

# Low Frequency Indoor Radiolocation

by

Matthew Stephen Reynolds

S.B., Massachusetts Institute of Technology (1998)  
M.Eng., Massachusetts Institute of Technology (1999)

Submitted to the Program in Media Arts and Sciences  
in partial fulfillment of the requirements for the degree of

Doctor of Philosophy

at the

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

February 2003

© Massachusetts Institute of Technology 2003. All rights reserved.

Author .....  
Program in Media Arts and Sciences  
January 10, 2003

Certified by.....  
Neil A. Gershenfeld  
Associate Professor of Media Arts and Sciences  
Thesis Supervisor

Accepted by .....  
Andrew Lippman  
Chairperson, Department Committee on Graduate Students

# Low Frequency Indoor Radiolocation

by

Matthew Stephen Reynolds

Submitted to the Program in Media Arts and Sciences  
on January 10, 2003, in partial fulfillment of the  
requirements for the degree of  
Doctor of Philosophy

## Abstract

This thesis concerns the application of electromagnetic wave propagation to the problem of indoor radiolocation. Determining the location of people and objects relative to their environment is crucial for asset tracking, security, and human-computer interface (HCI) applications. These applications may be as simple as tracking the location of a valuable shipping carton or detecting the theft of a laptop computer, or as complex as helping someone to find his or her way around an unfamiliar building.

Currently available technologies, such as GPS or differential GPS, can provide the position information to solve these problems as long as the people or objects to be tracked are outdoors, where the microwave radio signals from the 24 orbiting GPS satellites may be received, but there is an unmet demand for a similar system that works indoors, where the physics of microwave radio propagation results in greatly attenuated signals and correspondingly poor GPS reception.

This thesis suggests a novel means of solving these problems involving the precise measurement of signals whose wavelengths are comparable to the size of a building. It is shown that this “mid-field” frequency regime can provide useful propagation characteristics with very little fixed infrastructure. Using a wavelength of 150m, over 4000 amplitude and differential carrier phase measurements were taken in the Wiesner Building. Least-squares power law fits to that data over paths of up to 30m yield meter-class position estimates at 1KHz acquisition rates. The contributions of this thesis include detailed indoor propagation measurements, as well as candidate empirical and theoretical models for that data. Additionally, new types of high precision measurement instrumentation and high efficiency RF power amplifiers have been created to enable these measurements.

Thesis Supervisor: Neil A. Gershenfeld

Title: Associate Professor of Media Arts and Sciences

The thesis reader page goes here.

## Acknowledgments

I would like to begin by giving thanks for the unconditional love and support of my Mom, Dad, and brother Josh, who have taken such good care of me over the past near-decade at MIT. I would not have been able to survive such a long journey through the perils of academia without their support.

Neil Gershenfeld has been my friend and advisor for seven years. He took me in when I was a wretched undergraduate, gave me intellectual tools (and lab equipment) beyond my wildest dreams, polished my ego to a fine luster, and appropriately pushed me out several years later to do the same for others.

Neil, Joe Paradiso, and Richard Greenspan have given generously of their time and attention by serving on my thesis committee. In particular, Dr. Greenspan contributed many hours of experimental advice and to the editing and proofreading of this document.

Perhaps the best of Neil's (brilliant or crazy?) ideas was to assemble the current and former members of the Physics and Media research group. I hope to spend the rest of my life surrounded by such talented and thoughtful people. During my tenure in the group I have had the pleasure of working with Susan Bottari, Joe Paradiso, Josh Smith, Rich Fletcher, Bernd Schoner, Rehmi Post, Yael Magure, Ben Vigoda, Chris Turner, Chris Verplaetse, Femi Omojola, JP Strachan, Ed Boyden, Jason Taylor, Ben Recht, Manu Prakash, and Raffi Krikorian. Many other Media Lab friends have lent a supportive ear when times were tough, most notably Diana Young, Kelly E Dobson, and Wendy Ju.

I have had the pleasure of working alongside two very talented undergraduate/M.Eng researchers. The first was Joey Richards, extraordinary friend and extraordinary colleague, who deserves a medal for the many times he saved the Karamazov Battalion from certain doom. The second, Mike Krypel, deserves particular mention in these Acknowledgments as this research is as much his as mine. Mike's patience, dedication, and skill have made this thesis possible.

# Contents

<b>1</b>	<b>The Indoor Positioning Problem</b>	<b>12</b>
1.1	Introduction . . . . .	12
1.2	Limitations of GPS . . . . .	13
1.3	Non-GPS approaches to indoor positioning . . . . .	14
1.4	A new approach for indoor radiolocation . . . . .	19
1.5	Thesis Accomplishments . . . . .	20
1.5.1	Precision measurement apparatus . . . . .	20
1.5.2	Empirical propagation maps . . . . .	21
1.5.3	Propagation model . . . . .	22
1.6	Media Lab-specific context for this work . . . . .	22
<b>2</b>	<b>Prior Art: Position tracking for human-computer interfaces</b>	<b>24</b>
2.1	The importance of position data for HCI . . . . .	24
2.2	Existing 3D position tracking systems . . . . .	25
2.2.1	Infrared beacon systems . . . . .	25
2.2.2	Fine grained infrared tracking systems . . . . .	26
2.2.3	Ultrasonic position measurement devices . . . . .	27
2.2.4	Magnetic field motion capture devices . . . . .	28
2.3	The new approach . . . . .	29
2.3.1	Comparison with prior approaches . . . . .	31
2.4	Novel applications of the proposed system . . . . .	33
2.5	Conclusion . . . . .	35

<b>3</b>	<b>System Engineering</b>	<b>36</b>
3.1	Introduction . . . . .	36
3.2	Radio propagation and the choice of operating frequency . . . . .	36
3.3	Available bandwidth and waveform design . . . . .	38
3.4	Antenna design choice . . . . .	39
3.4.1	<i>E</i> -field antenna . . . . .	39
3.4.2	<i>H</i> -field antenna . . . . .	41
3.4.3	Radiated power . . . . .	43
3.4.4	Antenna radiation efficiency . . . . .	44
3.5	Noise . . . . .	44
3.5.1	Atmospheric noise . . . . .	45
3.5.2	Man-made noise . . . . .	46
3.5.3	Receiver noise figure . . . . .	47
3.6	Link budget . . . . .	48
3.6.1	Path loss . . . . .	48
3.7	Sources of timing uncertainty . . . . .	50
3.7.1	Clock jitter at receiver and transmitter . . . . .	50
3.7.2	SNR versus phase measurement accuracy . . . . .	52
3.7.3	Coherent averaging . . . . .	53
3.7.4	Effects of multipath propagation . . . . .	53
3.8	Geometric factors . . . . .	55
3.9	Conclusion . . . . .	56
<b>4</b>	<b>Design of Experiment and Apparatus</b>	<b>58</b>
4.1	Introduction . . . . .	58
4.2	Transmitter and transmitting antenna design . . . . .	60
4.2.1	Class A amplifiers . . . . .	60
4.2.2	Class C amplifiers . . . . .	62
4.2.3	Class E amplifiers . . . . .	64
4.2.4	Antenna design . . . . .	66

4.3	Receiver design . . . . .	67
4.3.1	Receiving antenna . . . . .	68
4.3.2	Filtering and downconversion . . . . .	69
4.3.3	IF-DSP: The software radio . . . . .	70
4.4	Conclusion . . . . .	71
<b>5</b>	<b>Signal processing methods</b>	<b>72</b>
5.1	Introduction . . . . .	72
5.2	Linear systems and estimators: the amplitude fit . . . . .	73
5.2.1	The linear least-squares estimator . . . . .	74
5.2.2	Bias and error covariance . . . . .	75
5.2.3	The Cramer-Rao bound . . . . .	76
5.2.4	Estimation, filtering, and smoothing . . . . .	78
5.2.5	Recursive filters . . . . .	78
5.2.6	Discrete Kalman filters . . . . .	79
5.3	Nonlinear systems and estimators: Generalizations . . . . .	82
5.4	The nonlinear estimator: phase estimation . . . . .	83
5.4.1	The sinusoidal estimation problem . . . . .	83
5.4.2	Cramer-Rao bounds for sinusoidal estimation . . . . .	84
5.4.3	A maximum-likelihood sinusoid estimator . . . . .	85
5.5	A time domain variant of the periodogram estimator . . . . .	88
5.6	Conclusions . . . . .	89
<b>6</b>	<b>Experimental Results and Analysis</b>	<b>90</b>
6.1	Introduction . . . . .	90
6.2	Calibration and performance of experimental apparatus . . . . .	92
6.2.1	Amplitude calibration and results . . . . .	93
6.2.2	Phase measurement characterization . . . . .	95
6.2.3	Calibration and characterization summary . . . . .	96
6.3	Channel noise measurement methodology . . . . .	96
6.3.1	Spatial and temporal analysis of noise data . . . . .	97

6.4	Amplitude and phase measurements . . . . .	100
6.4.1	Transmitter and receiver “truth” data . . . . .	100
6.5	Outdoor control experiment . . . . .	101
6.5.1	Raw amplitude and phase data . . . . .	102
6.5.2	Analysis of outdoor data . . . . .	103
6.6	Indoor experiments . . . . .	106
6.6.1	Limitations of these experiments . . . . .	107
6.6.2	Typical experiment I: Basement, west hallway . . . . .	109
6.6.3	Typical experiment II: 4th floor, south hallway . . . . .	112
6.7	Conclusions . . . . .	118
<b>7</b>	<b>A simplified indoor mid-field propagation model</b>	<b>120</b>
7.1	Introduction . . . . .	120
7.2	The wave solutions to Maxwell’s equations . . . . .	121
7.2.1	Reflection and refraction of plane waves . . . . .	122
7.3	Waves in structures . . . . .	124
7.3.1	TE and TM modes . . . . .	124
7.3.2	TEM modes and the parallel plate model . . . . .	125
7.4	The parallel plate model . . . . .	126
7.4.1	Qualitative fit to data . . . . .	127
7.5	Conclusions, and questions for the future . . . . .	128
<b>8</b>	<b>Conclusions</b>	<b>130</b>
8.1	Contributions of this thesis . . . . .	133
8.2	Further work . . . . .	134
8.2.1	Experiment design and operating frequency . . . . .	134
8.2.2	Testing in multiple buildings . . . . .	135
8.2.3	Propagation modeling . . . . .	135
8.2.4	System design and development . . . . .	136
8.3	Caution and Benediction . . . . .	136

# List of Figures

1-1	A taxonomy of indoor positioning systems . . . . .	15
2-1	Physical configuration of the new system, showing the transmitters in the corners of the building and a small number of receivers. . . . .	29
2-2	Block diagram of the Karamazov immersive musical system . . . . .	34
3-1	Comparison of positioning system operating wavelengths . . . . .	37
3-2	Effective circuit of a small loop antenna . . . . .	42
3-3	Atmospheric noise vs. frequency, from CCIR 322. . . . .	46
3-4	Man-made noise vs. frequency, from ITU Report 372. . . . .	47
4-1	Overview of experimental apparatus . . . . .	59
4-2	Transmitter clock and timing source . . . . .	61
4-3	Class A amplifier simplified schematic . . . . .	61
4-4	Class C amplifier schematic . . . . .	63
4-5	Class-E amplifier designed for this thesis . . . . .	64
4-6	Class-E drain voltage measured at $V_{dd} = 60V$ . . . . .	65
4-7	Class-E output voltage measured at $V_{dd} = 60V$ . . . . .	66
4-8	The author and the transmitting antenna . . . . .	67
4-9	Receiver signal flow, hardware . . . . .	68
4-10	Receiver signal flow, software IF processing . . . . .	70
5-1	FFT amplitude estimation performance with respect to the CRB for varying $N$ . Note that since amplitude and phase estimation are both $O(\frac{1}{\gamma N})$ the phase estimation plot is identical. . . . .	87

5-2	FFT frequency estimation performance with respect to the CRB for varying $N$ . . . . .	87
6-1	Typical amplitude calibration results . . . . .	93
6-2	Results of ten amplitude calibrations, taken over two months, showing apparatus gain drift . . . . .	94
6-3	Result of phase characterization showing conformance to DDS synthesizer phase . . . . .	95
6-4	Noise power, August 20, 2002, basement . . . . .	98
6-5	Noise power, July 25, 2002, fourth floor . . . . .	99
6-6	Outdoor measurement configuration . . . . .	102
6-7	A photograph of the outdoor test configuration . . . . .	103
6-8	Amplitude and phase versus transmitter-receiver distance, outdoors .	104
6-9	Amplitude fit to distance, outdoors . . . . .	105
6-10	Measurement coverage of the Wiesner Building . . . . .	107
6-11	A photograph of an indoor experiment in progress . . . . .	108
6-12	Schematic of basement experiment . . . . .	109
6-13	Amplitude and phase versus transmitter-receiver distance, basement .	110
6-14	Amplitude fit to distance, basement . . . . .	111
6-15	Fourth floor measurement configuration . . . . .	113
6-16	Amplitude and phase versus transmitter-receiver distance, fourth floor	114
6-17	Amplitude fit to distance, fourth floor . . . . .	115
7-1	Parallel plate waveguide properties . . . . .	127
7-2	Transmission line model . . . . .	128
7-3	Transmission line model results vs experiment . . . . .	129

# List of Tables

2.1	Comparison of position measurement systems . . . . .	32
3.1	System link budget . . . . .	49

# Chapter 1

## The Indoor Positioning Problem

### 1.1 Introduction

The subject of this thesis is the determination of the position of a person or object using electromagnetic radiation. Technological advances in this area have had a profound impact on society over the past 50 years, as radar systems, Loran, the Global Positioning System (GPS), optical surveying systems, and precise atomic timekeeping have affected almost every aspect of modern life by making it much easier to know where people and things are.

