define the user's unknown location to be on a hyperbola of constant range difference, with the foci at the base stations. For this reason *differential range type systems* are more often called *hyperbolic type systems*. To determine the user's location uniquely in two dimensions two differential range measurements must be used, with the user's location defined by the intersection between the two hyperbolas. In three dimensions the user's location is known to be on a hyperbolic surface of revolution and three differential range measurements are necessary to uniquely define the user's location. In this dissertation all systems of this class will be referred to as *hyperbolic systems*.

Most position location systems make use of some method of measuring range, whether in an absolute or differential sense. The range must be measured from characteristics of a radio signal transmitted between the user's location and the base stations at known positions. In the literature, systems have been proposed which measure either the time, phase, or frequency of the arriving signal or a combination of these parameters. *Time systems* generally transmit a pulse and measure its arrival time to determine path delay and hence range. For good time resolution, short pulses and the associated

confusing, because  
differential GPS,  
locus etc. do not  
work this way

large bandwidths are required. This is where spread spectrum approaches can be used to allow better temporal resolution for a given pulse width.

*Phase systems* measure the phase of a signal transmitted between the user's location and the base stations at known locations. The phase of the received signal is a function of the distance traveled by the wave and therefore range information can be determined. The phase information is measured modulo  $2\pi$  due to its cyclic nature, resulting in ambiguities in determining the integer number of cycles. These ambiguities can be resolved by measuring at low frequencies where wavelengths are large enough to ensure less than one phase cycle over the measurement area. Generally, a combination of low frequencies for determination of ambiguities and high frequencies for good resolution are utilized.

Systems which measure the change in frequency of a signal transmitted between the base station and the user are called *Doppler systems*. If the base stations are in motion, as would be the case if they were non-geosynchronous satellites, but their locations are continuously known then a frequency shift can occur as the signal travels. This frequency shift, due to the Doppler effect, is a function of the velocity with which the base station is approaching to or withdrawing from the user. By measuring the change in frequency, the rate of change of range between the user and the base station can be determined. If the trajectory of the base station is known then the user's location can be uniquely defined from the received Doppler frequency changes. The fact that the rate of change of range is measured, and not the range itself, means that *Doppler systems* are inherently *hyperbolic type systems*. check with Transit !

Table 2.2 gives the names and classifications of several commercial posi-

tion location and navigation systems. The system name is given along with the type of distance measure implemented (*ranging* or *hyperbolic*), the parameter measured (*time*, *phase*, or *Doppler*), the base station location (ground or satellite), and time period when the system was first introduced. As one can see from the table, many different systems have been proposed and implemented. These systems will be discussed in more specific terms in the rest of this Chapter; the mathematical foundations which relate them together are presented in Chapter 3. It may be helpful to refer back to the table as each system is discussed.

### 2.4.1 Tracking Radar

Radar tracking systems use a servo control loop and a directional antenna to mechanically follow a target; for the position location application the target could be a radar beacon at the user's location. The elevation and azimuth angles to the target are determined from the pointing direction of the antenna. The range to the target is determined by the time it takes the pulse to reach the target and be returned to the radar receiver. Angle accuracies of less than a degree and range accuracies of better than a meter are achievable with monopulse systems [48]. Tracking radar systems which use only a single antenna element are only able to track a single target at a time. Radars which use multiple antennas or phased arrays to track several targets are complex. At low angles reflections from the ground can cause large errors in elevation angle. The peak errors can be many times larger than the angular separation between the target and its mirror image.



Table 2.2: Position location and navigation systems with classifications

*not complete*

<i>SYSTEM NAME</i>	<i>DISTANCE MEASURE</i>	<i>PARAMETER MEASURED</i>	<i>BASE STATION</i>	<i>YEAR INTRODUCED</i>
Loran	Hyperbolic	Time	Ground	early-40's
Decca	Hyperbolic	Phase	Ground	early-40's
Omega	Hyperbolic	Phase	Ground	mid-50's
Transit	Hyperbolic	Doppler	Satellite	early-60's
Hazeltine	Hyperbolic	Time	Ground	early-70's
GE	Hyperbolic	Phase	Ground	early-70's
JTIDS	Range	Time	Ground	mid-70's
PLRS	Range	Time	Ground	mid-70's
Sarsat	Hyperbolic	Doppler	Satellite	early-80's
Accutrak	Range	Phase	Ground	mid-80's
GPS	Range	Time	Satellite	late-80's
GLONASS	Range	Time	Satellite	late-80's

### **2.4.2 Direction Finding**

Interferometer direction finding systems make use of an array of spaced antennas to measure the angle of arrival of an incoming wavefront. The phase differences between the signals from the antennas in the array are compared to determine the angle of arrival. The resolution of phase interferometers is better for systems with larger baseline distances between antenna elements but at the expense of ambiguities. For this reason, multiple baseline interferometers are often used with both short baselines for unambiguous readings and long baselines for high resolution. Angular accuracies of several degrees are possible with multiple baseline systems [49] but this is generally not good enough for precise position location applications. An angular resolution of one degree results in a position error of over 8 meters at a distance of 500 meters, therefore resolution to much better than a degree would be necessary.

### **2.4.3 Multilateration**

Multilateration systems can be divided into two broad classes: those with base stations located on or near the ground and those with base stations placed on satellites in orbit around the earth. Ground based multilateration systems ground can be further subdivided into three categories based roughly on their utility and area of coverage. The categories are: navigation systems intended for wide area use; vehicle monitoring systems primarily for metropolitan areas; and local area systems for much more confined areas.

## Satellite Based Methods

In the Free World there currently exist two satellite-based navigation systems that are in wide use [16, 17]. The first of these is the Navy Navigation Satellite System (Transit) which was designed for one-half nautical mile accuracy maritime navigation. The system uses Doppler frequency shift information from a complete pass of a single satellite to estimate the user's position. The system was made operational in 1964 and will be retired from service in 1996.

The idea behind the Transit system came from two Johns Hopkins Applied Physics Lab engineers who observed the Doppler effect from the Russian Sputnik I spacecraft at 20 MHz in 1957 [50]. Guier and Weiffenbach used the Doppler frequency versus time information from the spacecraft to determine its orbital parameters. The navigation system was actually invented when F. T. McClure, Chairman of the Applied Physics Lab, realized that the problem could be solved in the reverse order [51]; the satellite ephemeris and Doppler curve could be used to find the location of a receiver.

The Transit satellites, 5 in all, orbit at an altitude of 1,100 km and contain precise (by 1960's standards) frequency references. The satellites are in polar orbit and the time between passes is typically  $1\frac{1}{2}$  hours. The Transit system makes use of Doppler measurements to estimate the rate of change of range and is thus a hyperbolic type system. The observed Doppler shift at the user's location, along with satellite position and velocity information, are used in an iterative nonlinear least squares curve fitting procedure to solve for position. A dual frequency approach using 150 and 400 MHz is used to remove ionospheric effects from the position calculation.

Table 2.3 gives a rather optimistic 2-D RMS error budget for the Transit system [52]. Several important assumptions have been made in the table: at least one full satellite pass is used for computation; good user-satellite geometry exists; the user's receiver is stationary. The Transit system is capable of quite respectable error performance for surveying applications, with about 3 meters being absolute best case. The drawback here, of course, is the 15 minute measurement time and the hours between updates. If less than a full satellite pass is used for the position estimate then the error degrades dramatically. The error is also increased if the user is in motion during the satellite pass. A complete single pass, with the user moving at a rate of 5 km/hr, results in an accuracy of about 30 meters. The calculation of positions is quite computer intensive, requiring between 30 and 60 seconds on a small personal computer. *too optimistic*

The Global Positioning System (GPS) consists of a constellation of 18 satellites (plus 3 spares) at an altitude of 22,000 km. GPS uses time delay measurements to estimate the ranges, plus a user clock bias term, between the user and each of the satellites. The system can therefore be classified as a pseudo-range type system making use of time measurements. *incorrect*

The satellites transmit a direct sequence spread spectrum signal and a correlation receiver is used to determine the time-of-arrival at the user location. The satellites contain accurate timing standards which are updated regularly from the ground. The time-of-arrival of the signals from several satellites are then converted to range information and triangulation is used to estimate the user's position. The system is now able to provide continuous 2-D coverage but by 1993 GPS is scheduled to be able to provide continuous

Table 2.3: Error budget for Transit system with single pass and stationary user

<i>ERROR SOURCE</i>	<i>POSITION ERROR (METERS)</i>	
	Theoretical Limit	Best Case Practice
Satellite Position	0.11	2.00
Noise	0.05	0.50
Oscillator Drift	0.01	0.10
Timing	0.07	0.50
Ionosphere	0.02	2.00
Dry Troposphere	0.04	0.10
Water Vapor	0.12	0.25
RESULTANT	0.19	2.93

3-D world-wide coverage.

The satellites transmit two different phase shift keyed Pseudo Random Noise sequences. The precise code (P-Code) is a one week long sequence sent at a rate of 10.23 Mbit/sec. The course/acquisition code (C/A Code) is sent at one-tenth the rate of the precise code and lasts for 1 millisecond. The C/A and P codes are transmitted simultaneously and continuously. Civilian users have access only to the less accurate C/A codes.

GPS has the capability to provide the civilian user with near real-time accuracies of about 40 meters. Currently the available accuracy has been degraded, due to national security interests, to no better than 100 meters. User errors as high as 200 meters are being observed since this intentional reduction in accuracy, termed Selective Availability, was implemented by the Department of Defense in March 1990 [20]. Some users have concluded that GPS accuracy is now “considerably worse than the accuracy enjoyed by Loran-C users in many coastal U.S. locations” [53]. With Selective Availability implemented the Loran-C system is indeed able to provide more repeatable position measurements than GPS in many cases.

A relatively complex receiver capable of tracking 4 satellites is required for three-dimensional navigation with GPS. To obtain near real-time accuracies of 20 meters, differential techniques employing a reference receiver at a known location are required [54]. Better accuracies can be obtained by making use of the carrier phase information but ambiguity resolution can be a serious problem [55]. GPS information can be combined with an inertial navigation system to improve the accuracy during satellite blockages or poor geometry conditions [56]. Table 2.4 lists the 2-D position errors and costs, obtained

from the literature, for some of these GPS configurations.

Several studies have been done which indicate that multipath interference and satellite blockages are potential sources of problems for the system [19]. The results of research conducted in Cedar Rapids, Iowa, on satellite outage rates are listed in Table 2.5. From this information it is clear that serious degradations of performance will occur in hilly or dense metropolitan areas. Other researchers have predicted that in cities two or more GPS satellites will be visible 30% of the time in streets and 60% of the time at intersections [57].

The Soviet Union has a satellite navigation system called GLONASS, which operates in a similar manner to GPS. The main differences between GPS and GLONASS are in the areas of multiple access, code lengths, operating frequencies, and data formats [28]. GPS makes use of a Code Division Multiple Access (CDMA) scheme while GLONASS uses Frequency Division Multiple Access (FDMA) to separate the signals from each satellite. The code length for the GPS C/A Code is 1023 bits while the equivalent GLONASS code is only 511 bits long. The ratio of the two transmitted frequencies is 0.2% lower in GLONASS than in GPS. Dale has performed tests with the GLONASS system and has found  $3\sigma$  errors of 30 m for the Soviet C/A Code [58].

Other Low Earth Orbit Satellite (LEOS) systems have been successfully used to locate users on the ground. The Sarsat system is operated by a number of countries as an emergency location system mainly for aircraft and ships. The vehicles carry beacons, either activated manually or automatically upon impact, which transmit at 121.5, 243, or 406 MHz. The signals are received by low earth orbit satellites which relay the time history of the

Table 2.4: Position errors for several GPS configurations using carrier phase

<i>MODE</i>	<i>C/A CODE</i>	<i>P CODE</i>	<i>COST</i>
Normal	8 m	6 m	\$5k
Differential	5 m	3 m	\$15k
Diff-Inertial	???	2 m	\$30k



Table 2.5: Satellite outage probabilities for various environments

<i>TERRAIN</i>	<i>OUTAGE RATE</i>
Flat	1-2%
Hilly	15-18%
Residential	12-15%
Metropolitan	70-75%

measured Doppler shift to earth stations. The location of the beacon is calculated, using techniques similar to the Transit system, with a typical accuracy of 20 km. A new beacon design is predicted to improve the accuracy to about 5 km in the near future [28].

The question has been raised as to whether a network of Low Earth Orbit Satellites (LEOS) could be used for precise position location of moving vehicles. This is a complex question and no one has yet attempted to implement such a system. It is worth considering the possibilities of providing precise real time positioning capabilities.

A realistic network of LEOS's could probably only guarantee that one or two satellites at a time would be visible to the user. It is quite likely that with a LEOS system there would be times during which only a single satellite would be visible. This being the case, a system must be examined which can utilize only a single satellite to determine the user position.

For a LEOS navigation system to be useful the satellite orbits must be determined accurately. The user's position error will be larger than the uncertainty in the satellite position. This is particularly important for low satellite orbits because of the effects of orbit irregularities caused by the unequal distribution of mass on the Earth.

One simple method of determining the satellite position is the use of a GPS receiver on the LEOS. This method is simple but has the disadvantage of being subject to intentional accuracy degradation. One technique of removing the accuracy degradation is to track the satellite using the on board GPS receiver along with a network of ground-based reference receivers [59]. Researchers at the Jet Propulsion Lab have shown that sub-decimeter accu-

not true anymore,  
think of Iridium

not if you use the  
satellites just as  
transponders

does not make  
sense; why use  
GPS to locate  
satellites which  
then locate the  
object?

racies, down to even the lowest orbits, can be obtained with a network of six ground receivers [60, 61]. An open question still exists as to the resulting accuracy of such a differential network in the presence of Selective Availability but it is probable that accuracies of better than 10 meters could be obtained.

Another method to determine the satellite orbit is to make use of the received Doppler shifted signal from the satellite to calculate the orbit, as has been used with the Transit system [50]. Again a network of ground stations is required and the resulting accuracy being on the order of 5 meters.

In order for the user to be provided with an inexpensive navigation receiver the hardware must be simple and readily available. This brings up the issue of the timing accuracies which would be required both on the satellite and in the ground-based receiver. A GPS receiver tracks multiple satellites and is therefore able to remove clock bias from the position estimate. A system which evaluates information from a single satellite would not have such a luxury. Timing information may have to be broadcast from the satellite to the user's receiver to compensate for drift. The satellite timing will probably have to be referenced to a ground-based standard such as NBS broadcasts or GPS timing standards, the aim being to make the requirements on the ground-based receiver as flexible as possible. Assuming a satellite velocity of approximately 7000 m/s then timing errors of  $1 \mu\text{s}$  would result in less than 1 cm degradation in accuracy. This is quite reasonable and relatively inexpensive with technology available currently.

A LEOS navigation system which estimates position based on measurements from a single satellite is restrained in some ways compared to a system which tracks multiple satellites. Probably the most important question about

such a system is the time required to obtain a position fix of given accuracy from the satellite. This relates back to the question of how much of each satellite pass must be observed for a desired position accuracy. The Transit system makes use of an entire pass of the satellite.

The time required to obtain a fix will depend on the accuracy desired and the portion of the satellite pass received. Preliminary studies, using a computer simulation program, indicate that at least five minutes of data would be required for an accuracy of 50 meters with a satellite at 1,000 km altitude and a stationary user. This is because a nonlinear least square algorithm is used to determine the user's position from the satellite ephemeris and the Doppler curve. The accuracy of the iterative solution degrades rapidly as the number of input samples is reduced. The accuracy is a strong function of the portion of the Doppler curve which is measured. This represents a severe limitation to the objective of obtaining near real-time position updates from a LEOS system.

The previous discussion of the position accuracy obtained for a specific satellite observation time can be put into more quantitative terms by making use of the Geometric Dilution of Precision (GDOP) concept. The GDOP is a measure of the sensitivity of the position error to the satellite-user geometry during the measurement time. The GDOP specifies to what extent the measurement errors are magnified in the position estimate (for example,  $\text{GDOP} \approx 1$  indicates that the measured ranging errors are approximately equal to the position errors). A general technique for assessing the GDOP for any user-satellite geometry, developed by Lee [62], is presented in Chapter 3. By making some simplifying assumptions, this technique can be used to estimate

two completely  
different  
issues!

how sensitive the position estimate is to the satellite viewing time.

Assume the geometry shown in Figure 2.1 where the satellite passes directly overhead the user and only a portion of the satellite pass is observed by the user at point  $U$ . The portion of the pass observed by the user is defined by the angle  $\phi$  in the figure. If  $\phi = 0^\circ$  then none of the pass is observed and if  $\phi = 90^\circ$  the full pass is observed. The portion of the pass used for the position estimate is assumed to be symmetric about the point of closest approach, labeled point  $P$ . Based on these assumptions, a lower bound on the GDOP can be calculated as [62]

$$\text{GDOP} \geq \frac{1}{\sqrt{N} 4 \cos(\phi/2) \sin^2(\phi/4)} \quad (2.2)$$

where  $N$  is the number of measurements made during the viewing time and  $\phi$  is the angular portion of the pass used for the estimate.

The lower bound on GDOP can be plotted as a function of the satellite observation time by making some further assumptions about the satellite orbit. Assuming a satellite altitude of 1,000 km and a velocity of 7,000 m/s then the time for the full pass is approximately 15 minutes. The lower bound on GDOP is plotted versus the observation time in Figure 2.2, where 5 measurements were assumed during the observation time ( $N = 5$ ). Clearly, as the observation time decreases the GDOP, and hence the sensitivity to measurement errors, increases rapidly.

The GDOP is so large for satellite observation times of less than 5 minutes as to preclude accurate position estimates. The sensitivity predicted by the high GDOP is dominant in the direction perpendicular to the plane containing the satellite trajectory and the user. It is important to keep in mind

that this is only a lower bound based on a satellite pass directly overhead the user and assuming satellite observations symmetric about the point of closest approach. For a more realistic scenario, where these conditions may not be met, the GDOP could actually be much greater than predicted here.

In summary, the key characteristics of low earth orbit satellite navigation systems can now be mentioned. Accuracies of 5 meters for the LEOS ephemeris can be obtained by employing a network of ground stations. Drift in the timing at the user's receiver can be compensated for by transmitting precise timing information to the user, with receiver timing accuracies on the order of  $1\ \mu\text{s}$  being sufficient. Two frequencies would have to be transmitted to correct for ionospheric errors. A system such as this would have best case accuracies of approximately 10 meters for a complete single satellite pass and stationary user (using the same error analysis approach as Table 2.3). Computation time for a position estimate would be about 30 seconds on a personal computer, just as with Transit).

The length of time required to obtain an initial position fix from a single satellite will be on the order of five minutes. After initial acquisition position updates could be obtained at intervals of several minutes. It does not appear that near real-time updates can be obtained from a LEOS system which tracks only a single satellite. Errors would degrade during periods of poor satellite-user geometry.

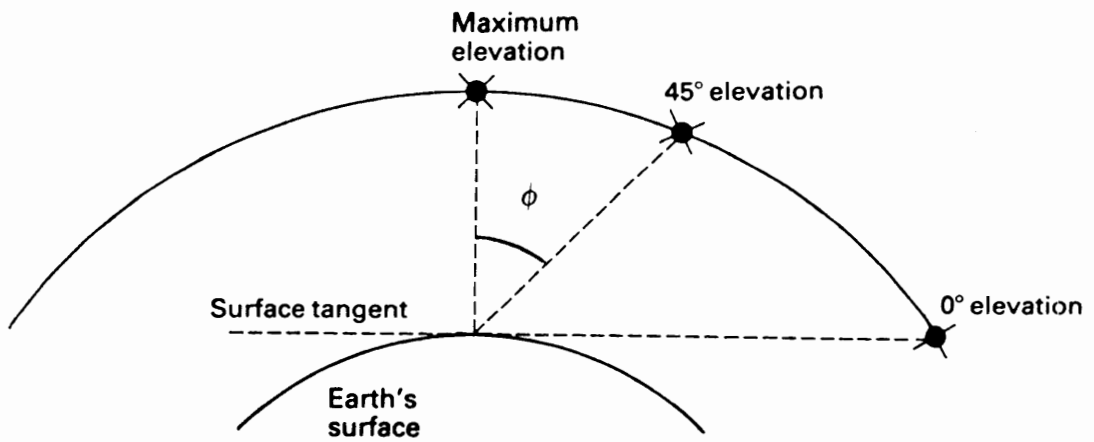


Figure 2.1: Low earth orbit satellite geometry for geometric dilution of precision calculation (from Tetley [17])

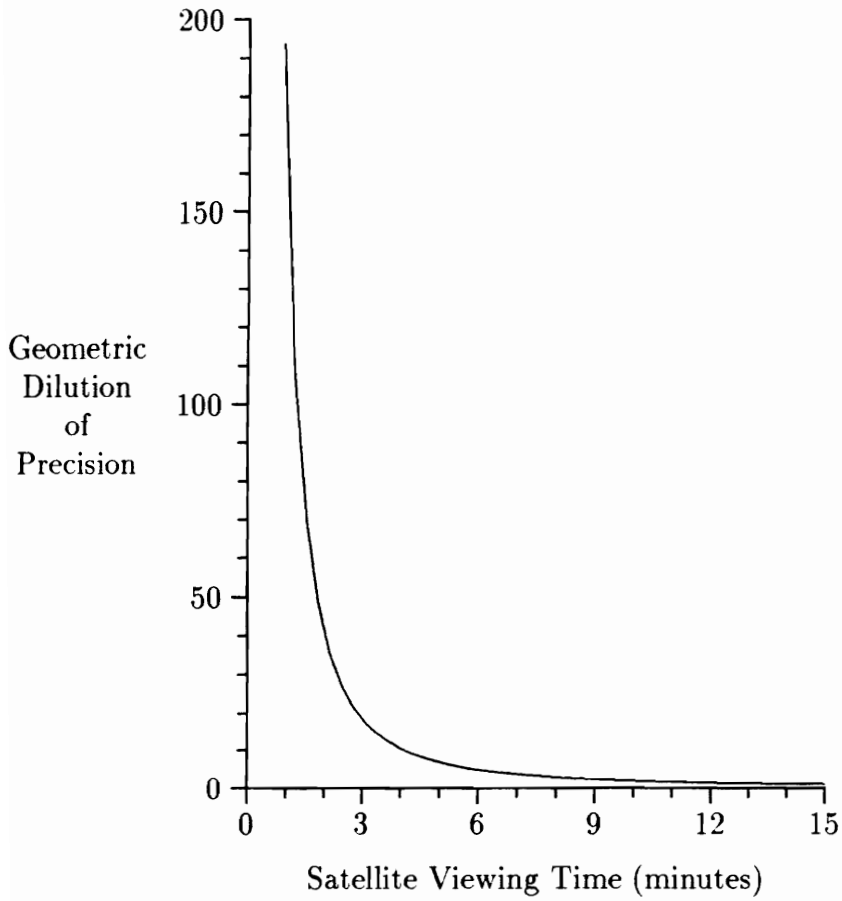


Figure 2.2: Geometric dilution of precision versus satellite viewing time



## Ground Navigation Systems

During World War II progress in electronic navigation developed at a feverish pace. Of the systems developed during the war effort, Loran is probably the most famous [23]. A modern version of this system, Loran-C, is still in use today for maritime and aircraft navigation [16, 17]. Loran-C is capable of absolute accuracies from 185 to 463 meters with a repeatability from 15 to 90 meters [17, page 114]. Loran-C is used for maritime navigation, automatic vehicle location in cities [63], and in an increasing number of small aircraft [28].

Loran makes use of differential measurements in time and is thus classified as a hyperbolic type system. Each Loran transmitter sends out a train of pulses with a width of  $250\ \mu\text{s}$  with a carrier frequency of 100 kHz. The pulses from three stations are observed at the user's receiver. Cycle matching is used to accurately determine the time delay between the arriving pulses. The difference in arrival times between the two pairs of transmitters locates the receiver on a pair of hyperbolas, the intersection of which gives the position estimate.

The Decca navigator is another of the early hyperbolic type systems but signal phase, rather than arrival time, is used as a measure of the difference in distance. In the Decca system one master transmitter in combination with three frequency-locked slave transmitters to form what is known as a *chain*. The stations transmit subharmonics of a specific reference frequency (340 kHz) and at the receiver frequency multiplication is used to make phase measurement at a comparison frequency. At regular intervals the chains

transmit on frequencies which result in much lower comparison frequencies and thus a longer wavelength. The longer wavelength results in a much larger unambiguous region and allows the correct lane to be determined. The system operated in the 100 kHz frequency range and typical accuracies are 100 meters RMS.

The Omega system was first conceived in 1947 and implemented in the mid-50's as a truly global navigation system. Omega uses differential phase measurements and is therefore a hyperbolic type system. Very low frequencies from 10 to 14 kHz are used to provide nearly global coverage with only eight transmitters. The primary frequency for the Omega system is at 10.2 kHz; two additional frequencies of 13.6 and 11.33 kHz are also transmitted for ambiguity resolution. By subtracting the phases measured for the primary and the ambiguity resolution frequencies lane widths of 24 and 72 miles can be obtained, as compared to only 8 miles for the primary frequency alone. With ambiguity lane widths of this size even relatively fast moving aircraft can resolve ambiguities by counting the phase cycles. In 1977 the 95% accuracies for the Omega system were found to be between 6.3 and 7.3 nautical miles.

### **Automatic Vehicle Location Systems**

During the late 1960's and early 1970's, a number of systems were developed for metropolitan vehicle location using phase or time-delay measurement techniques. An excellent synopsis of these developments is presented in the tutorial articles by Roth [24] and Riter and McCoy [25]. In 1968 Roth has

estimated that there were over 100 companies performing work in automatic vehicle monitoring. At that time, state of the art systems were limited to applications where accuracies of better than 300 m were not required [24].

As an example, one particular system was developed by the Hazeltine Corporation [64, 65, 66]. The system operated in the 900 MHz band and made use of a number of fixed receiver sites around the metropolitan area. The vehicles are equipped with pulse transceivers which cyclically transmit a pattern of  $0.1 \mu\text{s}$  rise time pulses [66]. The arrival times measured from the leading edge of pulses transmitted by the vehicle were calculated at each receiver site. Differences between the arrival times at each of the fixed receiver sites were used at a central fixed location to compute the location of the mobile transmitter. Because the system utilized differences in arrival times, it was a hyperbolic type system. Each vehicle transmitted in a predetermined time slot so that a maximum of 5000 vehicles could be tracked at a time with update rates of 10 s. The 95% accuracy of the system based on testing done by Hazeltine in Dallas was 83 m (after post-processing the data)[64]. Independent testing done by other researchers in Philadelphia found the 95% accuracy to be 503 m [65].

Another system, tested by General Electric, made use of a radio carrier at 150 MHz frequency modulated by a single audio tone at 3 kHz. At a set of base station receivers, the audio phase was used as a measure of the propagation delay; accuracies for 95% of the time were between 137 and 275 m for tests in the Schenectady area [26].

## Local Area Systems

A comprehensive survey of the commercially available positioning systems, as of 1985, is presented by Sampson [27]. Each system offers a different combination of range, accuracy, dynamics, operating modes, and inherent limitations. Horizontal accuracies for multiuser systems are typically in the range of  $\pm 1$  to  $\pm 5$  meters for both phase measurement and pulse timing techniques. Sampson concludes that phase comparison systems offer the best accuracy due to the reduced measurement bandwidth.

One system worth discussing in more detail is the Accutrak system developed by Palmer at the University of Regina [10, 67, 11]. The system was developed for position location use in farm fields. The system operates in the VHF band (148–174 MHz) using a frequency hopping technique where phase delays are measured at 192 frequencies within the band. A transceiver is placed in the vehicle whose position is to be determined. The transceiver sends out a pulse which is retransmitted by a series of transponders located at known locations on the periphery of the measurement area. The transceiver then performs a phase comparison on the returning pulses to obtain the differential delays. The vertical coordinate is determined by a group of barometric pressure sensors with a quoted accuracy of better than one meter. Horizontal accuracies are stated to be at about the half-meter level RMS.

Several military systems have been developed for land navigation and positioning. The Position Location and Reporting System (PLRS), developed in the mid-70's, operates in the 420 to 450 MHz band and position accuracy is stated to be 10 to 50 m [68]. The PLRS system makes use of time infor-

mation to measure absolute distances and is therefore a ranging type system. The system makes use of an elaborate spread spectrum implementation for low probability of intercept and anti-jamming features.

The Joint Tactical Information Distribution System (JTIDS) was developed in the mid-1970's and operates at L-band. Each user terminal has a precise clock that is updated periodically by a master reference clock. Interterminal messages are timed with respect to the local clock and relative positions determined by triangulation. The JTIDS system is a ranging type system which measures time delay information. Accuracies of between 30 and 150 m have been predicted [69, 70].

# Chapter 3

## Position Location by Radio

In this Chapter, radio position location techniques using tracking radar, direction finding, and multilateration will be discussed in detail and the mathematical foundations presented. The accuracy limitations and some system implementation issues will be covered.

### 3.1 Tracking Radar

By far the most common technique used for radar tracking is the *monopulse* method which derives angular error information on the basis of a single radar pulse. The monopulse method uses more than one antenna beam simultaneously; the relative phase or amplitude of the return pulse received in each beam is used to determine the angle of arrival. The discussion which follows describes the amplitude-comparison monopulse technique along with its accuracy limitations as presented by Skolnik [48].

Most radar tracking systems use a directional antenna to mechanically follow a moving target, for the tracking to be successful the servo-control loop must be updated with accurate information about the angular error between the target and the antenna position. In the amplitude-comparison monopulse technique two overlapping antenna patterns, as shown in Figure 3.1*a*, are used to obtain the angular error in one coordinate. The sum of the patterns is shown in Figure 3.1*b* and the difference in Figure 3.1*c*. The transmitter uses the sum pattern while the receiver uses both the sum and difference patterns. Signals received from the sum and difference patterns are converted to an intermediate frequency and a phase comparison used to determine the error voltage shown in Figure 3.1*d*. The block diagram for an amplitude-comparison monopulse system for determining one coordinate is shown in Figure 3.2. The output of the monopulse radar, specifically the angle-error signal, is used to drive a servo-control system to steer the antenna in the direction of the target.

In general, if tracking is required in both elevation and azimuth then two overlapping antenna patterns in each plane are required. The system is more complex than the previous description but the principles are the same. A monopulse radar which tracks in both elevation and azimuth planes can provide the three dimensional location of a target in spherical coordinates. The orientations of elevation and azimuth motors which position the antenna provide the two angles ( $\theta, \phi$ ) and the range signal from the radar provides the distance between the radar and the target ( $r$ ).

For the position location application the target could be a radar beacon which receives the radar pulse and then amplifies and retransmits the pulse,

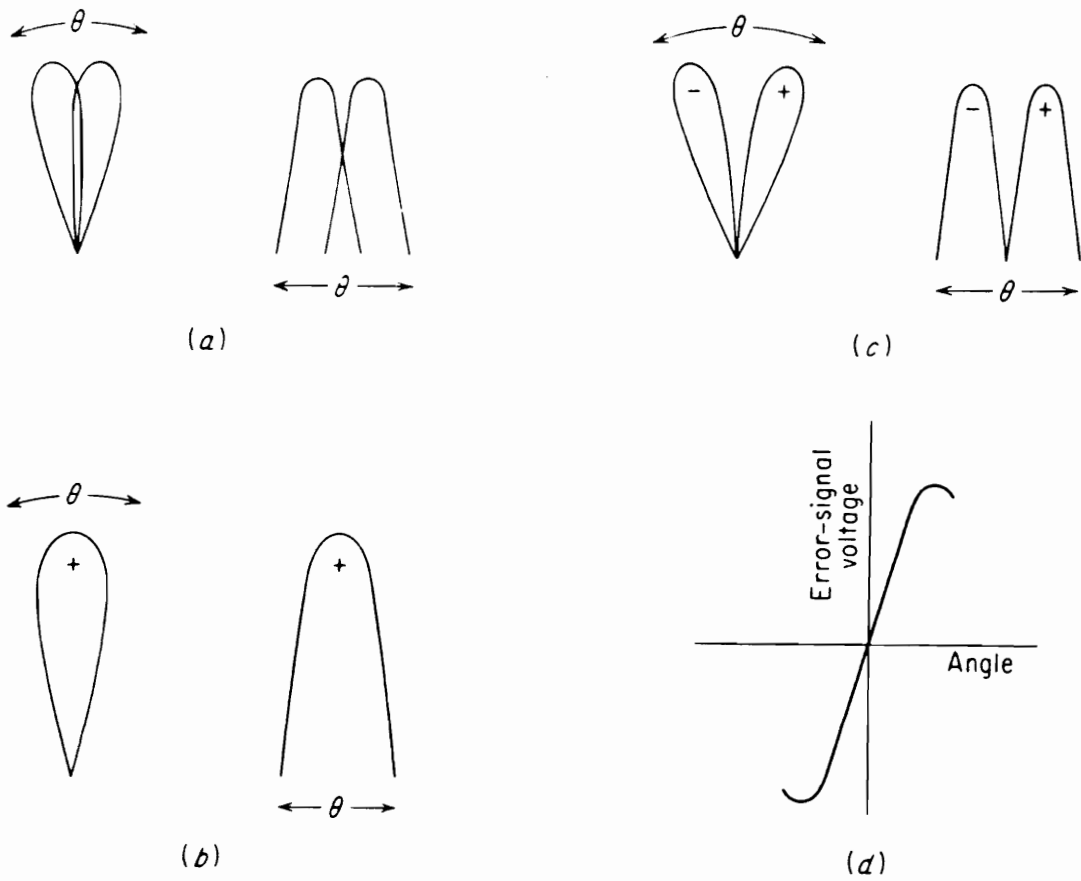


Figure 3.1: Monopulse antenna patterns and error signal (from Skolnik [48])



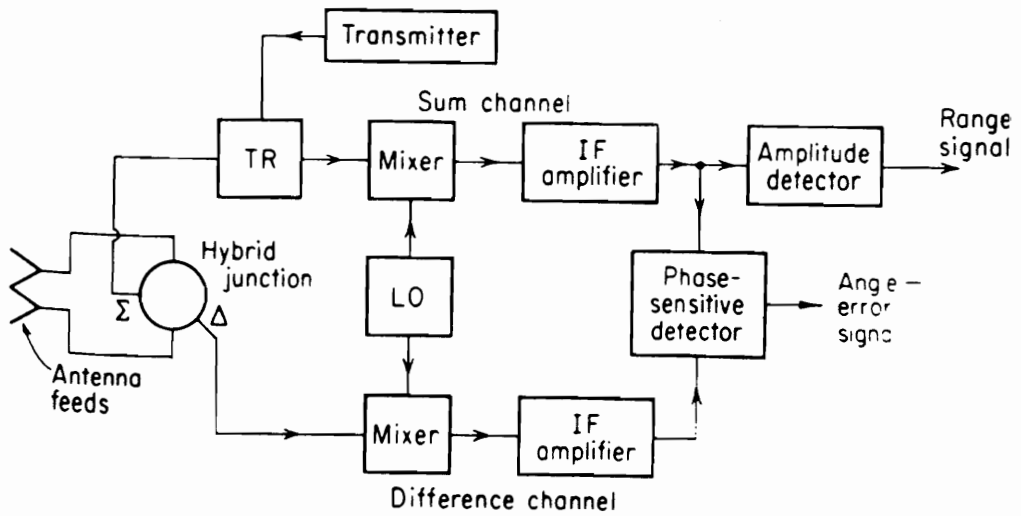


Figure 3.2: Block diagram of amplitude-comparison monopulse tracking radar (from Skolnik [48])

possibly performing a frequency conversion in the process. By easing the problem of target identification and increasing the signal-to-noise ratio, the radar beacon would mitigate several of the more common sources of tracking errors. Angle fluctuations due to target glint could be minimized if the radar beacon had a nearly omnidirectional radiation pattern. The signal-to-noise ratio at the receiver can be proportional to  $1/R^2$  with an active radar beacon as opposed to  $1/R^4$  with a passive reflective target. The servo noise due to the backlash and compliance in the servomechanism will serve as a lower limit on the angular tracking error. The servo noise is not a function of the received pulse and is therefore independent of the target range.