There are many important applications for position measurement systems. These applications range from the familiar geographic information system (GIS) applications, such as mapmaking or finding one's location on a map, to more complex applications like these:

- assistive navigation for disabled people
- personal navigation in new or unfamiliar surroundings
- emergency route finding for police, firefighters, and the military
- optimizing the delivery of goods or services
- logistics or supply-chain management

- robot navigation
- building more effective, large scale human-computer interfaces

For all of these applications, an accurate, wireless, automatic position measurement system is needed as the heart of a larger and more complex location aware information system. The current state of the art in electronic navigation systems is GPS, which over the last 30 years has already revolutionized the fields of surveying, GIS, and vehicle navigation, among others. GPS has been so successful because it is in effect a *utility*- an always-on, nearly ubiquitous system which can be freely used by anyone, anywhere outdoors, and at any time. In its current instantiation, GPS is essentially a public service for the world, provided free of charge to its users (but at considerable cost to the government of the United States).

Using a small and relatively inexpensive receiver, people and objects outdoors can determine their location anywhere on Earth with up to millimeter precision using complex systems that rely on almost all aspects of 20th century physics- from orbital dynamics, to the geodesy of the Earth, to the quantum mechanics describing atomic behavior, to the General Relativistic corrections applied to GPS's orbiting atomic clocks. Equally profound are the many societal benefits of this system, which include enabling more efficient agricultural production, rescuing lost hikers, surveying ancient archaeological sites, and assisting in ship and aircraft navigation in the most remote reaches of the Earth.

## 1.2 Limitations of GPS

Even with the incredible advances made possible by modern GPS technology, however, there is an unmet need for a reliable position measurement system that can work indoors, where the microwave radio signals used by GPS are greatly attenuated. This problem has taken on a new urgency since the Federal Communications Commission issued the Enhanced 911 (E911) directive requiring all cellular operators to provide automated handset location information to 911 call centers by 2005.

Because of the worldwide availability of GPS, and the maturity of GPS technology, GPS-based systems are currently the leading candidate to provide this indoor position information. However, the microwave radio signals used by GPS are weakened by absorption or distorted by multipath propagation when they are transmitted through ordinary building materials, such as steel, concrete, and energy efficient coated glass windows. The GPS signals become so weak that they cannot generally be received indoors without long duration coherent averaging and the use of additional information supplied out-of-band from an external data source. Recent developments in *augmented GPS*, for example those described in [29], suggest that augmented GPS may provide a partial solution to the indoor position measurement problem by increasing system processing gain by 15dB to 20dB, which would result in much better indoor performance.

Unfortunately, even augmented GPS systems are still subject to multipath propagation and signal coverage issues that will prevent them from solving all problems. Additionally, although GPS is a globally-available utility, it is controlled by the US Government, which retains the right to deny GPS as it pleases. Also, augmented GPS systems require additional infrastructure and may not be able to provide coverage deep within a concrete and steel building.

### **1.3 Non-GPS approaches to indoor positioning**

In addition to the GPS based approach mentioned above, many different non-GPS based approaches to the indoor positioning problem have been considered, both by this author and by many different commercial and research teams. None has yet shown good performance while being economically viable enough to be widely deployed. In the context of this thesis, Figure 1-1 provides an overall taxonomy of these approaches. Chapter 2 gives a detailed discussion of prior art in indoor position measurement for human-computer interface (HCI) applications. Additionally, journal references for the systems referred to in Figure 1-1 may be found in Chapter 2.

Evaluating the suitability of a particular position measurement system for a par-

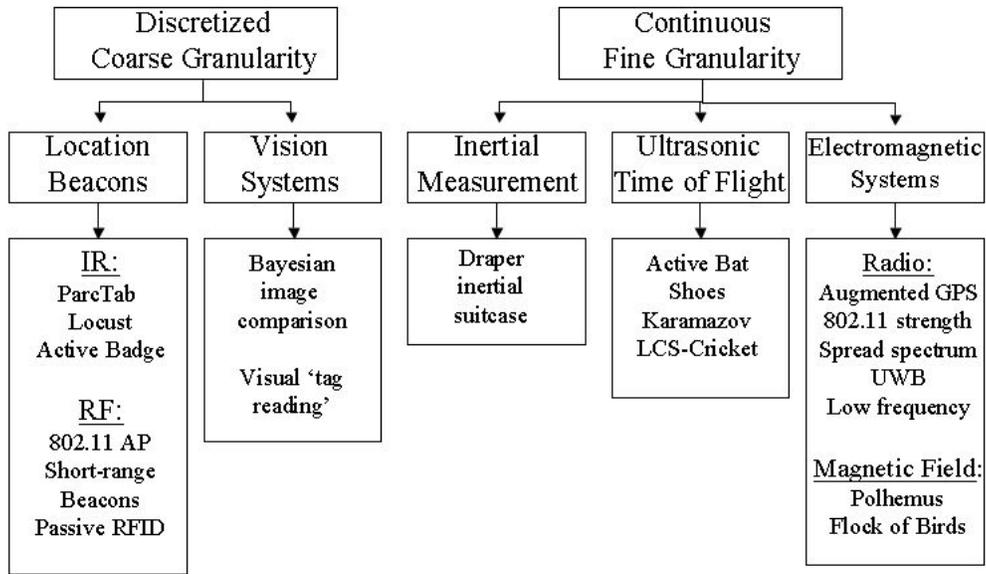


Figure 1-1: A taxonomy of indoor positioning systems

ticular application is a complex task. There are many criteria that can be used to evaluate candidate systems. Among the most important of these are the following:

### Granularity

The granularity of a position measurement system refers to the level of discretization of position solutions. Some systems (those in the left branch in Figure 1-1) produce results that are discrete, at the room level. Other systems (those in the right branch of Figure 1-1) have a continuous coordinate system and produce fine-grained position solutions. Note that a continuous-readout system is more general than a discretized system; one can always segment a continuous readout into an arbitrary discrete readout in a computationally efficient manner. The system proposed in this thesis is a continuous readout system.

## **Accuracy**

Accuracy is obviously a very important property of any position measurement system. There are three senses in which the word “accuracy” is generally meant, and to which it is usually imprecisely applied. The first refers to the absolute accuracy of the system; that is, the difference between the user’s actual position as determined by some “truth” data, and the reported result from the system under test. The second sense is precision, that is, to how many significant figures the result is reported. The third sense is “repeatability”, meaning the deviation between two measurements of a stationary person or object made at different times. Obviously the goal of any position measurement system is to be as accurate as possible in all senses of the word. The experiments presented in Chapter 6 indicate that meter-class position errors are achievable with the methods presented in this thesis.

## **Update rate**

The systems considered in this thesis are those capable of producing position solutions quickly enough to track motion on human time scales. It is generally agreed in the HCI community that update rates exceeding 30Hz to 60Hz satisfy this requirement (see Section 2.2.2 for more information). Each potential system has fundamental physical limits to update rate, which in turn may or may not be shared among multiple users in a particular region of space. Generally, optical and radio frequency systems are capable of very rapid update rates, well in excess of 30Hz, but there is a tradeoff between update rate and accuracy, especially when the user is moving. The system described here was designed around a 1KHz update rate, per user.

## **Device size, cost, and power**

Constraints on system size apply primarily to the user-side element of the system. The position measurement device must be small and unobtrusive in order to be well-accepted by potential users. Intimately connected with this notion of user acceptance is the requirement that the user-carried equipment should not consume excessive

power during operation; a battery powered unit should be able to operate for at least a full day on one battery charge.

Economic constraints apply to both the user-carried equipment and the required system infrastructure. In order for a system to achieve wide acceptance, user equipment cost should not exceed that of commonly used mobile computing equipment (under one or two thousand dollars, at 2002 prices) and would ideally be as low as possible.

### **Infrastructure requirements**

Infrastructure cost is a function both of equipment cost as well as installation and maintenance cost. Since some system designs would require infrastructure in every room of a large building, it is apparent that installation and maintenance costs could quickly exceed the fixed cost of equipment. It would be difficult and expensive to widely deploy a large amount of infrastructure in a large building, where there could be hundreds or thousands of different rooms or spaces requiring coverage. Additionally, system availability could be adversely affected if perfect functioning of all of the infrastructure were necessary for proper operation. For these reasons, the preferred systems are those with a few key pieces of infrastructure, which can serve a wide coverage area while being made to work very reliably. A basic system developed around the principles of this thesis would require only three transmitting stations, strategically placed in a building, to produce 3 dimensional position solutions. But due to signal strength and propagation considerations, it is almost always beneficial to overdetermine the problem by having redundant position information available from multiple transmitters.

### **Scalability for multiple users**

One of the key criteria for the viability of a positioning system is the ability to support multiple users simultaneously. Systems in which the fixed infrastructure transmits all the information required to produce a position solution, which each user receives and uses to perform the position solution computations independently, are preferable from

the scalability point of view. Clearly, if each receiving unit operates independently and does not uniquely consume any transmitting resource, an unlimited number of such receivers can be supported without penalty. This is the approach used in this thesis.

### **Bandwidth occupancy**

Another important constraint is system occupied bandwidth, regardless of the design particulars of any given system. Radio frequency spectrum is a precious resource, controlled by governments. Often these governments auction licenses for the use of radio spectrum; the cost for such licenses can reach into the billions of dollars. Therefore the least bandwidth required for the operation of a particular system should be used. It is also to be expected that if that system becomes widely deployed, interference between adjacent systems could become a problem. This interference can be minimized by a combination of spatial, temporal, and frequency separation between adjacent systems. Narrowband systems, like the one presented in this thesis, lend themselves to the use of frequency division multiple access (FDMA) or time division multiple access (TDMA) techniques. Wideband systems, like GPS, lend themselves to the use of code division multiple access (CDMA) and orthogonal frequency division multiplexing (OFDM). The choice of channel sharing methodology is intimately tied to the performance of the positioning system and must be considered during the initial system engineering work.

In addition to interference issues between adjacent positioning systems, interference with other systems making use of radio frequency spectrum may also arise. For this reason, the International Telecommunications Union (ITU) works as a global coordinating body among the regulatory agencies of different governments, such as the Federal Communications Commission (FCC) in the US, or Japan's Ministry of Posts and Telecommunication (MPT). Frequency bands are allocated to different services on an international basis, so the availability of a suitable frequency band for radionavigation has to be investigated. This issue is addressed in Chapter 3 of this thesis.

## Privacy

User privacy is an important consideration often overlooked by communication system engineers. Users may not accept a positioning system if they know that their whereabouts are always available to people or information systems that they do not trust. This is currently a topic of debate among cellular telephone users and telecom companies; because of the FCC's Enhanced 911 requirements, users will most likely be continually tracked, even when they are not calling 911. In addition to the obvious discomfort that users may feel if every movement is scrutinized, data mining and clustering techniques could be used to provide information beyond bathroom breaks and trips to the coffee machine- inferences about associations between people, their interactions, and other extremely personal details could be made on the basis of position information.

In the context of position measurement systems, there are two fundamental approaches to management of user position data. In the "centralized" approach, the position measurement infrastructure tracks each user independently, and makes this data available to other information systems from a central point. This type of system is most vulnerable to "Big Brother" privacy implications. The second type of system leaves the system infrastructure ignorant of user position; in these systems, users determine their own position relative to the fixed system infrastructure. They can then choose to keep this information to themselves, or to broadcast it on an out-of-band channel to a central system if they so choose. It is the latter type of system that is considered in this thesis.

## 1.4 A new approach for indoor radiolocation

This thesis approaches the problem of indoor position measurement by using what is believed to be a fundamentally different regime of electromagnetic wave propagation than existing indoor positioning systems. Instead of a "far-field" system, where the operating wavelength is very short in comparison to the dimensions of the building (like GPS, with its 19cm wavelength), or a "near-field" system with an operating

wavelength very long compared to the building (like the Polhemus type near-DC magnetic field systems), this thesis proposes the use of a “mid-field” system, where the wavelength is on the order of the dimensions of the building. The measurements presented in this thesis were performed at a wavelength of 150m, comparable to the longest dimension of the Wiesner Building, which is essentially a cube with sides of length 50m.

This mid-field frequency regime, the soft boundary between near-field and far-field propagation, is shown in this thesis to be suitable for a radio positioning system that can provide meter-class position solutions, at KHz update rates, with little fixed infrastructure. The primary goal of this thesis was to demonstrate the validity of this mid-field hypothesis by a combination of theory and experiment, thereby producing a “proof of concept” for a low frequency indoor radio positioning system based on the principle of narrowband amplitude and phase measurement.

## **1.5 Thesis Accomplishments**

This thesis has met three main goals: first, to produce a very accurate portable measurement apparatus of great precision for making propagation measurements; second, to make detailed empirical propagation maps in the Wiesner Building using that apparatus, and third, to produce a sketch of an explanatory model that can be used to guide further experimental work aimed at completely understanding indoor propagation at long wavelengths.

### **1.5.1 Precision measurement apparatus**

To accomplish the first goal, a test apparatus has been constructed that has been used to produce the first known indoor low frequency propagation map with amplitude and phase data accurate to well within 1dB and 1 degree over as much as 70dB of dynamic range, within 1 millisecond of acquisition time. This apparatus consists of a fixed-location transmitter unit based on a stable clock source and a high efficiency digital

power amplifier and integrated antenna which is of a novel high-efficiency Class E<sup>1</sup> design. In this experiment, a portable receiver was designed to record signal amplitude and phase with respect to a reference signal distributed by a fiber-optic cable trailed by the receiver. It is shown in Chapter 3 how this reference signal would be derived in a fully operational system without the use of the fiber-optic cable used in these experiments.

## 1.5.2 Empirical propagation maps

The second goal is to produce an empirical propagation map based on the data collected with the apparatus described above. The receiver has been wheeled around the Wiesner Building on a cart and the received signal parameters have been recorded along with the exact position of the receiver as manually surveyed with respect to a building blueprint. This has been done for six different transmitter locations and over 4000 different receiver locations. The signal amplitude and phase data, plotted with respect to the actual (surveyed) 3D position of the receiver, form a propagation map which can subsequently be used to determine the location of the mobile receiver. This is an important accomplishment because this propagation map can be used in conjunction with two other fixed location transmitters to obviate the need for a reference signal at the receiver, allowing the receiver to operate in an untethered mode. This map has been produced by post-processing the experimental data to produce an accurate map of signal propagation for a wide variety of locations in the building. Since the receiver is used in a transition region between the near-field and far-field propagation regions, and the expected variation in measured range is large and comparable to the wavelength, this map is necessarily complex because of propagation effects such as multipath and signal conduction along metallic objects.

---

<sup>1</sup>The “class” of an amplifier refers to the bias region of the active amplifying device (eg a power transistor). Typical Class A or Class B linear amplifiers are limited to efficiencies of about 50%-60%, while the nonlinear Class E amplifier designed for this experiment is capable of better than 90% efficiency. This design methodology results in a physically smaller, lower cost, and more spectrally pure transmitter in which all active devices operate in a digital mode. In so doing the phase shift through the amplifier may also be tightly controlled, which is absolutely necessary for the precise phase measurements made for this thesis.

The tradeoff between map complexity and position accuracy, position solution convergence, and computation load have been examined in the context of producing a small, self-contained, and inexpensive position determining receiver. Results presented in Chapter 6 show good signal quality at transmitter-receiver spatial separations of up to 30m, yielding amplitude-based range fits with a standard deviation of approximately 1-1.5m at 1KHz update rates.

### **1.5.3 Propagation model**

The third goal is to produce a candidate for a theoretical model that partially explains the measured data with respect to known obstructions in the signal path (eg steel beams or reinforcing metal in the building). The parallel plate waveguide model presented in this thesis accounts for certain features of the structure of the building, such as the metal reinforcing material used in the Wiesner Building’s floors and ceilings. It was not possible to find a first-principles analytical solution that was sufficiently general to be useful for such a complex propagation environment, while at the same time being sufficiently computationally tractable to be desirable for implementation in a hand-held device. However, in the future an analytical solution might be attempted on certain well-constrained subcases of the general problem, for comparison to the empirical data presented in this thesis.

## **1.6 Media Lab-specific context for this work**

There are many reasons why this work is uniquely relevant and well-suited for pursuit at the Media Lab. The most compelling of these are the many human-computer interface (HCI) applications for a generalized building-wide indoor positioning system. At the Media Lab, applications have already been proposed in the fields of tangible interfaces [25], wearable computing and augmented intelligence [38], and many others which require fine-grained position data with update rates suitable for building dynamic human-computer interfaces. The required precision and update rates rule out

all but infrared or radio frequency systems<sup>2</sup>. No existing system can, for the same amount of fixed infrastructure, provide the precision or update rate that a system derived from the principles shown here could provide.

This thesis has benefited both from the interdisciplinary culture of the Media Lab as well as from the extraordinary resources available here. In the five years that Prof. Gershenfeld has supported this author's research, an exceptional radio frequency engineering lab has been assembled that surpasses most facilities at MIT (and many in the commercial sector as well). This laboratory, combined with the mathematics and physical science knowledge and experience of the faculty have been the key elements required for successful completion of this thesis.

There is an additional expected benefit arising from this thesis that could only be realized at the Media Lab, where there is frequent interaction across many disciplines of art and science. The high efficiency transmitter built for the initial propagation experiments is a promising candidate to replace a large, expensive, and power hungry linear radio frequency generation system in nuclear quadrupole resonance (NQR) and nuclear magnetic resonance (NMR) systems. The integrated-antenna digital transmitter technique being developed for this thesis produces less noise than the amplifiers currently used in NMR and NQR research, and may result in an improvement in certain NMR and NQR experiments with broad application ranging from health care to explosives detection. This work is currently in preparation for separate publication.

---

<sup>2</sup>Since infrared signals do not penetrate ordinary building materials, a radio frequency system is much more attractive if a way can be found to reduce the amount of required infrastructure- hence this thesis.

# Chapter 2

## Prior Art: Position tracking for human-computer interfaces

### 2.1 The importance of position data for HCI

Many of the object identification and 3D tracking technologies that have previously been used in tabletop human-computer interfaces (HCIs) fail to scale to the much larger real world outside and away from the desktop computer. In many real world scale applications, including ubiquitous computing [48] and large scale tangible user interfaces (TUIs)[25], this capability to identify and track physical objects and people in a large 3D space is of great importance.

In this section we will begin with an overview of currently available 3D position tracking systems that can be used indoors on a scale larger than the tabletop. We will then proceed to describe a proposed indoor radio position tracking system that would provide fine grained 3D position solutions over a relatively large area. This would be accomplished with the added benefit that tagged objects or people are untethered by wires and unencumbered by large batteries. Finally we will suggest how the new system could find application as an element of tangible user interfaces that extend beyond the desktop into the space of ubiquitous computing.

## 2.2 Existing 3D position tracking systems

In this section we will identify several of the most common existing position tracking technologies, show examples of their application in the HCI field, and point out their shortcomings in specific application areas. Hightower and Borriello [24] present a useful taxonomy of these systems in the context of ubiquitous computing research.

In general these systems may be divided into two categories: (1) high position granularity, beacon type systems, and (2) continuous readout position tracking systems. The essential difference between these two types of systems is that the first type cannot be directly mapped on to a linear coordinate system (since the measurements are discrete and may be overlapping), while the second type of system can be mapped on to a linear coordinate system. This continuous readout capability is the more general of the two, since it is easy to provide artificial granularity in continuous measurements by maintaining a database of coordinate boundaries surrounding discrete reference points.

### 2.2.1 Infrared beacon systems

Infrared beacon systems use short range transmissions of modulated infrared light to transmit the identity of a mobile object to a fixed receiver in a particular known location. Because the IR light transmission range is limited to a few meters and is restricted to line-of-sight optical propagation, reception of a beacon message from a mobile object by a fixed receiving station is sufficient to establish proximity of the tracked object. An example of a simple system of this type is the Locust system, developed at the MIT Media Laboratory [38]. A more complex infrared system is used in the ParcTab system, developed at Xerox PARC [44].

One of the earliest and best known position tracking systems used in HCI research is the Active Badge system [43] developed at the Olivetti Research Laboratory. This system consists of two elements- a specially designed personnel badge, and a back-end processing infrastructure distributed throughout an office environment. The badge is equipped with infrared light emitting diodes (LEDs) used for short-range position

beacon communication, as well as two pushbuttons for signalling a user's intentions to the system. Additionally, each badge contains a small speaker and two visible LEDs with which the system can page the user.

Active Badges need not be expensive to produce, since they consist of only an inexpensive microcontroller and a few support components, and their batteries last about one year due to their 10 second beacon repeat timing. The back-end processing infrastructure is also inexpensive, but since the badge's low power infrared light signals travel only about 30m in line-of-sight, frequent replication of this infrastructure is necessary to ensure that a user's beacon signal is received by the infrastructure. This leads to a relatively high total installation and maintenance expense, since a typical office environment is full of obstructing walls and corridors. The granularity of the Active Badge system is low- it is not possible to determine the location of the badge at a higher resolution than the known locations of the receiving stations.

Harter describes the Active Badge system architecture in [22]. At one time, the Olivetti laboratory deployed more than 1500 badges and 2000 receiving stations in locations all over the world, including one installation at Cambridge University of 200 badges and 300 sensors. These systems were used in several experiments such as a location sensitive communications system that could patch communications to a user at the nearest telephone, videoconference station, or X Windows display.

### **2.2.2 Fine grained infrared tracking systems**

In augmented reality (and to a lesser extent, virtual reality) applications there is a need for very high update rate, low latency, and high accuracy. To prevent motion sickness, more than 60Hz update rate and latency less than 10msec is needed [6]. In addition, registration errors between an augmented reality scene and the user's local environment must be minimized at all costs since they can contribute to user disorientation.

Welch and Bishop describe a fine grained infrared tracking system in [50] which is designed to overcome these issues. The HiBall tracker uses a matrix of 3000 infrared LEDs mounted on the ceiling, which illuminate a wire-tethered tracking device

consisting of six lateral effect photodiodes. The LEDs are flashed in a predetermined sequence, and the tracking unit uses this pattern to estimate its position and orientation.

Welch and Bishop claim a 70Hz position update rate with RMS accuracies of about 0.2mm (position) and .03 degrees (orientation) over several meters of measurement distance. This is excellent performance but must be weighed against the very high cost and complexity of the system. Because of the extreme infrastructure requirements and the tethered nature of the tracking device the system will probably remain confined to research applications in augmented reality for which no other solution will work.

### **2.2.3 Ultrasonic position measurement devices**

There is a long history of the use of ultrasonic time of flight position measurement devices in HClIs, especially in the area of computer enhanced music environments. It is easy to use a microprocessor to measure the time of flight of a sound wave in air, because the speed of sound in air is relatively slow, at approximately 1msec per foot at room temperature. This, coupled with the inexpensive and widely available piezoelectric transducers, have led to a great deal of experimentation with ultrasonic position measurement systems. For example, Gelhaar [19] describes a computer aided music environment called *Sound=Space* which uses ultrasonic position measurement to create an interactive music space. More recent examples include this author's multi-user musical environment incorporating ultrasonic tracking [35] and Paradiso's dancing shoes [32] in which ultrasonic position measurement is used to track the position of a dancer's foot.

Perhaps the most general of these systems, however, is the Bat developed at AT&T's Cambridge research laboratory [45]. The Bat hardware is described in detail in [46], which is the definitive hardware reference for this system. The Bat is a small device about the size of a pager worn on a person's belt. This device contains a battery, microprocessor, and 418MHz radio transceiver as well as an ultrasonic transducer. Like the IR beacon systems mentioned previously, the Bat system requires a lot of infrastructure- at least three receiving stations whose locations are known must

be deployed in each room.

The room infrastructure communicates with the Bat via the 418MHz radio link. The system controller schedules a timeslice in which each Bat is to transmit a pulse of ultrasound. This pulse is received by the room infrastructure, and the time delay between the radio transmission and the reception of a sound pulse is used to determine the position of the Bat in three dimensions.

Ward presents data suggesting that the typical (95%) measurement accuracy is approximately 14cm for a single-pulse measurement, and 8cm with averaging over ten pulses. The system is limited to an aggregate update rate of about 50Hz for all users in a certain space, because in a typical office it takes approximately 20msec for reverberations to die out [46]. This limitation is severe, because large rooms often contain more than one person or object to be tracked. While a 50Hz update rate may be acceptable for HCI applications, a 5Hz or 10Hz update rate is very limiting.

#### **2.2.4 Magnetic field motion capture devices**

For HCI applications, specifically the capture of human motion for the purpose of animating computer generated characters, some prior work in the area of magnetic field based systems exists. These systems fall into two categories- those using the amplitude measurement of very low frequency magnetic fields, including the Polhemus system, or those using DC magnetic fields, for example Ascension Technology's Flock of Birds [5] family of motion capture devices.

Both of these systems consist of a single transmitter unit consisting of an antenna made from two orthogonal coils, and a large number of small receiving units which are attached to the clothing of the person whose motion is to be recorded. The associated computer system can then track the motion of as many as 50 points on the person's body. The receiving units contain little onboard circuitry; the majority of the signal processing is done in the box containing the transmitter. The person being tracked is therefore wired with a rather large and unwieldy umbilical tether which connects her to a large metal box. The maximum tracking range of this system depends on the size and placement of the transmitting antenna, but ranges of up to 15m are claimed

with a measurement precision of 5-7 cm [5]. The system claims to provide position updates at rates of 30-50 measurements per second.