Multipath signals reflected from the earth's surface can introduce errors in tracking radars. The problem of tracking targets at low elevation angles has been well documented and is illustrated in Figure 3.3. The radar receives a pulse via the direct path and an additional pulse via the surface reflected path. These two signals combine at the receiver to cause angular errors which can be many times larger than the angular separation between the target and its image.

Multipath reflections would be of particular concern for a ground based position location system where all of the targets would be near the surface of the earth. The most obvious method of eliminating this problem is using a narrow enough antenna beamwidth to assure that the earth's surface is not illuminated, thus requiring a large antenna or high frequencies. Another technique would be to elevate the target well above the radar so that only the sidelobes of the radar antenna would illuminate the ground. On a work site with large altitude variations it may not always be possible to have enough

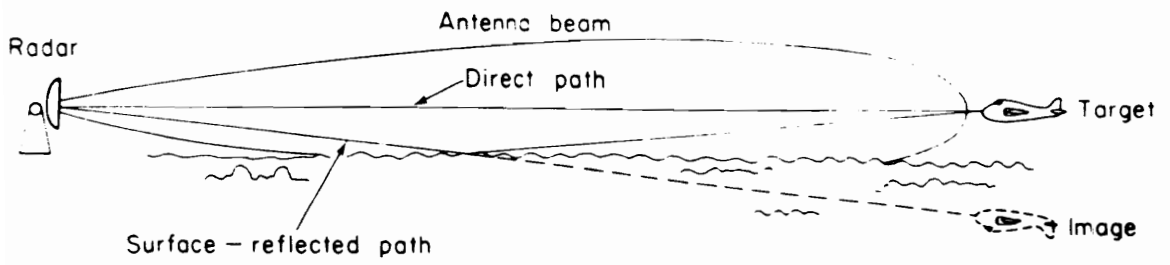


Figure 3.3: Low-angle tracking with surface reflection present (from Skolnik [48])

control over the radar/beacon geometry to assure that no surface multipath would exist.

The surface reflected signal travels a longer path length than the direct signal and with the proper range resolution the two signals can be separated. The range resolution required to separate the direct and reflected signal is [48]

$$\Delta R = \frac{2h_a h_t}{R} \quad (3.1)$$

where  $h_a$  is the radar antenna height,  $h_t$  is the target height, and  $R$  is the range to the target. For a 1000 meter site with radar and target heights of 10 meters the range resolution required is 0.2 meter. This range resolution corresponds to a pulse width of 1.3 ns and would require a receiver bandwidth of approximately 770 MHz. To achieve this sort of bandwidth would necessitate center frequencies above about 10 GHz to obtain realistic percentage bandwidths (less than 10%) for the receiver. These high operating frequencies are susceptible to fading due to rain and such a receiver would be costly to implement. While it is possible in principle to achieve this resolution, such a system would not always be practical due to cost and complexity constraints.

Another method of reducing the effect of multipath reflections is through the use of frequency diversity. A change in frequency changes the phase relationship between the direct and reflected signals. The angle of the target can be deduced from the behavior of the error signal versus frequency. It turns out that the bandwidths required to get rid of the multipath error are essentially the same as those necessary when using short pulses to separate the signals in range. The bandwidths necessary to eliminate multipath by

using range resolution or frequency diversity are generally quite large for most applications.

Radar tracking systems must acquire the target either at the start of operations or after the loss of tracking lock. The radar must scan an angular sector to find the target before tracking can begin. The time required to acquire the target depends on the antenna beamwidth and the size of the angular sector to be covered. If narrow beamwidths are implemented to reduce multipath interference then there will be a corresponding increase in the acquisition time.

Monopulse tracking radars are capable of tracking only a single target at a time. If a number of vehicles are to be tracked on a site, it may become prohibitively expensive to install a separate monopulse system for each vehicle.

## 3.2 Direction Finding

Direction finding through the use of directional antennas with deep nulls in their radiation patterns has been around for many years. However, modern high accuracy passive direction finding was an outgrowth of the development of monopulse radar systems. As receiver sensitivities improved, many radar systems used separated transmit and receive antennas (*bistatic radar*) as opposed to the collocated case (*monostatic radar*). The transmitter illuminated the target from one location while the receiver detected the angle of arrival from the return at another location. The receiver in this case is nothing more than a passive direction finding system.

The most common technique for high resolution direction finding is the interferometer method which makes use of the phase of the arriving signals at an array of antennas to determine angle of arrival. Interferometry has been used by radio astronomers for years to make precise measurements of the positions of radio stars. Generally, a linear or circular array of antennas is used and the spacing between antennas results in time of arrival differences and hence phase differences at the antennas. The spacing between the antennas is a critical factor in system performance; large spacings result in high resolution at the expense of ambiguities in the angle of arrival [49].

To determine the performance of interferometer direction finding systems, a simple two antenna system, as shown in Figure 3.4, can be analyzed. The phase difference between the two received voltages is a function of the path length difference ( $D \cos \phi$ ) and can be written as

$$\Delta\Psi = \frac{2\pi}{\lambda} D \cos \phi \quad (3.2)$$

where  $\Delta\Psi$  is the phase difference,  $2\pi/\lambda$  is the free space propagation constant, and  $D$  is the spacing between the antennas. Solving for the angle of arrival ( $\phi$ ) in terms of the other parameters gives

$$\phi = \cos^{-1}\left(\frac{\lambda \Delta\Psi}{2\pi D}\right) \quad (3.3)$$

where phase difference,  $\Delta\Psi$ , can only be measured modulo  $2\pi$  due to its cyclic nature.

The received phase difference gives information only about the fractional portion of the phase and provides no way to resolve the integer number of phase cycles. From equation 3.3 the maximum antenna spacing which allows

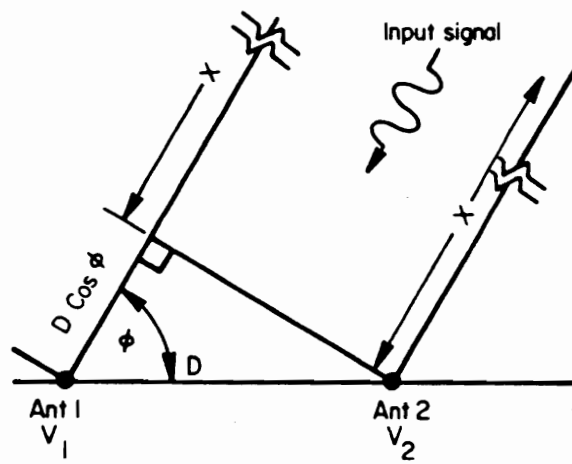


Figure 3.4: Single baseline interferometer geometry (from Lipsky [49])

unambiguous angle resolution can be calculated. Noting that  $\Delta\Psi \in (-\pi, \pi)$ , we can substitute the upper and lower bounds into equation 3.3 giving  $|(2D)/\lambda| \leq 1$ . Solving for the spacing gives the result  $D \leq \lambda/2$  for an unambiguous angle determination and for spacings larger than this ambiguities are present.

The resolution capability of the interferometer system is a linear function of the antenna spacing. From equation 3.2 one can see that large antenna spacings ( $D$ ) are required to obtain large changes in the measured phase ( $\Delta\Psi$ ) for a given angular change ( $\phi$ ). The interferometer can achieve high resolution with a choice of large  $D/\lambda$  ratios at the expense of solving the ambiguity problem that arises.

One method of resolving the ambiguities is known as multiple baseline interferometry. In such a scheme, the spacings between the antennas in the array are unequal. A wide spaced baseline pair can have its ambiguities resolved by another more closely spaced pair which has no ambiguities. The fine angular resolution is achieved by the longest baseline pair since it has the highest resolution.

A binary technique has been used to provide high resolution unambiguous coverage from boresight to the  $\pm 90^\circ$  field of view. In the binary interferometer a linear array of antennas with spacings starting at  $\lambda/2$  and progressing in a binary fashion ( $\lambda/2, \lambda, 2\lambda, 4\lambda, \dots$ ) up to the longest baseline, as shown in Figure 3.5. In this manner each baseline pair resolves the ambiguities for the next longer pair. A binary system is limited in its frequency bandwidth at the high end by the  $\lambda/2$  spacing and at the low frequency end by the physical dimensions required. Nonbinary spacings for linear arrays and circular



arrays have also been successfully implemented.

Lipsky claims that nonbinary spaced microwave interferometers for military applications are capable of between  $0.1\text{--}15^\circ$  RMS angular accuracies [49]. Interferometers have the potential for good accuracy but this is coupled with the finite probability of incorrect ambiguity resolution resulting in a gross error.

Interferometers are sensitive to multipath interference because the reflected signals alter the phase relationships between the signals received by the antenna array. Some of the ambiguity resolution and multipath problems can be improved by using frequency diversity and measuring signal amplitudes as well as phases; of course this increases the system complexity. Extensive work has been done recently in applying maximum likelihood and adaptive techniques to the angle of arrival estimation problem in the presence of specular and diffuse multipath. Haykin presents a review of the work which has been done and an assessment of the capability [71].

For a position location system based on direction finding arrays to provide a 3-D position estimate, three noncollocated arrays would be necessary. The three angle of arrival measurements from the arrays could then be used to solve for the position location. To provide accuracies of 1 meter on a 1000 meter site angular accuracies of approximately  $0.05^\circ$  would be required. This would be beyond the capabilities of most practical direction finding systems, particularly if multipath interference were present.

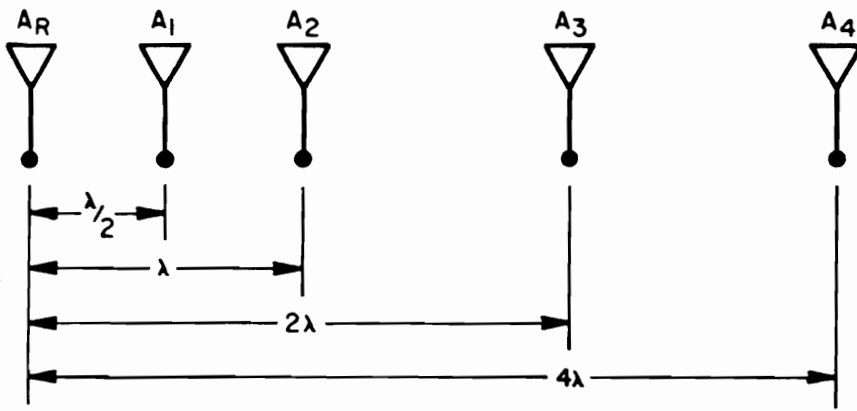


Figure 3.5: Unambiguous binary baseline interferometer (from Lipsky [49])

### 3.3 Multilateration

Multilateration systems are those which estimate the user's position based on the distance or difference in distance between the user and a set of base stations at known locations. The base stations need not be stationary, but their locations must be known at the instant when the range measurements are made. Those systems which measure distance directly are called ranging systems and those which measure difference in distance are called hyperbolic systems. The distances can be calculated by measuring the time, phase, or frequency of signals transmitted between the user's location and the base stations.

#### 3.3.1 Ranging Method

Ranging type systems measure the absolute distance between the user's location and a set of base stations. In a two dimensional system, each range measurement between a base station and the user defines the unknown position to be on a circle centered at the base station, as shown in Figure 3.6. In the figure the circles of constant range are denoted by the dashed curves. To determine the user's location uniquely in two dimensions the distances between the user and two base stations (labeled *A* and *B* in the figure) must be measured, with the intersection of the two circles defining the user position. In three dimensions the range between the user and a base station locates the user on a sphere centered at the base station.

In three dimensions the basic equations which define the measured ranges

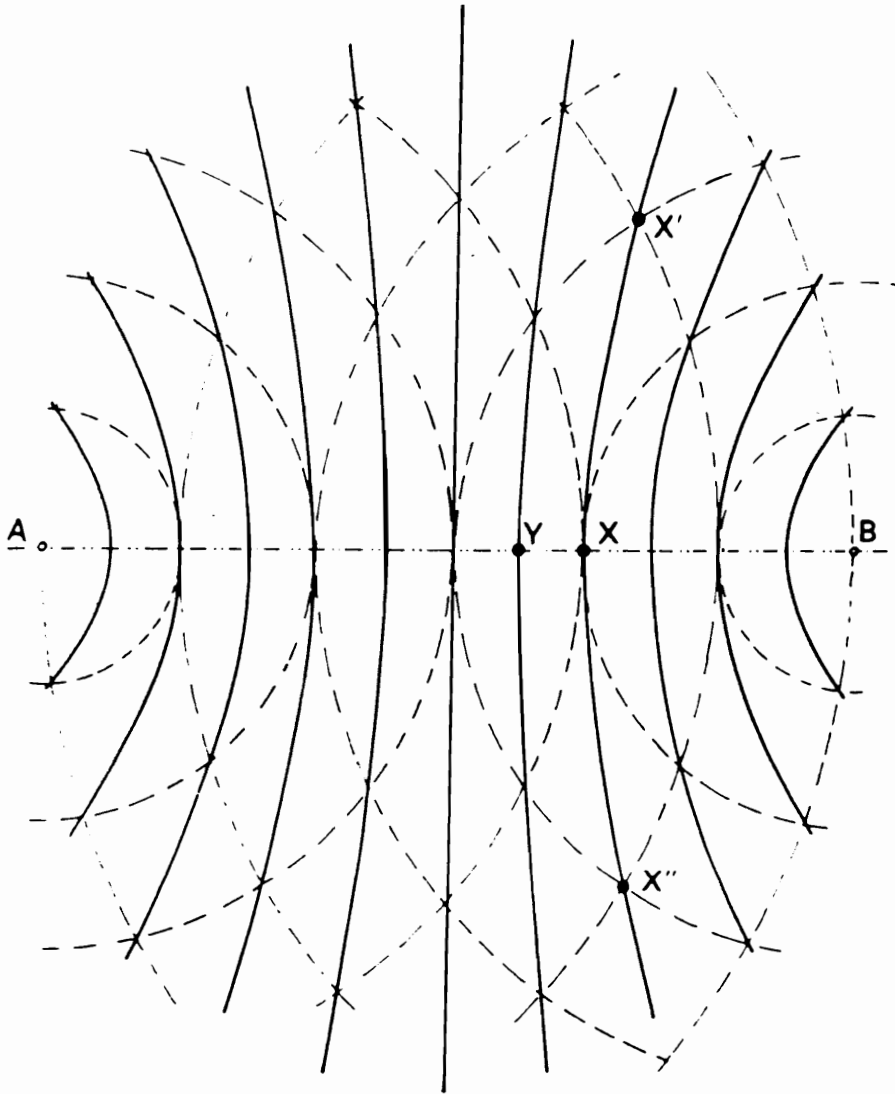


Figure 3.6: Ranging method of multilateration (from Tetley [17])

are

$$R_i = \sqrt{(x - x_i)^2 + (y - y_i)^2 + (z - z_i)^2} \quad (3.4)$$

where  $i$  corresponds to the base station number;  $x, y$ , and  $z$  are the user's unknown location;  $x_i, y_i$ , and  $z_i$  are the known base station locations; and  $R_i$  is the range from the user location to the  $i^{\text{th}}$  base station location. These equations are each in the cartesian center-radius form for a circle with radius  $R_i$  and center at the base station position  $x_i, y_i, z_i$ . The number of equations to be solved is equal to the total number of base stations ( $N$ ) because there is an equation for each base station ( $i = 1, 2, \dots, N$ ).

The set of equations 3.4 define a nonlinear  $N \times 3$  system which must be solved for the user's unknown location, generally requiring an iterative approach. If the number of base stations equals the number of unknowns ( $N = 3$ ), then the system is consistent and a unique solution can exist. If, on the other hand, the number of base stations exceeds the number of unknowns ( $N > 3$ ) then the system is inconsistent and no unique solution vector exists. In the inconsistent system case a criteria must be selected by which a solution can be found. In most cases the least squared error criteria is used to obtain a solution but many other alternatives exist.

Ranging type systems require that some form of synchronization be employed so that the transmitted and received waveforms can be compared using time or phase. This synchronization can be implemented by using stable clocks at both the user's location and the base stations, as in the Global Positioning System, the Position Location Reporting System, and the Joint Tactical Information Distribution System. The Global Positioning

System makes use of cesium standard clocks with a stability of about one part in  $10^{12}$  per day for timing synchronization. Another method of performing range measurements is to place a transponder, with known delay characteristics, at the base station locations and to compare the transmitter oscillator directly to the received signal. The transponder approach has been patented by Accutrak Systems Inc.; their implementation makes use of frequency translators with a fixed delay time [67].

In most actual system implementations there will be a clock or phase bias introduced into the range measurements. In this case an extra unknown is introduced in the system and four equations, rather than three, must be solved to determine the user's position. The new equations which include the bias term are

$$R_i = \sqrt{(x - x_i)^2 + (y - y_i)^2 + (z - z_i)^2} + T \quad (3.5)$$

where  $T$  is the unknown user bias. In this case the  $R_i$  values would be called pseudo-ranges because of the presence of the bias term.

The above equations are nonlinear and can be solved directly using an iterative approach. Many users, in order to simplify the calculations, opt to employ a set of linearized equations based on incremental relationships [72]. These linearized equations introduce some error into the calculation but the increase in computational speed is often worthwhile. The linearized equations also allow a more tractable error analysis approach to be followed.

To perform the linearization, let  $\hat{x}, \hat{y}, \hat{z}, \hat{T}$  be nominal or estimated values of the user location  $(x, y, z)$  and bias  $(T)$ . The nominal values can be updated or corrected by the terms  $\Delta x, \Delta y, \Delta z, \Delta T$ . In a similar manner  $\hat{R}_i$  is the

nominal value for the pseudo-range from the  $i^{\text{th}}$  base station and  $\Delta R_i$  is the difference between the actual and nominal pseudo-range. In other words, the following relations have just been defined

$$\begin{aligned} x &= \hat{x} + \Delta x \\ y &= \hat{y} + \Delta y \\ z &= \hat{z} + \Delta z \\ T &= \hat{T} + \Delta T \\ R_i &= \hat{R}_i + \Delta R_i \end{aligned} \quad (3.6)$$

where the nominal pseudo-range value is

$$\hat{R}_i = \sqrt{(\hat{x} - x_i)^2 + (\hat{y} - y_i)^2 + (\hat{z} - z_i)^2} + \hat{T} \quad (3.7)$$

Substituting the incremental equations back into the basic ranging equations 3.5 gives

$$\begin{aligned} \hat{R}_i + \Delta R_i - \hat{T} - \Delta T &= \\ \sqrt{(\hat{x} + \Delta x - x_i)^2 + (\hat{y} + \Delta y - y_i)^2 + (\hat{z} + \Delta z - z_i)^2} & \end{aligned} \quad (3.8)$$

where  $i$  ranges from 1 to the number of base stations. By using a Taylor series expansion and neglecting second order terms the equations become

$$\begin{aligned} \hat{R}_i + \Delta R_i - \hat{T} - \Delta T &= \\ \sqrt{(\hat{x} - x_i)^2 + (\hat{y} - y_i)^2 + (\hat{z} - z_i)^2} + \frac{(\hat{x} - x_i)\Delta x + (\hat{y} - y_i)\Delta y + (\hat{z} - z_i)\Delta z}{\sqrt{(\hat{x} - x_i)^2 + (\hat{y} - y_i)^2 + (\hat{z} - z_i)^2}} & \end{aligned} \quad (3.9)$$

Substituting equation 3.7 and solving for the difference between the actual and nominal ranges gives

$$\Delta R_i = \frac{(\hat{x} - x_i)}{\hat{R}_i - \hat{T}} \Delta x + \frac{(\hat{y} - y_i)}{\hat{R}_i - \hat{T}} \Delta y + \frac{(\hat{z} - z_i)}{\hat{R}_i - \hat{T}} \Delta z + \Delta T. \quad (3.10)$$

The set of equations defined in 3.10 are linearized equations which relate the pseudo-ranges measurements to the user's position and bias. The known quantities on the left hand side of the equations are actually incremental pseudo-ranges which represent the difference between the actual pseudo-ranges and the previous update values. The quantities to be computed,  $\Delta x, \Delta y, \Delta z, \Delta T$ , are corrections to the current estimate of the user's position and bias. The coefficients on the ~~left~~<sup>right</sup> hand side are the direction cosines of the vector between the user's location and the base stations.

The linearized equations can be presented more compactly in matrix form. The matrix notation for the set of  $N$  equations and 4 unknowns defined in equation 3.10 is

$$\begin{bmatrix} \alpha_{11} & \alpha_{12} & \alpha_{13} & 1 \\ \alpha_{21} & \alpha_{22} & \alpha_{23} & 1 \\ \vdots & \vdots & \vdots & \vdots \\ \alpha_{N1} & \alpha_{N2} & \alpha_{N3} & 1 \end{bmatrix} \times \begin{bmatrix} \Delta x \\ \Delta y \\ \Delta z \\ \Delta T \end{bmatrix} = \begin{bmatrix} \Delta R_1 \\ \Delta R_2 \\ \vdots \\ \Delta R_N \end{bmatrix} \quad (3.11)$$

where  $\alpha_{ij}$  is the direction cosine of the angle between the vector to the  $i^{\text{th}}$  base station and the  $j^{\text{th}}$  coordinate. The direction cosines are then

$$\begin{aligned} \alpha_{i1} &= \frac{\hat{x} - x_i}{\hat{R}_i - \hat{T}} \\ \alpha_{i2} &= \frac{\hat{y} - y_i}{\hat{R}_i - \hat{T}} \\ \alpha_{i3} &= \frac{\hat{z} - z_i}{\hat{R}_i - \hat{T}}. \end{aligned} \quad (3.12)$$



Letting

$$\mathbf{A} \equiv \begin{bmatrix} \alpha_{11} & \alpha_{12} & \alpha_{13} & 1 \\ \alpha_{21} & \alpha_{22} & \alpha_{23} & 1 \\ \vdots & \vdots & \vdots & \vdots \\ \alpha_{N1} & \alpha_{N2} & \alpha_{N3} & 1 \end{bmatrix}$$

$$\mathbf{x} \equiv \begin{bmatrix} \Delta x \\ \Delta y \\ \Delta z \\ \Delta T \end{bmatrix}$$

$$\mathbf{r} \equiv \begin{bmatrix} \Delta R_1 \\ \Delta R_2 \\ \vdots \\ \Delta R_N \end{bmatrix}$$

then the matrix equation for pseudo-ranging type systems becomes

$$\mathbf{Ax} = \mathbf{r}$$

and solving for the user's location vector ( $\mathbf{x}$ )

$$\mathbf{x} = \mathbf{A}^{-1}\mathbf{r}. \quad (3.13)$$

This equation represents the relationship between the pseudo-range measurements made by the system and the user's position and bias.

Because the incremental equation is linear, it can be used to express the relationship between the pseudo-range errors and the user's position errors in a simple form. The error relationship is

$$\boldsymbol{\epsilon}_x = \mathbf{A}^{-1}\boldsymbol{\epsilon}_r \quad (3.14)$$

where  $\epsilon_r$  is an array containing the pseudo-range measurement errors and  $\epsilon_x$  contains the corresponding errors in the estimated user position. It is important to note that the relationship between the measurement errors and the position errors is only a function of the geometry between the user location and the base stations. All the geometric relationships are contained within the direction cosine matrix  $\mathbf{A}$ . The accuracy of ranging systems depends to a large extent on whether good geometric conditions exist between the base stations and the user's position.

From equation 3.14 one can see that the system geometry, contained within the matrix  $\mathbf{A}$ , magnifies the measurement errors. The measure of how the system geometry degrades accuracy is often referred to as *dilution of precision*. The Geometric Dilution of Precision (GDOP) is a standard concept for quantifying this relationship. Jorgensen has shown that the GDOP for ranging type systems, assuming unity pseudo-range errors, is

$$\text{GDOP} = \sqrt{\text{TRACE}[(\mathbf{A}^T \mathbf{A})^{-1}]} \quad (3.15)$$

where the TRACE is the sum of the diagonal elements of the matrix. The GDOP can be used as a criteria for designing the system geometry or for selecting among the base stations if more than the required number are present.

### 3.3.2 Hyperbolic Method

Hyperbolic systems are those which measure the difference in distance between the user's location and a pair of base stations. The difference in distance between the user and a pair of base stations defines the user's unknown location to be on a hyperbola of constant range difference, with the foci at

the base stations. This is illustrated graphically in Figure 3.7 with the two base stations labeled  $M$  and  $S$ . A user located anywhere on one of these hyperbolas will exhibit the same difference in distance to the two base stations.

To determine the user's location uniquely in two dimensions two differential range measurements must be used, with the user's location defined by the intersection between the two hyperbolas, as shown in Figure 3.8. In the figure the base stations are labeled  $A$ ,  $B_1$ ,  $B_2$  and the point  $P$  is located at the intersection of the  $s$  and  $u$  hyperbolas. In three dimensions the user's location is known to be on a hyperbolic surface of revolution and three differential range measurements are necessary to uniquely define the user's location.

The hyperbolic method uses only relative measurements and therefore does not require special synchronization between the user and the base stations. Signals transmitted from or received by pairs of base stations are compared to each other to determine the difference in delay. The hyperbolic method exhibits a more nonuniform error distribution than an equivalent ranging system. This geometrical characteristic of a ranging system results in a larger portion of the measurement area, compared to an equivalent hyperbolic system, over which a location fix of given accuracy can be obtained [74].

The equations which describe the user's position in relation to the base station locations and the measured differential ranges are presented below. Let  $\Delta R_{jk}$  denote the difference in distance between the user location and the  $j^{th}$  and  $k^{th}$  base stations (that is,  $\Delta R_{jk} = R_j - R_k$ ). The equations to be

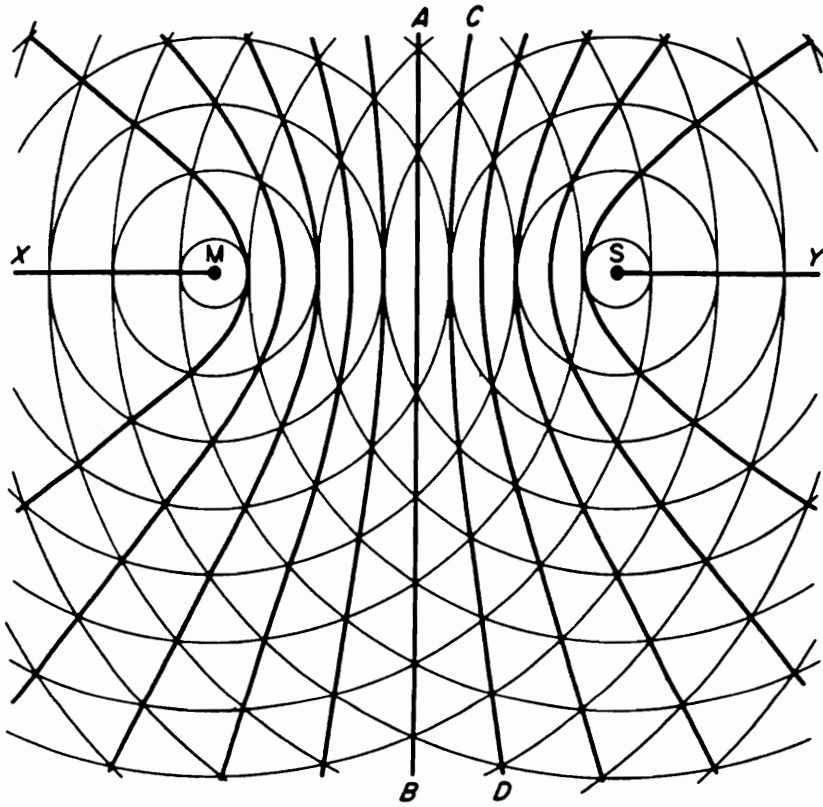


Figure 3.7: Hyperbolas of constant differential distance for a pair of base stations (from Powell [73])

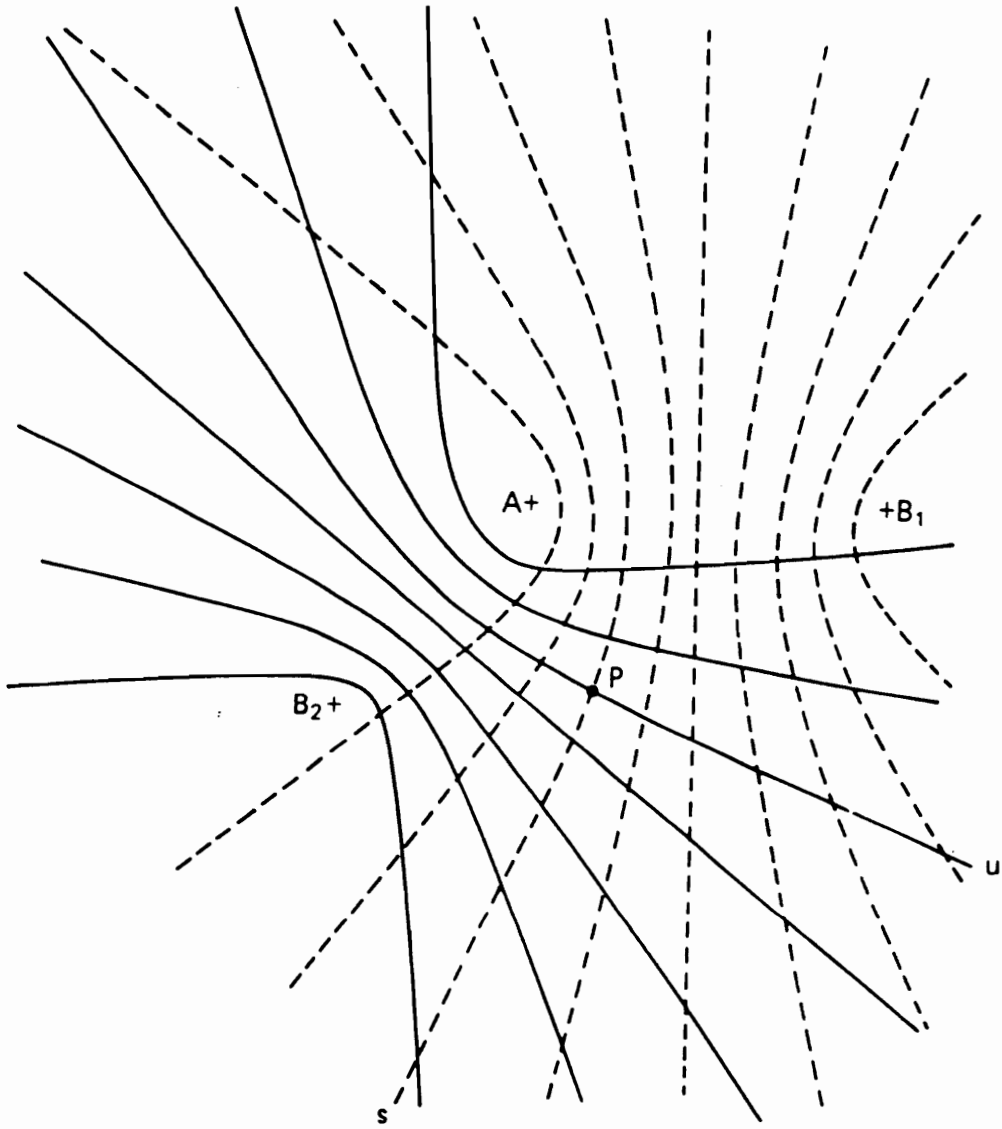


Figure 3.8: Example of position location using a pair of hyperbolas of constant differential distance for two pairs of base stations (from Sonnenberg [16])

solved are then

$$\Delta R_{jk} = \sqrt{(x - x_j)^2 + (y - y_j)^2 + (z - z_j)^2} - \sqrt{(x - x_k)^2 + (y - y_k)^2 + (z - z_k)^2} \quad (3.16)$$

where  $x, y$ , and  $z$  are the user's unknown location; and  $x_j, y_j, z_j$  are the locations of the  $j^{\text{th}}$  base station. The equations each describe a hyperbola with foci at the base station locations  $x_j, y_j, z_j$  and  $x_k, y_k, z_k$ . These hyperbolic equations, like those for the ranging type system, are nonlinear and can be solved using an iterative solution.

The hyperbolic equations in 3.16 can be put into polynomial form by an algebraic manipulation giving the resulting system

$$c_1^n x^2 + c_2^n y^2 + c_3^n z^2 + c_4^n x + c_5^n y + c_6^n z + c_7^n xy + c_8^n xz + c_9^n yz + c_{10}^n = 0 \quad (3.17)$$

where  $n$  ranges from one up to the number of base stations minus one ( $n = 1, 2, \dots, N - 1$ ). The coefficients are all functions of known quantities, the base station locations and the measured differential range ( $c_m^n = f(x_j, y_j, z_j, x_k, y_k, z_k, \Delta R_{jk}) \forall m = 1, 2, \dots, 10; n = 1, 2, \dots, N - 1$ ). From this form of the equations one can notice the nonlinearity and cross-coupling between the  $x, y, z$  coordinate terms. This indicates that a measurement error in a pair of base stations located along one particular coordinate axis will adversely affect the other two coordinates as well.

The system of equations 3.16 is an  $(N - 1) \times 3$  system where  $N$  is the number of base stations. Because only differences in range are used there is always one less equation than the number of base stations. In many cases one base stations is selected as a reference and all range measurements are

made relative to that station. In the above equations this means that, for example,  $j$  would be fixed at the first base station ( $j \equiv 1$ ) and then  $k$  would span the other base stations ( $k = 2, 3, \dots, N$ ) for a total of  $N - 1$  equations.

If there are four base stations ( $N = 4$ ) then a  $3 \times 3$  consistent system of equations results and there is a unique solution for the user's position  $(x, y, z)$ . If there are more base stations ( $N > 4$ ) then an inconsistent system with more equations than unknowns results. As mentioned previously, some criteria must be determined for selecting an 'optimum' solution. The differences between the various criteria which can be used for determining a solution are discussed more fully in Section 3.3.3 and in Appendix B.

As with the ranging equations, attempts have been made to linearize the hyperbolic equations. Unfortunately, the increased complexity of the hyperbolic equations results in a more troublesome linearized representation than was obtained for the ranging equations. Linearized solutions using the Taylor expansion or similar approaches have been presented for two dimensional systems (that is, those which measure only horizontal position) in several references [75, 26, 76]. The hyperbolic equations can be linearized in three dimensions but the resulting equations are quite cumbersome. Because of an inherent squaring operation the linearization process creates an ambiguity in sign which causes multiple solutions to the system of equations [77, 26]. In practice, multipath and other measurement errors have been found to complicate the choice of the proper sign resulting in the possibility of large position errors.

Nicholson has done an excellent study of the error penalty induced by linearization of the hyperbolic equations [75, 77]. He has developed sev-

eral error tolerant linearization approaches using Weighted Least Squares (WLS), Sequentially Weighted Least Squares (SWLS), and Adaptive Sequentially Weighted Least Squares (ASWLS) methods. These techniques assign weighting coefficients based on measures of the residual errors and in the latter case using adaptive testing criteria. The linearized methods were found to have a two to one speed advantage over the nonlinear solutions but at the expense of an increase in errors. At the 95% confidence level the LLS algorithm increased position errors by 117% and the SWLS and ASWLS algorithms resulted in a 57% error increase over the nonlinear solution. At the one standard deviation error level the differences between the four techniques were much less dramatic.

Ultimately, Nicholson selected the nonlinear approach as better than the linearized versions because it resulted in improved accuracy, provided a satisfactory update rate, and avoided the mathematical complexities involved with the sign ambiguity [77]. The conclusion which can be drawn from these results indicates that if accuracy is an important concern then a nonlinear solution of the hyperbolic equations is mandatory.

Error analysis for hyperbolic type systems can be performed by analyzing the geometry of the hyperbolas. A hyperbolic pattern for two base stations is shown in Figure 3.9, where the base stations are  $M$  and  $S$ . One can note that away from the baseline the hyperbolas diverge meaning that a given change in the position of the user results in a smaller change in the differential distance. This loss in sensitivity away from the baselines can be measured by an *expansion factor* that is equal to the cosecant of half the angle subtended by the base stations (the angle  $\gamma/2$  in the figure) [73]. Contours of



constant expansion factor are plotted for values of 1 to 5 in the figure where the contours are circles with the baseline as a chord. The sensitivity of the system decreases as the distance between the user and the baseline is long compared with the baseline length. This is particularly a problem for systems which are constrained to use a short baseline length. The expansion factor concept will be utilized later in the analysis of the distribution of position errors.

In a two dimensional system the position estimate is obtained as the intersection between two hyperbolic lines of position. The effects of errors on this position estimate can be seen in Figure 3.10 which shows a magnified view of the intersection of two hyperbolas, labeled 1 and 2. The ‘angle of cut’ between the two hyperbolas is labeled as the angle  $\beta$  in the figure. Errors in the hyperbolas due to measurement errors cause the position estimate, ideally at the point  $P$ , to be displaced.

If errors cause the hyperbolas to be displaced by  $\sigma_1$  and  $\sigma_2$  then the estimated position will be moved from point  $P$  by a distance  $d$  found from the geometric relation

$$d^2 \sin^2 \beta = \sigma_1^2 + \sigma_2^2 + 2\sigma_1\sigma_2 \cos \beta. \quad (3.18)$$

By now assuming that the errors are random rather than deterministic an estimate of the RMS error can be made. Assume that the errors are normally distributed with zero mean and variances  $\sigma_1^2, \sigma_2^2$  and a correlation coefficient  $\rho$  between the two distributions. Taking the mean of both sides of equation 3.18 gives

$$\overline{d^2 \sin^2 \beta} = \sigma_1^2 + \sigma_2^2 + 2\rho\sigma_1\sigma_2 \cos \beta. \quad (3.19)$$



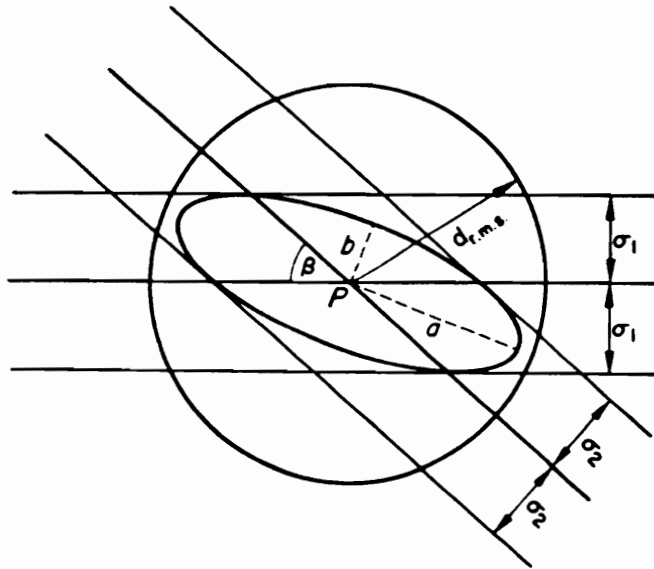


Figure 3.10: Magnification of hyperbolic intersection and error analysis geometry (from Powell [73])

The RMS error  $d_{RMS}$  is then simply equal to  $\sqrt{d^2}$ . In Figure 3.10 geometrically  $d_{RMS} = \sqrt{a^2 + b^2}$  where  $a$  and  $b$  are the axes of the ellipse enclosed by the parallelogram around point  $P$ . The length of the lines  $a$  and  $b$  each represent the RMS error along one dimension and therefore the error in two dimensions is the square root of the sum of the squares. The standard deviation  $\sigma$  corresponds to a probability of 68.26%, therefore the parallelogram formed by the shifted hyperbolas would contain  $68.26\% \times 68.26\% = 46.6\%$  of the position estimates at the point  $P$ . Of course, the RMS error circle would include 68.26% of the position estimates at the point  $P$ .

Solving equation 3.19 for  $d_{RMS}$  gives the result

$$d_{RMS} = \sqrt{d^2} = \csc \beta \sqrt{\sigma_1^2 + \sigma_2^2 + 2\rho\sigma_1\sigma_2 \cos \beta}. \quad (3.20)$$

From this equation the expected RMS errors at any point can be determined based on the Gaussian error assumption, estimates of the variances and correlation for the hyperbolas, and the angle of cut.

Several important observations can be made regarding equation 3.20. First, it should be noted that the equation is a quite conservative estimate of the error at any particular location because it does not take into account the direction in which the error displaces the position estimate from the true position. The actual error distribution is represented as the ellipse in Figure 3.10 and one can see the definite directional orientation to the errors. This means that at any given point the position error may be considerably less in certain directions than the RMS value for that point from equation 3.20 would indicate.

The RMS error equation 3.20 can be used to compute the error contours

necessary to analyze the coverage area of any particular configuration of base stations. The standard deviation of the hyperbolic errors  $\sigma$  can be calculated from the following relationship

$$\sigma = \sigma_D \csc \frac{\gamma}{2} \quad (3.21)$$

where  $\sigma_D$  is the standard deviation of the differential distance errors for the pair of base stations. The  $\csc \gamma/2$  term is simply the expansion factor which was previously discussed and shown in Figure 3.9. The angle  $\gamma/2$  is the angle subtended by the baseline at the observation point. The quantity  $\sigma_D$  is relatively easy to estimate or it can be measured directly from experimental data. Equation 3.21 shows the magnification effect the geometry can have on the measured distance errors.

A more generalized approach to determining the sensitivity of position errors to user-base station geometry has been developed in three dimensions for hyperbolic systems [78, 62]. The errors can be analyzed for a least squared error solution by forming a covariance matrix of the form

$$\mathbf{\Gamma} = \begin{bmatrix} \Gamma_{xx} & \Gamma_{xy} & \Gamma_{xz} \\ \Gamma_{xy} & \Gamma_{yy} & \Gamma_{zy} \\ \Gamma_{xz} & \Gamma_{zy} & \Gamma_{zz} \end{bmatrix} \quad (3.22)$$

where the  $\Gamma_{ij}$  are magnification factors which indicate how much the measurement ranging error is magnified by the user-base station geometry. The

diagonal terms of the matrix are

$$\begin{aligned}\frac{\sigma_x^2}{\sigma_m^2} &= \Gamma_{xx} \\ \frac{\sigma_y^2}{\sigma_m^2} &= \Gamma_{yy} \\ \frac{\sigma_z^2}{\sigma_m^2} &= \Gamma_{zz}\end{aligned}\tag{3.23}$$

where  $\sigma_m$  is the mean squared measurement ranging error and  $\sigma_x, \sigma_y, \sigma_z$  are the standard deviation of the errors in the  $x, y, z$  coordinates. The Geometric Dilution of Precision (GDOP), which measures the ratio of the RMS measurement ranging error to the RMS position error, is

$$\begin{aligned}\text{GDOP} &= \frac{\sqrt{\sigma_x^2 + \sigma_y^2 + \sigma_z^2}}{\sigma_m} \\ &= \sqrt{\Gamma_{xx} + \Gamma_{yy} + \Gamma_{zz}}\end{aligned}\tag{3.24}$$

where the  $\Gamma$  terms may be calculated solely from the user-base station geometry.

The calculation of the  $\Gamma$  matrix is a rather involved procedure requiring that the user-base station geometry be transformed mathematically into an equivalent mechanical system. The basic procedure is to project the vectors between the user and the base stations unto a sphere of unit radius with the user's location at the center. Equivalent 'masses', which are inversely proportional to the variances expected from each base station, are placed at the points where the vectors intersect the unit sphere. Lee has rather ingeniously shown that the  $\Gamma$  matrix can be found as the inverse of a matrix which contains the mechanical moments and products of inertia of the configuration of equivalent 'masses' [78].

### 3.3.3 Solving Nonlinear Systems of Equations

Both the ranging and hyperbolic methods require the solution of a system of nonlinear equations. In many cases the number of base stations will exceed the minimum number required to obtain a position estimate. The extra base stations are used for redundancy, diversity improvement of error performance, and to maintain improved user/base station geometries. The information from the extra base stations causes the system of nonlinear equations to be inconsistent because the number of equations exceeds the number of unknowns. In this case no unique solution for the system of equations exists, instead a criteria must be selected for choosing an 'optimum' solution. Optimum is actually a subjective term here because one must decide in what sense one solution is better than another.

#### Ranging Method

In Section 3.3.1 it was shown that by an appropriate incremental process the ranging equations could be linearized without undue loss of accuracy. The position estimate for the ranging problem then requires only the solution of the linear system defined in equation 3.13. If the system is consistent, as will be the case if three base stations are used, then only a  $3 \times 3$  matrix inversion and multiplication are required to obtain the position estimate. Four base stations and a  $4 \times 4$  system are required if pseudo-ranges with bias terms are used. If information from more than three or four base stations is used then the system is inconsistent and in this case a linear least squares approach will work. Other techniques such as least absolute deviation or robust methods

can be used and since the equations are linear, little difficulty is encountered in obtaining a solution.

### Hyperbolic Method

The ranging equations can in most cases be solved in a quite straight forward manner, while the hyperbolic equations present more of a challenge. As was discussed in Section 3.3.2, the ranging equations can be linearized but at a considerable loss of accuracy. For this reason it is usually best to resort to a direct nonlinear solution to the hyperbolic equations at the expense of extra computational time. If the system is consistent, meaning that information from four receivers is utilized, then a unique solution exists and a simple iterative approach can be used [79]. For the case of an inconsistent system the problem is a bit more involved because no unique solution exists.

No matter which set of equations we choose to solve, the requirement is to obtain a position estimate from equations of the form

$$f_i(x, y, z; x_1, y_1, z_1; x_{i+1}, y_{i+1}, z_{i+1}); \quad i = 1, 2, \dots, (N - 1) \quad (3.25)$$

where once again  $x, y, z$  is the estimated position;  $x_j, y_j, z_j$  is the location of the  $j^{\text{th}}$  base station;  $\Delta R_{jk}$  is the range difference between the user and the  $j^{\text{th}}$  and  $k^{\text{th}}$  base stations; and  $N$  is the number of base stations. To simplify notation, let us drop the arguments and simply refer to the functions in equations 3.25 as  $f_i()$ . The hyperbolic functions  $f_i()$  can be either in the difference-range form of equations 3.16 or the polynomial form of equations 3.17. One would like for  $f_i() = 0$  for  $i = 1, 2, \dots, (N - 1)$  but this cannot be the case since in general no unique solution will exactly satisfy all



the equations. This is due to the measurement errors which will invariable be present. We are left with that fact that for any estimated position each equation will have a residual  $\delta_i$  and thus

$$f_i() - \Delta R_{1(i+1)} = \delta_i \neq 0; \quad i = 1, 2, \dots, (N - 1).$$

The problem then becomes one of finding suitable solutions which are in some manner optimum.

The classic way of attacking the problem is to adopt the criteria of minimizing the sum of the squares of the residuals [80, 81]. The resulting *least squares* function

$$\mathcal{S} = \sum_{i=1}^{(N-1)} [\delta_i]^2 = \sum_{i=1}^{(N-1)} [f_i() - \Delta R_{1(i+1)}]^2 \quad (3.26)$$

is to be minimized. This requires a multidimensional minimization over the solution vector  $x, y, z$ . Often this type of minimization is difficult but an alternative is to find the derivatives of  $\mathcal{S}$  and set them equal to zero. Requiring that

$$\frac{\partial \mathcal{S}}{\partial x} = \frac{\partial \mathcal{S}}{\partial y} = \frac{\partial \mathcal{S}}{\partial z} = 0$$

defines a  $3 \times 3$  system of nonlinear equations. These equations

$$\begin{aligned} \frac{\partial \mathcal{S}}{\partial x} &= \frac{\partial}{\partial x} \left( \sum_{i=1}^{(N-1)} [\delta_i]^2 \right) = 0 \\ \frac{\partial \mathcal{S}}{\partial y} &= \frac{\partial}{\partial y} \left( \sum_{i=1}^{(N-1)} [\delta_i]^2 \right) = 0 \\ \frac{\partial \mathcal{S}}{\partial z} &= \frac{\partial}{\partial z} \left( \sum_{i=1}^{(N-1)} [\delta_i]^2 \right) = 0 \end{aligned} \quad (3.27)$$

are known as the *normal equations*. What is known as “Hobson’s choice” must then be made between two problems: solve the set of three nonlinear normal equations 3.27 or minimize the function 3.26 in three dimensional space. Methods of solving the consistent system of equations numerically are described in Appendix C.

Another classic approach to the inconsistent system solution is to minimize the sum of the residuals rather than their squares. This is known as either the *minimax*, *minimum deviation*, or *least absolute deviation* approach [82, 83]. In this case the function to be minimized becomes

$$\mathcal{S} = \sum_{i=1}^{(N-1)} [\delta_i] = \sum_{i=1}^{(N-1)} [f_i() - \Delta R_{1(i+1)}] \quad (3.28)$$

leading to the following normal equations

$$\begin{aligned} \frac{\partial \mathcal{S}}{\partial x} &= \frac{\partial}{\partial x} \left( \sum_{i=1}^{(N-1)} [\delta_i] \right) = 0 \\ \frac{\partial \mathcal{S}}{\partial y} &= \frac{\partial}{\partial y} \left( \sum_{i=1}^{(N-1)} [\delta_i] \right) = 0 \\ \frac{\partial \mathcal{S}}{\partial z} &= \frac{\partial}{\partial z} \left( \sum_{i=1}^{(N-1)} [\delta_i] \right) = 0 \end{aligned} \quad (3.29)$$

One possible advantage of the least absolute deviation approach is that it is not as easily influenced by a few large measurement errors which can sometimes occur.

At this point the question of whether either the least squares or least absolute deviation approaches are in any sense optimum solutions should be examined. The answer to this questions depends on the definition of an optimum solution and upon the statistical nature of the measurement

errors. One measure of an optimum solution is the maximum likelihood criteria which effectively maximizes the probability that a particular position estimate is correct given the measured range differences and base station locations (that is,  $MAX [p(x, y, z | \Delta R_{11}, \Delta R_{12}, \dots, \Delta R_{1(N-1)})]$ ).

If the range difference errors are uncorrelated and Gaussian distributed with zero mean and equal variances then the least squares solution is the maximum likelihood estimate [84]. If the errors are uncorrelated and exponentially distributed with zero mean and equal variances then the least absolute deviation solution is the maximum likelihood estimate [83]. A more likely scenario is that the errors would be uncorrelated zero mean Gaussian but with unequal variances because the different base station pairs may experience different conditions. In this case the *weighted least squares* solution with weighting coefficients inversely proportional to the respective variances is the maximum likelihood estimate [84]. The function to be minimized becomes

$$\mathcal{S} = \sum_{i=1}^{(N-1)} \frac{1}{\sigma_i^2} [\delta_i]^2 = \sum_{i=1}^{(N-1)} \frac{1}{\sigma_i^2} [f_i() - \Delta R_{1(i+1)}]^2 \quad (3.30)$$

where  $\sigma_i^2$  is the variance for the difference-range errors in the  $(1, i + 1)$  base station pair. This method is also known as chi-squared fitting. The problem in this case is that the individual variances are not known *a priori* and may be difficult to estimate.

In many cases experimental errors may be much greater than the tails of the Gaussian distribution might predict. Errors such as this are often called *outliers*. The least squares techniques are very sensitive to these outlier points as they are actually emphasized in the solution process. The concept of *robust*

*estimators* is one that has developed recently to overcome this limitation. The idea behind this approach is to develop an estimator which is not sensitive to small departures from the idealized distribution for which the estimator is optimized [85]. In this case the term ‘small departures’ denotes fractionally large departures for a small number of data points. This means that a very few points (that is, *outliers*) may exist with a higher probability than the statistical model would predict.

The particular class of robust techniques called *M-estimates* are based on maximum likelihood techniques and can generally be applied to the parameter estimation problem at hand [86]. The objective in this method is to minimize the expression

$$\sum_{i=1}^{(N-1)} \rho(f_i(x, y, z; x_1, y_1, z_1; x_{i+1}, y_{i+1}, z_{i+1}) - \Delta R_{1(i+1)})$$

over three dimensional space. In this equation  $\rho()$  is a function selected to match the particular statistical distribution which the errors are most likely to follow.

The normal equations to be solved in the case of robust M-estimators are

$$\begin{aligned} \sum_{i=1}^{(N-1)} \frac{1}{\sigma_j} \psi(\zeta_i) \frac{\partial}{\partial x} (f_i()) &= 0 \\ \sum_{i=1}^{(N-1)} \frac{1}{\sigma_j} \psi(\zeta_i) \frac{\partial}{\partial y} (f_i()) &= 0 \\ \sum_{i=1}^{(N-1)} \frac{1}{\sigma_j} \psi(\zeta_i) \frac{\partial}{\partial z} (f_i()) &= 0 \end{aligned} \tag{3.31}$$

where  $\psi(\zeta_i)$  is a weighting function which is specialized to the particular distribution for which the solution is optimized. The function  $\psi(\zeta_i)$  is found

from the relation  $\psi(\zeta_i) \equiv \frac{d\rho(\zeta_i)}{d\zeta_i}$ . The parameter  $\zeta_i$  is functionally dependent on the base station locations and the measured range difference but not on the user's unknown coordinates  $x, y, z$ . In many cases the weighting function  $\psi()$  can be chosen in such a way as to minimize the influence of outlier points. Unfortunately, the tradeoff is an increased computational burden. The normal equations 3.31 are often tricky to solve because of the presence of discontinuities [85, 87]. Depending on the choice of  $\rho()$ , the convergence of the position estimate can be a strong function of the initial estimate [83]. The theoretical derivation of robust estimators from maximum likelihood theory is presented in Appendix B. Methods of solving the resulting system of equations numerically are discussed in Appendix C.

### 3.3.4 Time Delay Measurement

The time of arrival of signals transmitted between the user and the base stations can be used as a way to measure the range. If a pulse is transmitted between the user and the base station then the range can be determined from the relationship  $R = cT_R$  where  $R$  is the range,  $c$  is the speed of light, and  $T_R$  is the propagation time. In general, this requires the use of pulses transmissions with narrow pulse widths for good resolution, resulting in large bandwidths. There are practical as well as regulatory limitations placed on the amount of spectrum available to such a system. The challenge then becomes one of obtaining the best possible temporal resolution within the constraint of limited bandwidths.

The most common method of measuring the time of arrival for a pulse is

to note when the rising or falling edge crosses some threshold. Noise is one of the limiting factors in the threshold measurement because it perturbs the pulse shape and distorts the threshold crossings. For the high signal-to-noise ratio case with rectangular pulses Skolnik gives an equation for the RMS timing error ( $\delta T_R$ ) in such an operation as

$$\delta T_R = \frac{t_r}{\sqrt{2\frac{S}{N}}} \quad (3.32)$$

where  $t_r$  is the pulse rise time and  $S/N$  is the signal-to-noise power ratio [48]. Assuming the pulse has been bandlimited by a receiving filter of bandwidth  $B$  then  $t_r \approx 1/B$ . Substituting  $S = E/\tau$  and  $N = N_0 B$  then

$$\delta T_R = \sqrt{\frac{\tau}{2BE/N_0}} \quad (3.33)$$

where  $E$  is the signal energy,  $N_0$  is the noise power per unit bandwidth, and  $\tau$  is the pulse width.

These relationships places a limit on the resolution which can be obtained for a specified signal bandwidth; for example, to resolve  $R$  at the centimeter level assuming that  $S/N = 20$  dB then a rise time ( $\tau$ ) of  $5 \times 10^{-10}$  seconds and a bandwidth ( $B$ ) of approximately 2 GHz is required. For a higher power ground based system, a  $S/N = 50$  dB might be feasible to achieve. This would result in a rise time requirement of  $5 \times 10^{-7}$  seconds and a bandwidth of only about 2 MHz. If the link budget can guarantee the necessary  $S/N$  ratio, then such a system is a candidate for high accuracy positioning applications.

Spread spectrum techniques may be used to improve the time resolution of time delay measurements to better than one pulse width without the high transmit power otherwise required. This increase in resolution allows spread

spectrum time delay measurement systems to be feasible for centimeter level accuracy position location system. The increase in resolution for spread spectrum systems is due to the additional information added to the signal during the spreading process. An autocorrelation operation can be performed to detect the time of arrival of the pulse to better than a tenth of a pulse width.

The particular method of spreading the pulse has an impact on the resolution which can be achieved. The most popular methods of spread spectrum involve phase-shift-keyed (PSK) for frequency-shift-keyed (FSK) coded sequences. The properties of the code which is to be transmitted is an important consideration. The tradeoffs involved will be presented in Chapter 6 which gives details about a spread spectrum time measurement system and computer simulations of performance.

### **3.3.5 Phase Delay Measurement**

Phase systems measure the phase of a signal transmitted between the user's location and the base stations at known locations. The phase of the received signal is a function of the distance traveled by the wave and therefore range information can be determined. The phase information is measured modulo  $2\pi$  due to its cyclic nature, resulting in ambiguities in determining the integer number of cycles. These ambiguities can be resolved by measuring at low frequencies where wavelengths are large enough to ensure less than one phase cycle over the measurement area. Generally, a combination of low frequencies for determination of ambiguities and high frequencies for good resolution are

utilized.

The phase measurement involved in this type of system is inherently a narrowband operation. This is an advantage in the signal-to-noise ratio area because very narrow bandwidths can be used in the receiver. Even if frequency hopping techniques are used the dwell time will be long enough for relatively narrowband filtering.

The phase of the received signal can be related to the distance traveled by noting that  $\phi = 2\pi(R/\lambda)$ , where  $\phi$  is the received phase,  $\lambda$  is the wavelength, and  $R$  is the distance traveled by the wave. The phase can only be measured modulo  $2\pi$  and thus the integer multiple must be determined by other means. The complete phase  $\phi'$  is the fractional part plus the integer multiple ( $\phi' = \phi + K360^\circ$ ). The integer number of cycles ( $K$ ) is unknown causing the measured phase readings to be ambiguous because they repeat periodically; each  $360^\circ$  cycle of phase is called a 'lane'. This is illustrated in Figure 3.11 where the hyperbolic lines of constant phase difference are shown for one base station pair. In the figure one lane exists between the two hyperbolas labeled  $H_1$  and  $H_2$ . The lane widths are one-half wavelength along the baseline between the base stations ( $A, B$ ) and are magnified by the expansion factor (see Section 3.3.2) away from the baselines. To determine the beacon's position uniquely, methods of resolving these ambiguities must be employed [88].

There are three methods in the literature for resolving the ambiguities which are present in a phase measurement system. The first, and probably the simplest technique, is to start at a known location and to track the phase variations as the vehicle moves through the phase lanes. This technique,



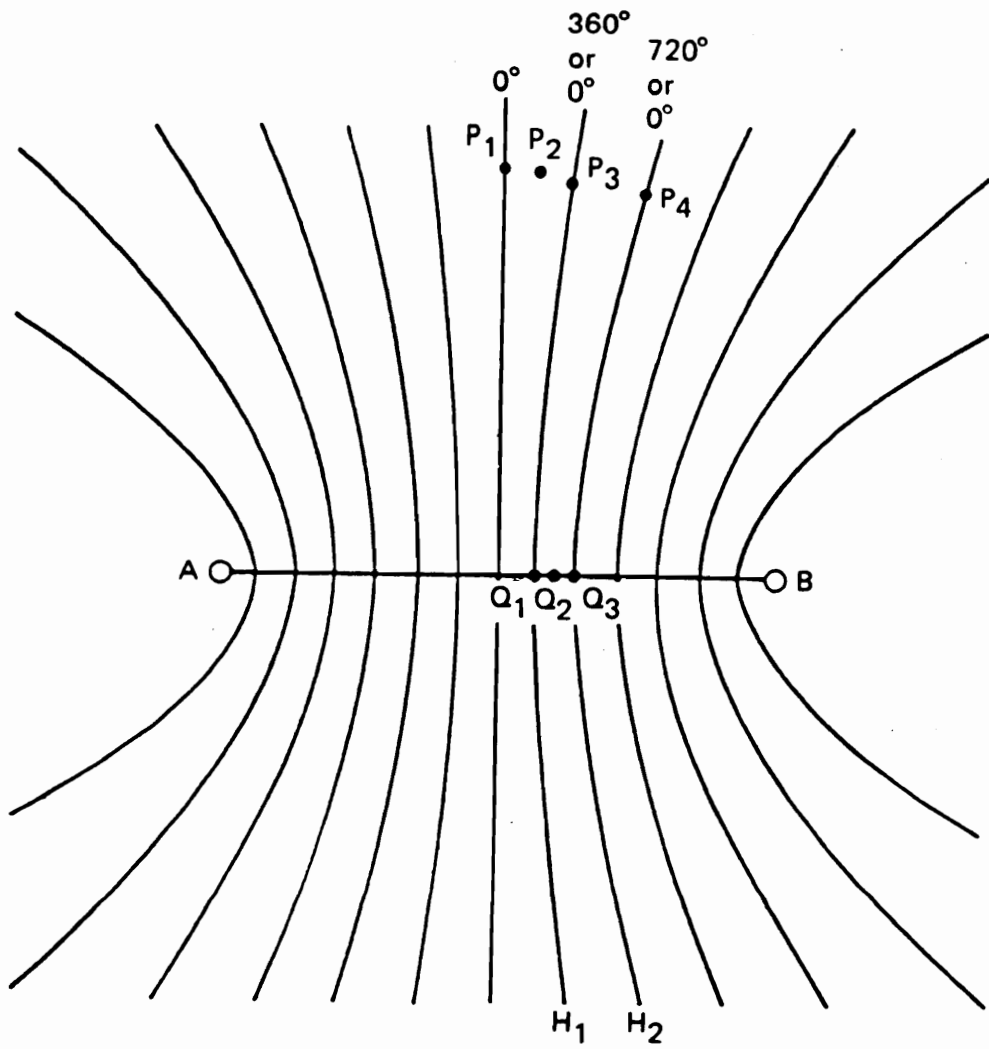


Figure 3.11: Hyperbolic lines of constant phase difference for one base station pair (from Sonnenberg [16])

known as lane counting, is used for the Decca Navigator system and several other low frequency ocean navigation systems. For a low frequency type system the lanes are quite large with respect to the speed of the moving vehicle, making it easy to track the relatively slowly changing phase. Many of the older navigation systems actually used mechanical counter wheels to track the number of phase cycles from a known starting point.

For a high resolution position location system the frequency of operation must be significantly higher than those used in oceanic navigation systems and consequently the lanes are smaller in size. Also many of the vehicles whose position must be determined on a local area site will be moving with velocities much greater than those encountered in ocean navigation. The land based vehicles will also have a much greater range of vehicle dynamics which may make determining the lane transitions difficult. In view of these factors, lane counting is not a reliable technique for ambiguity resolution for high frequency position location systems.

The second method of ambiguity resolution is to measure ambiguity digits based on discrete frequencies [88]. The principle here is very similar to the method of resolving the angle ambiguities in interferometers by using different baseline spacings as discussed in Section 3.2. To resolve the phase ambiguities multiple frequencies are used.

The idea can be illustrated with a simple example using a low frequency tone to modulate a high frequency carrier and measuring the phase of the modulation and the carrier. In this manner the low frequency modulation, which has a longer wavelength than the carrier, can be used to determine the position with a correspondingly large unambiguous area. This is shown

graphically in Figure 3.12 for a single pair of receivers with a  $5\lambda$  spacing between them. The curves shown in Figure 3.12a are the hyperbolic lanes for the higher frequency, while Figure 3.12b is for the low frequency. The integer number of phase cycles ( $K$ ) can be determined for the high frequency carrier based on the measurements of the low frequency modulation tone.

In general if high resolution is required over large areas, then multiple tones must be transmitted to resolve the ambiguities at all levels. The relationship between the frequencies can be binary or some other sequential spacing. There is no need for the multiple frequencies to be phase coherent. This means a frequency hopping scheme can be used with the spacing between frequencies selected to resolve the ambiguities at each level of resolution.

By measuring the slope of the phase versus frequency curve, a third method of resolving ambiguities can be developed. Figure 3.13 illustrates the phase versus frequency curve and its relationship to the distance between the transmitter and receiver. From the figure, the slope of the phase-frequency curve is  $d\phi/df = \pi R/\lambda$ . If enough points are measured to accurately calculate the slope then an unambiguous estimate of position can be obtained. This technique lends itself to use in a frequency hopping type spread spectrum system where phase measurements are continuously made at a number of frequencies. The slope information from the most recent measurements can be used to resolve the ambiguity in phase. The frequency spacing between phase samples must be small enough to assure at least two samples per cycle of phase, otherwise aliasing will occur. The frequency spacing must be selected so that the sampling criteria is met at the largest range ( $R$ ) for

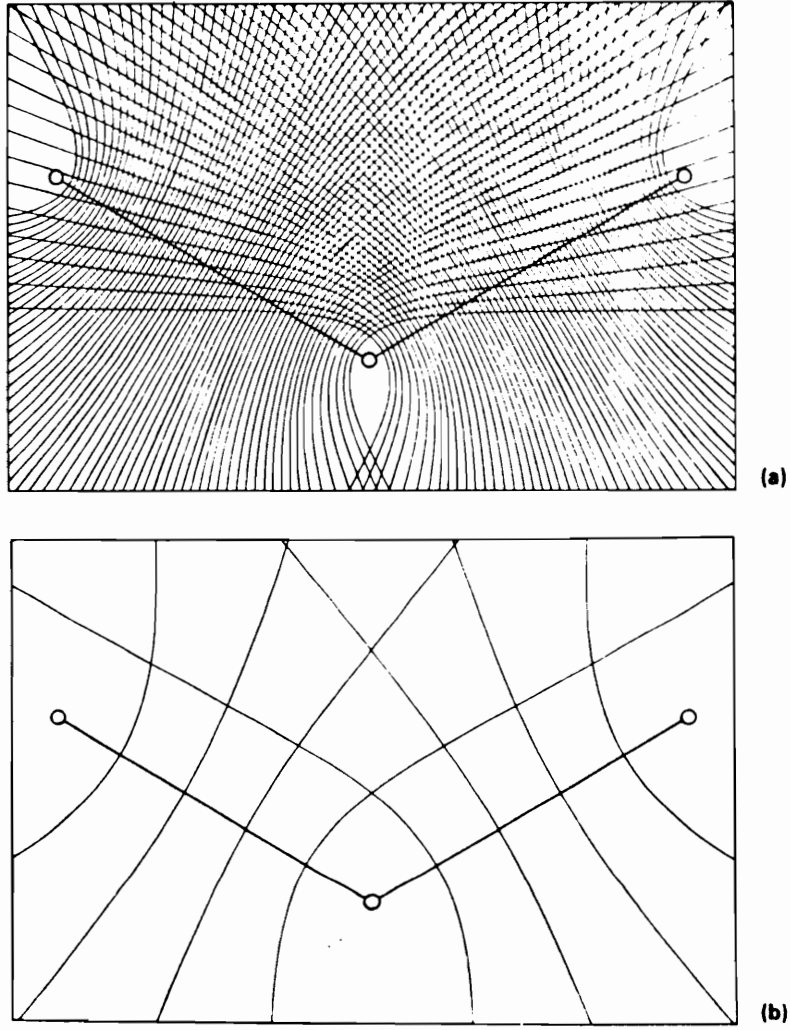


Figure 3.12: Determination of ambiguity digits in two dimensions using multiple frequencies (from Sonnenberg [16])

which the system is to operate.

### 3.3.6 Frequency Shift Measurement

Systems which measure the change in frequency of a signal transmitted between the base station and the user are called Doppler systems. If the base stations are in motion, as would be the case if they were non-geosynchronous satellites, but their locations are continuously known then a frequency shift can occur as the signal travels. This frequency shift ( $\Delta f$ ), due to the Doppler effect, is a function of the velocity with which the base station is approaching to or withdrawing from the user ( $d\rho/dt$ ) thus

$$\Delta f = -\frac{f_T}{c} \frac{d\rho}{dt} \quad (3.34)$$

where  $f_T$  is the transmitted frequency,  $c$  is the speed of light, and  $\rho$  is the range between the user and the base station. The actual frequency measured by the receiver is  $f = f_t + \Delta f$ . By measuring the change in frequency from  $f_T$ , the rate of change of range between the user and the base station can be determined. If the trajectory of the base station is known then the user's location can be uniquely defined from the received Doppler frequency changes. The fact that the rate of change of range is measured, and not the range itself, means that Doppler systems are inherently hyperbolic type } systems.