While this level of performance is often sufficient for the system's intended use in motion capture from a performer on a stage, it is not sufficient for many other applications in which the person or object being tracked must be free to move over a wide distance. Because all signal processing is done centrally at the transmitter, these systems are not suited to spatial tiling with other such systems to cover a wide operational space. Also, both of these systems are vulnerable to interference from nearby metallic objects, since these objects distort the magnetic fields emitted by the transmitter units. There is a large body of work, for example [8] and [21], on augmenting magnetic tracker data with input from inertial and ultrasonic systems, and carefully curve fitting the measured response of these systems to achieve good performance in practical systems. However, Azuma [6] shows that even these extreme measures do not yield acceptable results in many real world HCI applications.

## 2.3 The new approach

The author's proposed low frequency position tracking system has several inherent advantages over prior art.

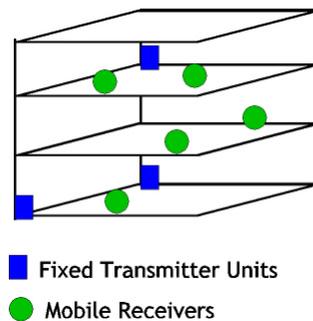


Figure 2-1: Physical configuration of the new system, showing the transmitters in the corners of the building and a small number of receivers.

In the proposed system, either three or four fixed transmitters are located at the corners of a building or another similarly large space (see Figure 2-1). The

transmitters (represented here by blue rectangles) are interconnected with a fiber optic cable in order to permit timing signal exchange between them. A control unit is located at one of the transmitter sites; this control unit serves to synchronize signals among the transmitters and produce the navigation signal.

A portable receiver (represented in Figure 2-1 by a green circle) is collocated with the person or object to be tracked. Any number of receivers may be used, since they passively receive the signals sent by the transmitters and are therefore not in competition for any centralized resources. No other infrastructure is needed, as the radio frequency signals from the fixed transmitters have greater useful range indoors than microwave radio or optical systems. Each receiver is able to deduce its own position with respect to the fixed transmitter sites. As their locations are surveyed with respect to the building, the fixed transmitters serve to anchor the receiver's position with respect to the real world.

First, little infrastructure is required to support a virtually unlimited number of simultaneous users. It is shown in this thesis that as few as three noncoplanar transmitting antennas can provide a useful position solution to meter-class precision at 1KHz update rates.

Second, since position data is provided to each person or object being tracked, privacy issues are mitigated because the person or object being tracked must choose to publish that information to a central server in order to enable centralized location databases. If it is desirable for a central system to track people or objects that carry receivers, an out of band signalling scheme (eg an ordinary wireless network such as 802.11) could be used to collect position measurements from the receivers and store them in a centralized database. The inherently decentralized approach used here has two main benefits: first, there is no limit to the number of receivers that can be active at any time, and second, when humans carry receivers, they must choose to volunteer their position to the central system.

Third, as it is a radio frequency system, the inherent update rates are very high; position solution rates are limited by averaging time, which trades against system accuracy, and computing power at the receiver, which affects power consumption at

the receiver. This system is expected to enable a wide variety of applications that have been contemplated in the ubiquitous computing space but have been previously physically or economically infeasible with the position tracking systems present in the prior art.

### **2.3.1 Comparison with prior approaches**

The proposed approach differs from the aforementioned prior work in several ways. First, unlike the beacon type systems, this approach yields a continuous position, fine grained readout, on coordinate axes that can be mapped into a user coordinate space. This has many advantages in applications for which a fine grained position estimate is needed, for example in applications in which data from a computer system is to be overlaid with real-world people or objects.

Second, the proposed approach makes use of electromagnetic wave propagation in a fundamentally different frequency regime than magnetostatic field motion capture systems like the Polhemus(tm) or Ascension(tm) devices. Using amplitude information alone, meter-class position solutions are possible at KHz update rates, with useful ranges from a single transmitter of up to 30m.

Third, unlike all but the extremely expensive, complex infrared system described in [50], a system based on this principle would exhibit very high update rates and low latency, limited only by the transmitted signal bandwidth. Since the transmitter signal frequency is in the MHz region, thousands of RF cycles are transmitted per millisecond. If we assume that the upper limit on desired system update rate is perhaps 5 milliseconds (comparable to the fastest human reflex reaction) there is still an opportunity to average over thousands of transmitter cycles.

These aspects are summarized and compared with the other approaches mentioned in Table 2.1.

Table 2.1: Comparison of position measurement systems

<i>Technology</i>	<i>Reference</i>	<i>Range</i>	<i>Granularity</i>	<i>Accuracy</i>	<i>Update rate</i>	<i>Cost (per bldg.)</i>
GPS (example)		outdoors only	world	1-10m	10Hz	n/a
IR beacon	Locust [38]	10m	room	room	10Hz	medium
IR beacon	ParcTab [44]	10m	room	room	30Hz	high
IR beacon	ActiveBadge [43]	30m	room	room	30Hz	high
fine grained IR	HiBall [50]	2m	2m	0.2mm	70Hz	extremely high
Ultrasonic	FKB [35]	10m	10m	$\approx$ 2cm	30Hz (aggregate)	high
Ultrasonic	Bat [45]	10m	10m	8-14cm	50Hz (aggregate)	high
Magnetic field	Flock of Birds [5]	15m	15m	5-7cm	30-50Hz	extremely high
Mid-field	Proposed here	150m <i>est</i>	building	meter-class <i>est</i>	1KHz <i>est</i>	medium

## 2.4 Novel applications of the proposed system

There are many human-computer interface applications that cannot be practically realized with conventional position measurement systems like those described in Table 2.1. These systems typically have modest cost but very coarse granularity, or fine granularity but very high cost.

Since cost is of importance both to researchers and in the deployment of a commercial system, most published work in the ubiquitous computing community has concentrated on systems that can be built at reasonable cost. Generally, these fall into two categories: first are the beacon-type systems that provide coarse granularity; they can indicate that a user is present in a certain room, but cannot provide a true position in 3D space. The second category are ultrasonic systems that can produce 3D positions but at a very low update rate, and supporting only a few simultaneous users. Applications therefore fall into two related categories. The proposed system allows the user interface designer to use fine granularity position data without the constraints of update rate or extreme cost.

Systems with coarse granularity can be used to produce a location-aware system that allows room-to-room changes in contextual information. For example, the Location Oriented Communications and Transportable Desktops applications use the Active Badge to allow users to communicate using the nearest telephone or X Windows display, respectively [22]. A user may be located within a building to room scale precision, and that location can be made available to others via a location database server. Other applications of these systems include the context aware wearable computing applications presented by Starner [38].

Fine grained systems, such as the Bat [46] and the Karamazov ultrasonic systems [35] take advantage of their position accuracy to provide the user with an immersive environment. For example, in the Karamazov application (see Figure 2-2), a wearable sensor pack integrates gesture sensing and precise multiuser position tracking, as well as large scale projection display, to produce an interactive musical environment that could not exist without such a fine grained system. The environment is rich and

immersive in that the user's position and gesture data are integrated on human space and time scales to produce music and graphics in a visually interesting manner. In essence the entire stage becomes a playable musical instrument that happens to encompass the performer's entire body. Also, the spatial awareness is used to produce a visually interesting display on illuminated juggling clubs whose colors vary depending on user position.

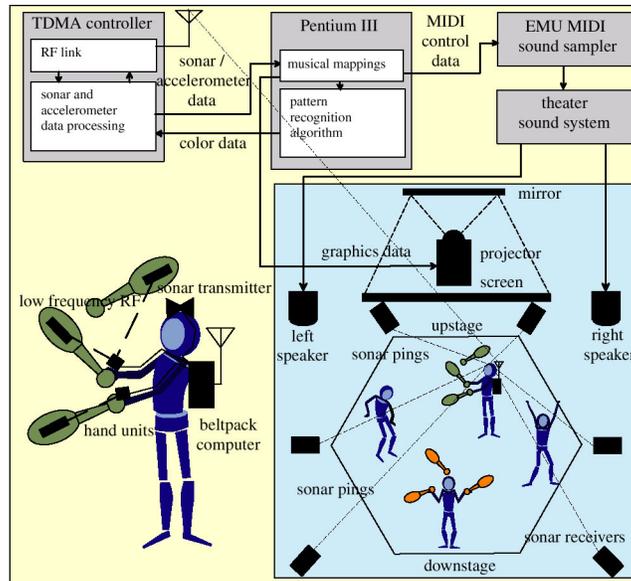


Figure 2-2: Block diagram of the Karamazov immersive musical system

One can imagine, if the proposed system is successful, updating the active office systems as well as the musical stage environment to take advantage of an unlimited number of users as well as the ability to use spatial position as an input device (similar to the “virtual button” produced by clicking the Bat device at a pre-defined spatial position). This would allow for a much richer environment by allowing ordinary devices, such as pens, erasers, ordinary musical instruments, and other objects with a well defined use metaphor, to have a spatial context that could activate virtual context, thus allowing the device to have meaning in a tangible, ubiquitous computing application.

However, the most intriguing applications may derive from the Luminous Room concept presented in [40]. The Luminous Room concept allows the display of a

computer to expand beyond the traditional CRT into an entire space. Based on spatial information measured from tangible user interface objects, such as small, graspable models of lenses or buildings, systems such as the Urp urban planning application may be constructed. However, the current Urp prototype is limited to application on a single, well defined surface. The proposed position tracking system may be able to provide a true *boundarylessness*, where a tangible user interface object can be placed anywhere in the space of an entire building. Assuming the existence of a suitable display (part of the *IO Bulb* concept in [40]), these tangible user interface objects will work anywhere in a suitably equipped building, thus merging the concepts of ubiquitous computing and the tangible user interface into a truly ubiquitous, tangible interface that would work over a wide area, seamlessly. This would revolutionize interface design by allowing for the true boundarylessness that is currently impractical with current systems.

## 2.5 Conclusion

It has been shown that there is a significant gap in the field of indoor position sensing technology. Existing systems cannot support a wide variety of position-aware applications that require high update rates, building-wide coverage, and good position accuracy. No existing system matches the proposed system's attributes of minimal infrastructure, unlimited user support, fast update rates, and good precision over building scale size ranges. The proposed system may find application in a number of applications ranging from ubiquitous computing applications (eg making a device like the ParcTab or a system like Urp economically feasible in a commercial deployment) to entirely new tangible user interfaces that extend the human-computer interface beyond the desktop scale on to the spatial scale of ordinary human activities.

# Chapter 3

## System Engineering

### 3.1 Introduction

In this chapter we identify and analyze the system engineering issues that affect the measurements conducted in the course of this research. Factors such as antenna design, transmitter field strength, receiver noise figure, phase measurement performance, man-made and natural sources of channel noise, and geometry issues are considered with the goal of determining the baseline specifications and fundamental limits for the experiments that are described in this thesis.

### 3.2 Radio propagation and the choice of operating frequency

The radio propagation tests conducted in this thesis were carried out under assumptions appropriate for a single-frequency hyperbolic system operating at 2.0 MHz. This system was intended to exploit the gap between systems employing near-field propagation, like the Omega [13] navigation system, and far-field propagation like that used by GPS [1]. This is shown graphically in Figure 3-1.

This operating frequency was chosen because of its expected indoor propagation characteristics. In this “mid-field” frequency regime, where the signal wavelength

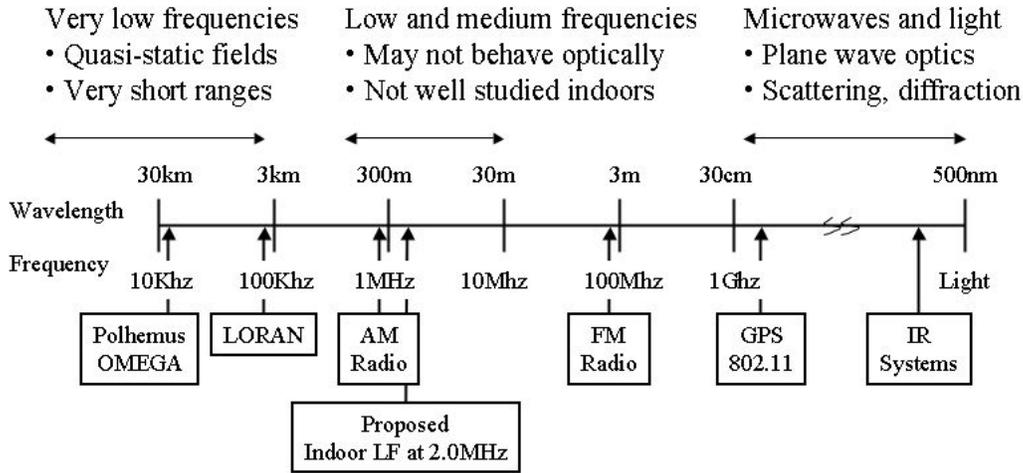


Figure 3-1: Comparison of positioning system operating wavelengths

is comparable to the size of the building, favorable propagation characteristics were expected.