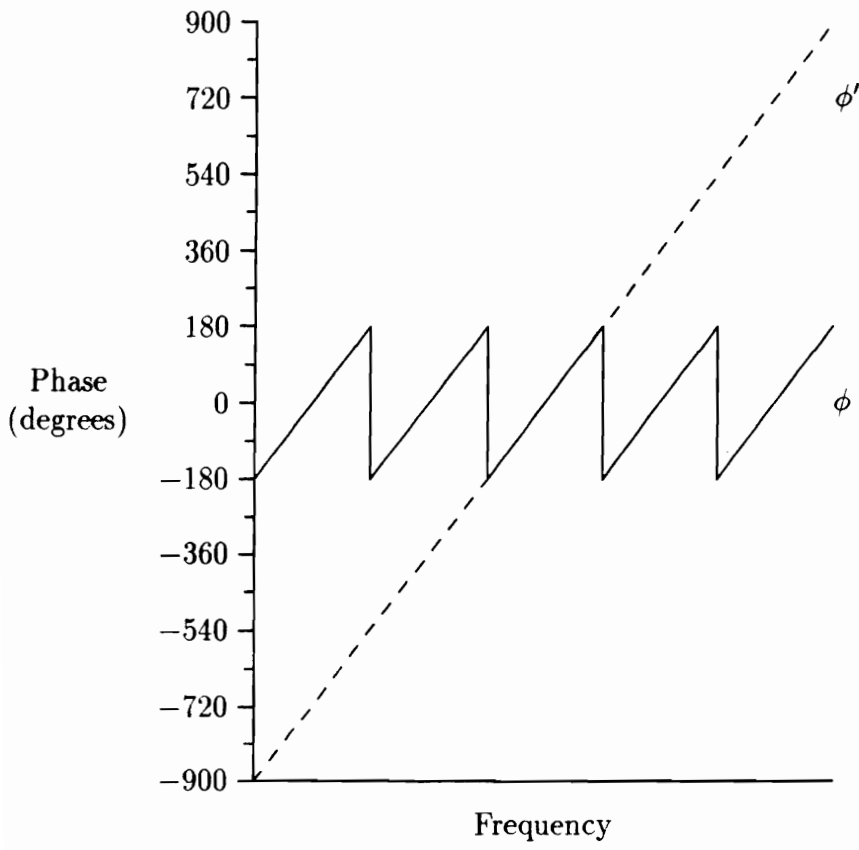


Figure 3.13: Ambiguity resolution using slope of the phase versus frequency curve

## Chapter 4

# Phase System Theory and Simulation

The basic aim of this research was to investigate the technologies which would be most capable of meeting the requirements of the construction automation program, as outlined in Chapter 1. The desire was to begin by testing the simplest approaches which appeared to be theoretically capable of meeting these requirements. The research could then proceed to more complex implementations if theoretical analysis, simulation, or experimental results indicated it was necessary. In this manner, the simplest and most cost effective system which would attain the desired results could be achieved, without effort being spent investigating unduly complex systems.

With this goal in mind, the entire range of options discussed in Chapter 2 were considered. Multilateration radio frequency methods were chosen for ease of implementation, all weather capability, and ability to simultaneously

track multiple vehicles. Ground based systems were found to be able to overcome several of the drawbacks of satellite systems when used in local area applications; namely, more precise knowledge of base station locations, less chance of signal blockages, lack of most ionospheric and tropospheric errors, lower path loss, and constant base station geometry. The currently existing radio position location systems were found to fall short of meeting the requirements for construction automation. For this reason, new approaches of performing radio frequency multilateration were proposed and considered.

The proposed systems were analyzed theoretically, with computer simulation techniques, and in a series of experimental tests. The prototypes of the proposed systems were first tested as a proof of concept at a small area site that was 4.2 m by 4.2 m with no altitude variation. The final experimental tests were conducted at a site approximately the size of a construction project: a rectangular farm field about 110 m by 70 m with about 5 m of altitude variation across the site. So that all results could be compared directly, the analysis and simulations were performed assuming the same geometric configuration as the full scale experimental test site. Unless stated otherwise, all comments in the discussion which follows pertain to the full scale test site.

## **4.1 System Description**

The first decision which had to be made was about the structure of the proposed multilateration system. As mentioned in Chapter 1, one goal was to connect the position location system to a CAD database for the construction site. For this reason the actual position information was needed at the off-site



data processing point rather than at the user's location. It then made sense to perform all the analysis and position computation off-site, instead of the more traditional approach of doing the calculation at the user's location. Another goal was to keep the user's equipment as simple, rugged, and inexpensive as possible because of the harsh environment on the construction site. The most logical method of meeting these objectives was to place a beacon transmitter at the point to be located and to use receivers on the periphery of the site to measure the signal and to calculate estimated position. In this way, all the signal processing and position calculation would be done at a central location where the CAD database could be maintained.

The type of measurement which was to be made on the received signal then had to be selected. As discussed in Chapter 3, the receivers could use time delay, phase delay, or frequency shift information to calculate ranges to the user's location. Since the base stations are stationary for a ground based multilateration system, the use of Doppler frequency shift was ruled out as a possible alternative. Swept frequency chirp approaches could be used but the problems of obtaining the appropriate sweep rates with sufficient accuracy were found to be considerable. Time delay measurement was a possible candidate but if conventional pulse transmission was used then large signal-to-noise ratios (SNR) or wide bandwidths would be required to achieve the necessary resolution. This would require relatively high transmit power to assure adequate SNR at the maximum operating range. Spread spectrum techniques could have been used to overcome the requirement for high transmit power at the expense of a more complex transmitter and correlation receiver. Phase delay systems were theoretically capable of high accuracy

assuming the problem of resolving ambiguities could be solved adequately. Phase systems were found to be much simpler to implement than either conventional pulse or spread spectrum time delay systems. In addition, phase delay measurement systems, due to their essentially continuous wave nature, allowed extremely narrow measurement bandwidths and consequently high SNR's. For these reasons, the first system to be considered was one using phase delay measurements.

Once the commitment to phase delay techniques was made, the choice between the various ambiguity resolution methods had to be considered. Lane counting was seen as useful for testing purposes but unreliable for use in an actual system operating in the frequency ranges proposed for the position location system. Frequency differencing and the slope method of resolving ambiguities, both of which involve the measurement of phase delay at a number of frequencies, were found to be more appropriate than lane counting. In the frequency differencing method the measured frequencies are selected to have specific relationships to resolve the different levels of ambiguities while in the slope method the frequencies are generally uniformly spaced. In the limit as the number of sampled frequencies increases the two methods become indistinguishable.

The sampled frequency approach to ambiguity resolution lends itself to use in a frequency hopping mode. This results in a more spectrally efficient scheme because many users can share the same frequency band [89]. Another advantage of the frequency hopping method is the inherent anti-jam capability which it possesses; any jammers or continuous wave interferers will have little effect on system performance due to the small dwell time on any

one frequency. The frequency may be randomly chosen based on a pseudo-random code generator. At the receiver the incoming signal is mixed with a locally generated replica which is offset in frequency to create an intermediate frequency (IF). A continuous wave signal present at the receiver's input would then be spread at the IF by the same amount as the local reference. A signal with the same bandwidth as the local reference but which is asynchronous would be spread to twice the bandwidth at IF. Because the IF filter bandwidth is a small fraction of the bandwidth of the spreading, the receiver can reject most continuous wave or non-synchronous wideband signals.

The type of synchronization employed in the system depends on whether the ranging or hyperbolic method is used. As discussed previously ranging systems, when compared to equivalent hyperbolic systems, provide a larger measurement area of a specific accuracy. This larger area comes at the expense of the more sophisticated synchronization that is required for ranging systems. Theoretically, ranging allows one less base station to be used but in practice the numbers of base stations are equal for the two methods due to inevitable timing biases. For a phase measurement system which implements the ranging method, either a synchronization link is required or highly stable system clocks must be employed. The other alternative, which is covered by a United States Patent, is to use transponders with known time delays. None of the previously mentioned synchronization approaches seemed feasible for the construction automation application and therefore hyperbolic methods were selected. Hyperbolic techniques require no special synchronization because only phase differences are measured and not absolute phases. The geometric inferiority which hyperbolic methods suffer compared to ranging

methods can be overcome by proper placement of the base stations; this can be accomplished by using longer baselines for a given measurement area.

The key features of the proposed multilateration system can now be summarized. The user places a beacon transmitter at the position whose coordinates are to be estimated. Phase measurements are made in a differential mode to arrive at phase-differences for each pair of base station receivers. The transmitter continuously hops in frequency and the receivers synchronously follow the hopping. The slope of a number of phase versus frequency readings are used to develop an ambiguity resolution estimate to determine the integer number of phase cycles. The whole phases, consisting of the integer and fractional parts, are then converted into range differences using the speed of propagation. The hyperbolic equations are then solved using the unambiguous differential ranges to provide a position estimate.

Once the basic design was selected, several details about the implementation of the system had to be considered. The operating frequency and bandwidth had to be chosen so that the system could be tested experimentally. Some consideration also had to be given to the possible licensing of the system should it go beyond the experimental prototype stage. The general configuration of the base station receivers had to be selected to provide good user geometries and at the same time be physically realizable.

A number of factors went into the selection of an operating frequency band for the experimental prototype system. Assuming that phase can be measured to a fixed resolution then higher operating frequencies result in higher resolution position estimates. There is thus a lower limit on the operating frequency which can be expected to achieve any specific range difference

resolution.

The relationship between the measured phase resolution ( $\phi$ ) in degrees and the resolution of the range difference estimate ( $\Delta R$ ) is

$$\Delta R = \frac{\phi}{180} \chi \frac{\lambda}{2} \quad (4.1)$$

where  $\chi$  is the lane expansion factor as discussed in Chapter 3 and  $\lambda$  is the wavelength. Let us for the moment neglect any channel induced errors such as multipath and ducting; errors in the measurement system would then determine the maximum achievable phase resolution. It is reasonable to assume that phase can be measured to a resolution of about  $5^\circ$  taking into account typical drift and calibration errors in the measurement system. For typical user-base station geometries an expansion factor of  $\chi = 2$  is reasonable.

Using the available surveying techniques (either laser electronic distance measuring stations or optical theodolites), the base station locations can be known to approximately the centimeter level [47]. The centimeter level is therefore the best range difference resolution that the proposed system could realistically hope to achieve. If range difference estimates at the centimeter level ( $\Delta R = 0.01 \text{ m}$ ) are required then from equation 4.1 frequencies of greater than 417 MHz ( $\lambda \leq 0.72 \text{ m}$ ) must be used.

Before selecting an operating frequency a study of the existing literature on propagation measurements in various bands was made. Frequencies of less than 3 GHz were targeted due to the increased cost of commercial hardware at the higher microwave frequencies. The question was raised early in the research as to whether operating at higher frequencies in any way reduced

the multipath error problems. A number of studies have been performed in the 450 MHz band [90, 91], the 800-900 MHz band [91, 92, 93], the 1300 MHz band [94], and the 1500 MHz band [95]. This is a representative but by no means a comprehensive listing of references on propagation studies. It is often difficult to compare the findings of different researchers for the various frequency bands due to the wide variety of measurement systems and techniques. No general trend in the severity of multipath interference versus frequency could be inferred from these single frequency band experiments.

Several studies have been done by individual researchers comparing propagation at different frequencies using the same measurement locations and similar measurement apparatus at each frequency. From these studies it is possible to gain insight into the statistical behavior of multipath as a function of frequency. Patsiokas and others have compared propagation at 150 MHz, 450 MHz, and 850 MHz in multistory office buildings and found that multipath severity increased with frequency [96]. Turin and others compared 488 MHz, 1280 MHz, and 2920 MHz in urban channels and found the statistical results for the three frequencies to be very similar [97]. Turin and others also performed a phase-ranging simulation for an urban vehicle monitoring system at 488 MHz and 1280 MHz and found no substantial difference in the performance of the system at these two frequencies [98]. In addition, LaBel [99] has found no direct relationship between multipath intensity and frequency based on outdoor measurements at 145 MHz, 434 MHz, and 862 MHz. Based on these research findings, there is no indication that using higher frequencies would provide any reduction in multipath severity for a phase measurement position location system.

One possible advantage of the using higher frequencies is the more pronounced Brewster angle effect for the vertical polarization Fresnel reflection coefficient. For both vertical and horizontal polarization at grazing angles approaching  $90^\circ$  the magnitude of the reflection coefficient ( $\rho$ ) is lower as the frequency is increased. These two effects are apparent in the Figure 4.1 which plots the magnitude ( $\rho$ ) and phase ( $\phi$ ) of the reflection coefficients versus grazing angle ( $\psi$ ) for typical ground conditions. The smaller reflection coefficients may make some types of multipath (for example, ground reflection) less severe at higher frequencies for some grazing angles. This effect does not appear to be evident in any of the experimental propagation studies which were previously discussed.

A disadvantage of moving to higher frequencies is the more stringent requirements placed on the ambiguity resolution processing. As the operating frequency increases the ambiguity lane width decreases linearly, making it more difficult to determine the integer number of phase cycles reliably. For the slope method of ambiguity resolution a more accurate estimate of  $d\phi/df$  would be necessary. If we require that the range difference error for the ambiguity be less than one-half of the lane width then the slope error,  $(d\phi/df)_e$ , must satisfy the relationship

$$\left(\frac{d\phi}{df}\right)_e \leq \frac{2\pi\lambda_t\chi}{4c} \quad (4.2)$$

where  $\lambda_t$  is the wavelength of the transmitted signal,  $\chi$  is the lane expansion factor, and  $c$  is the speed of light. Therefore, the required slope error is inversely proportional to the transmit frequency. In general, to get a more accurate slope estimate more frequency samples over a larger bandwidth

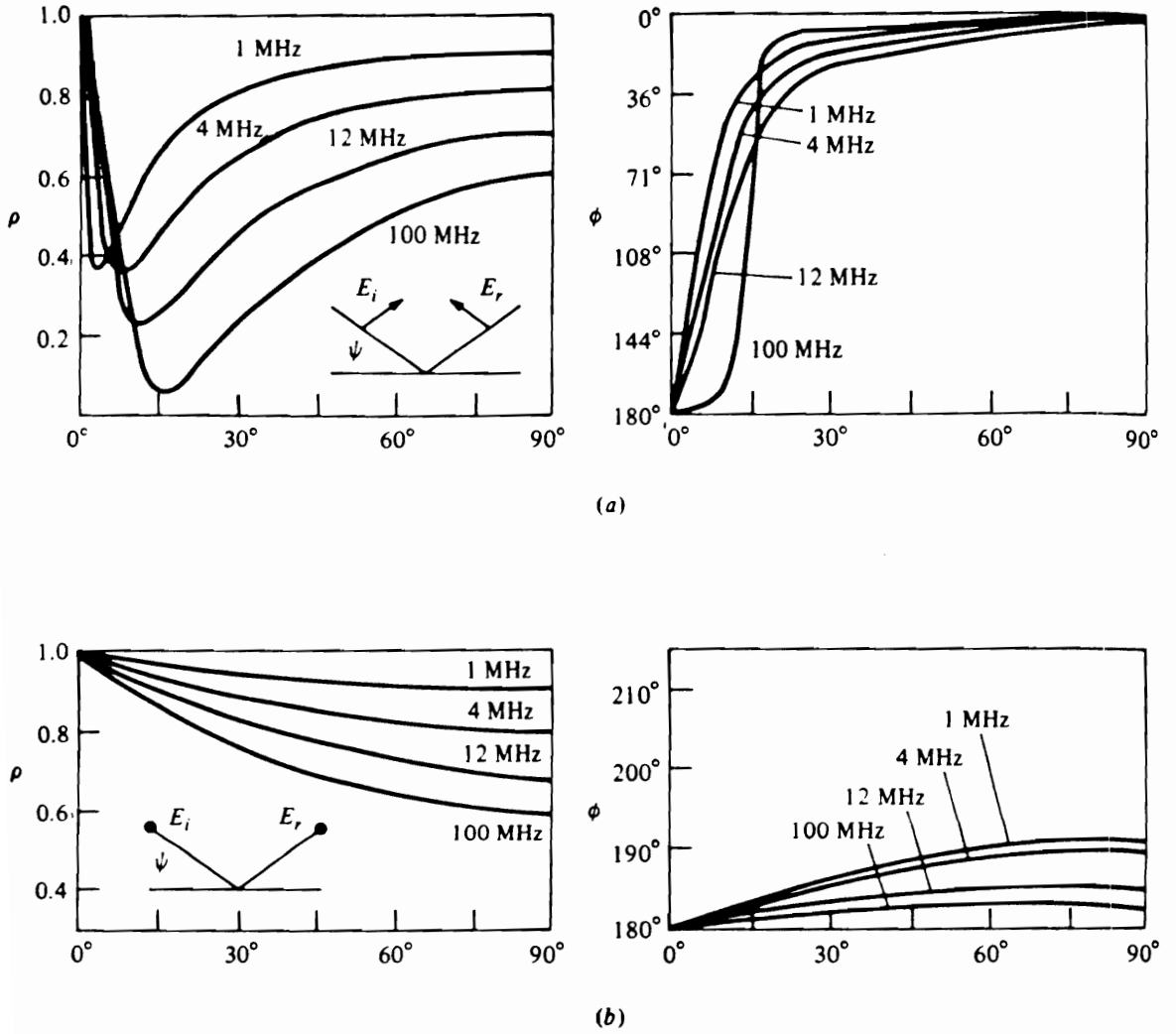


Figure 4.1: Reflection coefficients as a function of grazing angle for (a) vertical polarization and (b) horizontal polarization (from Collin [100])



would be required. As the operating frequency is increased at some point it will no longer be possible to determine the slope with sufficient accuracy to resolve the lane ambiguity. The point at which insufficient accuracy for ambiguity resolution occurs will depend on the measurement system errors and channel induced errors.

From the previous calculation using equation 4.1, 417 MHz represents the lowest frequency which can achieve centimeter level position estimates based on the assumption of no channel induced errors. The 420-450 MHz UHF band is allocated for radiolocation in the United States and was therefore a viable choice for the experimental operating frequency band. Commercial receivers were available at a reasonable cost and some hardware to operate in this band was already on hand. Since the propagation studies in the literature indicated no particular advantage to using higher frequencies, the 420-450 MHz UHF band was chosen for the experimental prototype test system.

The highly modified commercial receivers which were acquired for the prototype system were able to cover from about 390 MHz to 455 MHz. At the band edges there was excessive rolloff of the receiver response due to a fixed tuned front end filter centered at 425 MHz. For this reason the system operating bandwidth was limited to approximately 30 MHz from 410 MHz to 440 MHz.

The prototype system used a frequency hopping spread spectrum approach to cover the entire 410-440 MHz band. The spacing between the hopping frequencies was chosen to assure that at least two samples would be achieved for each cycle of phase versus frequency at the maximum range dif-

ference. If a smaller frequency sampling increment were used aliasing would result and the slope could not be determined. The slope of the phase versus frequency curve is  $d\phi/df = 2\pi D/c$ , where  $\phi$  is phase,  $f$  is frequency,  $D$  is the range difference, and  $c$  is the speed of light. Solving for  $df$  and enforcing the minimum sampling criteria the resulting requirement for the frequency spacing is  $df \leq c/(2D)$ . For the experimental test site the maximum range difference ( $D$ ) is about 125 m resulting in a maximum frequency spacing of 1.2 MHz. To assure an adequate safety margin the frequency hopping step was chosen to be 1.005 MHz. The extra 5 kHz portion of the frequency step size was required because of the fixed divide ratios in the phase locked loop synthesizer used to derive the signals. The result was 30 possible hopping frequencies between 410 MHz and 440.015 MHz.

The full scale experimental testing was performed at a rectangular site approximately 110 m by 70 m. By analyzing the error contours for hyperbolic multilateration systems, as discussed in Chapter 3, it was determined that a baseline orientation of  $90^\circ$  optimized the user-base station geometry within the measurement area. For this reason, base station towers were located at the corners of the rectangular site. Two antennas were placed on each tower, one at a height of 3 m and the other at 10 m. A physical layout of the measurement area and the system components is shown in Figure 4.2.

## 4.2 Computer Simulation

A computer simulation program has been developed as a tool to help determine the proposed system's error performance under a variety of possible

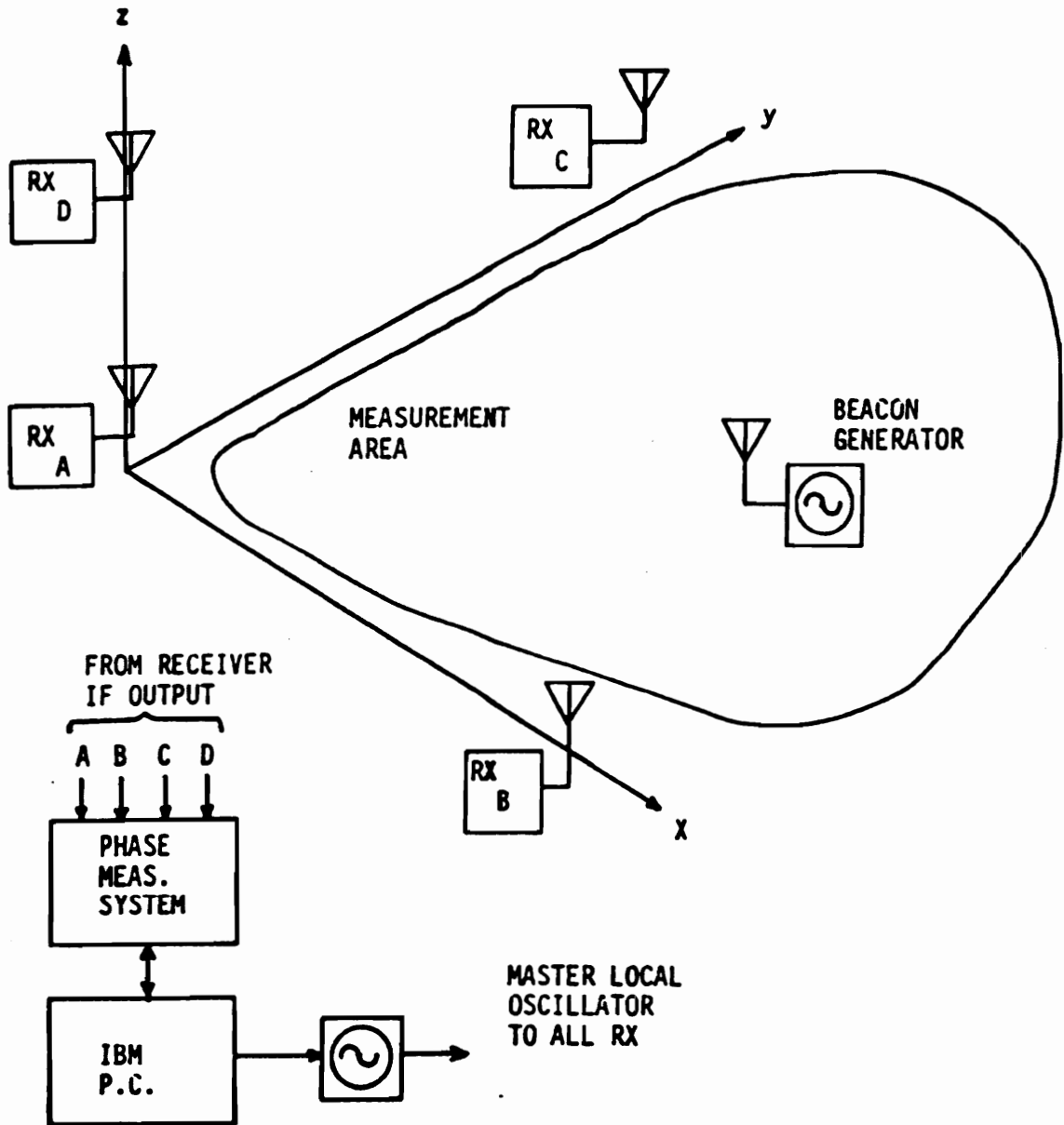


Figure 4.2: Physical layout of measurement area and position location system (from Purdy [101])

operating conditions. Techniques of improving the system's accuracy can be analyzed without the costly and time consuming effort required for hardware implementation and experimental testing. The simulation is statistical in nature and therefore is not meant to predict system performance at any one specific site. The simulation gives an estimate of the average system performance at a large ensemble of potential sites.

The simulation technique is based on Monte Carlo analysis [102] and is similar in principle to the phase ranging simulation discussed by Turin [98]. Errors in both phase and position are analyzed at randomly distributed locations within the measurement area. The measurement system errors as well as the effect of the radio channel are modeled in the simulation.

The position error  $\delta_i$  is evaluated at a sequence of points  $p_i$  where the iteration number  $i$  ranges from 1 up to the maximum number of points in the simulation  $P$ . The position error  $\delta_i$  is simply the Euclidean distance between the actual position  $p_i$  and the estimated position  $e_i$  for each point. At each step of the simulation the random point  $p_i$  is selected as  $p_i = (x_i, y_i, z_i)$  where  $x_i, y_i, z_i$  are uniformly distributed random variables within the measurement volume of the system.

A block diagram illustrating the main components of the simulation program is shown in Figure 4.3. At the beginning of the program all variables are initialized and a random beacon position which is within the measurement volume is selected. The direct path distance between the beacon position and each of the base station receivers is calculated geometrically. The path loss of the arriving direct line-of-sight electric field component is calculated based on a  $1/R$  spherical spreading loss for the far field. The phase of the

received direct path signal is calculated based on the free space wavelength of the transmitted signal and the path length.

The simulation uses a channel model to introduce the effects of multipath interference due to scatterers in and around the measurement area. The most complicated multipath environment in which the system is expected to operate is a congested factory. Research has been performed on multipath modeling in UHF factory channels by Rappaport [103] and he has suggested that a hybrid geometric-statistical approach may be the most appropriate method of modeling such a channel [94]. In this type of model large scattering objects are represented geometrically using ray tracing and random multipaths are simulated using a statistical model that is generally based on empirical data.

The computer simulation uses a ray tracing approach to model the geometric reflections from the ground and any other large reflectors whose dimensions exceed a wavelength. The complex reflection coefficient of the ground is calculated based on an assumed conductivity and permittivity [100] as

$$\rho e^{j\phi} = \frac{(\epsilon_r - j\sigma/(\omega\epsilon_0)) \sin \psi - \sqrt{(\epsilon_r - j\sigma/(\omega\epsilon_0)) - \cos^2 \psi}}{(\epsilon_r - j\sigma/(\omega\epsilon_0)) \sin \psi + \sqrt{(\epsilon_r - j\sigma/(\omega\epsilon_0)) - \cos^2 \psi}} \quad (4.3)$$

for vertical polarization and

$$\rho e^{j\phi} = \frac{\sin \psi - \sqrt{(\epsilon_r - j\sigma/(\omega\epsilon_0)) - \cos^2 \psi}}{\sin \psi + \sqrt{(\epsilon_r - j\sigma/(\omega\epsilon_0)) - \cos^2 \psi}} \quad (4.4)$$

for horizontal polarization, where  $\psi$  is the grazing angle of incidence, and the permittivity  $\epsilon$  is the product of the relative permittivity  $\epsilon_r$  and the permittivity of free space  $\epsilon_0 = 10^{-9}/36\pi$  farads/meter.

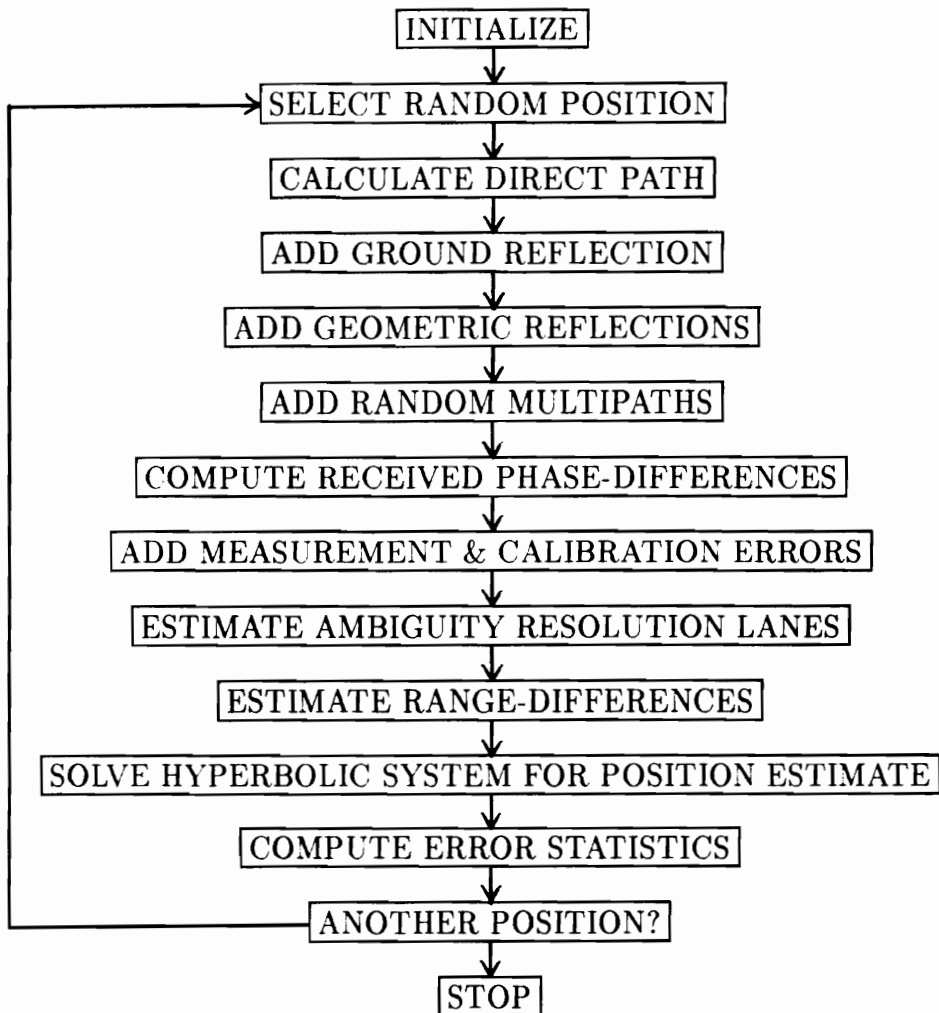


Figure 4.3: Block diagram of simulation program main routines

The ground at the measurement site is assumed to be planar with constitutive parameters typical for dry soil. The reflection coefficient of the ground is calculated, assuming a conductivity ( $\sigma$ ) of 0.01 Semens/meter and a relative permittivity ( $\varepsilon_r$ ) of 15. The values of permittivity and conductivity were obtained from Table 4.1 as typical values for soil with high moisture content [104]. The path length for the ground reflection is determined using ray tracing from the geometry between the transmit and receive antennas and the ground plane. Any additional large reflectors are modeled in a similar manner to that just discussed for the ground reflection.

The random scatterers in the channel are simulated using a statistical model similar to that developed by Turin [97] for urban multipath propagation. The model was later employed by Hashemi [105] for urban channels and Rappaport [94] for indoor channels. The general form of the channel model is that of a linear filter with a complex valued impulse response

$$h(t) = \sum_{j=0}^N \alpha_j \delta(t - t_j) e^{j\theta_j} \quad (4.5)$$

where  $\alpha_j$  is a real attenuation factor,  $t_j$  is the time delay and  $\theta_j$  is the phase delay for the  $j^{\text{th}}$  path. The path number  $j$  ranges from 0, which represents the direct line-of-sight path, to the maximum number of paths  $N$ . The phase shift  $\theta_j$  includes the linear phase shift due to propagation delay and also any phase shifts caused by the reflection coefficients of the scatterers. Complete details of the statistical aspects of the channel model are given in Section 4.2.1 on multipath channel modeling.

The simulation program generates the random multipath components according to the model described in equation 4.5. Phasor addition is used to

sum all the arriving signal components at the receive antenna. The direct line-of-sight path along with all geometric and random multipaths combine to create a single sinusoid from which the phase information is extracted. The phase-difference information from pairs of receivers is found by subtracting the individual received phases, which mimics the operation of the phase detector in the hardware system.

Once the phase-differences have been calculated for the receiver pairs, the measurement and calibration errors are included as additive noise. Errors in the phase measurement system have been modeled as zero mean Gaussian with a variance of  $4^\circ$ . The phase error distribution of the measurement system was inferred from laboratory testing done with the phase detector subsystem. The calibration errors of the phase measurement system have been included as uniformly distributed in the range  $\pm 15^\circ$ .

At this point the simulation program has now generated phase data for each receiver pair as a function of frequency. The process of ambiguity resolution can then be performed on these phase data. Figure 4.4 is an expanded block diagram of the ambiguity resolution process which takes place in the simulator.

The first step in the ambiguity resolution is to median filter the phase versus frequency data. Median filtering is particularly effective for smoothing these phase data because points which differ from the median are removed while the slope information is retained. Median filtering is a commonly used nonlinear filtering technique for data whose error distribution is broad [83]. A complete description of the median filtering technique and its implementation are given in Appendix A. Briefly, at each frequency a median window is



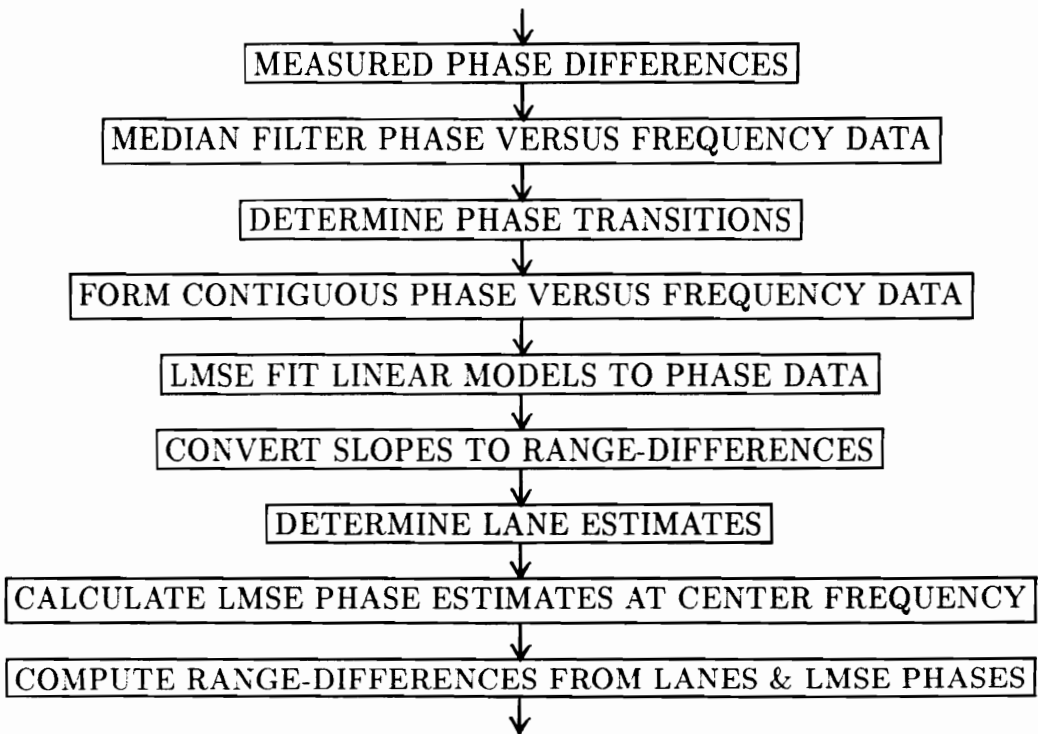


Figure 4.4: Block diagram of simulation program ambiguity resolution routines

created which includes the  $M$  frequencies on either side. A special median function which takes into account the cyclic nature of the phase information is used in the filter. This prevents incorrect filter outputs near the frequencies where the phase is in transition from one cycle to the next. The median value of these  $2M + 1$  points is then selected as the output of the filter.

For the 30 frequencies used in the simulation and experimental work, a value of  $M = 2$  has been found to be optimum for data smoothing and enhancing the phase transitions without loss of slope information. Figure 4.5 shows typical experimental phase data before and after being passed through the  $M = 2$  median filter. The filter removes wild points and sharpens the phase transitions while maintaining the slope information.

To determine the slope of the phase versus frequency the data must be made contiguous (that is, the cyclic phase transitions must be removed). The filtered data are in the interval from  $-\pi$  to  $\pi$ . We can therefore look for phase transitions by examining the abrupt changes in slope which coincide with the these transitions. After the transitions are located, appropriate  $2\pi$  radian multiples may be added to the phase in order to make the resulting phase data free of the cyclic behavior. As explained in Chapter 3, this is essentially determining the integer number of phase cycles  $K$  and recreating the whole phase  $\phi'$  from the measured fractional phase  $\phi$  as  $\phi' = \phi + K2\pi$ .

Once the phase versus frequency data is made contiguous, a linear curve fit is made to these data. Least mean squared error (LMSE) techniques, least absolute deviation (LAD), and robust fitting techniques have all been tested as possible fitting criteria. In certain cases the LAD technique produced improved results in the presence of points which did not fit the linear

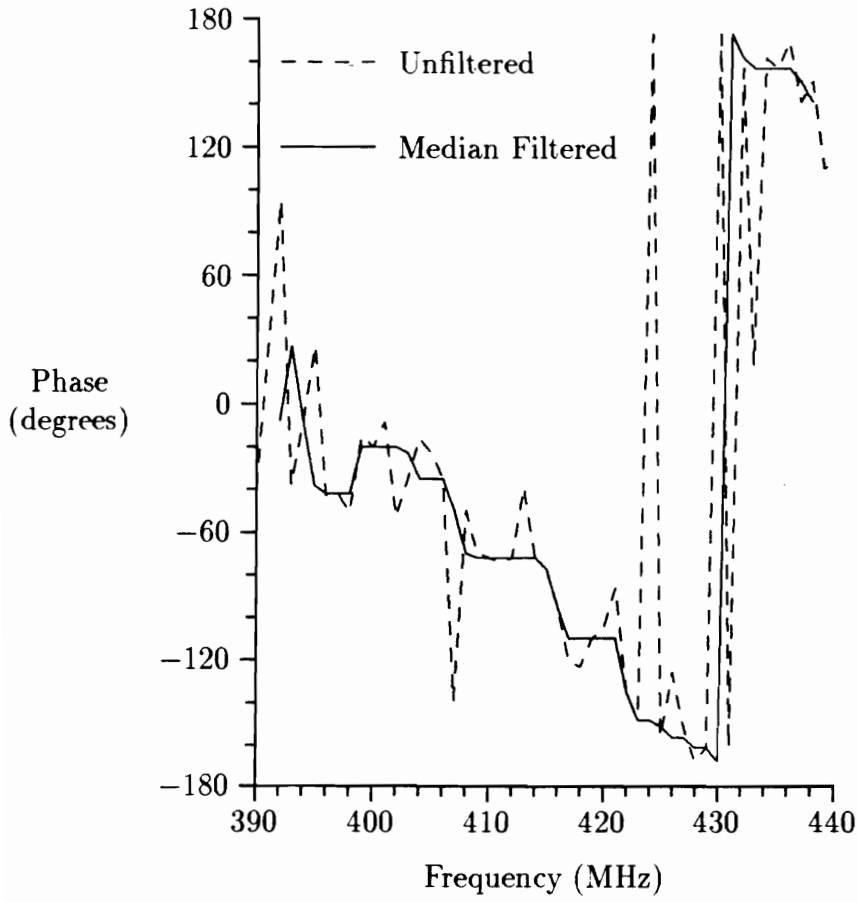


Figure 4.5: Phase versus frequency data before and after median filtering (window width = 5)

model well but overall the differences between the three techniques were not significant. For this reason most of the simulation work was performed using the more classic LMSE fitting criterion.

The two parameters in the linear model (the slope  $m$  and intercept  $b$ ) can be used to determine a range difference lane estimate and a fractional phase value for each receiver pair. The lane estimate can be found by first extracting the range difference from the slope as  $\Delta R = (d\phi/df)\lambda_c/2\pi$  where  $\lambda_c$  is the wavelength at the center frequency for the frequency hopping system ( $f_c = \{f_u - f_l\}/2 = \{440 - 410\}/2 = 425$  MHz). The lane estimate  $L$  is then the integer portion of the ratio of  $\Delta R/\lambda_c$ . The fractional phase is determined by evaluating the linear model fit at the center frequency and if the LMSE fit is used then this can be referred to as the LMSE phase estimate.

Based on the lane and LMSE phase estimates the range differences for each receiver pair may be calculated. The range differences are determined from the fractional phase  $\phi$  and the integer lane  $L$  as  $\Delta R = (\phi/2\pi + L)\lambda_c$ .

The range differences are then used to solve the hyperbolic system of equations to obtain a single position estimate. LMSE, LAD, and robust techniques have been tested as criteria for solving the system of equations. Nonlinear iterative Newton-Raphson matrix solutions to the resulting normal equations and three dimensional minimization of the equivalent objective function have both been evaluated. Techniques for solving the system of equations have been presented in Chapter 3; a more complete comparison of the theory and implementation for the methods is given in Appendices C and B.

Error statistics are calculated after each position estimate has been ob-

tained. Cumulative phase error, lane error, and position error statistics are compiled for the entire simulation run. The position error statistics have empirically been found to converge for approximately  $P = 100$  simulation points. The phase error statistics take longer to converge and approximately  $P = 10,000$  points are necessary. Phase statistics require more points for convergence than position statistics because the latter statistics represents a composite error based on phase at many frequencies and receivers. For example, in the current system 30 frequencies and 7 receiver pairs are used resulting in a single position estimate based on 210 phases. The  $P = 100$  points required for position error convergence are thus based on  $100 \times 210 = 10,500$  phases; this agrees quite well with the empirical value just quoted for convergence of the phase statistics.

#### 4.2.1 Multipath Channel Modeling

The channel model presented in equation 4.5 has several parameters whose statistics must be defined for the model to be complete. The distributions of the number of paths  $N$ , path delay times  $t_j$ , phase delays  $\theta_d$ , and the path amplitudes  $\alpha_j$  all must be considered. The goal when defining these parameters for the phase multilateration simulation was to maintain as simple a model as possible while still emulating real world conditions.

Many multipath studies have been performed in office buildings [92, 93, 95], homes [106], factories [94, 107, 103, 108], urban environments [97, 105, 109] and in irregular terrain [110]. Unfortunately, the multipath environment for the typical construction site does not fall neatly into any one of

the previously mentioned categories. The construction site channels will include effects of local topographic features (similar to the irregular terrain studies) and also many local scatterers such as equipment and building materials (similar to the factory studies). At the latter stages of construction projects involving buildings the multipath channels may resemble those of office buildings or homes. The approach taken in determining the statistical parameters of the channel has been an attempt to synthesize the results presented in the literature into a single model which is representative of what might be encountered on a typical construction site.

Turin has found experimentally that the probability distribution of the number of paths is modeled reasonably well as a Poisson process for urban propagation [97]. Saleh and Valenzuela found the Poisson model to apply very well to indoor propagation also [95]. Rappaport has found that for factory buildings the number of path components follows a Gaussian distribution [107]. In an earlier publication, Rappaport found that for high receiver threshold values the Poisson distribution fit the empirical data quite well [108]. The number of random multipaths ( $N$ ) in the channel is modeled as a Poisson distribution in the phase multilateration simulation program. The Poisson parameter ( $\lambda_m$ ) of 15 paths was selected based on the experimental results presented by Rappaport [108] for factory buildings. The Poisson distribution is a discrete probability function of the form

$$p(x) = \frac{\lambda^x e^{-\lambda}}{x!} \quad (4.6)$$

where the mean  $E\{X\}$  and variance  $\sigma^2$  are both equal to the Poisson parameter  $\lambda$  [111].

The path delay times are reasonably modeled as a Poisson sequence for urban environments based on Turin's results [97]. A modified Poisson process which takes into account the clustering of arrivals was developed by Suzuki [109]. Similar clustering of arrival times was noted by Hashemi [105] for urban propagation and by Saleh and Valenzuela [95] for office environments. Rappaport has suggested that multipath components may arrive independently (that is, not in clusters) for open plan offices and factories because they contain reflecting objects spaced uniformly in the work area [107]. The computer program simulates the excess delay ( $t_j$ ) of each path as determined from a Poisson distribution with parameter ( $\lambda_d$ ) of 35 nsec, where the Poisson parameter was determined from the empirical measurements of Rappaport [108].

Turin has suggested, as have many others, that the phases of the multipaths are mutually independent and uniformly distributed over the interval  $(-\pi, \pi)$  [97]. This is quite reasonable because the phase is very sensitive to path length and changes by  $360^\circ$  for each wavelength of path length change. Because the path lengths are generally many hundreds or thousands of wavelengths at the operating frequencies in question, the resulting phase components can be assumed to be uniformly distributed. The phase of the random multipaths ( $\theta_j$ ) in the computer simulation is therefore modeled as uniformly distributed on  $(-\pi, \pi)$ .

The literature on the distribution of path amplitude suggests that over local geographic areas the path strengths have Rayleigh or Rician distributions and over larger areas log-normal distributions exist [97, 110]. Rappaport has found that for factories the individual multipath components are

log-normally distributed about the mean value [107]. The mean value for the multipath components was found to follow a  $d^{n(t_j)}$  path loss law where the exponent  $n(t_j)$  is only a function of path delay [107]. In general, it was found that the path loss is greater for components that arrived later in the profiles [107]. The total power contained in the multipath profile was found to also follow a power law with the exponent having average values from 1.8 to 2.8 [107]. Suzuki has found that distributions of the near line-of-sight components over local areas follow a Nakagami distribution and that later paths are log-normal [109]. In the simulation program the path amplitude ( $\alpha_j$ ) is modeled as Rayleigh distributed about the mean because of the local topographic area over which the system operates. The Rayleigh distribution is a continuous distribution given as

$$f(x) = \frac{x}{\alpha^2} e^{-x^2/2\alpha^2} \quad (4.7)$$

with the mean  $E\{X\} = \alpha/\sqrt{\pi/2}$  and variance  $\sigma^2 = (2 - \pi/2)\alpha^2$ . The mean was made to follow a  $d^n$  power law where a value of  $n = 2$  was selected; to keep the model as simple as possible no explicit dependence on time delay was assumed for the power law exponent.

The mean amplitude of the multipaths was scaled by a constant parameter  $\eta$  which represents the ratio between the multipath amplitudes and that of the direct line-of-site path. LaBel has called this parameter the LOS/STD ratio (line-of-sight intensity to multipath intensity) and has found that ratios are typically between 1.9 and 4.1 [99]. LaBel's test results are based on measurements in several areas (wooded, half-wooded, open, and rural) during Summer, Autumn, and Winter. He found that there was no simple



relationship between the LOS/STD and frequency, environment type, or season. The position location simulation uses a value of  $\eta = 6.25$  determined empirically from the small test site experimental results (see page 135 for more details).

### 4.3 Simulation Results

The computer simulator was used to determine the expected distribution of differential phase errors in the small site tests of the position location system. A 10,000 point simulation produced the phase error distribution shown in Figure 4.6. The solid curve in the figure represents a least mean squared error curve fit of a Gaussian distribution. A number of different computer trials were performed and for all cases the phase error distribution has been found to be an excellent fit to a Gaussian distribution with a mean in the range of  $\pm 15^\circ$  and a variance of less than  $10^\circ$ . These results and their significance are discussed in more detail in reference [112], which represents the only study of phase-difference errors in the literature.

The approximate Gaussian nature of the phase error distribution justifies the use of position error distribution analysis of Cooper [113] or Marchand [114], both of which are based on the Gaussian error assumption. The Gaussian error hypothesis has been found to be reasonable for time-of-arrival systems but has never been investigated for a local area phase measurement system. The results also indicate that a simpler model for the position location simulation could be developed by modeling the phase errors directly. The idea of developing a simple model for the simulation results is referred

to as meta-modeling. A meta-model would be much easier to use than the current simulation which employs the rather complex channel model in equation 4.5 to generate phase errors and then indirectly obtains position errors.

It is important to note that although the phase error distributions can be modeled as Gaussian, in some respects this representation may be inadequate. Both computer simulations and experimental results have yielded phase error distributions which consistently have a characteristic 'dip' near the mean phase error. Of course, this dip is not in agreement with the behavior of the Gaussian model near the mean. Also, in general the tails of the simulated and experimental distributions are slightly larger than the Gaussian model would predict. These points indicate that an improved meta-model could be developed which would more accurately conform to the actual phase error distributions. The improved meta-model might be based on a modified Gaussian type distribution function.

The statistical distribution of the differential phase errors becomes significant when more than the minimum number of receivers needed to resolve the position are employed. The minimum number of receivers for a hyperbolic type system is the number of spatial dimensions plus one. If more than the minimum number of receivers is employed then, in general, no single unique solution will exist and a criterion will have to be selected upon which to choose an optimized solution. In many cases, this criteria will be based on a maximum likelihood estimate of the position location. It has been shown that the most probable value of the unknown position for the case of independent Gaussian distributed errors of differing variances is determined using the weighted least squares method [84].

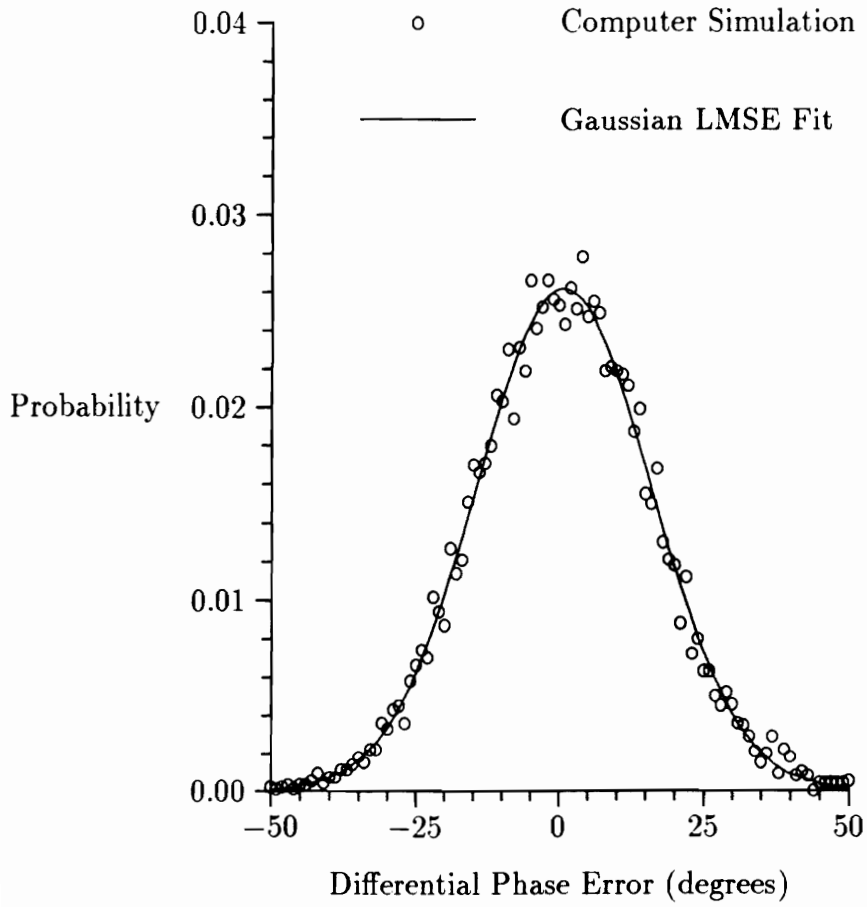


Figure 4.6: Simulated and model phase error probability distribution for small test site

These results suggest that the weighted least squares method may be the optimum method for estimating the beacon position. The weighting coefficients are based upon the individual variances of phase errors between each pair of receivers. It is a difficult problem to estimate *a priori* the proper weighting coefficients with any degree of certainty.

The simulator has also been used to determine the effect of the number and placement of receivers on the measurement accuracy. An 8 receiver array, with the receivers located at the corners of the small test site measurement area, produced the position error distribution shown in Figure 4.7. Using the classic least mean squared error (LMSE) estimate more accurate position determination was possible than with the 4 receiver array. It has been found that the LMSE estimate is easily skewed by measurement inaccuracies which lie in the tails of the error distribution.

Estimators which deweight highly deviant points have proven to be more effective than LMSE methods. The curve exhibiting the best performance in Figure 4.7 was produced using a majority vote algorithm to resolve ambiguities. The ambiguity in the X,Y and Z coordinate of the position is determined by sampling the predicted estimate from each set of receivers and selecting the one which occurs most frequently. In this manner the highly deviant points, which are most likely to be corrupted by multipath errors, are not used at all in the position calculation. This provides a significant improvement in system performance. Other classes of robust estimators, which provide a more gradual deweight of outliers, have also been tested [85, 86, 82] and results are presented in the next chapter.

The computer simulator has been used extensively to evaluate the poten-

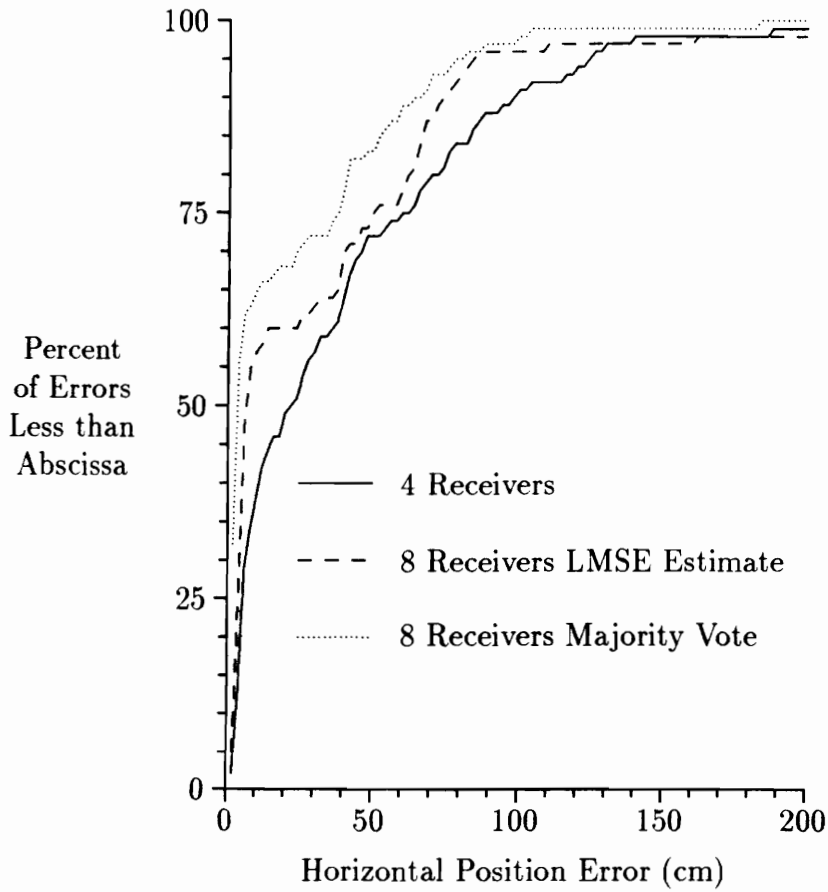


Figure 4.7: Simulated position error cumulative probability distribution for 4 receivers, 8 receivers LMSE and majority vote

tial improvement due to geometric changes in the receiver locations. For the experimental tests the simulator was used to determine optimized placements for the receivers. For more details on how the simulated and experimental test results compare, the reader is referred to Chapter 5 on the experimental measurements. For a real world position measurement system, the simulator could be used to tailor the receiver geometries to obtain increased performance.

Table 4.1: Soil dielectric constant and conductivity versus frequency and moisture content (from Njoku and Kong [104])

Frequency (MHz)	Moisture Content by Volume					
	10%		20%		30%	
	$\epsilon_r$	$\sigma$	$\epsilon_r$	$\sigma$	$\epsilon_r$	$\sigma$
300	6.0	0.0075	10.5	0.0125	16.7	0.020
3000	6.0	0.0667	10.5	0.183	16.7	0.333

## **Chapter 5**

### **Phase Measurement**

### **Experimental Tests**

A computer controlled experimental version of the phase multilateration system has been tested on a small scale site and also a full scale site. The small scale experimental test was conducted as a proof of concept and as a test of the prototype hardware. The small scale test was also used to calibrate a single parameter of the computer simulation program. Later, a full scale test was conducted at a site approximately the size of a typical construction project. The full scale site experiment was to determine the accuracy of the prototype system in a realistic operating environment. A comparison between the results obtained from the full scale tests and those predicted by computer simulation has served to verify the validity of the simulation program.



## 5.1 Hardware Description

Complete details of the hardware portions of the system and the design tradeoffs have been presented by Purdy [101], and for this reason only a brief description will be given here. A block diagram of the electrical portions of the prototype position location system is shown in Figure 5.1 and a simplified electrical diagram of the measurement subsystem is shown in Figure 5.2. The system is controlled by a main computer (labeled *CPU* in the figure) which synchronizes the entire measurement process. The computer controls the timing of the frequency hopping transmitter and receivers so that they are always tuned to the same frequency. In the prototype system the synchronization link between the beacon transmitter and the computer is accomplished via coaxial cable; in a finalized version of the system this would be a radio link.

The beacon transmitter receives the frequency tuning information from the computer and commands a phase-locked loop synthesizer with the proper divide ratios to generate the correct frequency. A preamplifier, filter, and final amplifier then condition the transmit signal before it is applied to the antenna. The antenna used for the beacon transmitter must have a nearly omnidirectional radiation pattern in the horizontal plane to assure adequate signal levels at all receiver sites. The receiver antennas do not necessarily have to be omnidirectional in the horizontal plane but must provide a wide enough main beamwidth to assure an adequately large measurement area. The experimental test system used center-fed vertically polarized half-wave dipole antennas for the beacon and receivers. The dipoles used small

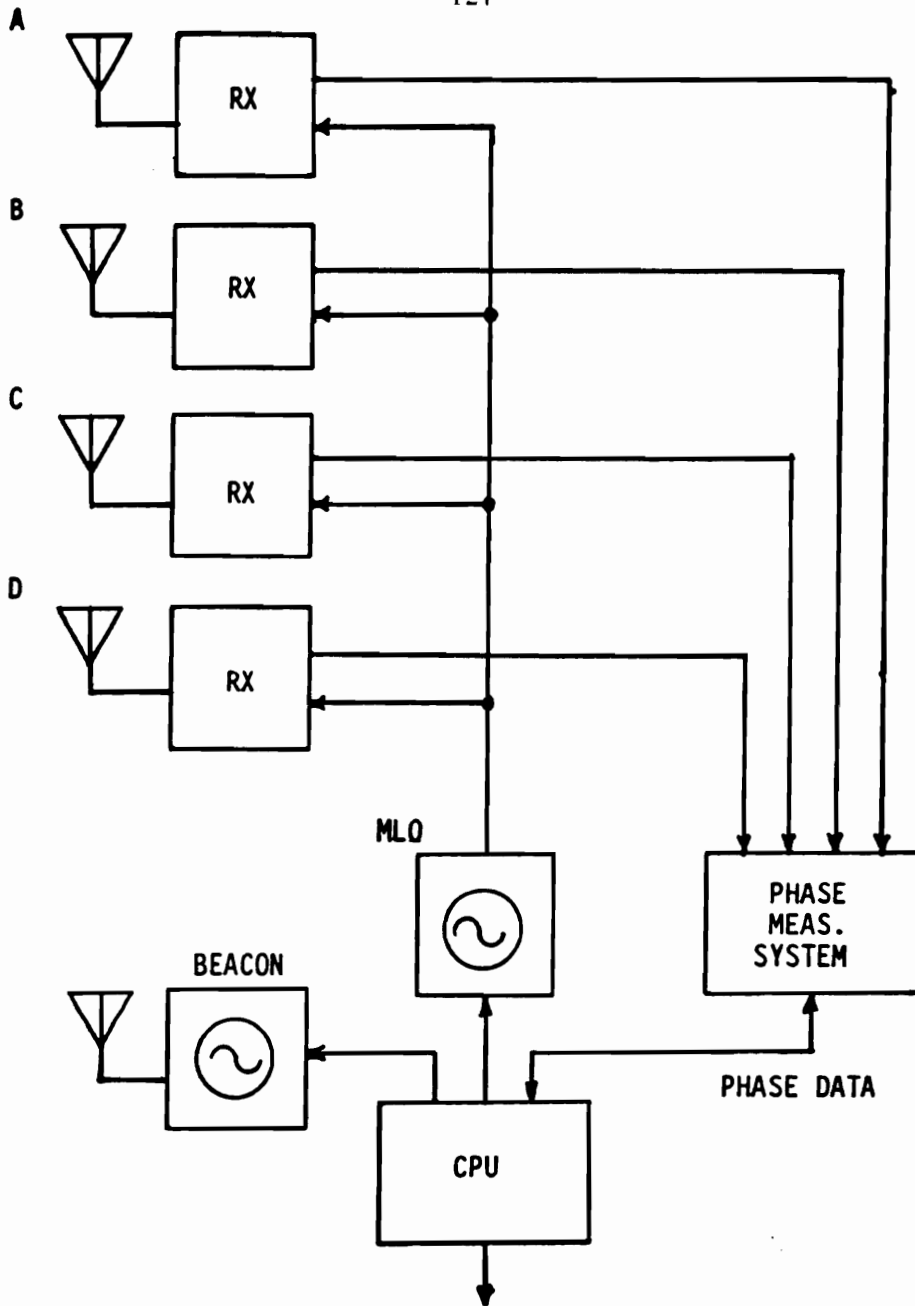


Figure 5.1: Electrical block diagram of prototype phase multilateration system (from Purdy [101])

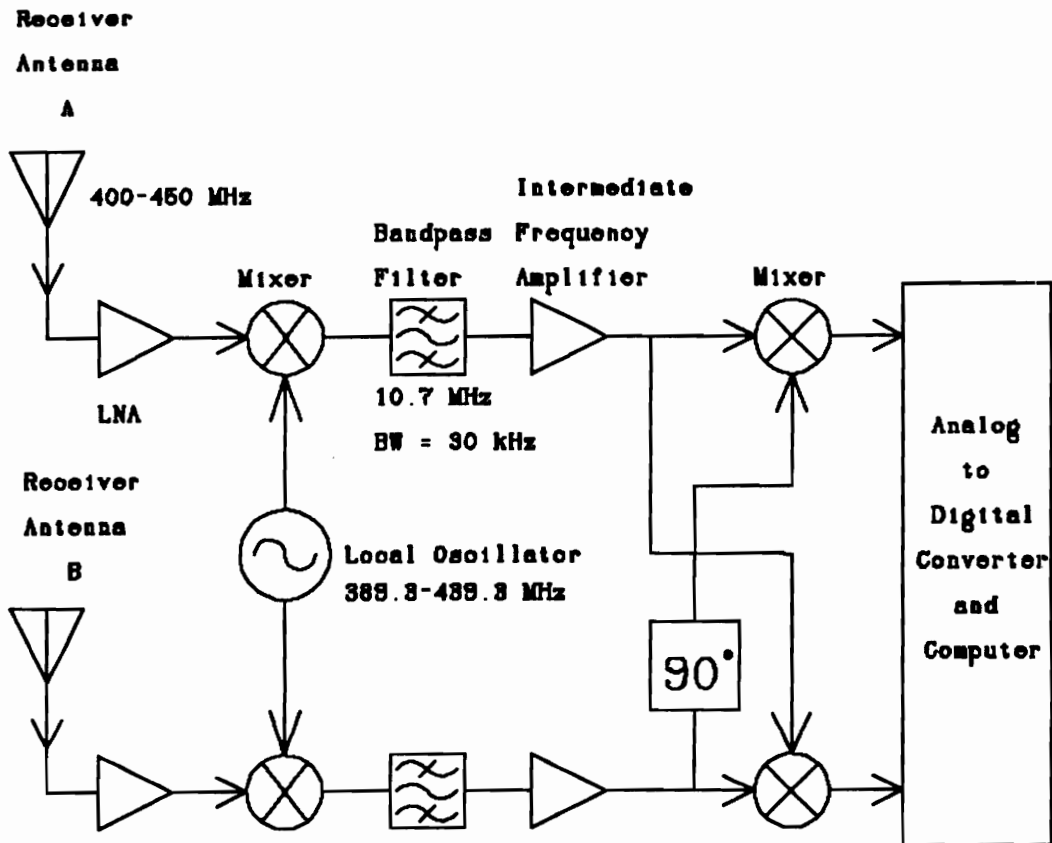


Figure 5.2: Simplified electrical diagram of the phase measurement system

length to diameter ratios to obtain the necessary bandwidth requirement. The vertical polarization provides a significant reduction in ground reflection at grazing angles close to the Brewster angle (see the discussion pertaining to Figure 4.1).

A master local oscillator (labeled *MLO* in the figure) is also connected to the computer via a serial link. The local oscillator gets tuning information from the computer and controls a phase-locked loop synthesizer to generate the appropriate offset frequency so that the base station receivers are tuned to the transmit frequency. The same local oscillator signal is supplied via coaxial cable to each of the base station receivers; in this manner a coherent frequency downconversion takes place at the receivers. The coherent downconversion is necessary to maintain the phase relationships between the receivers.

The receiver is a superheterodyne for the 410 to 440.015 MHz frequency range using a 10.7 MHz intermediate frequency with a RF bandwidth of 10 MHz and an IF bandwidth of 30 kHz. The local oscillator is tuned in steps from 399.3 to 429.315 MHz at a rate synchronous with the transmitter frequency hopping. The receivers chain begins with a low noise preamplifier and a fixed-tuned front end filter whose center frequency is 425 MHz. The front end filter is followed by a mixer and a series of intermediate frequency amplifiers and narrowband filters. The intermediate frequency outputs from the receivers are sent via coaxial cable to the phase detector subsystem located with the main computer.

The phase measurement system uses quadrature detection to determine the relative phase relationship between a pair of intermediate frequency signals. The signals are first conditioned through a limiting amplifier to remove

any amplitude variations and then a narrowband filter is used to remove as much noise as possible from the signals. The signals are then applied to a pair of analog multipliers (balanced mixers) fed in phase quadrature. The low pass filtered direct current signals from the mixers are then conditioned and digitized. A microprocessor analyzes the inphase and quadrature components of the phase detector outputs to determine the phase relationship. The resolution of the phase detector system is  $0.4^\circ$  and after calibration the system is capable of accuracy exceeding  $\pm 5^\circ$  [101].

Several limitations in the prototype hardware which have an influence on the accuracy should be discussed. The most serious problem with the hardware was an intermittent spurious oscillation in the local oscillator portion of the receivers. The receivers are supplied with a synchronization signal whose frequency is one-third that of the desired local oscillator signal; a frequency tripler is used to create the correct local oscillator signal. The tripler circuit was a part of the commercial UHF receiver used in the prototype system.

It was discovered late in the experimental testing process that the tripler circuit exhibited a spurious inband oscillation which was extremely temperature dependent. In some cases it was possible for the spurious oscillation to override the local oscillator signal and create false phase readings. These false readings were intermittent and could generally be eliminated by making a series of measurements at each test location. For all of the experimental test results presented here, a series of three measurements were taken at each test location. If each of the measurements did not agree then the entire dataset for that test location was discarded.

Another system limitation worth discussing is the relationship between

the signal amplitude and the corresponding measured phase. Unfortunately, it is quite difficult to build a measurement system which provides a constant phase shift independent of signal amplitude. The commercial receivers used in the prototype system exhibited a 1.13 degree/dB amplitude sensitivity and the phase detector subsystem had a sensitivity of 0.77 degree/dB in the normal operating range. One method of compensating for this phase shift would be to measure the signal amplitude and use a lookup table of correction factors. Another method would be to use special amplifiers, mixers, and associated receiver components which provide a low phase/amplitude sensitivity.

The phase-locked loop synthesizers used to generate the transmit and local oscillator frequencies were found to have relatively poor frequency stability. The frequency output of the synthesizers exhibited a slow periodic drift at intervals of about 5 seconds. The drift was about 8 kHz at 450 MHz resulting in a short term stability of about  $2 \times 10^{-5}$ . The frequency stability caused problems in two ways: it placed a lower limit on the measurement bandwidth in the system and it induced errors into the calibration process. Had the frequency stability been better, a much narrower measurement bandwidth and hence higher signal-to-noise ratio could have been achieved. The drift caused the local oscillator frequency to shift around in the passband of the intermediate frequency filters. The measurement system would be calibrated with the intermediate frequency at a particular spot in the passband; as the frequency drifted subsequent errors would be introduced due to the changing phase shift imposed by the filter.

One advantage of using a differential type phase measurement system is

that the requirements for frequency stability are less stringent than those for a system which measures absolute phase. The phase measurements are integrated over a much longer time period than any possible time delays between the two signals being compared. The longest differential distance present in the current prototype system is approximately 125 m and the resulting time delay is about  $0.42 \mu\text{s}$ . The phase measurements are integrated over a period of approximately 50 ms; this gives a worst case ratio of 120,000:1 between the measurement period and the propagation delay time. The significance of this result is that any perturbations in phase will appear at both phase detector inputs almost simultaneously (when compared to the length of the measurement window). As long as the perturbations in phase occur at a rate much less than the maximum delay time in the system ( $1/0.42 \mu\text{s} = 2.4 \text{ MHz}$ ) then they will not have a significant effect. Since the system measures difference in phase, any changes in phase that are common to the two signals being compared are not measured.

The measured phase data are analyzed using techniques identical to those used in the computer simulation described in Chapter 4.

## 5.2 Small Scale Site

The small scale site was a level area 4.2 m by 4.2 m located adjacent to a brick building with several parabolic dish antennas and pedestals nearby. One edge of the measurement area was paralleled by a wire mesh fence approximately 10 m away. It was anticipated that the presence of reflective objects near the test area would create a realistic multipath environment

in which to test the system. Because this was primarily a proof of concept test, only the minimum number of base stations which could resolve three-dimensional position estimates were used (that is, 4 base stations). One of the base stations was selected as the origin of the reference coordinate system; the other three base station antennas were placed along the cartesian coordinate axes as shown in Figure 5.3. The reference antenna, which is labeled antenna A in the figure, was located at a height of 2 m above the ground.