We will define mid-field such that  $d \approx \lambda/2\pi$ , or about 23m at the 2.0MHz candidate operating frequency. This marks the transition in dominant propagation mode between near-field and far-field modes[39]. In free space, in the near field, magnetic dipole coupling has a  $d^{-3}$   $H$ -field decrease with distance. This corresponds to a  $d^{-6}$  power dependence, or 60dB attenuation per decade in distance. In the far field, electromagnetic waves propagate with a  $d^{-1}$   $H$ -field decrease, resulting in  $d^{-2}$  power dependence, or 20dB attenuation per decade in distance. There have been no prior studies on indoor radio propagation at these wavelengths, but it was surmised that this transition region would provide smooth amplitude and phase variation with distance, with little distortion due to the metal structural elements in the building. It is the surprising result of Chapter 6 that this initial assumption is not true; phase is not

a well behaved function of distance at these wavelengths, but amplitude is reasonably well behaved.

### 3.3 Available bandwidth and waveform design

In a single frequency hyperbolic system, a fixed base station consists of a transmitter  $T$  which emits an electromagnetic carrier wave given simply by

$$T = A \sin(\omega t + \phi) \tag{3.1}$$

where  $A$  is the carrier amplitude ( $A^2$  is its power),  $\omega$  is its angular frequency in radians per second, and  $\phi$  is its phase. The candidate operating wavelength is  $\lambda = c/(2 \cdot 10^6) = 150m$ . This wavelength is comparable to the major dimension of a typical building and therefore meets our definition of a mid-field system.

It is a desirable property of a single frequency hyperbolic system that the position solution's precision is not directly related to the transmitted signal bandwidth[16] as in spread-spectrum systems, for example, and therefore we may proceed in our analysis to neglect the modulation that is inevitable for channel sharing purposes when considering propagation and antenna design effects. We will return to consider this channel sharing modulation when considering issues of update rate and maximum averaging time during signal processing.

One undesirable result of using a narrow band transmitted signal is a range ambiguity which occurs at each *lane width* of  $\lambda/2$  [16], resulting in a shorter maximum useful range without a means of range disambiguation (eg the *multipulse* method used by the Decca Navigator system). Another unpleasant consequence of using a narrowband carrier signal is the inability of a receiving station to use a simple time of first arrival method for reducing the effects of multipath interference. Since a given transmitter is active for much longer than the time required for a multipath signal to reach the receiver, a relatively large amount of the received signal energy is contaminated by multipath. This effect will be described further in Section 3.7.4.

## 3.4 Antenna design choice

The signal defined in Eq. 3.1 is fed either to an electrically small vertical monopole  $E$ -field antenna, one whose length  $l \ll \lambda$ , or to a small loop  $H$ -field antenna whose diameter  $d \ll \lambda$ . These are the two most likely choices for an indoor antenna system that is subject to space constraints; a full size antenna is neither practical, for physical size reasons given the wavelength of 150m, nor desirable, because of the difficulty of locating its phase center when the receiver is in close proximity to such a large antenna.

We will first define an *effective height*  $h_e$  which is the height of a hypothetical Hertzian monopole conductor above its ground plane [47]. This effective height will be used for all subsequent antenna calculations as it provides an equal basis for comparison among antenna types.

### 3.4.1 $E$ -field antenna

In the case of an electrically short vertical monopole  $E$ -field antenna, we can assume that the current distribution along the vertical conductor is uniform [47], and we are primarily concerned with the emitted electric field:

$$E_z = \frac{Ih_e}{2\pi\epsilon_0} \left\| \left[ \underbrace{\frac{1}{\omega d^3}}_{e\text{-static}} + \underbrace{\frac{j}{cd^2}}_{m\text{-static}} - \underbrace{\frac{\omega}{c^2 d}}_{\text{radiation}} \right] \right\| \quad (3.2)$$

where  $E_z$  is the electric field in the  $\hat{z}$  direction, along the long axis of the conductor.  $I$  is the antenna current in amperes,  $\epsilon_0$  is the permittivity of free space,  $h_e$  is the antenna's effective monopole height,  $d$  is the distance from transmitter to observer,  $c$  is the velocity of light,  $j$  is  $\sqrt{-1}$ , and  $\omega = 2\pi f$  is the operating frequency. It should be noted that the second, magnetostatic term arises from the current carried by the monopole conductor. There is no corresponding electrostatic term in the  $H$ -field antenna described below.

Since the power transferred by an electromagnetic wave is  $E^2/120\pi$ , it is apparent

from Eq. 3.2 that since  $h_e$  appears as a multiplier, the monopole length must be an appreciable fraction of the operating wavelength to achieve reasonable antenna efficiency. This is because the effective height is dependent on  $\lambda$ . If we assume that our short monopole antenna has a length  $l < 0.01\lambda$ , there is a related and more pressing issue, which is that the feedpoint impedance of a short monopole antenna is very high and almost purely reactive. The feedpoint impedance is:

$$Z_{ant} = R_r + R_l + \frac{1}{j\omega C} \quad (3.3)$$

where  $R_r$  is the radiation resistance of the antenna,  $R_l$  is the equivalent loss resistance of the antenna, and  $C$  is the antenna's capacitance. Because the radiation resistance of such a short monopole is low [7]:

$$R_r \approx 160\pi^2 \left( \frac{h_e}{\lambda} \right)^2 \quad (3.4)$$

$$\frac{h_e}{\lambda} \approx 0.01$$

$$R_r \approx 0.16\Omega$$

and loss resistance must be minimized for highest antenna current (and therefore efficiency), the antenna's impedance is dominated by the reactive term  $\frac{1}{j\omega C}$ . A large matching inductor and very high feed voltages are therefore needed to result in a reasonable antenna current  $I$ . At 2.0MHz, for example, assuming a reasonable value of antenna capacitance of 5pF, the monopole's input impedance is about  $0.16 - 3980j$ , which is very hard to match to a typical transmitter's  $50\Omega$  output impedance.

Another issue affecting the use of an electrically short monopole antenna is grounding. A short monopole is generally fed in an unbalanced mode; that is, the feedpoint is driven against a ground plane. Because an effective ground plane at  $\lambda = 150m$  is quite large, the deployment of a monopole antenna would require a number of ground radials or a large amount of wire screening to provide a good ground for the antenna. If an inferior ground is used, RF current will flow on the shield of the coaxial feedline as well as flowing unequally on the ground plane, thus distorting the phase center

(the location of an imaginary point source radiating the RF power) of the antenna. This in turn will result in position error because the apparent radiative source, often referred to as the *electrical center* of the antenna, will be spread out over a large physical extent.

### 3.4.2 *H*-field antenna

An *H*-field antenna may be the best choice for an HF indoor radio positioning system. In the case of an electrically small loop *H*-field antenna,

$$H_\phi = \frac{Ih_e}{2\pi} \left\| \left[ \underbrace{\frac{j}{d^2}}_{m\text{-static}} - \underbrace{\frac{\omega}{cd}}_{\text{radiation}} \right] \right\| \quad (3.5)$$

where  $H_\phi$  is the tangential magnetic field. There is no electrostatic coupling term in this expression because of the balanced property of the small loop, discussed below.

Like an electrically short monopole, a small loop also has a low radiation resistance  $R_r$  [7]:

$$R_r = 31k\Omega \left( \frac{nA}{\lambda^2} \right)^2 \quad (3.6)$$

A typical small transmitting antenna might have an area  $A$  of  $1m^2$  and 5 turns, resulting in  $R_r = 0.0015\Omega$ . Resistive loss will clearly play a dominant role in the efficiency of this antenna, because the very low radiation resistance is on the order of the loss resistance that can be expected from even a thick copper conductor.

A small loop transmitting antenna is attractive for at least three reasons. First, the loop antenna lends itself to easier matching to a  $50\Omega$  transmitter output via a tapped capacitor or tapped coil arrangement, as shown in Figure 3-2. Secondly, it can exhibit a higher efficiency in a smaller space than a correspondingly sized monopole antenna. Finally, a small loop antenna can be built with an inherently balanced feedpoint, greatly reducing the possibility of feedline radiation and allowing the antenna's phase center to be the geometric center of the loop. This simplifies transmitter feed arrangements because a balanced feedpoint can always be found

on the loop as part of the impedance matching process. Unlike a short monopole antenna, there is therefore no need for a good RF ground for an  $H$ -field antenna because there is no reference plane to feed against. A DC ground will still be needed for safety reasons.

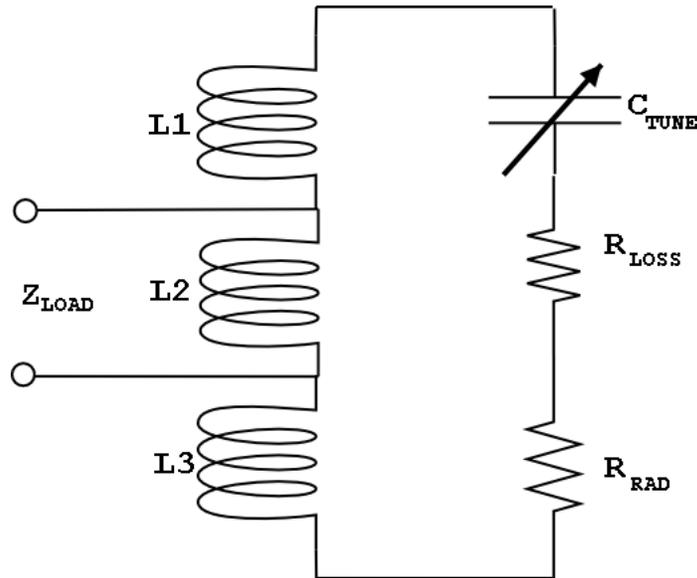


Figure 3-2: Effective circuit of a small loop antenna

One disadvantage of the small loop antenna is that it must be operated in a resonant mode to achieve high efficiency. The loop is therefore included in either a series or parallel resonant circuit. The quality factor  $Q$  of this resonant circuit must be very high in order to achieve a low loss resistance, so the tank circuit voltage is correspondingly high. In an ideal case, assuming a resistive loss smaller than the radiation resistance and a reasonable antenna inductance of perhaps 7uH,

$$Q = \frac{2\pi fL}{R_r} \tag{3.7}$$

$$Q_{desired} \approx 60,000$$

This exceptionally high  $Q$  is unachievable in practice. The  $Q$  of a small loop transmitting antenna is generally limited to between about 1000 and 10000, which results in a very high voltage appearing across the small loop's resonating capacitance. This

can lead to dielectric breakdown or arcing, since the voltage across the tuning capacitor is proportional to applied power. For this reason, a high  $Q$ , high voltage vacuum variable capacitor is used in the prototype  $H$ -field antenna that has been constructed. This type of capacitor is unfortunately rather expensive but has the benefits of temperature stability, high voltage handling capability, and small size. A lower cost approach would be to build a custom high  $Q$  capacitor using parallel copper plates with a Teflon sheet dielectric, but this capacitor would not be as easily adjustable as the vacuum variable used in the prototype.

### 3.4.3 Radiated power

Electrically small antennas (like the  $E$  and  $H$ -field antennas described above) are generally inefficient. This is because of their low radiation resistance (Eq. 3.6). Because the power that is actually radiated is that “dissipated” in  $R_r$ , while power dissipated in  $R_l$  is wasted as heat, the antenna designer must minimize  $R_l$  at all costs to maximize the power actually radiated by the antenna.

Another way to express the effective radiated power of a given antenna is to look at the emitted  $E_z$  field (or the corresponding  $H_\phi$  as this is the dual of  $E_z$ ). Looking at it this way, radiated power  $P_r$  is:

$$P_r = \frac{4\pi d^2 E_z^2}{120\pi G} = I^2 R_r \quad (3.8)$$

where  $G$  is the antenna’s gain (a function of the radiation pattern of the antenna; this value is 3 for a short monopole or small loop [47]) and  $\sqrt{\frac{\mu_0}{\epsilon_0}} \approx 120\pi$  is the impedance of free space. It is this radiated power that is the starting point for the link budget given in Section 3.6.

### 3.4.4 Antenna radiation efficiency

In general we define antenna efficiency  $\eta$  as the ratio of radiated power to antenna input power. This is to be distinguished from transmitter efficiency<sup>1</sup>. Assuming that all losses in the antenna and its matching circuit are accounted for in the antenna's  $Q$ ,

$$\eta = \frac{QR_r}{2\pi fL} \quad (3.9)$$

For example, given the design parameters  $Q = 1000$  and  $L = 7\mu H$ , then  $\eta = 0.0174$ . We can treat  $\eta$  as a power attenuation in dB:

$$\begin{aligned} \text{gain} &= 10 \log \eta \\ &= -17.59\text{dB} \end{aligned} \quad (3.10)$$

Thus a relatively high power transmitter will be needed to produce a reasonable radiated power; for example, a 100W transmitter using this antenna will radiate only 1.74W.