### 5.2.1 Experimental Procedure

The phase was measured for the three pairs of receivers at frequencies of 400 MHz and 450 MHz and differencing techniques used to obtain an equivalent phase at 50 MHz to resolve ambiguities. No frequency diversity was used for the testing; two frequencies are the minimum number necessary to resolve the phase ambiguities present for the given size of the measurement area. A total of 50 points randomly selected and evenly distributed throughout the measurement area were measured for each test sequence. The true location of the test transmitter was known to an accuracy of  $\pm 1$  cm using conventional survey techniques.

Calibration was done before each sequence of measurements by applying equal phase signals to each of the receivers; the equal phase signals were known to be within  $5^\circ$  of each other at the operating frequency. Calibration coefficients at each frequency and for each pair of receivers could be obtained from the measured phases with the equal phase signals applied.



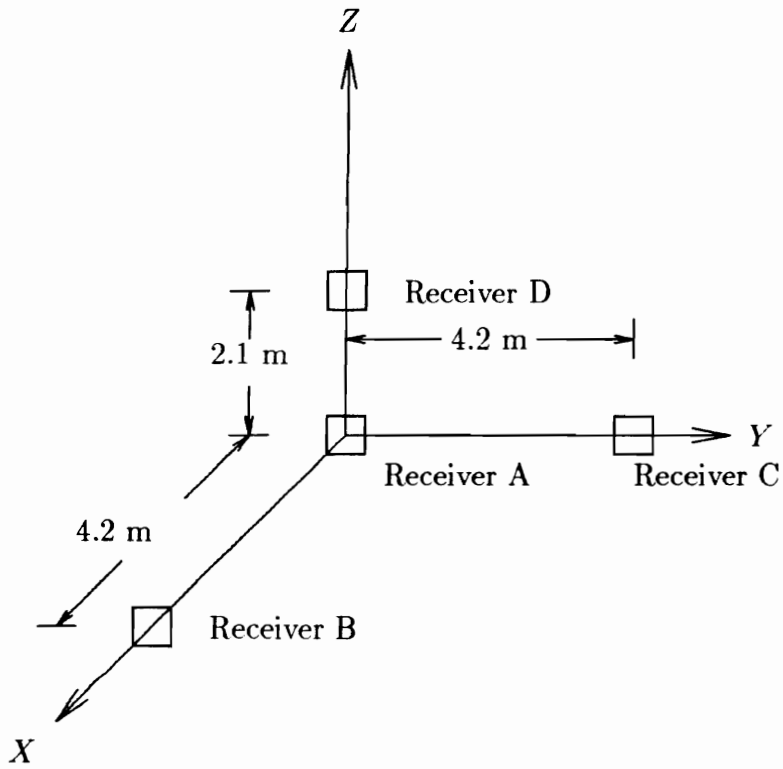


Figure 5.3: Small scale site base station geometry

### 5.2.2 Experimental Results

A complete sequence of measurements (50 points each) were made for dry ground conditions and for wet conditions. For the wet conditions the floor of the measurement area was flooded to a level of 2 cm of water to create a uniform reflective surface. The reason for performing the dry ground and over water tests was to determine the relative influence of ground reflection on the accuracy of the position estimates.

The estimated positions were compared to the true surveyed locations and the resulting cumulative error distributions are shown in Figure 5.4. For the dry ground conditions the resulting RMS horizontal error was 47 cm. For wet conditions, where the ground multipath is more severe, the system performance is degraded to an RMS horizontal error of 53 cm. At the 90% confidence level the two curves both have an RMS error of about 120 cm. Both curves exhibit a characteristic ‘knee’ between 30 and 35 cm, which corresponds to a single lane width. The knee represents the transition point beyond which ambiguity resolution errors dominate. Another knee is present at about 70 cm, indicating the presence of position estimates in error by two lanes. The small test site experimental results are discussed in more detail in reference [115].

The small test site experiment was also used to scale the direct line-of-sight to multipath ratio parameter ( $\eta$ ) in the computer simulation channel model (see page 117). The parameter  $\eta$  basically shifts the position error curve up or down by altering the amount of multipath present in the channel. In general, the shape of the position error curve is not affected by small

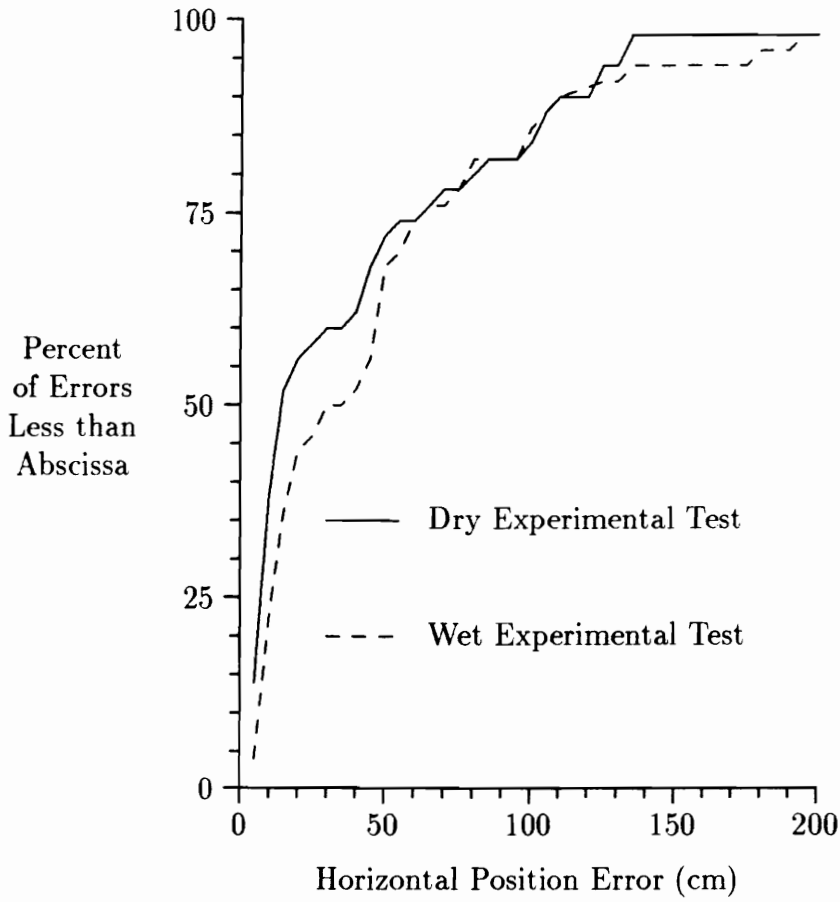


Figure 5.4: Small scale site position error cumulative probability distributions for dry and wet conditions

changes in the line-of-sight to multipath ratio. A value of  $\eta = 6.25$  was found to properly scale the simulation to the level of multipath present on the small scale site; this can be seen in Figure 5.5 which compares the experimental and simulation curves for dry ground test conditions. The shape of the curves are in good agreement indicating that the simulation may be properly modeling some of the features of the experimental system.

### 5.3 Full Scale Site

The full scale tests for the system were conducted on a rectangular field that was approximately 110 m by 70 m and had about 5 m of altitude variation across the site. Two corrugated metal buildings were located adjacent to the site and a wood frame building was about 20 m away. The field containing the test area was enclosed by barbed wire and in some cases metal mesh fencing. It was expected that the system would receive multipath reflections from the buildings and clutter near the measurement area as well as from the local topographic features present in the valley where tests were conducted.

Base station antenna towers were located at the corners of the rectangular measurement area. Each tower contained a lower antenna at 3 m above the base and an upper antenna at a height of 10 m. The roving beacon antenna was placed on a 5 m high movable pedestal.

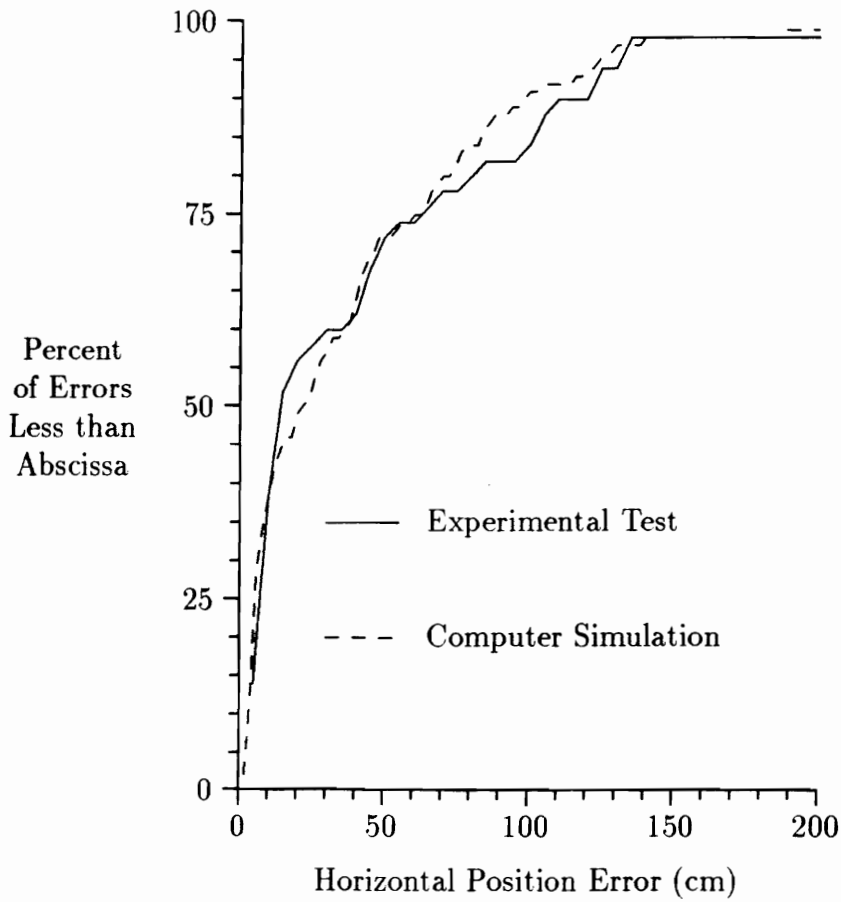


Figure 5.5: Small scale site position error cumulative probability distributions for dry conditions and computer simulation

### 5.3.1 Experimental Procedure

The system frequency hopped through a total of 30 frequencies, spaced at 1.005 MHz intervals, to obtain a sequence of phase measurements at each test point. A total of 124 position measurements were conducted at randomly distributed locations around the test area. The measurement locations in relation to the base station antenna positions are shown in Figure 5.6.

Calibration was done by making phase measurements at a series of surveyed locations. The calibration coefficients could then be obtained by using the median values of the measurements from the calibration points.

A source of error in the calibration process is the conventional survey methods used for reference positions. Initial survey techniques did not measure the positions of the receive and transmit antennas directly, but instead measured the location of a prism placed closed to the antennas. The difference in position of the prism and antenna introduced an error into the calibration process. This error was found to be on the order of 7 to 10 centimeters which introduces corresponding phase errors of about 50 to 80 degrees, depending upon the exact location. These errors are due to differences between the actual and measured locations of the calibration points and base station antennas. If the calibration and base station antenna locations are in error then the calibration coefficients determined by the system will be incorrect.

A later survey technique, which measures the positions of the antennas directly, was implemented for the test results presented here. A grid of control points was placed on the measurement site using an electronic-distance-

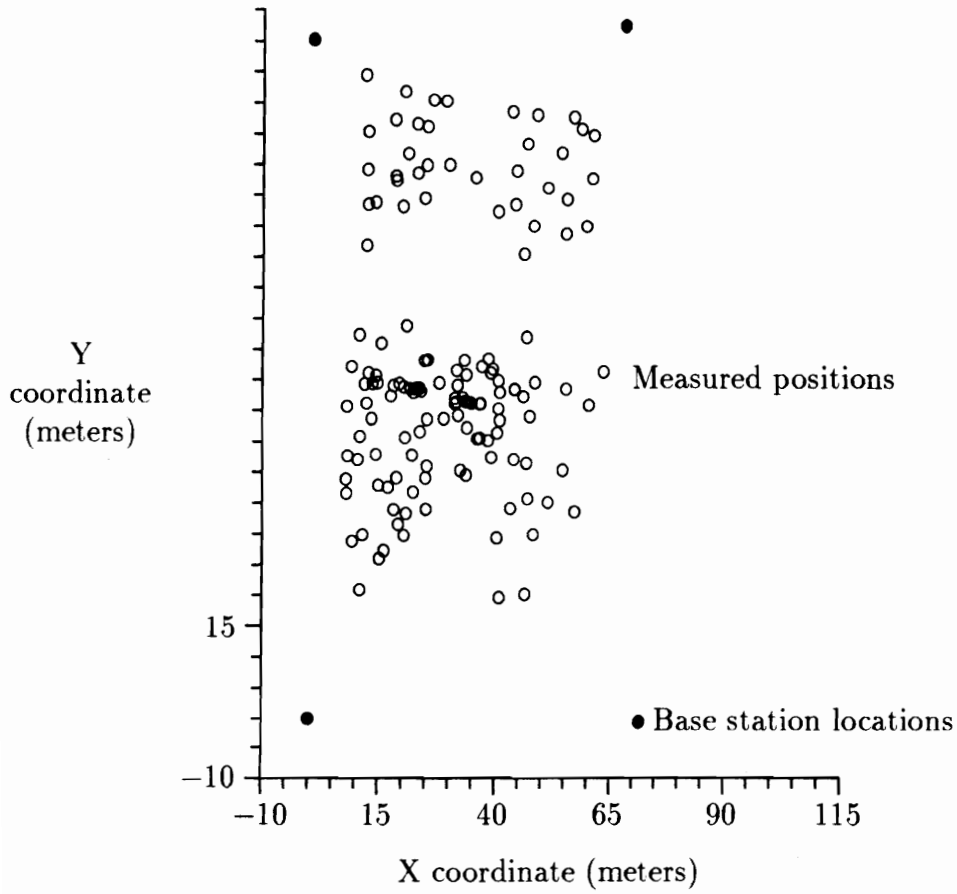


Figure 5.6: Full scale site measurement positions and base station locations

measuring (EDM) device. These control points were then used as locations from which to make theodolite angle measurements to each of the antennas. The resulting position accuracy, which appears to be at the 2 to 3 centimeter level, produces much lower phase errors.

### 5.3.2 Experimental Results

The estimated beacon positions were calculated using the data from all 8 receivers (resulting in 7 differential measurements) and the 30 frequencies using standard nonlinear least squares techniques [84]. Standard least squared error techniques were used to obtain a base curve for comparison with some of the non-standard techniques to be tested. The resulting position error distribution is shown in Figure 5.7 for all valid measurement points. The RMS horizontal position error was 1.8 meters; the three-dimensional error of 6.22 meters was much larger than the horizontal error due to geometric limitations (refer to the discussion in Section 5.3.4 for more details).

Different techniques, other than least squared error estimation, have been used to determine the estimated position. Robust estimators, which provide a more gradual deweighting of outliers, have been able to cut the RMS errors almost in half. Unfortunately, this improved error is at the expense of increased computational complexity (see Chapter 3 and Appendix B for a detailed description of the robust estimators which have been used). Figure 5.8 illustrates the full scale test site position error distributions for the position error which has been achieved using median filtering and robust processing methods compared to classic least squared error analysis. An error



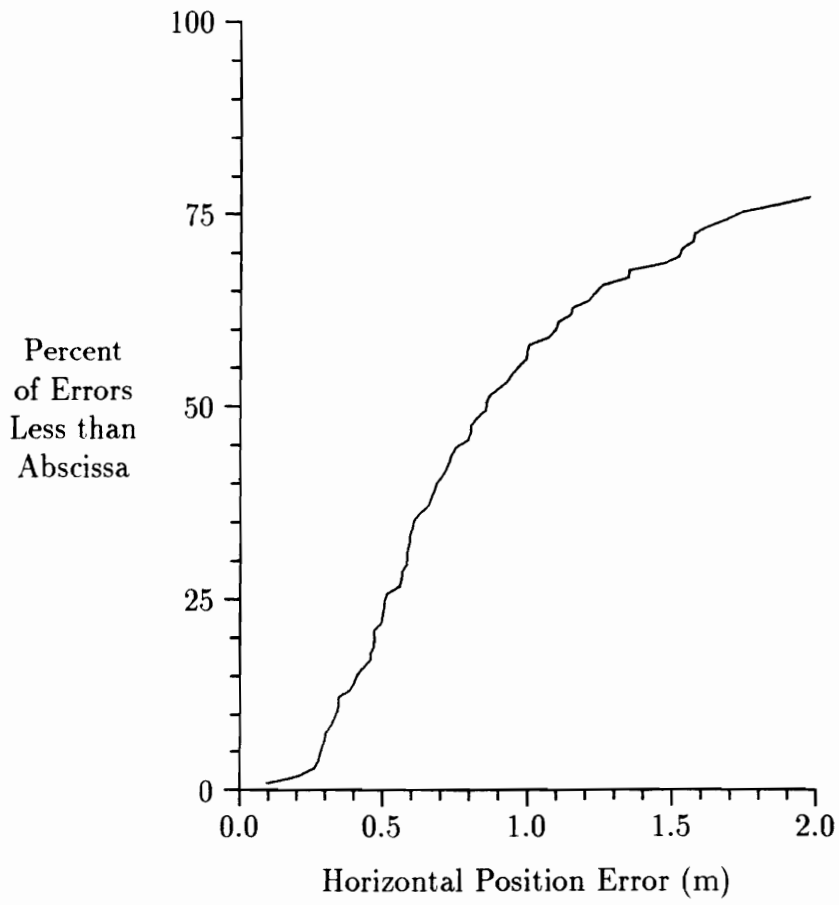


Figure 5.7: Full scale site test position error cumulative distribution

improvement of better than 45% has been obtained with resulting horizontal RMS errors that are at the 95 centimeters level. A dramatic reduction in the has been acheived in the number of points with errors exceeding about 75 centimeters; comparatively little improvement has been obtained for the points with errors less than this value.

### **5.3.3 Comparison of Simulation and Experiment**

The computer simulation was developed as a tool to assist in comparing the errors produced by various configurations of the positioning system. A statistical simulation of this nature can only give indications about how different system configurations compare over a large sample of test sites. No single physical test site may be expected to agree identically with the simulated performance. However, the general shape and features of the error distribution curves for the simulated and experimental systems should match up. Changes in antenna geometry, number of receivers, and other such factors should cause similar changes in the error curves for the simulation and experiment.

The general shape of the cumulative position error distributions agree quite closely for both the small and full test sites. This is illustrated in Figure 5.9 for the full scale test site and has been shown previously for the small scale site.

The phase error probability distributions have similar features between the simulated and experimental full scale site results, as can be seen from Figure 5.10. The simulated phase errors exhibit a narrower peak and conse-

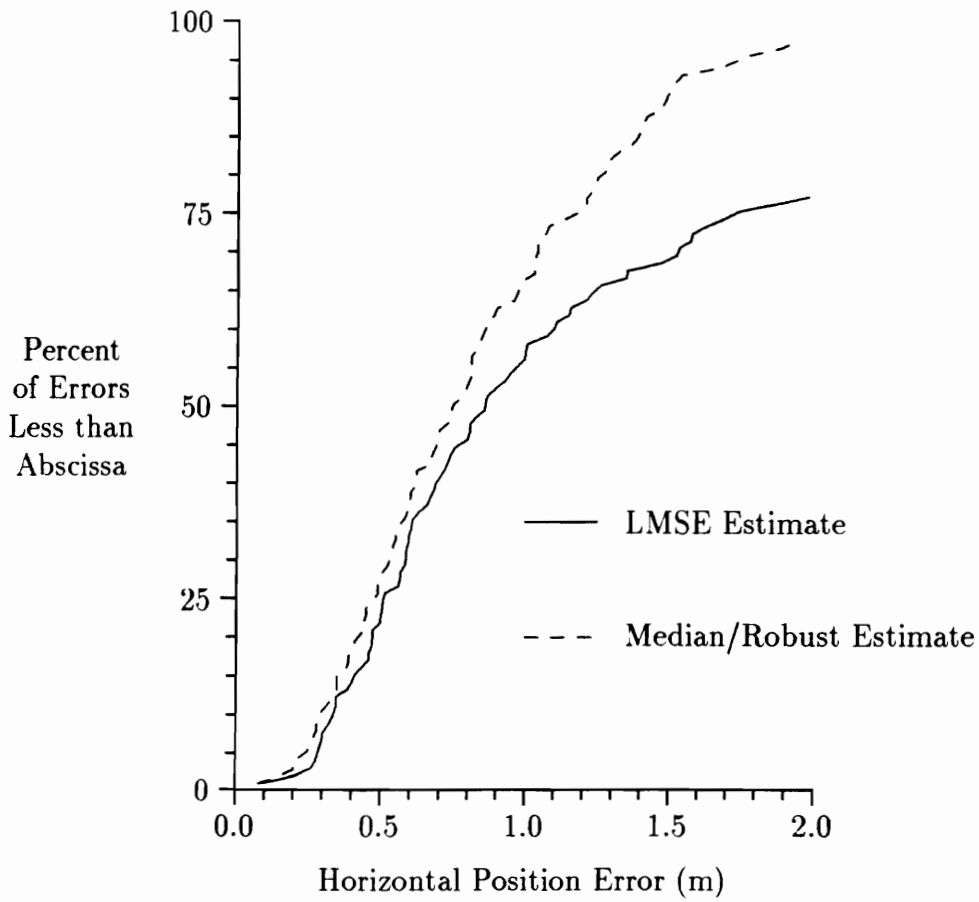


Figure 5.8: Full scale site experimental position error with median filtering and robust estimation compared to least mean squared error analysis

quently lower tails than the experimental results. A careful inspection of the figure will reveal the characteristic ‘dip’ near the center of the phase error distribution curve. This dip is a deviation from the Gaussian model which has been shown to fit the rest of the curve rather well. No similar phase error comparison can be made for the small test site because of the limited number of test points and measurement frequencies.

Care should be taken when comparing the phase error distributions to the position error distributions. The phase error distributions are calculated over all frequencies and receiver pairs. The distribution of the phase errors in frequency and in receiver pairs influences strongly the resulting position errors.

Ideally, one would like to have enough phase measurements to separate the phase error distributions for each frequency and receiver pair. Unfortunately, many thousands of phase measurements are required to obtain a phase error probability distribution. The resulting distribution curves converge slowly to their final values. If insufficient points are used to determine the distributions then large variances will be present in the curves; valid comparisons between curves cannot be made in this case.

### **5.3.4 Vertical Coordinate Accuracy**

It is well known that the vertical errors made by ground-based hyperbolic systems substantially exceed horizontal errors [73, 62]. This effect has been studied by Lee [62], and occurs because of the relatively short baseline distances present in the vertical direction. The horizontal baseline distances

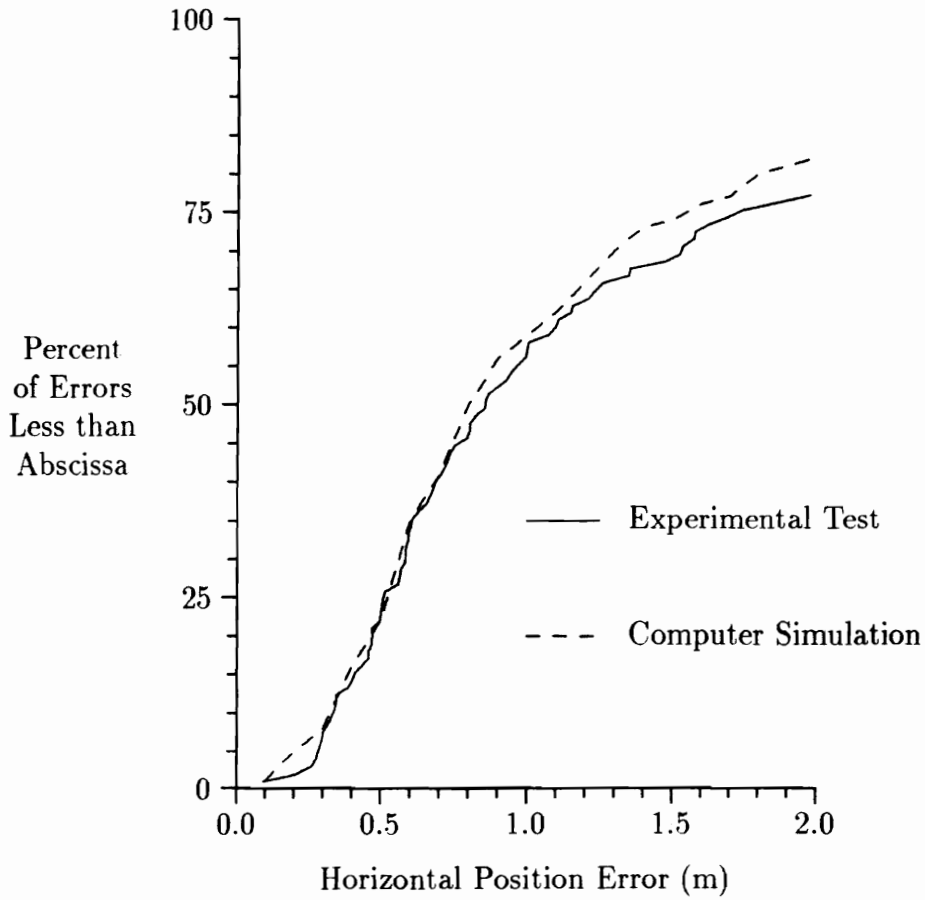


Figure 5.9: Full scale site simulated versus experimental cumulative position error distribution

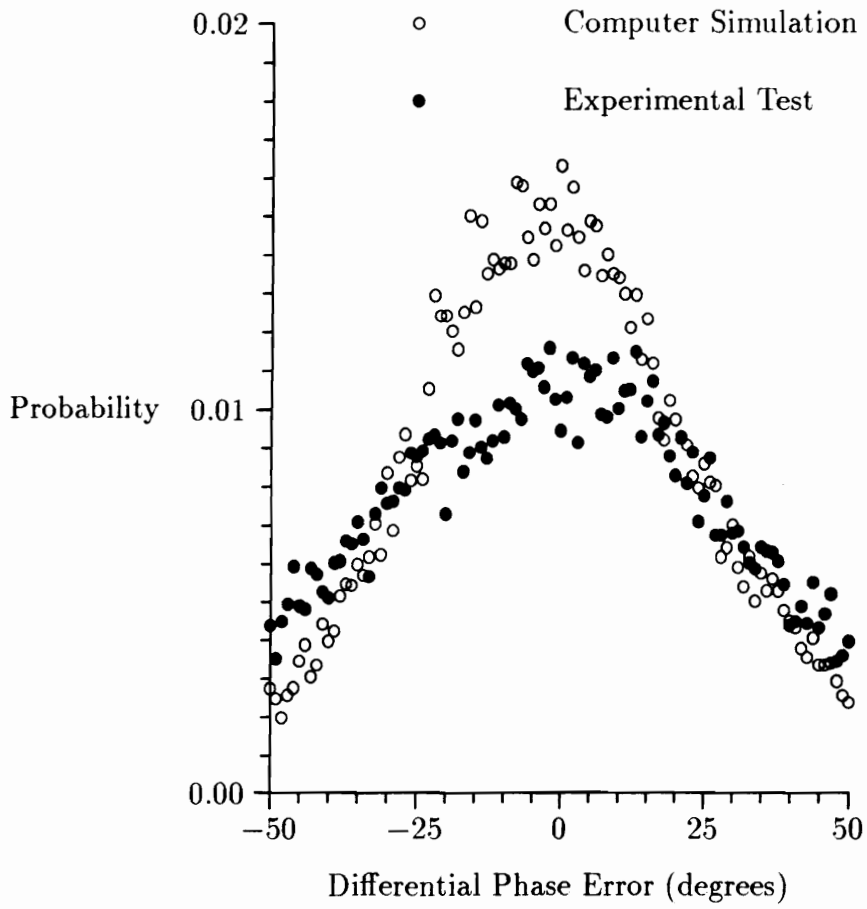


Figure 5.10: Full scale site simulated versus experimental phase error distribution

for the experimental system were 156 and 99 wavelengths while the vertical baselines were only 10.5 wavelengths. The effect of the shortened baselines is an increased divergence of the lanes away from the baselines. This is illustrated in Figure 5.11 where a baseline of 7 wavelengths is shown in (a) in comparison with a 25 wavelength baseline in (b). The lines of position are much closer to being parallel in the long baseline configuration while the short baseline geometry results in extreme sensitivity to distance errors in the vertical dimension (in other words, the lane expansion factor  $\chi$  is large).

Using the equivalent mass technique of analyzing geometric errors in a hyperbolic type system, a lower bound on the ratio of the vertical error to the ranging error can be obtained [78]. Using the method discussed in Chapter 3 for the configuration of base stations in the full scale site, the lower bound on the vertical to ranging errors is found to be

$$\frac{\sigma_z}{\sigma_r} \geq \frac{1}{\sqrt{2} \sin \theta} \quad (5.1)$$

where  $\theta$  represents the angle between the beacon and the nearest base station upper antenna. For a point located midway between the base stations and the center of the site, the lower bound on the ratio of the vertical to ranging errors is  $\sigma_z/\sigma_r \geq 3.2$ . Using typical values of the horizontal to ranging error  $\sigma_{xy}/\sigma_r = 2$ , the lower bound on the vertical to horizontal errors is found to be  $\sigma_z/\sigma_{xy} \geq 1.6$ . The simulated and experimental results show that vertical errors exceed horizontal errors by a factor of about  $\sigma_z/\sigma_{xy} = 3.5$ . Realistic changes in receiver geometry, due to limitations on receiver height, can only reduce the vertical error slightly. To reduce the vertical error more significantly, other methods of determining altitude must be investigated.

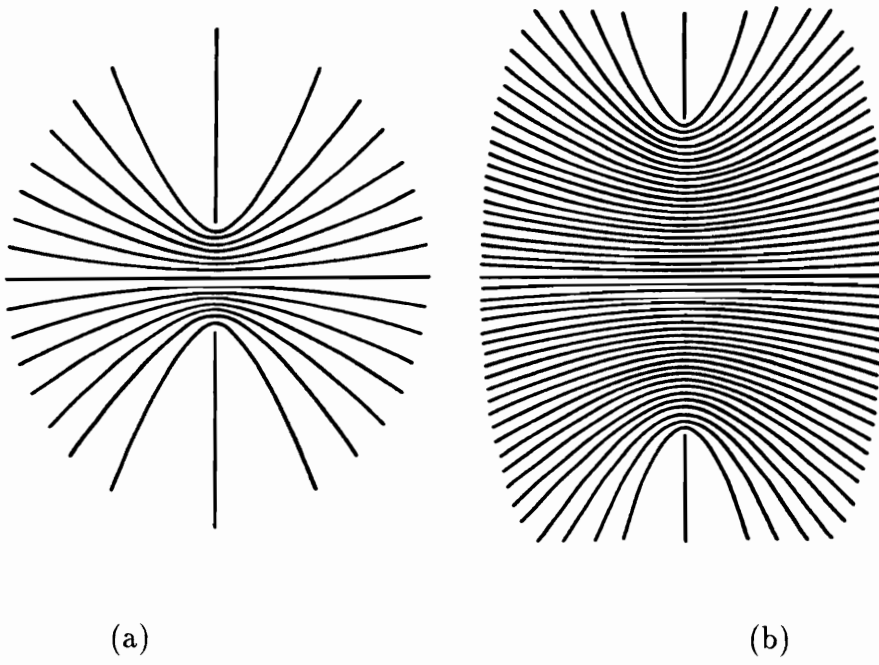


Figure 5.11: Divergence of phase lanes for baseline lengths of (a) 7 wavelength and (b) 25 wavelengths



Laser plane techniques are currently used to monitor the height of equipment quite accurately. In these systems a rotating laser beam strikes a photosensitive strip on the vehicle. By determining which section of the strip is illuminated, vertical positions can be estimated to several centimeters. These methods could be used in conjunction with phase measurement systems to provide accurate 3-D coverage from a hybrid laser/radio system.

Palmer has described a system using phase measurement techniques for horizontal position combined with barometric pressure readings to provide altitude [10]. The temperature, humidity, and pressure at each receiver and the beacon are measured. The pressures from the receivers are used to compute a pressure gradient plane from which the beacon's elevation can be determined. Published results indicate that elevation accuracies of better than one meter can be obtained with this method.

## 5.4 System Improvements

A number of methods have been investigated for improving the accuracy beyond that of the basic system which was tested experimentally. The effect of multipath errors can be reduced by several methods: antennas suited to rejecting multipath may be employed; knowledge of the limits on vehicle dynamics can help to remove large errors; specialized techniques to analyze phase information can be used.

The half-wave dipoles are center-fed using a balun transformer perpendicular to the antenna elements. The feed introduces some asymmetry into the radiation pattern of the antenna and changes the effective phase center. This

has been found to cause errors in the position estimates as the orientation of the beacon antenna is changed relative to the receive antennas. The received phase has been noted to shift by as much as  $35^\circ$  due to changes in the horizontal orientation of the antenna. One method of overcoming this limitation is to feed the dipole from below with a feed line which passes through the lower antenna element. A decoupling technique using a lossy dielectric can be employed to eliminate the interaction between the feedline and the lower antenna element.

Another antenna which is broadbanded and provides a nearly omnidirectional horizontal radiation pattern is the disccone antenna. A properly designed disccone antenna would easily cover the bandwidth required by the phase measurement system. Some further investigation would be required to determine the phase center of the disccone and how it changes with frequency. The disccone antenna is fed directly with an unbalanced transmission line from below the antenna; this should eliminate any position errors caused by horizontal orientation of the antenna.

Another method of reducing multipath is to use directional antennas at the receivers to eliminate some of the undesired reflected signals. The amount of directionality which can be employed is limited due to the desire not to mechanically track the beacon with the receive antenna. This places a limit on the amount of suppression of multipath which can be obtained. Studies by Tranquilla and Best [116, 117] have shown that Yagi antennas are not suitable for uncorrected use in UHF radio positioning systems due to significant phase center movement which can occur even within the mainlobe. They also found that log-periodic antennas, due to their broad front lobe, had little

phase center movement over the front hemisphere [117]. Unfortunately, log-periodic arrays generally exhibit relatively poor front-to-back ratios, reducing their effectiveness at eliminating multipath interference.

Rhee [118] has tested the relative performance of omni and directional antennas in urban environments. He found that in urban environments directional antennas did not provide the interference rejection that theoretical front-to-back ratios would predict. In one of the tests, antenna range data predicted a 25 dB front-to-back ratio while urban measurements showed only 10 dB.

More exotic techniques employing multiple antennas exist for reducing multipath: vertical antenna redundancy for removing ground reflections [100]; adaptive nulling phased-arrays [119]; close-spaced antenna diversity [120]; and energy-density antennas [121]. The improvement in position accuracy obtained from these methods is hard to judge theoretically or with computer simulation; experimental testing is probably the only method of evaluating the benefits.