## 3.5 Noise

As we will see later in Section 3.7.2, the phase measurement performance of a hyperbolic radiopositioning system is governed fundamentally by the signal to noise ratio (SNR) at the receiver. Thus it is very important to be able to predict noise power at the receiver in order to provide bounds on the expected performance of the system as a whole.

There are four main contributors to the noise power expected in the receiver. These include factors that cannot be controlled, such as atmospheric noise and man-made channel noise from electrical and electronic devices, as well as design parameters such as antenna thermal noise voltage and receiver noise figure. Assuming uncorrelated, additive white Gaussian noise (AWGN), these noise sources combine at the

---

<sup>1</sup>With modern switching power amplifier techniques, such as those described in Chapter 4, power amplifier efficiency can exceed 90%.

receiver to produce the SNR that will be the ultimate limit to phase measurement performance. It is shown in this section, and later confirmed by experiment in Chapter 6, that man made noise is the most important contributor to signal to noise ratio at the measurement receiver.

### 3.5.1 Atmospheric noise

Atmospheric noise, which is defined as noise of natural origin created in the atmosphere, arises at long wavelengths from a wide variety of mechanisms, including lightning, the interaction of the Earth’s magnetosphere with the solar wind, and many other factors. Thus, the characterization of atmospheric noise is an inherently empirical process. Generally, a calibrated spectrum analyzer with a calibrated  $E$  or  $H$ -field antenna is used to sample noise at many intervals during the day and night. Because HF frequencies, such as the 2.0 MHz carrier of the present system, exhibit skywave propagation at night, a diurnal variation is expected and in fact observed in the experiments described in Chapter 6. Atmospheric noise at 2.0 MHz is generally lowest during the daytime, and increases by as much as 10dB at night. Surveys of 2.0 MHz atmospheric noise and groundwave and skywave propagation characteristics were made as early as the 1940s by Pierce and others [34] as 2.0 MHz was the operating frequency of the original Loran system, Loran-A.

More modern measurements of atmospheric noise values versus frequency are found in CCIR Report 322 [42]. These results are summarized in Figure 3-3.

It is apparent from this data that the atmospheric noise can be expected to be at least 45dB greater than thermal noise. Assuming a 1 KHz measurement bandwidth  $\Delta f$ , and that  $R_l$  is the dominant factor in the real part of the antenna’s complex impedance, thermal noise may be found as:

$$\begin{aligned}
 P_{thermal} &= 4kT\Delta f & (3.11) \\
 &\approx 1.59 \cdot 10^{-17} \text{ W} \\
 &\approx -137 \text{ dBm}
 \end{aligned}$$

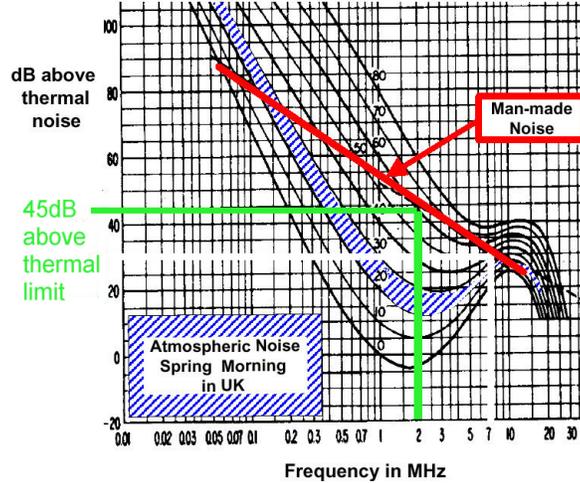


Figure 3-3: Atmospheric noise vs. frequency, from CCIR 322.

Thus we expect from Figure 3-3 that the atmospheric noise will produce a noise power of  $-137\text{dBm} + 45\text{dB} = -92\text{dBm}$  at the receiver input terminals.

### 3.5.2 Man-made noise

Man-made noise is hard to characterize because of the wide variety of electrical and electronic devices that can generate electromagnetic noise. This noise may be broadly characterized into two classes: wideband noise, for example motor noise arising from arcing, and narrowband interfering carriers, arising for example from a microprocessor clock oscillator. These signals are generated by a wide variety of electronic equipment; the FCC's Part 15 regulations[2] for unlicensed intentional radiators specify a field strength limit of  $100\mu\text{V}/\text{m}$  measured at 30m for emissions in the band 1.705-10MHz.

As in the case of atmospheric noise, empirical observation is probably the only way to arrive at a meaningful bound on man-made noise. The International Telecommunications Union (ITU) has published a set of aggregate man-made noise measurements in Report 372[41]. These measurements appear in Figure 3-4.

It is apparent from Figure 3-4 that man-made noise levels may be expected to have a field strength  $E_n$  of at least  $10\text{dB}\mu\text{V}/\text{m}$ . Knowing the antenna factor of our receiving antenna, which relates the incident field strength in  $\text{V}/\text{m}$  to power in dB

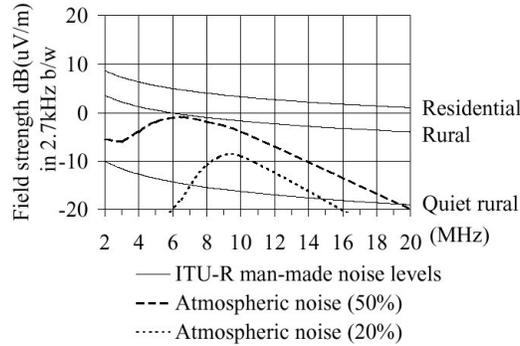


Figure 3-4: Man-made noise vs. frequency, from ITU Report 372.

across  $50\Omega$ , we may estimate the noise power at the receiver's input terminals. For example, the EMCO Model 6502  $H$ -field receiving antenna exhibits an antenna factor  $F_a$  of about  $11\text{dB}\mu\text{V}/\text{m}$  [14]. Since

$$V_r = E_n + F_a \text{ in dB} \quad (3.12)$$

$$V_r \approx 21 \text{ dB}\mu\text{V}$$

$$P_r \approx -86 \text{ dBm}$$

Thus we expect man-made noise to be about 10dB greater than atmospheric noise at 2.0 MHz. This is expected to be an overestimate of the actual noise power at the phase measurement receiver because the EMCO antenna is an active antenna and has a correspondingly high antenna factor. Also, the 2.7KHz measurement bandwidth for the ITU measurements in Figure 3-4 exceeds the effective measurement bandwidth of the receiver. However, it was expected that man-made noise will be the dominant factor in the noise power at the receiver antenna terminals. This is shown to be true by experiment in Chapter 6.

### 3.5.3 Receiver noise figure

Noise figure is the noise contribution of the receiver's active circuitry itself, measured in dB of output noise above that expected if the only noise contributor were thermal noise. A well designed receiver at these frequencies exhibits a noise figure of about

2dB. Since we expect from Eq. 3.12 that thermal noise power will be about -137dBm, noise contribution from receiver noise figure will be far lower than the expected man-made noise. This is the case for the apparatus built for this thesis.

Because of the low noise figure of the receiver, the expected noise power in the measurement bandwidth will be taken to be the contribution from man-made noise (-86dBm in the receiver bandwidth) for the purpose of subsequent analysis, because it is the dominant contributor to total noise power.

## 3.6 Link budget

When designing a communication system it is usual to prepare a link budget to show the expected link margin, the amount of “excess” signal above the targeted SNR. In the present case, there is no such thing as an excess SNR, since SNR is directly related to phase measurement accuracy and therefore geometrically related to position measurement accuracy. However, preparation of a link budget with a target SNR exceeding 10dB is instructive as a design aid as it indicates the amount of attenuation that can be tolerated while still producing a strong signal at the receiver.

### 3.6.1 Path loss

To prepare a link budget one must account for as many sources of signal and noise power as possible. The most important variable not yet addressed is path loss over the expected distances of operation. A lower bound can be placed on path loss by using the Friis free space loss equation[16].

$$\begin{aligned}
 L &= 10 \log \left( \frac{4\pi d}{\lambda} \right)^2 & (3.13) \\
 d &\approx \lambda \\
 L &\geq 21 \text{ dB}
 \end{aligned}$$

The Friis path loss expression is very obviously an underestimate of actual system path loss since it fails to account for any obstructions in the path; it assumes the

far-field propagation limit where  $d \gg \lambda$  and signal power varies as  $d^{-2}$ , while we are applying it when  $d \approx \lambda$ . In the magnetostatic limit that applies in that case, the signal strength would vary as  $d^{-6}$ , assuming dipole-dipole coupling between the fields of the transmitting and receiving antenna loops. We shall therefore revise the Friis path loss estimate, taking into account the magnetostatic limiting case:

$$\begin{aligned}
 L &= 10 \log \left( \frac{4\pi d}{\lambda} \right)^6 & (3.14) \\
 d &\approx \lambda \\
 L &\approx 63 \text{ dB}
 \end{aligned}$$

It is this value that will be used for the link budget analysis in Table 3.1. As with the man-made noise limit derived above, the comparison of actual path loss to the result of Eq. 3.14 was subsequently measured. Typical measured values indoors are on the order of 60dB to 70dB for the experiments described in Chapter 6. The link budget presented here is therefore representative of the observed system behavior.

Table 3.1: System link budget

<i>Contributor</i>	<i>Value</i>	<i>Units</i>
Transmitter output power	50.0	dBm
Antenna radiation gain (loss)	-17.6	dB
Path gain (loss)	-63.0	dB
Receiver ant gain (EMCO 6502)	-40.0	dB
Signal power at receiver	-70.6	dBm
Receiver noise bandwidth	1.0	KHz
System temperature	290	Kelvin
Thermal noise power	-137	dBm
Atmospheric noise power	-117	dBm
Man-made noise power	-86.0	dBm
Noise power at receiver	$\approx$ -86.0	dBm
Signal to noise ratio	15.4	dB

## 3.7 Sources of timing uncertainty

The next set of issues under consideration affect the system phase measurement uncertainty. These issues include time and frequency uncertainty of the receiver and transmitter clocks, as well as the fundamental limit to phase measurement precision given a certain signal to noise ratio. Radio propagation delays are mentioned as an unknown quantity prior to actual measurement. The actual values vary from location to location and can be determined from the phase plots in Chapter 6. Finally, limits to averaging imposed by the time division multiple access (TDMA) scheduling of signals from the fixed transmitters are considered.

### 3.7.1 Clock jitter at receiver and transmitter

A hyperbolic radio positioning system requires precise frequency and time synchronization between transmitting stations. This is because the relative phases of pairs of received signals are used to produce an estimate of the difference in distance between pairs of transmitters. Since an uncharacterized phase offset between pairs of transmitters will result in a position uncertainty, these offsets must be well characterized.

There are three methods that have been used to accomplish this goal. The first, used in the original Loran-A system[34], is to designate one transmitter as the “master” and the remaining transmitters as “slaves”. The slaves receive the master’s navigation signal in-band, and phase lock to it. This method has the advantage of not requiring any hard wired interconnection between the master and the slaves. Before the advent of precise atomic timekeeping[28] and satellite time transfer[33] this was the only method suitable for operation over long inter-transmitter baselines. However, propagation uncertainties are significantly increased, since the propagation delays between distant sites are unknown, especially when ionospheric (sky wave) propagation is utilized.

The second method is to use satellite time transfer and atomic clocks at the transmitter sites. This is the method used to synchronize clocks at Loran-C transmitter sites[16]. In effect, it is also the method used to synchronize GPS clocks. In both

cases, radio communication between an orbiting atomic clock on a GPS service vehicle (satellite) is used to synchronize the orbiting clock to ‘GPS time’ which is related to universal coordinated time (UTC). UTC itself is derived from an aggregate measure of atomic clocks. In the case of GPS, the service vehicle’s onboard clocks are set by calibrating service vehicle clocks as they overfly five ground stations around the world[1].

The third method, which is the one chosen for the present system, is direct cable interconnect of the transmitter sites. This method has the advantage of simplicity and low cost for short inter-transmitter baselines. In the present system, fiber optic interconnection is used to avoid the problem of unwanted radiation from signal leakage of coaxial cables carrying RF signals. The transmission delay over fiber does depend on temperature and mechanical stress, which could be significant over long fiber runs, but this can be compensated for by minimizing stress and using equal lengths of fiber to feed all transmitter sites on the assumption that building temperatures are reasonably well controlled.

With a careful application of the third method, the timing of all transmitted signals can be maintained within tight tolerances, with jitter limited by that of the transmitter’s oscillator. For the initial test system, a quartz crystal oscillator was chosen. Quartz oscillators have poor long term stability; a typical number for an ovenized quartz oscillator is an aging factor of 1 part in  $10^{-9}$  per day in contrast to an atomic fountain clock demonstrating a long term stability of 1 part in  $10^{-15}$ [28]. But quartz oscillators have excellent short term stability, as demonstrated by their excellent phase noise characteristics.

Since all transmitter sites are driven from a common clock at the master, any long term transmitter clock variation can be rendered irrelevant by phase locking the receiver’s clock frequency to the transmitter. Since a typical quartz oscillator operating at 2.0 MHz has an RMS jitter specification in the picosecond range, clock stability at the receiver and transmitter were expected to have far less impact on the position solution than radio propagation effects. This has been verified by experiment; location to location phase differences are on the order of tens of degrees (surprisingly),

so clock jitter is not a dominant contributor to phase accuracy.

### 3.7.2 SNR versus phase measurement accuracy

The accuracy of a phase measurement between two signals of the same frequency is limited by the signal to noise ratio of the two signals. In this analysis, following Forsell[16], we determine the RMS phase error between two such signals.

We first assume that at a given SNR we may write the noise power  $N$  as a time varying noise voltage  $n(t)$  corresponding to a phase deviation  $\Delta\varphi$  (where  $\varphi = \omega t$ ) of a noiseless sinusoid of amplitude  $\sqrt{2S}$ .

$$\sqrt{2S} \cos \varphi \Delta\varphi = n(t) \quad (3.15)$$

We then take the time average of the two sides of Eq. 3.15 and relate  $\Delta\varphi$  to SNR.

$$\sqrt{(\Delta\varphi)^2} = \frac{1}{\sqrt{\frac{S}{N}}} \quad (3.16)$$

Since the noise present on the two signals is assumed to be uncorrelated, the noise induced errors will add quadratically, resulting in an RMS error:

$$\Delta\varphi_{1,2} = \sqrt{\frac{1}{(S/N)_1 + (S/N)_2}} \quad (3.17)$$

Thus, if we assume from the result of Table 3.1 that the signals from two transmitters both exhibit an SNR of 15.4dB, we expect  $\Delta\varphi_{1,2} \approx 0.24$ . Since  $\Delta\varphi_{1,2} = 2\pi ft$ , the RMS time error is 19 nanoseconds at  $f = 2.0\text{MHz}$ , corresponding in an error in the difference in distance between the two transmitters of 5.7 meters.

This is a reasonable result; since  $\Delta\varphi_{1,2} \propto 1/\text{SNR}$ , each 6dB increase in SNR via an increase in transmitter power or reduction in noise level by shielding or signal averaging will reduce RMS distance difference error by a factor of 2.

### 3.7.3 Coherent averaging

Coherent averaging can greatly improve the effective SNR of the received signal by providing *processing gain*. After averaging coherently over  $n$  measurements, the signal power will be proportional to  $n^2$  while the noise power grows as only as  $n$  because it is uncorrelated. The SNR therefore increases as  $n$ . However, the measurement time has also increased by a factor of  $n$ . If we are willing to tolerate an increase in measurement time by a factor of 4 (for 6dB processing gain), we will decrease RMS distance difference error by a factor of 2.

The primary limitation on the benefits of coherent averaging is that caused by user motion within the effective bandwidth of the positioning system; without additional information user motion is indistinguishable from noise as it can be modelled no better than random noise. If, however, additional information about the user's motion is available, for example from an inertial measurement unit, uncertainty caused by user motion may be greatly reduced. If the user's path is unknown, but the general characteristics of human motion are known, an estimator may be built to attempt to remove its effects.

If the user remains stationary, and all noise sources are random and zero-mean, averaging may theoretically be carried out up to the limits of transmitter and receiver clock jitter, resulting in an arbitrarily good position solution. However, the atmospheric noise described in Section 3.5.1 exhibits predictable diurnal and seasonal behavior, and the man-made noise described in Section 3.5.2 can be expected to be influenced by human habits. Therefore this noise cannot be completely eliminated by indefinitely long averaging periods, as the long term behavior will in effect supply a bias to any measurement made on a useful time scale. Instead a more complex estimation system must be constructed that takes this predictability into account.

### 3.7.4 Effects of multipath propagation

Radio propagation delays are the most important unanswered question in the development of any indoor radiopositioning system. The indoor environment is full of

conductive objects, some of which may act as scatterers if they are large compared to  $\lambda$ . Some building materials may also act as a Faraday cage, screening the receiver and therefore decreasing the SNR.

At very short wavelengths, and well in to the microwave region, indoor radio propagation has been shown to exhibit specular reflection and refraction effects, yielding behavior similar to that observed at optical wavelengths [4] [30] [23]. At these frequencies we may model the received signal  $R(t)$  as the sum of the direct path and multipath components:

$$R(t) = S(t) + N(t) + \sum_{i=1}^M \alpha_i s(t - \tau_i) + n_i(t - \tau_i) \quad (3.18)$$

where  $\alpha_i$  is the amplitude of the  $i$ -th multipath component and  $\tau_i$  is its delay with respect to the direct path signal  $S(t)$ .

Given this model, Friedlander [17] treats the effect of multipath on a hyperbolic passive sonar system, and presents a detailed analysis leading to a Cramer-Rao lower bound (CRB) on error in range and depth estimation. Abel and Smith [3] present an approximate maximum-likelihood estimator used in prefiltering range differences to produce source position estimates; in other words making explicit use of multipath to improve the source position estimate. They show, using a received signal model similar to that in Eq. 3.18 and a maximum-likelihood equivalent estimator, that the bound variance is approximately inversely proportional to the number of multipath reflections present in the received signal.

However, the mid-field system proposed in this thesis operates at a much lower frequency, and at wavelengths much longer than those for which the specular reflection assumptions implicit in Eq. 3.18 apply. Therefore a different type of channel model is needed; Chapter 7 presents one such possible model. Regardless of the underlying model, given a linear channel model we may treat the problem as a multiple source model in the position estimation process. Because of the limited bandwidth of the transmitted signal, almost all of the transmission time (and thus most of the received signal power) will be contaminated by multipath (or multiple source) interference. A

reasonable multipath model, perhaps including an autocorrelation based multipath separation mechanism [16] may significantly improve immunity to multipath, if a suitable channel model can be found.

### 3.8 Geometric factors

As shown in the previous section, SNR translates directly into phase measurement precision. We will now finish this survey by considering without definitive result the hyperbolic position solution process. This process is more complicated than the Loran-C position solution[16] because in this system the receiver is not so far away from the transmitting stations that the incoming signal can be modelled as a plane wave. However, the following argument, using the plane wave assumption, will be used to show that the expected RMS phase error derived above is in fact a weak upper bound on position error in a plane.

Any two phase differences derived from three transmitting stations serve to locate a point on the plane by intersection of the corresponding hyperbolas. The error surface is then an ellipse on that plane[16]. In an actual system, the three transmitting stations would not be coplanar (either by necessity, or, more likely, by design) and the error volume would be the combination of the corresponding ellipsoids.

This problem is similar to the pseudorange problem in GPS. A pseudorange may be assumed to have the following form:

$$\rho = \|s - u\| + b + \epsilon \tag{3.19}$$

where  $\rho$  is the pseudorange vector,  $s$  is the satellite position,  $u$  is the user position,  $b$  is the user's clock offset from GPS time, and  $e$  is a zero mean Gaussian random variable encapsulating noise.

The estimation of a position solution from measured pseudoranges is covered in [10], [15], [36], and [20]. As the hyperbolic position solution is nonlinear, these methods generally involve either a Taylor-series linearization about a position guess, a

least-squares fit to the nonlinear solution, or the closed form method of [11] which is noniterative.

Chaffee and Abel show in [9] that the traditional GDOP matrix (the root sum square of variance in the position estimate) is the Cramer-Rao bound on estimates of position, as long as pseudorange errors have Gaussian distribution. They show further in [10] that the Bancroft solution to the GPS pseudorange solution may not be optimal in the overdetermined case. This result is applicable to the present system as position solutions may easily be overdetermined if additional transmitters are added to fill in gaps in coverage.

Without recourse to the specific solution methods described above, a simple estimate of position error can be made geometrically. In the three transmitter, coplanar case, start by assuming that the RMS phase error is small and applies to one phase difference at a time. Additionally, assume that the transmitter pairs are located at right angles. The error then serves to displace the line of position formed by the intersection of one fixed hyperbola with another, variable focus hyperbola to which the error is assumed to apply. The length of this line is simply the RMS error of one difference in distance. Since the same argument applies with the error applied to the second hyperbola, the error surface is bounded by a square with sides whose length is the RMS error of the phase measurement.

Thus, in the simple coplanar case, the area of the error surface is bounded by the square of the RMS phase difference error. For the result derived in Section 3.7.2, this error surface is about 6 meters on a side, without the use of signal averaging.

### 3.9 Conclusion

In this chapter we have identified the primary expected sources of error in an indoor hyperbolic radiopositioning system. This preliminary analysis was used to design the experimental apparatus presented in this thesis.

The relationship between signal to noise ratio (SNR) and phase measurement error has also been derived. This relationship drives the radio engineering portion of the

project. Several techniques for increasing the SNR are evident from this analysis, including an increase in transmitter power or antenna efficiency, or an increase in SNR by using coherent averaging to provide a significant processing gain.

# Chapter 4

## Design of Experiment and Apparatus

### 4.1 Introduction

The preparatory system engineering work presented in the previous chapter summarizes the initial planning that was undertaken in order to decide the manner and direction of the thesis work. We now show how that feasibility analysis was translated into the implementation of the experimental apparatus that was designed and built for this thesis.

The fixed transmitting equipment, shown on the left in Figure 4-1, is designed to produce a stable, monochromatic signal at 2.0MHz, while simultaneously producing an identical reference signal modulated on an optical-fiber carried,  $\approx 800\text{nm}$  modulated laser optical signal. This is the “fixed equipment” shown in Figure 4-1. The transmitter output power can be adjusted easily and controlled within  $\pm 1\text{dB}$ , while the aggregate transmitted phase noise is better than  $-120\text{dBc/Hz}$  at  $10\text{KHz}$ .

The receiving unit, shown on the right in Figure 4-1, is capable of simultaneously receiving the over-the-air signal from the transmitter unit as well as the modulated optical signal from the fiber optic reference receiver. The receiver is implemented with a combination of analog signal processing, consisting of amplification, filtering, and downconversion to an intermediate frequency (IF), and digitization and subsequent

digital signal processing which was performed in MATLAB. In a given experiment,

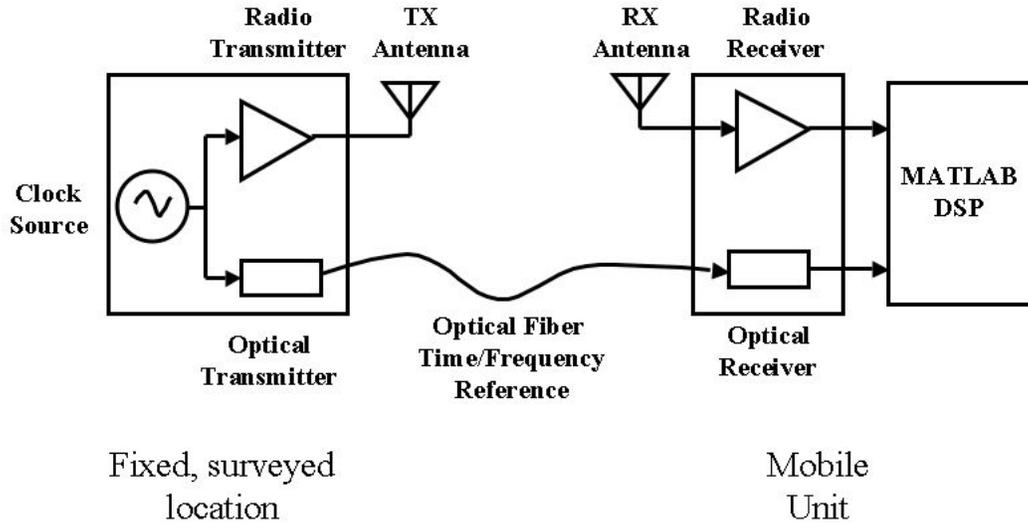


Figure 4-1: Overview of experimental apparatus

the location of the transmitting antenna was surveyed with respect to the Wiesner Building blueprints using a tape measure. These measurements were accurate to within a few cm, mainly due to inaccuracies in the building blueprints. The transmitter and receiver were then calibrated *in situ* to avoid any problems caused by moving the equipment. Then the receiver was moved by hand along a similarly surveyed path while simultaneously digitizing and recording the over-the-air signal and the reference signal. This process produces the desired plots of amplitude and phase versus receiver position.

## 4.2 Transmitter and transmitting antenna design

The first module to be described is the transmitter unit. The design goals to be met, based on the preliminary results of Chapter 2, are:

- Stable, monochromatic signal generation at a fixed frequency of 2.0MHz
- Exceptional phase and amplitude stability both in the short term (phase and amplitude noise) and in the long term (day to day apparatus variation)
- Variable output power from +30dBm to +50dBm
- Fiber optic reference signal at  $\approx 800\text{nm}$ , carrying a reference signal from the transmitter to the receiver for use in phase estimation
- Simple and rugged design

The transmitter unit was built in three sections: the clock, timing, and reference source, the transmit power amplifier, and the transmitting antenna.

The implementation selected to meet these requirements consists of a master oscillator, based on a temperature compensated 8.00000MHz crystal oscillator (TCXO) stable to within  $\pm 2\text{ppm}$ , as well as a digital divider chain, level converters, and power MOSFET gate drive circuitry. This circuitry is shown schematically in Figure 4-2.