One signal processing technique which may be applied is linear prediction. The Kalman filter, a recursive linear filter which minimizes the mean-square error of an estimate, is used extensively in navigation and tracking systems; such as air traffic control, the Global Positioning System, and sonar tracking [122, 123]. A model for the vehicle motion dynamics can be used to eliminate unrealistic position estimates.

Channel sounding methods can be used to determine the number and time delay of multipath components [124]. Unfortunately, bandwidths on the order of 500 MHz are required to obtain sufficient resolution to remove

the effects of the multipaths. Skolnik also concludes that using frequency diversity to resolve multipath components requires bandwidths approaching those that would necessary using conventional pulse techniques to separate the signals [48].

## Chapter 6

# Time System Theory and Simulation

Experimental testing and computer simulations have shown that a frequency hopping phase measurement system with a bandwidth of 30 MHz using 30 frequency samples is capable of RMS horizontal errors of about one-half meter. While accuracies of this level are useful for construction applications, many operations require accuracy much greater than this. The limitation on the accuracy of phase measurement systems appears to be to multipath interference which causes errors in resolving the phase ambiguities. To remove the effect of the multipath errors by using phase measurement techniques requires unrealistic bandwidths given the current demands for spectrum space. For this reason, time delay spread spectrum techniques were investigated for use in the local area position location problem.

For the phase measurement system proposed in Chapters 4 and 5 the slope

of the phase versus frequency curve was used to resolve the phase ambiguities. Many of the large position errors in the system were due to errors in this ambiguity resolution process. There is no way to know by examining the phase versus frequency curve (at least over small bandwidths) that multipath errors have caused slope errors. In other words, there is not any way to know that multipath errors have corrupted a measurement. Obviously, this is one weakness of the phase measurement approach if multipath errors are significant.

As discussed in Chapter 3, time delay measurement can be used to estimate differential ranges. A pulse time delay system has the potential of removing multipath components if the path delay times are sufficiently large. Even if the delay times are not large enough to remove the effect of the multipath, in many cases the mere presence of the multipath can be detected. This is a major advantage because corrupted measurements can be deweighted or even discarded.

Conventional pulse timing systems require large bandwidths and/or high signal-to-noise (SNR) ratios to achieve the kind of temporal resolution necessary to resolve most multipath components (see Chapter 3). Spread spectrum techniques can be used to obtain most of the benefits of a short pulse system without the power or SNR constraints. A long pulse with a large bandwidth is transmitted and by using correlation techniques the arrival time of the pulse can be determined to much greater resolution than a conventional pulse. The increased resolution is due to the wideband modulation which has been included in the pulse.

In the direct sequence spread spectrum method a long pulse of duration

$T$  is divided into  $S$  subpulses of width  $\tau$  where the subpulses are coded to form a pseudo-noise sequence. If correlation is used at the receiver to detect the pulses then the correlator output will appear as shown in Figure 6.1. A peak of width  $\tau$  and amplitude  $S$  will be present and time sidelobes will be present out to a time  $T$  on either side of the main peak. The peak can generally be resolved to within one-tenth of the pulse duration  $T$ .

For a conventional pulse containing no modulation the correlator output would be a triangular function which starts at  $-T$ , ramps up to a value of  $S$  at time 0, and then ramps down to amplitude 0 at time  $T$ . The correlator output is then one wide triangle rather than the sharp spike and sidelobes created by the spread spectrum pulse. The key advantage of the spread spectrum pulse is this much narrower correlation peak. In the presence of noise or interference, the peak of the spread spectrum correlation function can be detected with much greater resolution than the conventional pulse. Therefore, spread spectrum techniques can be used to achieve temporal resolution without the associated fast risetimes or high transmit powers that would otherwise be necessary.

Spread spectrum techniques were conceived during World War II as a method of encrypting signals. Some of the earliest work done with spread spectrum techniques was on ranging systems for aerospace vehicles [125] and much of the development for the Global Positioning System followed. Other early researchers realized the potential of spread spectrum as a technique for combating multipath in communications channels [126]. By using spread spectrum pulse trains excellent temporal resolution can be achieved without the high transmit powers and/or large bandwidths required for a conventional

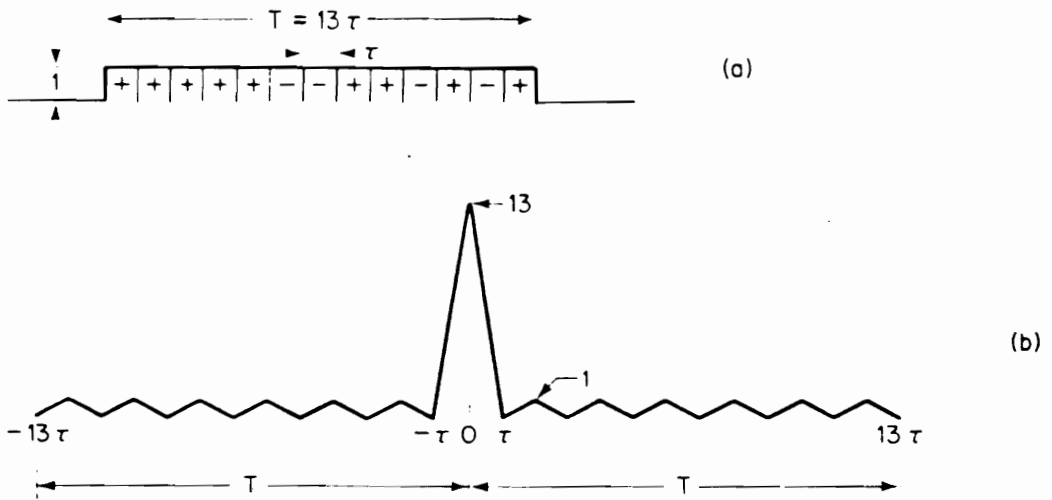


Figure 6.1: Example of (a) PSK pulse and (b) corresponding correlator output (after Skolnik [48])



pulse system with equivalent resolution [127]. More recently research has been performed using direct sequence spread spectrum methods to combat multipath in the urban radio channel [128].

By using this spread spectrum approach it is possible to achieve sub-pulse resolution to better than one-tenth of a pulse width. The time of arrival data contains information about the multipath signal components. Late arriving multipaths can be ignored and the accuracy of the resulting ambiguity resolution estimate improved. In addition to improving the accuracy, a confidence factor can be assigned to the corresponding position estimate.

If multipath is ignored, the theoretical accuracy of a phase measurement system exceeds that of a time of arrival spread spectrum system, when compared for equivalent power and bandwidth. The measurement of signal time of arrival using direct sequence spread spectrum techniques is more complex than the phase measurement system but it offers several advantages. A hybrid system combining the two technologies can provide both high accuracy and robustness in the presence of multipath. The time of arrival system can be used for ambiguity resolution and confidence evaluation and the phase system used for fine position calculation.

## 6.1 System Description

A block diagram of the proposed spread spectrum time delay system is shown in Figure 6.2. The direct sequence system transmits a Barker, Gold, or other pseudo-random (PN) coded phase shift keyed (PSK) pulse train. The Barker code sequence results in equal time sidelobes as noted by Skolnik [48] and

was chosen for the initial simulation work. The equal time sidelobes are important because close-in sidelobes of multipath signals can make estimating the maximum correlation delay difficult. With a 25 MHz modulation rate ambiguity position estimates can be determined with a resolution of approximately 30 cm.

The phase modulated signal is picked up by the antennas, with a time delay dependent on the path length between the beacon and each antenna. The receivers then downconvert the signals to an intermediate frequency and send them to a time delay correlator. The correlator can be an analog type which introduces a known delay ( $\tau_d$ ) into one signal path and then multiplies the two signals together. The analog multiplier output is then integrated over the period of the pulse to derive a correlation value ( $R_{ab}(\tau_d)$ ) corresponding to the delay value ( $\tau_d$ ). The resulting correlation output is then

$$R_{ab}(\tau_d) = \frac{1}{T} \int_T A(t) B(t - \tau_d) dt \quad (6.1)$$

where  $A$  is the undelayed input signal and  $B$  is the delayed one.

The correlation can also be implemented digitally if a high enough sampling rate is used; in this case the time delay, multiplication, and integration functions can all be done in software. A discrete correlation using the digital samples of the signals produces the resulting function

$$R_{ab}(m) = \frac{1}{N} \sum_{n=0}^{N-|m|-1} A(n) B(n + m) \quad (6.2)$$

where  $A$  and  $B$  are the sampled signals. In practice this type of digital correlation can be performed more rapidly in the frequency domain using the Fast Fourier Transform (FFT).

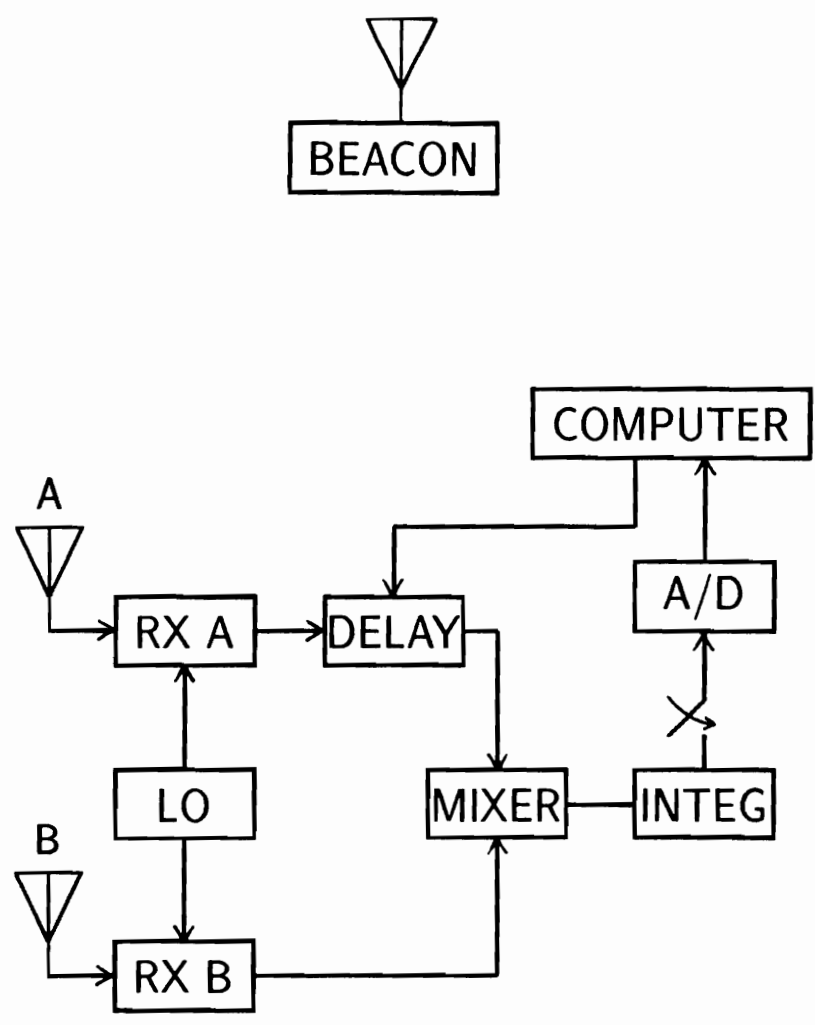


Figure 6.2: Block diagram of time domain spread spectrum system

An analog correlator is one way to estimate the delay times between the signals at each pair of receivers. An integrate and dump correlation receiver of this form is one realization of a matched filter receiver [129]. A switchable delay line is used to delay one of the signals at the correlator. The switchable delay line could simply be a relay or PIN diode matrix which routes the signal through a set of cables of different lengths. If the cables have a sequential relationship (for example, binary or octal) then by appropriate switching a whole range of time delays can be implemented.

After one of the signals passes through the switchable delay line, the two signals being compared are multiplied together over the period of the burst and the resulting product summed. When the two signals are exactly in time synchronization the correlator output will be maximum. More details about the time delay system are presented in reference [130].

## 6.2 Computer Simulation

Simulations of the direct sequence time of arrival technique have been performed using a modified version of the phase multilateration simulation. The direct sequence simulation operates within the frame work of the simulation program discussed in Chapter 4 and shown in Figure 4.3. The direct sequence routines in Figure 6.3 directly replace the ambiguity resolution routines using the phase slope technique in Figure 4.4. The rest of the simulation, including the channel modeling and position estimation routines, are exactly the same. Since most of the simulation program has already been discussed, only the features unique to the direct sequence simulation will be presented here.

First, the PSK pulse is generated as a sequence of sinusoids of phase either 0 or  $\pi$ . The transmit pulse is then sampled at  $B$  samples per cycle and stored in a buffer. The geometric reflections are then added by scaling the transmitted pulse by the path amplitude  $\alpha$  and delay  $t_d$ . The random multipaths are added to the buffer by using the same scaling and shifting approach. The receive buffer is then digitally correlated with the stored version of the received pulse from the reference receiver. This cross-correlation operation mimics an analog time delay correlator. Once the correlator output is obtained, the peak is determined. The time delay of the peak of the correlator output is used as the differential path time delay estimate. The range differences are then computed based on the time delay estimates and the speed of propagation.

### 6.3 Simulation Results

An 11-bit Barker coded sequence was used to phase shift key a 425 MHz carrier at a 10 MHz modulation rate. The 10 MHz rate was chosen for simulation because it is the largest intermediate frequency bandwidth that could be accommodated by the receivers used for experimental testing. The solid curve in Figure 6.4 depicts the correlator output for various differential distances, or equivalently different time delay taps on the analog delay line. The curve exhibits a peak at zero differential distance which is the true differential distance between the beacon transmitter and the two receivers.

When multipath components are added to the simulation the dashed and dotted curves in Figure 6.4 result. The case of 10 multipaths with an average

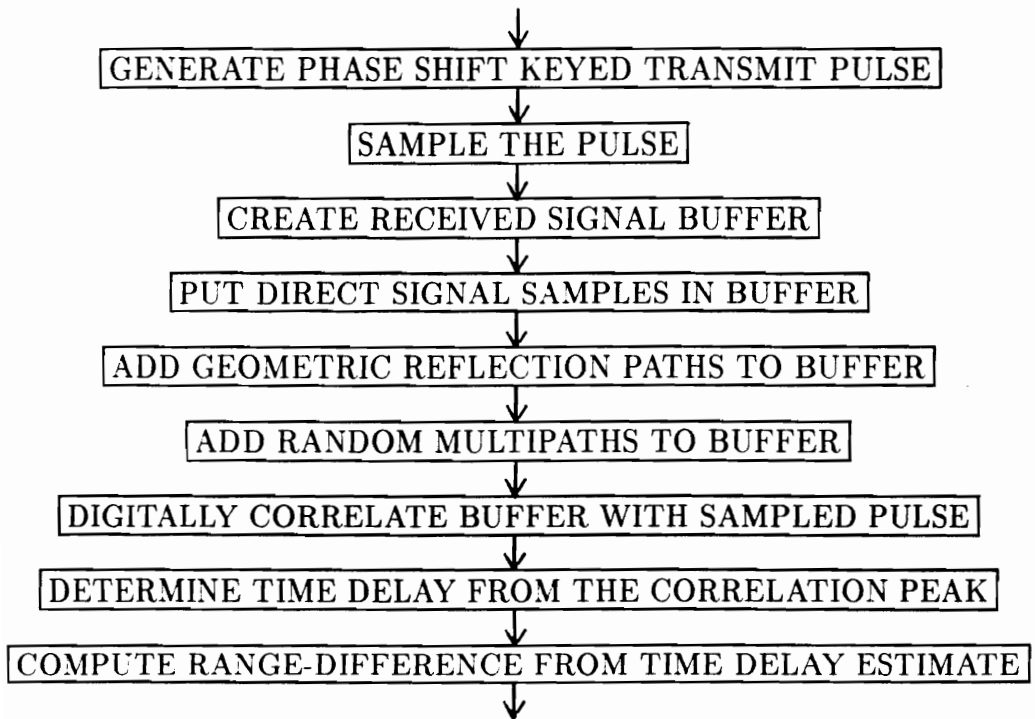


Figure 6.3: Block diagram of direct sequence spread spectrum simulation

delay of 67.6 nsec is shown as the dashed curve labeled multipath case A. The peak of the multipath curve is shifted from the no multipath case. This results in a position error of 3 meters which is much less than the ambiguity errors that occur in a phase measurement system operating under identical conditions. Another case is illustrated as the dotted curve labeled multipath case B which is for 10 multipaths with an average delay of 38.63 nsec. In this case the correlation mainlobe has been broadened due to the multipath components. In both cases A and B the multipath interference can at least be detected and in case B the effect can be removed.

A number of methods of resolving the direct component of the correlator output have been investigated. Since the close in multipaths can cause spreading as well as shifting of the main correlation peak, simply selecting the maximum value is not always sufficient. Techniques which use the area distribution of the mainlobe, such as measuring the centroid of the area, can sometimes overcome the shifting of the peak caused by multipath. The mean value of the 3 dB points can also be used to combat the same effect.

Simulation of the position errors for a time delay spread spectrum version of the position location system has shown great promise for removing some of the detrimental effects of multipath errors. Computer results indicate that the time delay system would be capable of decimeter level horizontal root-mean-square errors. Typical results for two different signal-to-noise ratios (SNR) are shown in Figure 6.5. Calculations indicate that at least a 20 decibel SNR, as indicated by the solid curve in the figure, could be maintained in a realistic system implementation. The results are much more encouraging than those for the phase measurement system. This can be seen in Figure 6.6,

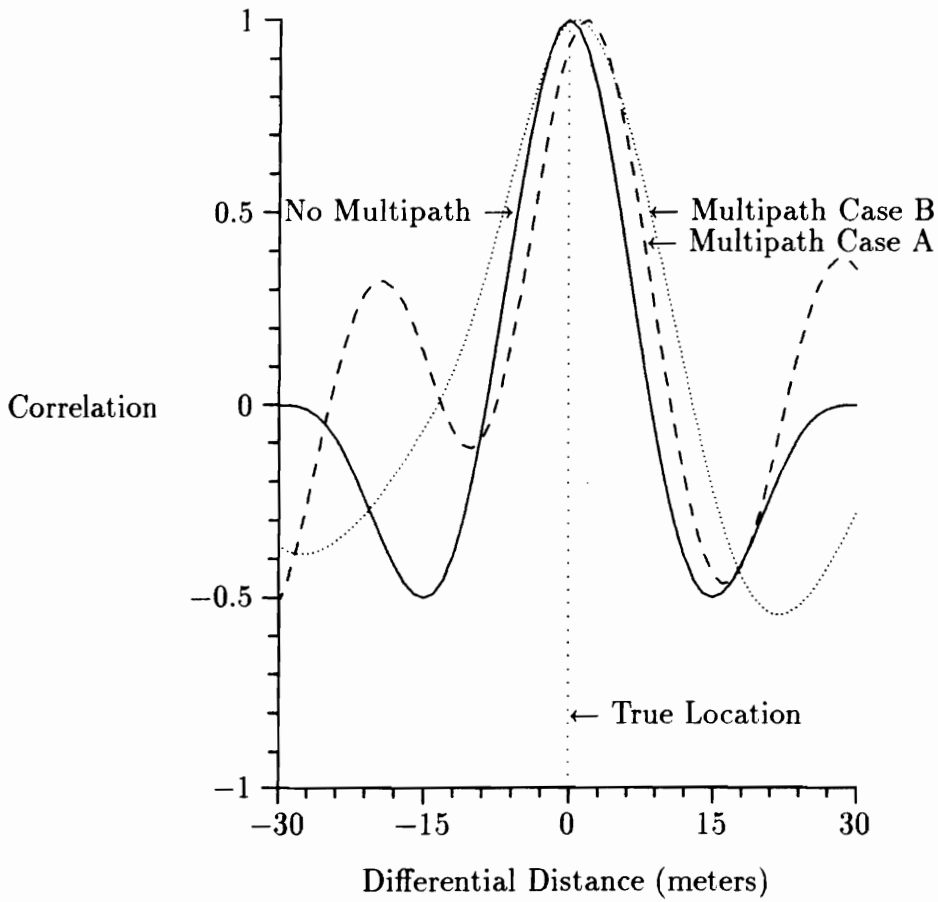


Figure 6.4: Output from time of arrival correlator with no multipaths and 10 multipaths



where the phase measurement and time delay system position errors are compared. The figure compares the time measurement and phase measurement system for computer simulations of the identical multipath and noise conditions using least squared error processing. The time delay approach is more resilient in the presence of multipath interference.

The computer simulation data for the direct sequence spread spectrum system has a RMS horizontal error of approximately 25 cm. This is about a 50% improvement over the frequency hopping phase measurement system. Keep in mind that the frequency hopping system used 30 MHz of bandwidth while the direct sequence simulation used only 10 MHz. If the direct sequence bandwidth were increased to 30 MHz then the resulting accuracy would be close to the decimeter level. This represents a significant improvement over the frequency hopping phase measurement system.

## 6.4 System Implementation

In an actual system the time delay information could be used to determine the range differences directly. However, if the range-differences are sufficiently accurate to resolve the phase ambiguities at the carrier frequency, then a hybrid approach can be used. The time delay information can be used to obtain ambiguity estimates to resolve the number of phase cycles. The carrier phase can then be combined with the lane estimates to provide a high resolution position estimate.

One advantage of this direct sequence time delay approach is that it could be more easily licensed by the Federal Communications Commission

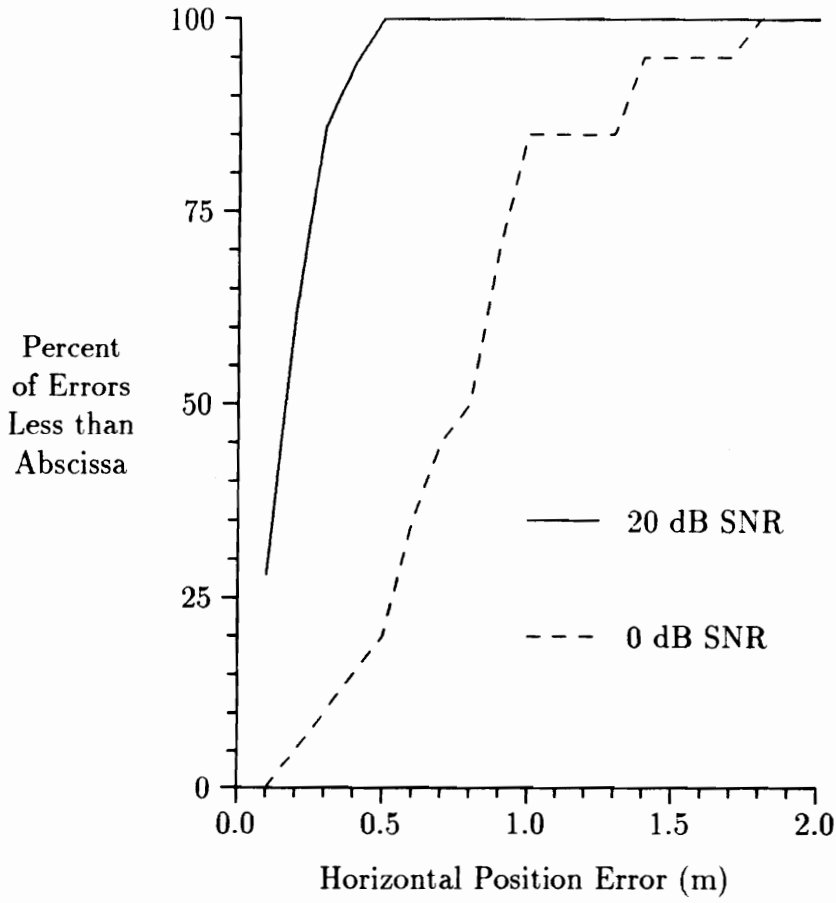


Figure 6.5: Time delay system position error distribution for signal-to-noise ratios of 0 and 20 dB

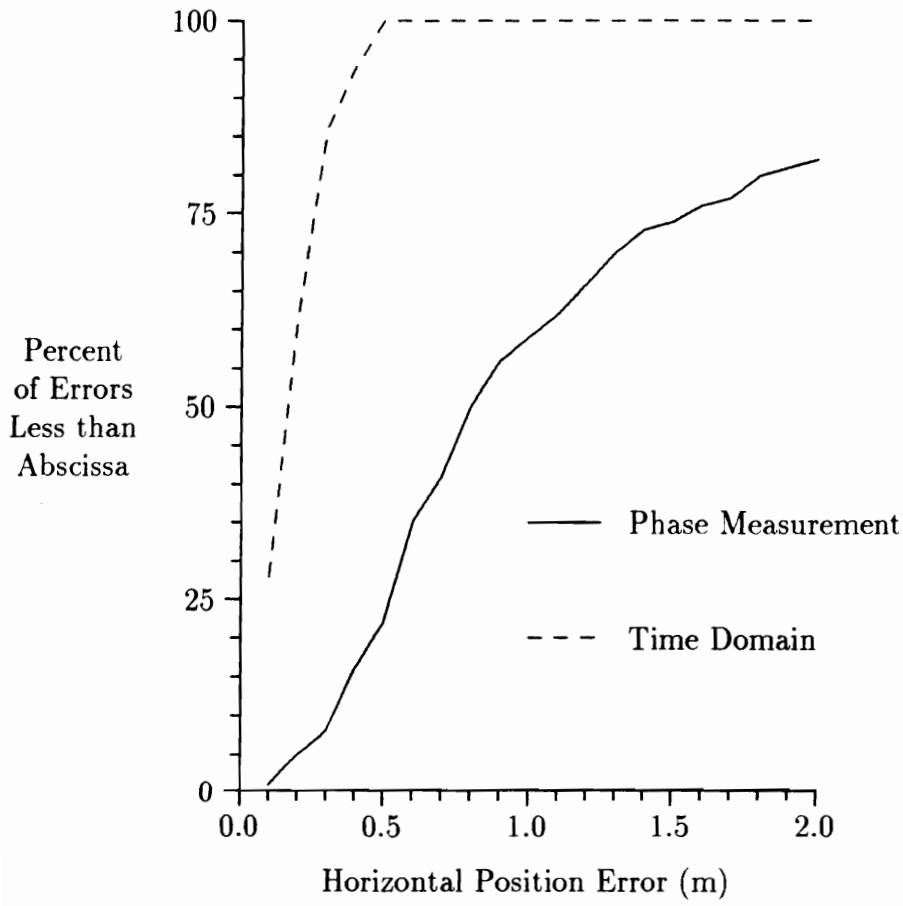


Figure 6.6: Time delay and phase measurement system position error distribution (computer simulation for identical multipath and noise conditions with LMSE estimation)

than the frequency hopping phase measurement system. By using different transmit PN codes for each user, a number of users could be accommodated at the expense of an increase in background noise. Theoretically, the frequency hopping technique could also be used in the same way but the dwell time necessary to make phase measurements is prohibitively large.

# Chapter 7

## Conclusions

The performance of two spread spectrum position location systems was examined using analytical, experimental, and computer simulation methods. The first implementation was a frequency hopping system relying on phase measurement to calculate position. The second was a direct sequence system using time delay measurement to estimate position.

Phase measurement type radio frequency position location systems have been discussed in general and in particular a frequency hopping implementation has been presented. A statistical computer simulation program has been developed which models quite effectively the position and phase errors in the experimental system. A series of experimental tests have been performed which serve to verify the validity of the computer work and provide an estimate of system accuracy.

The accuracy limitation for the system was found to be the characteristics of the radiowave propagation channel. Methods of improving the accuracy of

the system, using both signal processing and hardware approaches, have been studied. By median filtering the phase data and employing robust techniques for estimating the position, RMS errors can be improved by a factor of two over conventional least mean squared error solutions.

Phase measurement systems, with the restricted bandwidth available in the UHF range, are limited to accuracies of approximately one-half meter RMS horizontal errors. Accuracies in this range are suitable for a number of tasks in construction automation; for example, lay-down yard control, earth-moving quality control, and equipment optimization. Applications in other industries, such as open-pit mining, farming, and airport ground control, could also benefit from such a system.

Geometric limitations on the placement of receivers cause the vertical errors to exceed horizontal errors by a factor of between 2 and 4 times. Alternate methods, such as laser-planes or barometers, can be used to determine the vertical coordinate to accuracies of better than a meter.

For the first time, phase error distributions of phase multilateration systems operating in multipath channels have been characterized and modeled. The errors can be roughly approximated by a Gaussian distribution but caution must be used. The experimental and simulated phase errors differ from Gaussian behavior in two significant ways: the tails of the distribution exceed those predicted by the Gaussian distribution and there is a dip present near the mean value which is not present in the Gaussian distribution.

Theoretical as well as computer simulations have shown that direct sequence systems which use time delay measurement can greatly improve the final accuracy of a radio frequency position location system. Accuracies of

one-quarter meter horizontal RMS are possible due to the increased multipath immunity of the pulse spread spectrum approach. A hybrid system which combines time delay for course position and phase information for fine position has the potential for excellent accuracy even in the presence of multipath interference.

There are several areas in which further research is needed to answer questions that remain. First, the computer simulation program can be expanded to more accurately model the real world environment. The spatial correlation of the multipath channel can be included in the model to determine how the position error varies over small distances. This would allow the simulation to analyze the effectiveness of close-spaced antenna diversity and geometric beacon arrays. The simulation can be expanded to include the Doppler shift influence for moving transmitters or moving reflectors.

Further experimental work is necessary to test the direct sequence system and verify the error performance predicted by the simulation work. The simulation of the direct sequence system can be expanded to include the errors induced by the hardware, such as timing and sampling errors in the analog correlator.

Other topics can be investigated if the operating band of the system is moved up into the microwave or millimeter wave portion of the spectrum. In this case the system could feasibly occupy the 500 MHz of spectrum necessary to resolve the direct and multipath components. The ability to resolve more accurately the direct path is necessary if further improvements in the position estimate are to be achieved.

# Bibliography

- [1] Y. J. Beliveau, E. J. Lundberg, M. J. Feuerstein, and T. Pratt, "Applications of local area positioning systems for construction operations," in *Record of Land Navigation and Information Systems Int. Conf.*, (Coventry, United Kingdom), September 1990.
- [2] E. J. Lundberg and Y. J. Beliveau, "Automated lay-down yard control system," *J. of Construction Engineering and Management*, vol. 115, pp. 535–544, December 1989.
- [3] E. E. Halmos, Jr., "Operatorless equipment," *Constructor*, August 1986.
- [4] B. C. Paulson, Jr., "Automation and robotics for construction," *J. of Construction Engineering and Management*, vol. III, pp. 190–207, September 1985.
- [5] B. C. Paulson, Jr., "Automated control and robotics for construction," in *Small Computers in Construction*, pp. 9–19, 1984.
- [6] A. Warssawski and D. A. Sangrey, "Robotics in building construction," *J. of Construction Engineering and Management.*, vol. III, pp. 260–280, September 1985.
- [7] G. J. Geier, A. Cabak, and L. Sieh, "Design of an integrated navigation system for robotic vehicle applications," *NAVIGATION: J. of the Institute of Navigation*, vol. 34, pp. 325–336, Winter 1987–88.
- [8] C. D. McGillem and T. S. Rappaport, "A beacon navigation method for autonomous vehicles," *IEEE Trans. on Vehicular Technology*, vol. 38, pp. 132–139, August 1989.



- [9] J. H. Solberg and C. D. McGillem, "Factory transport using free-roaming robot vehicles," in *Proc. of Third Int. Conf. on Automated Materials Handling*, (Birmingham, United Kingdom), pp. 343-350, March 1986.
- [10] R. J. Palmer, "Test results of a precise, short range, RF navigational/positional system," in *Record of the Vehicle Navigation & Information Systems Conf.*, (Toronto, Ontario, Canada), pp. 151-155, September 1989.
- [11] R. J. Palmer, "Techniques of navigating in a farm field," *NAVIGATION: J. of the Institute of Navigation*, vol. 36, pp. 337-344, Winter 1989-90.
- [12] K. J. Bye, "Handover criteria and control in cellular and microcellular systems," in *Record of Fifth Int. Conf. on Mobile Radio and Personal Communications*, (Coventry, United Kingdom), pp. 94-98, December 1989.
- [13] E. Peralta, "Intelligent vehicle highway systems programs," in *Virginia Tech Intelligent Vehicle Highway Systems Conf.*, (Blacksburg, Virginia), April 1990.
- [14] U.S. Department of Transportation, "Report to congress on intelligent vehicle-highway systems," Document Number DOT-P-37-90-1, March 1990.
- [15] R. Ahola and R. Myllyla, "A time-of-flight laser receiver for moving objects," *IEEE Trans. on Instrumentation and Measurement*, vol. M-3S, pp. 216-221, June 1986.
- [16] G. J. Sonnenberg, *Radar and Electronic Navigation*. London: Butterworths, sixth ed., 1988.
- [17] L. Tetley and D. Calcutt, *Electronic Aids to Navigation*. London: Edward Arnold, 1986.
- [18] *Global Positioning System: Papers Published in NAVIGATION*. Washington: The Institute of Navigation, 1980-1986. Vols. I,II,III. ←

- [19] E. O. Frye, "GPS signal availability in land mobile applications," *NAVIGATION: J. of the Institute of Navigation*, vol. 36, pp. 287-301, Fall 1989.
- [20] G. T. Kremer, R. M. Kalafus, P. V. W. Loomis, and J. C. Reynolds, "The effect of Selective Availability on differential GPS corrections," *NAVIGATION: J. of the Institute of Navigation*, vol. 37, pp. 39-52, Spring 1990. α
- [21] D. Perlstein and M. C. Fernandez, "Designing and implementing automatic vehicle location," *Mobile Radio Technology*, January 1989.
- [22] M. R. Foster, "Vehicle location for route guidance," in *Record of the Vehicle Navigation & Information Systems Conf.*, (Toronto, Ontario, Canada), pp. 11-16, September 1989.
- [23] J. A. Pierce, A. A. McKenzie, and R. H. Woodward, *LORAN: Long Range Navigation*. Vol. 4 of *Radiation Laboratory Series*, New York: McGraw-Hill Book Company, 1948.
- [24] S. H. Roth, "History of automatic vehicle monitoring," *IEEE Trans. on Vehicular Technology*, vol. VT-26, pp. 2-6, February 1977.
- [25] S. Riter and J. McCoy, "Automatic vehicle location—an overview," *IEEE Trans. on Vehicular Technology*, vol. VT-26, pp. 7-18, February 1977.
- [26] W. T. Warren, J. R. Whitten, R. E. Anderson, and M. A. Merigo, "Vehicle location system experiment," *IEEE Trans. on Vehicular Technology*, vol. VT-21, pp. 92-101, August 1972.
- [27] S. R. Sampson, "A survey of commercially available positioning systems," *NAVIGATION: J. of the Institute of Navigation*, vol. 32, pp. 139-148, Summer 1985.
- [28] M. Kayton, *Navigation Land, Sea, Air, & Space*. New York: IEEE Press, 1990. α
- [29] J. R. Huddle, "Historical perspective on estimation techniques for position and gravity survey with inertial systems," *AIAA J. Guidance*, vol. 9, pp. 257-267, May-June 1986.

- [30] L. A. Whitcomb, "Using low-cost magnetic sensors on magnetically hostile land vehicles," in *Rec. IEEE Position Location and Navigation Symp.*, pp. 34–38, 1988.
- [31] J. K. Roberts and A. R. Miller, *Heat and Thermodynamics*. New York: Interscience, 1960.
- [32] D. Ensminger, *Ultrasonics: The Low and High Intensity Applications*. New York: Marcel Dekker, Inc., 1973.
- [33] J. R. Frederick, *Ultrasonic Engineering*. New York: John Wiley & Sons, Inc., 1965.
- [34] L. E. Kinsler and A. R. Frey, *Fundamentals of Acoustics*. New York: John Wiley & Sons, Inc., 2nd ed., 1962.
- [35] L. J. Sivian *J. Acoustical Society of America*, vol. 19, p. 914, 1947.
- [36] J. H. Shapiro, "Burst-mode atmospheric optical communication," in *Proc. of IEEE National Telecommunications Conference*, pp. 27.5.1–27.5.7, 1980.
- [37] J. H. Shapiro, "Optical communications through low-visibility weather conditions," in *Proc. of IEEE National Telecommunications Conference*, pp. 27.6.1–27.6.6, 1980.
- [38] R. S. Kennedy, "Communications through optical scattering channels: an introduction," *Proc. of the IEEE*, vol. 58, p. 1651, 1970.
- [39] N. S. Kopieka and J. Bordogna, "Background noise in optical communications systems," *Proc. of the IEEE*, vol. 58, p. 1571, 1970.
- [40] J. H. Shapiro, "Imaging and optical communication through atmospheric turbulence," in *Laser Beam Propagation in the Atmosphere*, (J. W. Strohbehn, ed.), Berlin: Springer-Verlag, 1978. Chapter 6.
- [41] Y. Castanet, "High speed infrared tracking," in *High Precision Navigation*, (K. Linkwitz and U. Hangleiter, eds.), pp. 410–418, Berlin: Springer-Verlag, 1989.

- [42] W. Mohlenbrink, "Realtime surveying in close range area with inertial navigation systems and optical target tracking techniques," in *High Precision Navigation*, (K. Linkwitz and U. Hangleiter, eds.), pp. 443–455, Berlin: Springer-Verlag, 1989.
- [43] G. Bayer, "Motorized electronic theodolites—high precision measurement robots for surveying," in *High Precision Navigation*, (K. Linkwitz and U. Hangleiter, eds.), pp. 395–408, Berlin: Springer-Verlag, 1989.
- [44] P. Krzystek, "High precision surveying of moving objects by electronic cameras," in *High Precision Navigation*, (K. Linkwitz and U. Hangleiter, eds.), pp. 419–431, Berlin: Springer-Verlag, 1989.
- [45] G. Bayer, P. Krzystek, and W. Mohlenbrink, "Realtime positioning of moving objects by dynamic target tracking," in *Record 16th International Congress of Photogrammetry and Remote Sensing, Com. V*, (Kyoto, Japan), July 1988.
- [46] F. H. Moffitt and H. Bouchard, *Surveying*. New York: Harper & Row, 8th ed., 1987.
- [47] T. J. M. Kennie and G. Petrie, *Engineering Surveying Technology*. New York: John Wiley & Sons, Inc., 1990.
- [48] M. I. Skolnik, *Introduction to Radar Systems*. New York: McGraw-Hill Book Company, second ed., 1980.
- [49] S. E. Lipsky, *Microwave Passive Direction Finding*. New York: John Wiley & Sons, 1987.
- [50] W. H. Guier and G. C. Weiffenbach, "A satellite Doppler navigation system," *Proc. of the IRE*, pp. 507–516, April 1960.
- [51] H. D. Black, "Satellites for earth surveying and ocean navigating," *Johns Hopkins Applied Physics Lab Technical Digest*, pp. 3–13, November 1981.
- [52] R. R. Newton, "The near-term potential of Doppler location," *Johns Hopkins Applied Physics Lab Technical Digest*, pp. 16–31, November 1981.

- [53] W. R. Miller, "Protest Selective Availability," *GPS World*, p. 12, July/August 1990.
- [54] E. G. Blackwell, "Overview of differential GPS methods," *NAVIGATION: J. of the Institute of Navigation*, vol. 32, pp. 114–125, Summer 1985. α
- [55] B. W. Remondi, "Performing centimeter-level surveys in seconds with GPS carrier phase: initial results," *NAVIGATION: J. of the Institute of Navigation*, vol. 32, pp. 386–400, Winter 1985–86.
- [56] R. V. C. Wong, K. P. Schwarz, and M. E. Cannon, "High-accuracy kinematic positioning by GPS-INS," *NAVIGATION: J. of the Institute of Navigation*, vol. 35, pp. 275–286, Summer 1988.
- [57] V. Held and K. D. Kricke, "GPS satellite navigation in the urban environment," *J. Royal Institute of Navigation*. August 1985.
- [58] S. A. Dale, I. D. Kitching, and P. Daly, "Position-fixing using the USSR's GLONASS C/A Code," in *Rec. IEEE Position Location and Navigation Symp.*, pp. 13–20, 1988. ^
- [59] S. C. Wu and V. J. Ondrasik, "Orbit determination of low-altitude earth satellites using GPS RF Doppler," in *Rec. IEEE Position Location and Navigation Symp.*, pp. 85–91, December 1982.
- [60] W. I. Bertiger and T. P. Yunck, "The limits of direct satellite tracking with GPS," *NAVIGATION: J. of the Institute of Navigation*, vol. 37, pp. 65–79, Spring 1990.
- [61] T. P. Yunck, S. Wu, and J. Wu, "Strategies for sub-decimeter satellite tracking with GPS," in *Rec. IEEE Position Location and Navigation Symp.*, pp. 122–128, November 1986.
- [62] H. B. Lee, "Accuracy limitations of hyperbolic multilateration systems," *IEEE Trans. on Aerospace and Electronic Systems*, vol. AES-11, pp. 16–29, January 1975. α
- [63] L. J. Cortland, "Loran-C vehicle tracking in Detroit's public safety dispatch system," *NAVIGATION: J. of the Institute of Navigation*, vol. 36, pp. 223–233, Fall 1989.

- [64] A. H. Riccio, "Automatic vehicle location Dallas test results," in *Rec. of Papers Presented at IEEE Vehicular Technology Conf.*, (Orlando, Florida), pp. 144-152, March 16-18 1977.
- [65] J. S. Ludwick, Jr., "Analysis of test data from an Automatic Vehicle Monitoring (AVM) test," in *Rec. of Papers Presented at IEEE Vehicular Technology Conf.*, (Denver, Colorado), pp. 23-33, March 22-24 1978.
- [66] E. N. Skomal, "Comparative analysis of six commercially available Automatic Vehicle Location (AVL) systems," in *Rec. of Papers Presented at IEEE Vehicular Technology Conf.*, (Denver, Colorado), pp. 34-45, March 22-24 1978.
- [67] R. J. Palmer and L. Fischer, "Short range ranging system," United States Patent, May 1989. Number 4,833,480.
- [68] J. A. Kivett and R. E. Cook, "Enhancing Position Location Reporting System (PLRS) user-to-user data capability," in *Rec. IEEE Position Location and Navigation Symp.*, pp. 154-161, 1986.
- [69] W. R. Fried, "Principles and simulation of JTIDS relative navigation," *IEEE Trans. on Aerospace Systems*, vol. 14, pp. 76-84, January 1978.
- [70] W. S. Widnall and J. F. Kelley, "JTIDS relative navigation with measurement sharing: design and performance," *IEEE Trans. on Aerospace and Electronic Systems*, vol. 22, pp. 146-154, March 1986.
- [71] S. S. Haykin, "Radar array processing for angle of arrival estimation," in *Array Signal Processing*, ch. 4, Englewood Cliffs, New Jersey: Prentice-Hall, Inc., 1985.
- [72] P. S. Jorgensen, "Navstar/Global Positioning System 18-satellite constellations," in *Global Positioning System: Papers Published in NAVIGATION*, Washington: The Institute of Navigation, 1984. Vol. II.
- [73] C. Powell, "Hyperbolic navigation," in *NAVIGATION SYSTEMS: A Survey of Modern Electronic Aids*, (G. E. Beck, ed.), ch. 4, London: Van Nostrand Reinhold, 1971. ~

- [74] C. Powell and A. R. Woods, "Lambda: a radio aid to hydrographic surveying," *J. of British IRE*, pp. 479–492, June 1963.
- [75] D. L. Nicholson, "Multipath sensitivity of a linearized algorithm used in time-difference-of-arrival location systems," in *Digest of Int. Electrical and Electronics Conference and Exposition*, October 1973. Paper No. 73252.
- [76] R. C. Hatch, "Hyperbolic positioning is per se passe," in *Rec. IEEE Position Location and Navigation Symp.*, (San Diego), pp. 51–58, 1978.
- [77] D. L. Nicholson, "Multipath and ducting tolerant location techniques for automatic vehicle location systems," in *Rec. of Papers Presented at IEEE Vehicular Technology Conf.*, (Washington, DC), pp. 151–154, March 24–26 1976.
- [78] H. B. Lee, "A novel procedure for assessing the accuracy of hyperbolic multilateration systems," *IEEE Trans. on Aerospace and Electronic Systems*, vol. AES-11, pp. 2–15, January 1975.
- [79] W. C. Rheinboldt, *Methods of Solving Systems of Nonlinear Equations*. Philadelphia: Society for Industrial and Applied Mathematics, 1974.
- [80] J. L. Jaech, *Statistical Analysis of Measurement Errors*. New York: John Wiley & Sons, 1985.
- [81] A. A. Clifford, *Multivariate Error Analysis*. London: Applied Science Publishers Ltd., 1973.
- [82] P. Bloomfield and W. L. Steiger, *Least Absolute Deviations*. Boston: Birkhauser, 1983.
- [83] W. H. Press, B. P. Flannery, S. A. Teukolsky, and W. T. Vetterling, *NUMERICAL RECIPES: The Art of Scientific Computing*. Cambridge: Cambridge University Press, 1987.
- [84] B. M. Shchigolev, *Mathematical Analysis of Observations*. London: Iliffe Books Limited, 1965.
- [85] P. J. Huber, *Robust Statistics*. New York: John Wiley & Sons, 1981.

- [86] P. J. Huber, *Robust Statistical Procedures*. Philadelphia: Society for Industrial and Applied Mathematics, 1977.
- [87] W. J. J. Rey, *Robust Statistical Procedures*. Berlin: Springer-Verlag, 1978.
- [88] A. Dunworth, "The resolution of ambiguities in electronic navigational aids," *Proc. of IEE Australia*, pp. 223–232, 1966.
- [89] R. C. Dixon, *Spread Spectrum Systems*. New York: John Wiley & Sons, 1976.
- [90] W. R. Young, Jr. and L. Y. Lacy, "Echoes in transmission at 450 Megacycles from land-to-car radio units," *Proc. IRE*, vol. 38, pp. 255–258, March 1950.
- [91] J. Shefer, "Propagation statistics of 900 MHz and 450 MHz signals inside buildings," in *Microwave Mobile Radio Symp.*, (Boulder, Colorado), 1973.
- [92] R. J. C. Bultitude, "Measurement, characterization and modeling of indoor 800/900 MHz radio channels for digital communications," *IEEE Communications Magazine*, vol. 25, pp. 5–12, June 1987.
- [93] D. M. J. Devasirvatham, "Time delay spread and signal level measurements of 850 MHz radio waves in building environments," *IEEE Trans. on Antennas and Propagation*, vol. AP-34, pp. 1300–1308, November 1986.
- [94] T. S. Rappaport, "Characterization of UHF multipath radio channels in factory buildings," *IEEE Trans. on Antennas and Propagation*, vol. 37, pp. 1058–1069, August 1989.
- [95] A. A. M. Saleh and A. J. Valenzuela, "A statistical model for indoor multipath propagation," *IEEE Journ. Selected Areas in Communications*, vol. SAC-5, pp. 138–146, February 1987.
- [96] S. J. Patsiakos, B. K. Johnson, and J. L. Dailing, "Propagation of radio signals inside buildings at 150, 450, and 850 MHz," in *Rec. of Papers Presented at IEEE Vehicular Technology Conf.*, pp. 66–71, 1986.



- [97] G. L. Turin, F. D. Clapp, T. L. Johnston, S. B. Fine, and D. Lavry, "A statistical model of urban multipath propagation," *IEEE Trans. on Vehicular Technology*, vol. VT-21, pp. 1-9, February 1972.
- [98] G. L. Turin, W. S. Jewell, and T. L. Johnston, "Simulation of urban vehicle-monitoring systems," *IEEE Trans. on Vehicular Technology*, vol. VT-21, pp. 9-16, February 1972.
- [99] J. LaBel, "Mobile radio signal statistics in non-urban environments," in *Rec. of Papers Presented at IEEE Vehicular Technology Conf.*, pp. 131-136, 1987.
- [100] R. E. Collin, *Antennas and Radiowave Propagation*. New York: McGraw-Hill Book Company, 1985.
- [101] D. S. Purdy, *An Application of the Hyperbolic Navigation Radio System for Automated Position and Control*. Master's thesis, Virginia Polytechnic Institute, Blacksburg, Virginia, August 1989.
- [102] W. Cheney and D. Kincaid, *Numerical Mathematics and Computing*. California: Brooks-Cole Publishing Company, 1980.
- [103] T. S. Rappaport and C. D. McGillem, "UHF fading in factories," *IEEE J. on Selected Areas in Communications*, pp. 40-48, January 1989.
- [104] Njoku and Kong, "Theory for passive microwave remote sensing of near-surface soil moisture," *J. Geophysics Research*, vol. 82, pp. 3108-3118.
- [105] H. Hashemi, "Simulation of the urban radio propagation channel," *IEEE Trans. on Vehicular Technology*, vol. VT-28, pp. 213-225, August 1979.
- [106] D. C. Cox, R. R. Murray, and A. W. Norris, "800 MHz attenuations measured in and around suburban homes," *AT&T Bell Laboratories Tech. Journ.*, vol. 6, pp. 921-954, July 1984.
- [107] T. S. Rappaport, "Wireless information links for the factory of the future," in *Society for Integrated Manufacturing Conf. Proc.*, (Atlanta, Georgia), pp. 643-650, November 1989.

- [108] T. S. Rappaport, "Radio channel modeling in manufacturing environments." Intermediate Report of Research, Virginia Polytechnic Institute and State University, Blacksburg, Virginia, December 1988. Part I.
- [109] H. Suzuki, "A statistical model for urban radio propagation," *IEEE Trans. on Communications*, vol. COM-25, pp. 673-680, July 1977.
- [110] H. F. Schmid, "A prediction model for multipath propagation of pulse signals at VHF and UHF over irregular terrain," *IEEE Trans. on Antennas and Propagation*, vol. AP-18, pp. 253-258, March 1970.
- [111] G. R. Cooper and C. D. McGillem, *Probabilistic Methods of Signal and System Analysis*. Fort Worth: Holt, Rinehart, and Winston, Inc., second ed., 1986.
- [112] M. J. Feuerstein, Y. J. Beliveau, T. S. Rappaport, and T. Pratt, "Distribution of phase errors in a UHF position location system," *Electronics Letters*, vol. 25, pp. 1086-1088, 3rd August 1989.
- [113] D. C. Cooper, "Statistical analysis of position-fixing general theory for systems with gaussian errors," *Proc. of the IEE*, vol. 119, pp. 637-640, June 1972.
- [114] N. Marchand, "Error distributions of best estimate of position from multiple time difference hyperbolic networks," *IEEE Trans. on Aerospace and Navigation Electronics*, pp. 96-100, June 1964.
- [115] M. J. Feuerstein, T. Pratt, and Y. J. Beliveau, "A precision automatic vehicle location system for use in constuction automation," in *Record of the Vehicle Navigation & Information Systems Conf.*, (Toronto, Ontario, Canada), pp. 202-205, September 1989.
- [116] J. M. Tranquilla and S. R. Best, "Antenna phase center measurement in UHF radio positioning systems," *Canadian Electrical Engineering J.*, vol. 12, pp. 11-18, January 1987.
- [117] J. M. Tranquilla and S. R. Best, "Approach to a suitable directional antenna for UHF radio positioning applications," *Canadian Electrical Engineering J.*, vol. 12, pp. 112-115, March 1987.

- [118] S. B. Rhee, "Relative performance of omni and directional antenna in an urban environment," in *Rec. IEEE Position Location and Navigation Symp.*, pp. 62–65, 1986.
- [119] T. Lo and J. Litva, "Adaptive beam space nulling of multipath signals," *IEEE Trans. on Antennas and Propagation*, vol. 38, pp. 129–134, January 1990.
- [120] R. G. Vaughan and J. B. Andersen, "Antenna diversity in mobile communications," *IEEE Trans. on Vehicular Technology*, vol. VT-36, pp. 149–172, November 1987.
- [121] E. N. Gilbert, "Energy reception for mobile radio," *Bell System Technical J.*, pp. 1779–1803, October 1965.
- [122] M. Schwartz and L. Shaw, *Signal Processing*. New York: McGraw-Hill Book Company, 1975.
- [123] R. A. Singer and K. W. Behnke, "Real-time tracking filter evaluation and selection for tactical applications," *IEEE Trans. on Aerospace and Electronic Systems*, vol. AES-7, pp. 100–110, January 1971.
- [124] S. Z. Hsue and S. S. Soliman, "Multipath delay estimation using parallel processing," in *Proc. of IEEE Conf.*, pp. 199–204, 1988.
- [125] B. M. Horton, "Noise modulated distance measuring systems," *IRE Proc.*, pp. 821–828, May 1959.
- [126] R. Price and P. E. Green, Jr., "A communications technique for multipath channels," *IRE Proc.*, pp. 555–570, March 1958.
- [127] A. W. Rihackek and R. M. Golden, "Resolution performance of pulse trains with large time-bandwidth products," *IEEE Trans. Aerospace and Electronic Systems*, pp. 677–685, July 1971.
- [128] G. L. Turin, "Introduction to spread-spectrum anti-multipath techniques and their application to urban digital radio," *Proc. IEEE*, vol. 68, pp. 328–353, March 1980.
- [129] R. E. Ziemer and R. L. Peterson, *Digital Communications and Spread Spectrum Systems*. New York: Macmillan Publishing Company, 1985.

- [130] M. J. Feuerstein and T. Pratt, "A local area position location system," in *Record of Fifth Int. Conf. on Mobile Radio and Personal Communications*, (Coventry, United Kingdom), pp. 79–83, December 1989.
- [131] J. P. Fitch, E. J. Coyle, and N. C. Galligher, "The analog median filter," *IEEE Trans. on Circuits and Systems*, vol. CAS-33, pp. 94–102, January 1986.
- [132] J. W. Tukey, "Nonlinear (nonsuperposable) methods for smoothing data," in *Cong. Record EASCON*, p. 673, 1974.
- [133] L. R. Rabiner, M. R. Sambur, and C. E. Schmidt, "Applications of a nonlinear smoothing algorithm to speech processing," *IEEE Trans. on Acoustics, Speech, and Signal Processing*, vol. ASSP-23, pp. 552–557, December 1975.
- [134] N. C. Gallagher and G. L. Wise, "A theoretical analysis of the properties of median filters," *IEEE Trans. on Acoustics, Speech, and Signal Processing*, vol. ASSP-29, pp. 1136–1141, December 1981.
- [135] P. D. Wendt, E. J. Coyle, and N. C. Gallagher, "Some convergence properties of median filters," *IEEE Trans. on Circuits and Systems*, vol. CAS-33, pp. 276–286, March 1986.

# Appendix A

## Median Filtering Theory

Median filters have found wide use in image processing, speech processing, and communications in impulsive noise environments [131]. For certain types of signals, median filters can achieve results which are impossible for other filtering techniques. The chief advantage of median filtering is the ability to preserve signal edges while filtering out impulses. Median filters are particularly useful for performing operations on signals which contain sharp discontinuities (for example, rectangular pulses and sawtooth waveforms). These discontinuities contain high frequency energy and are indistinguishable spectrally from many types of noise components. Any linear filtering would perform a low pass smoothing operation and consequently smear the sharp edges as well as removing the noise.

In the mid-70's Tukey proposed a nonlinear smoothing algorithm which was capable of preserving sharp discontinuities and still able to remove superimposed noise [132]. The algorithm was based on a combination of a

running median and conventional linear smoothing. Rabiner, Sambur, and Schmidt later proposed median filtering in combination with linear smoothing for use in certain speech processing applications [133]. In 1981 Gallagher and Wise performed the first theoretical analysis of the properties of median filters [134] and since then applications of median filters have become widespread [131].

The phase versus frequency data collected by the phase measurement system is in the form of a sawtooth wave. This is because the measurement system can only respond to the phase residue (that is, the fractional phase) which is the result of dividing by the wavelength. Each ramp of the sawtooth waveform corresponds to a  $2\pi$  phase cycle with abrupt discontinuities occurring between adjacent phase cycles. The sawtooth nature of the phase data lends itself to use with a median filter for removing impulsive types of phase noise and for delineating the discontinuities.

Wendt has indicated that the sawtooth like waveforms often present in television signals are invariant to median filtering (that is, they are not distorted by the filter) [135]. This is important because it indicates that the slope of the phase versus frequency data will be preserved in the filtering operation. The slope is extremely critical as it is used to calculate the ambiguity resolution estimates from the phase data set.

Discrete median filters are a special class of ranked-order digital filters used for smoothing signals. A median filter is a nonlinear filter consisting of a window with length  $2M + 1$  that slides over a sampled signal of length  $L$ . The basic idea is to rank the samples in the window in numerical order and select the median value as the filter output. Assume that the input to the

filter is a sequence of length  $L$  which takes on the value  $a(i)$  at position  $i$  where  $1 \leq i \leq L$ . The output of the filter at the  $i^{\text{th}}$  point would be

$$b(i) = \text{MEDIAN}\{a(i-M), \dots, a(i), \dots, a(i+M)\}$$

One property of a median filter is that, unlike a linear filter, it is not completely characterized by the impulse response. In fact, the response of a median filter to a Dirac delta function is identically zero [131].

The selection of the window length  $M$  for the median filter is the only parameter which must be considered. By a very loose analogy, one can think of the median filter as a low pass filter with a passband which gets narrower as the window size increases [134]. The general trend is the longer the median window the more it smooths out changes in the input. The size of the running median window is dependent on the minimum duration of any discontinuities which must be preserved [133]. The chief disadvantage of the longer median windows is the ‘clipping’ or ‘saturation’ effect which occurs at signal transitions smaller than the window width [133]. The choice of window width is then a tradeoff between the amount of smoothing which is necessary and the desire to avoid clipping of the signal.

The undesired clipping of the signal at transitions smaller than the window width can be avoided by taking advantage of prior knowledge of the signal characteristics. The transitions in the phase versus frequency signal always occur as discontinuities from  $-\pi$  to  $\pi$  or conversely from  $\pi$  to  $-\pi$ . With this knowledge, a special median operator has been developed which does not clip the transitions at these points. The modified median operator sorts the data within the window in a different order than the usual median

operator. The data within the window are first forced to be contiguous (that is, forced to be within the same phase cycle) by the appropriate addition or subtraction of  $2\pi$  radians. If the window contains only data from the same phase cycle (no transition is present) then no addition or subtraction is performed. A ranking operation is then performed and the median of these ranked data is extracted as the midpoint. The fractional phase is then recovered by dividing by  $2\pi$  and retaining only the residue.



# Appendix B

## Maximum Likelihood Estimation

In Chapter 3 it was mentioned that each of the hyperbolic range equations has a residual defined as

$$\delta_i = f_i(x, y, z; x_1, y_1, z_1; x_i, y_i, z_i) - \Delta R_{1i} \quad (\text{B.1})$$

where  $f_i$  are nonlinear functions of  $x, y, z$ ;  $\Delta R_{1i}$  are the measured range differences;  $i = 2, 3, \dots, N$ ; and  $N$  is the number of base stations. The question of obtaining an optimum solution to this system of equations boils down to the following: Given a set of measured range differences  $\{R_{1i}\}$  and known base station locations  $\{x_i, y_i, z_i\}$  then what is the probability that the estimated parameters  $(x, y, z)$  actually correspond to the true position? If this question can be answered then it may be possible to maximize the probability that the estimated parameters correspond to the actual position given a particular set of measured data. This is the concept behind the

*maximum likelihood* form of parameter estimation.

Suppose that each residual  $\delta_i$  is independently random with some as yet undefined probability distribution  $p(\delta_i)$ . For the moment, assume that the residuals have a zero mean value. It has been found experimentally and through simulation that this is not a bad assumption and the data can be detrended should a non-zero mean exist [112, 64].

The joint probability for the data set  $p(\delta_2, \dots, \delta_N)$  in some small neighborhood  $[\epsilon, -\epsilon]$  is the quantity which must be maximized. Because the individual probabilities are independent the joint probability is the product of the individual probabilities for each base station pair (that is,  $p(\delta_1, \delta_2) = p(\delta_1)p(\delta_2)$ ). Therefore

$$\underset{x, y, z}{MAX} p(\delta_2, \dots, \delta_N) = \prod_{i=2}^N \int_{-\epsilon}^{+\epsilon} p(\delta_i/\sigma_i) d\delta_i \quad (B.2)$$

for convenience the residuals have been scaled by weight factors  $\sigma_i$  which are assigned to each base station pair. Maximizing the joint probability  $p(\delta_2, \dots, \delta_N)$  is equivalent to minimizing the negative of the logarithm of the joint probability or

$$\underset{x, y, z}{MAX} p(\delta_2, \dots, \delta_N) \Leftrightarrow \underset{x, y, z}{MIN} \sum_{i=2}^N -\log \left\{ \int_{-\epsilon}^{+\epsilon} p(\delta_i/\sigma_i) d\delta_i \right\}. \quad (B.3)$$

To simplify the notation, let the integration of the probability density over the small  $\epsilon$  region be implicitly assumed. If the negative of the logarithm of the individual probability is defined as  $\rho(\delta_i/\sigma_i) = -\log\{p(\delta_i/\sigma_i)\}$  then the maximum likelihood criterion becomes

$$\underset{x, y, z}{MIN} \sum_{i=2}^N \rho(\delta_i/\sigma_i) \quad (B.4)$$

where a single equation must be minimized. As discussed in Chapter 3, if equation B.4 is differentiated with respect to  $x$ ,  $y$ ,  $z$  and each equation set equal to zero then a set of three equations and three unknowns results.

The results that have just been developed pertain to residuals  $\delta_i$  with a general probability distribution  $p(\delta_i/\sigma_i)$ . Therefore, at least theoretically, a maximum likelihood estimate can be found for any specific probability distribution  $p$ . For the specific case of  $\delta_i$  being Gaussian distributed then

$$p(\delta_i/\sigma_i) \sim e^{-\delta_i^2/2\sigma_i^2} \quad \rho(\delta_i/\sigma_i) \sim \frac{\delta_i^2}{2\sigma_i^2}. \quad (\text{B.5})$$

By referring back to equation B.4, the maximum likelihood estimator for Gaussian distributed residuals involves minimizing a function of the squared residuals ( $\rho$ ). If the variances are all equal (that is,  $\sigma_i = \sigma$ ,  $i = 2, 3, \dots, N$ ) then the maximum likelihood estimate is the least squared error criterion. If the variances are unequal then the weighted least squares is maximum likelihood optimum with the weighting coefficients equal to the inverse of the variances.

If the residuals are instead distributed as a double-sided exponential then

$$p(\delta_i/\sigma_i) \sim e^{-|\delta_i/\sigma_i|} \quad \rho(\delta_i/\sigma_i) \sim |\delta_i/\sigma_i| \quad (\text{B.6})$$

and by examining equation B.4 the maximum likelihood estimator minimizes the mean absolute residual. The double-sided exponential distribution has tails which are exponentially decreasing but at a rate asymptotically much slower than the Gaussian distribution. For this reason the least absolute deviation estimate often performs better than the least squared error criterion in cases where the tails of the residual distribution exceed that predicted by the Gaussian distribution [82].

The idea behind robust estimation is to choose a probability distribution  $p$  with a corresponding  $\rho$  for distributions with tails far more extensive than Gaussian. For experimental work, as has been discussed, the extensive tails are often more realistic [83]. An estimator which is maximum likelihood for such a distribution may be suboptimal for the actual experimental residual distribution but it may perform better than the classic least squares estimate. A number of different candidate distributions have been proposed in the literature [83, 82, 86] and have been tested with the experimental and simulation data discussed in Chapters 4, 5, and 6. The function to be minimized  $\rho$  does not necessarily have to correspond to any specific probability distribution  $p$ ; this leaves almost infinite flexibility in choosing  $\rho$ .

One of the best candidates is that proposed by Bloomfield and Steiger [82], namely

$$\rho_c(\delta_i/\sigma_i) = \begin{cases} \delta_i^2/2\sigma_i^2 & |\delta_i| \leq c \\ c|\delta_i/\sigma_i| - c^2/2 & |\delta_i| > c \end{cases} \quad (\text{B.7})$$

where  $c$  is a constant. If  $c$  is large then  $\rho_c$  treats the residuals quadratically, while if  $c$  is small the residuals are treated linearly. Therefore, the least squares and least deviation estimates are limiting cases for the Bloomfield–Steiger distribution. For large values of  $c$  the estimator converges rapidly to a solution given almost any initial estimate. Small values of  $c$  have been found to produce better position estimates but a much more accurate initial estimate is required. For small values of  $c$  a two-pass operation is required: a first course estimate of position is used as an initial guess to the second pass. This two-step process takes more computational time than otherwise would be required but a more accurate estimate results.

Another distribution which provides useful robust estimates is the Cauchy distribution. For the Cauchy distribution the resulting function to be minimized is of the form

$$\rho_c(\delta_i/\sigma_i) = \log \left[ 1 + \frac{1}{2}(\delta_i/\sigma_i)^2 \right]. \quad (\text{B.8})$$

The large residuals are treated logarithmically in this case resulting in an estimator which is less sensitive to outliers.

Many of the robust estimators present more challenging problems to solve numerically than the least squares estimator does. This occurs because the derivatives of  $\rho$  are often discontinuous (for example,  $d\rho/d\delta_i = \text{sgn}(\delta_i)$  for the double-sided exponential distribution), which presents problems for any iterative or minimization routines [83]. The objective function to be minimized may often contain a number of local minima requiring that a good initial estimate be used if the global minimum is to be found [87].

# Appendix C

## Solving Nonlinear Systems

The discussions in Section 3.3.3 and in Appendix B have developed the techniques necessary to transform the system of  $(N - 1)$  hyperbolic range equations into a problem which can be handled with standard computational techniques. The transformation makes use of a specific optimality criterion such as least squares, least deviation, or other robust methods. The result is a single objective function which must be minimized in the three-dimensional solution space (for example, equations 3.26 or 3.28). Equivalently, a nonlinear  $3 \times 3$  system of normal equations (such as equations 3.27, 3.29, or 3.31) can be found by differentiating the objective function with respect to  $x$ ,  $y$ , and  $z$ .

Many different computational methods exist for solving  $M$  nonlinear equations in  $M$  unknowns. Also a number of standard numerical techniques have been developed for multidimensional minimization. There are certainly a number of important tradeoffs which must be made in terms of conver-

gence rates, initial estimates, and complexity. Since these topics are treated in many references on numerical techniques [83, 102], only a brief discussion will be given for those techniques which have been most useful in the experimental and simulation work.

The system of normal equations can be solved by adapting the familiar Newton's method to multiple dimensions. In this case the equations to be solved simultaneously are

$$f_i(x, y, z) = 0 \quad i = 1, 2, 3 \quad (\text{C.1})$$

where all the functions  $f_i$  are nonlinear in  $x$ ,  $y$ , and  $z$ . The functions  $f_i$  can be expanded in a Taylor series as

$$f_i(x + \Delta x, y + \Delta y, z + \Delta z) = f_i(x, y, z) + \frac{\partial f_i}{\partial x} \Delta x + \frac{\partial f_i}{\partial y} \Delta y + \frac{\partial f_i}{\partial z} \Delta z + O(\Delta x^2, \Delta y^2, \Delta z^2) \quad (\text{C.2})$$

If second and higher order terms are ignored then the set of linear corrections  $(\Delta x, \Delta y, \Delta z)$  move the functions  $f_i$  simultaneously closer to zero.

In matrix form this Newton iteration can be written as

$$\begin{bmatrix} \frac{\partial f_1}{\partial x} & \frac{\partial f_1}{\partial y} & \frac{\partial f_1}{\partial z} \\ \frac{\partial f_2}{\partial x} & \frac{\partial f_2}{\partial y} & \frac{\partial f_2}{\partial z} \\ \frac{\partial f_3}{\partial x} & \frac{\partial f_3}{\partial y} & \frac{\partial f_3}{\partial z} \end{bmatrix} \begin{bmatrix} \Delta x \\ \Delta y \\ \Delta z \end{bmatrix} = \begin{bmatrix} -f_1 \\ -f_2 \\ -f_3 \end{bmatrix} \quad (\text{C.3})$$

which is in the familiar linear system form  $\mathbf{Ax} = \mathbf{b}$ . The matrix  $\mathbf{A}$  can be inverted using Gaussian elimination or Lower-Upper (LU) decomposition. At each step in the iterative process the new estimates are formed from the incremental updates as  $x^{new} = x^{old} + \Delta x$  and similarly for  $y$  and  $z$ .

There is approximately a 2.5 to 1 speed advantage in performing the iterative matrix inversions rather than performing the three-dimensional minimization. The Newton-Raphson algorithm which has just been presented converges to better than 1 cm in approximately 100 ms on an IBM PS-2 computer with a 10 MHz 80286 and the associated math coprocessor. The update rate could be increased if the iterative portions of the code were converted to assembly language and optimized for execution speed.



# Vita

Martin Feuerstein was born in Memphis, Tennessee on December 1, 1962. He received the B.E. degree in Electrical Engineering and Mathematics from Vanderbilt University in 1984. In 1987 he obtained the M.S. degree in Electrical Engineering from Northwestern University, where he was a Research Assistant with the Microwave Research Lab. He is currently pursuing the Ph.D. degree at VPI&SU, where he has been a Research Assistant and Project Assistant with the Satellite Communications Group.

During 1981 he was employed by WKNO Television and Radio as a video engineer. He was a programmer/analyst with Business Partners, Inc. between 1983 and 1984. From 1984 to 1985 he was an engineer with the Advanced Communications Terminals Division of Northern Telecom. He is a member of Tau Beta Pi, Eta Kappa Nu, and Phi Kappa Phi. He belongs to the Institute of Electrical and Electronics Engineers (Antennas and Propagation, Vehicular Technology, and Communications Societies), the Audio Engineering Society, and the Institute of Navigation. In his spare time he enjoys camping, whitewater rafting, and gardening.