Several different power amplifier and antenna configurations were tried experimentally to determine their suitability for these experiments. These designs are characterized by their class of operation. Class A, C, and E amplifiers were designed, built, and tested in the course of transmitter development.

### 4.2.1 Class A amplifiers

In a Class A amplifier, the active element (MOSFET or BJT) is biased at half the required peak operating current. This results in the best linearity, because the transistor's bias current ensures that the transistor is always in its linear region. Since the device can be modelled as a linear element, the amplifier gain is not dependent on the level of the input signal, unlike the nonlinear Class C and Class E amplifiers.

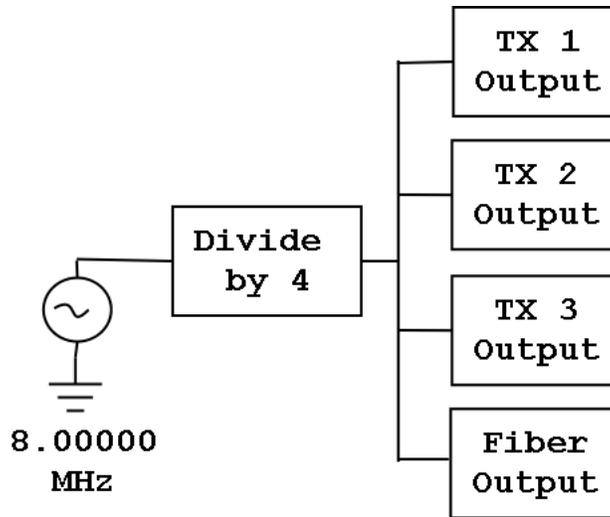


Figure 4-2: Transmitter clock and timing source

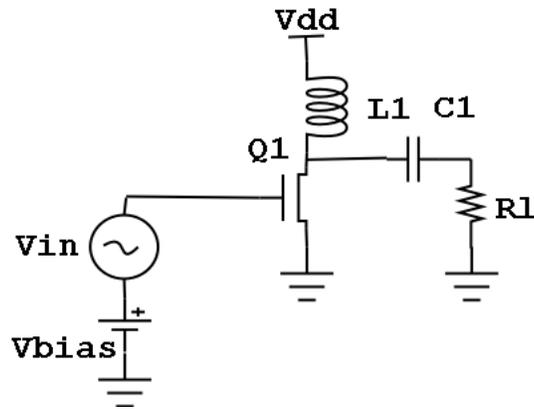


Figure 4-3: Class A amplifier simplified schematic

The simplified operating circuit of the Class A amplifier developed for this thesis is shown in Figure 4-3.

To analyze the operation of the Class A amplifier, start by determining the DC input power  $P_0$  of the amplifier by calculating the voltage and current flowing in the load resistor  $R_l$  during an entire sine wave cycle:

$$I_0 = \frac{V_{dd}}{2R_l}$$

$$P_0 = V_{dd}I_0$$

$$P_0 = \frac{V_{dd}^2}{2R_l} \quad (4.1)$$

Then consider the amount of that power that is actually transferred to the load  $R_l$ , as opposed to that which is dissipated in the active device as heat:

$$P = \frac{V_{pp}I_{pp}}{8} = \frac{V_{dd}^2}{8R_l} \quad (4.2)$$

The major drawback of the Class A amplifier is then obvious. Because the amplifier's efficiency  $\eta$  is defined as

$$\eta = \frac{P}{P_0} \quad (4.3)$$

and since  $P = P_0/4$  (due to the sine wave output signal), efficiency  $\eta \approx 25\%$ . Low efficiency, high heat dissipation, and attendant device size, cost, and power ramifications are the chief limitations of the Class A amplifier. The primary advantages of a Class A amplifier are those accompanying its inherent linearity. A Class A amplifier can be used to amplify any type of input signal, with any type of modulation. This is in contrast to the nonlinear amplifiers in the following sections- nonlinear amplifiers cannot reproduce signals which have a time varying amplitude (also referred to as a "non constant envelope") or multiple frequency components (polychromatic signals) without resorting to rather complex techniques of envelope restoration and attendant compromised linearity.

### 4.2.2 Class C amplifiers

Since a nonlinear transfer function is acceptable as the transmitted signal will be monochromatic and constant envelope (a continuous wave or CW signal), a nonlinear amplifier can be considered for application to this problem. Let us first consider the simplest nonlinear amplifier, which operates in Class C. The Class C amplifier schematic appears as Figure 4-4.

The active device in the amplifier, which may again be either a MOSFET or BJT, conducts only on the positive going half cycle of the input sine wave. So the device

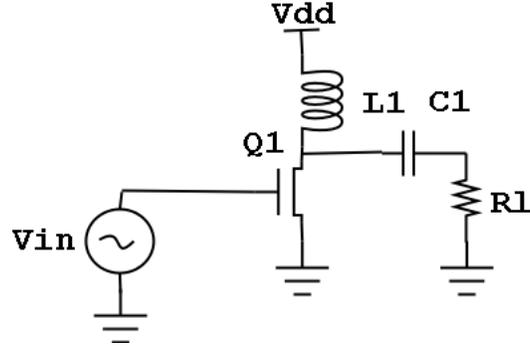


Figure 4-4: Class C amplifier schematic

has two states- on, and in the active region, and off, or inactive. The analysis is therefore carried out in piecewise fashion, first by determining the voltage across the load in the on and off periods:

$$V_s(t) = \begin{cases} V_{on} + V_m \cos \omega t & \text{N-ch off} \\ V_{on} & \text{N-ch on} \end{cases} \quad (4.4)$$

Then the effective voltage  $V_m$  across the load is calculated:

$$\begin{aligned} V_{dd} &= V_{on} + \frac{V_m}{\pi} \\ V_m &= \pi (V_{dd} - V_{on}) \end{aligned} \quad (4.5)$$

and the DC input power  $P_0$  and output power  $P$  are found:

$$\begin{aligned} P_0 &= V_{dd} I_0 \\ P_d &= V_{on} I_0 \\ P &= P_0 - P_d = I_0 (V_{dd} - V_{on}) \end{aligned} \quad (4.6)$$

With this information the efficiency  $\eta$  can be calculated:

$$\eta = \frac{P}{P_0} = \frac{V_{dd} - V_{on}}{V_{dd}} \quad (4.7)$$

From this expression it is clear that highest efficiency occurs when  $V_{on}$  is zero. No practical active device can exhibit this property; typical efficiencies for Class C amplifiers are in the neighborhood of  $\eta \approx 0.7$ .

### 4.2.3 Class E amplifiers

The Class E amplifier is the most efficient type of amplifier known, and was the type eventually selected for use in this thesis. It is a nonlinear amplifier routinely capable of efficiencies in the 90% range. The Class E amplifier was invented in 1964 by Gerald Ewing, and was identified and characterized by Alan Sokal in 1975 [37]. Its origin stems from the realization that the limiting factor in amplifier design is often active device power dissipation. In a Class C amplifier, the active device dissipates power when the switching device is on, because the switching device has a voltage drop  $V_{on}$ .

Referring to Figure 4-5, the Class E amplifier mitigates the problem of power dissipated by the switching device due to its voltage drop by adding an additional resonant circuit formed by L2 and C2 to the load resistance  $R_l$ . This network adds a phase delay such that the active device switches only when the voltage across the active device, and the derivative of the voltage across the active device, are both zero. This minimizes loss by eliminating the need to discharge the device and circuit capacitances. The Class E amplifier design used in this thesis is shown in Figure 4-5.

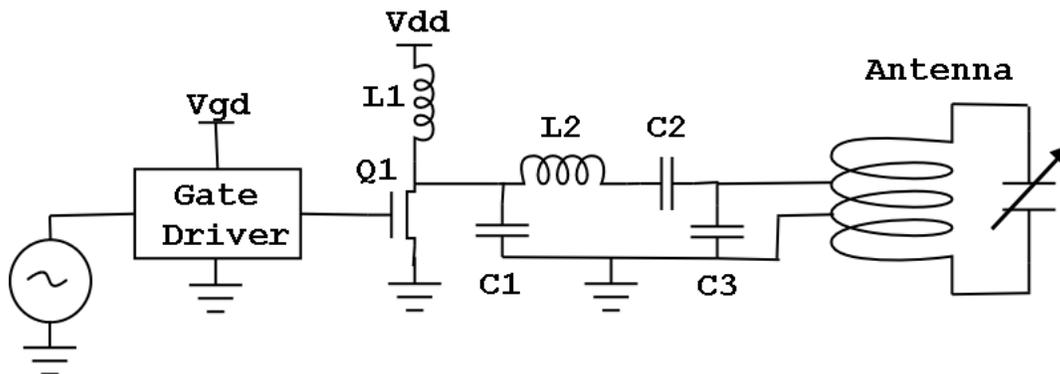


Figure 4-5: Class-E amplifier designed for this thesis

This amplifier has been carefully characterized to ensure that it is operating prop-

erly. For example, Figure 4-6 shows Q1's drain voltage on the same axes as the gate drive voltage. Note that the drain voltage rises just after the gate of the MOSFET has been turned off. This demonstrates that the amplifier is operating in Class E. In Figure 4-7 we see the output voltage at the antenna terminals of the amplifier.

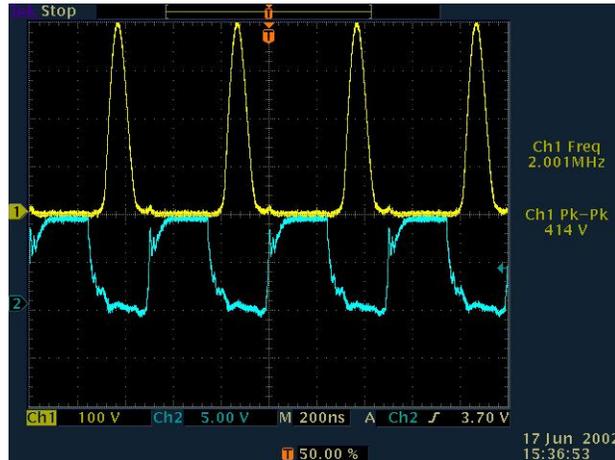


Figure 4-6: Class-E drain voltage measured at  $V_{dd} = 60V$ .

The amplifier is operating with a drain voltage of 60V and is producing a peak-peak output voltage of 278V across 50 ohms. The amplifier is therefore delivering 193.2W to the load. At that time the DC input current to the amplifier was 3.43A, resulting in a DC input power of  $60 \times 3.43 = 205.9W$ . The efficiency  $\eta = 193.2/205.9 = 0.938$ . This is very respectable performance for a Class E amplifier, and as a result of careful design only 12.7W were dissipated as heat.

The Class E amplifier is also unique in terms of phase control. Because the drive signal to the active device is a square wave of 50coherent copies of a master oscillator. This is the approach that would be used in a practical multi-transmitter system. A central oscillator would be fed to each transmitting station via equal length fiber optic cables, thus providing several widely separated transmitters with a source of phase coherence. The signal processing between the fiber optic receiver and the gate (or base) of the power amplifier MOSFET (or BJT) can be all digital, with controlled delays that are within hundreds of picoseconds of each other. This is much more straightforward than individually tuning and phase-calibrating linear amplifiers as

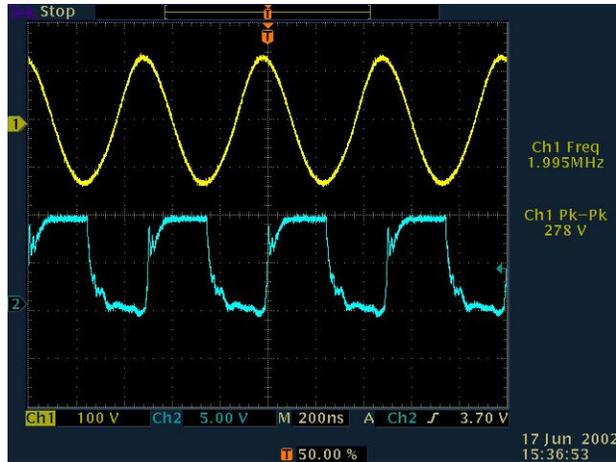


Figure 4-7: Class-E output voltage measured at  $V_{dd} = 60V$ .

would be required if a Class A design were used. Also there would be virtually no way that this phase alignment could drift over time, except for temperature related variations in propagation velocity in the fiber optic cable. This is not anticipated to be a problem in the relatively controlled environments inside most buildings.

#### 4.2.4 Antenna design

The antenna developed for the measurement apparatus is based on the small loop design presented in Chapter 3. It is made of 7.6cm wide copper strapping wound in a spiral around a wooden form. This coil has an inductance of approximately 7uH. It is tapped in two places at the 50Ω points, and presents a return loss of better than -20dB at 2.0MHz. The parallel resonating capacitor  $C_{tune}$  is a vacuum variable type capable of being adjusted between 100pF and 1000pF. Since the coil requires a 906pF capacitor to resonate at 2.0MHz, the capacitor is adjusted close to that value with a capacitance bridge, and then final alignment is performed as an  $S_{11}$  measurement on an HP 8714ES network analyzer.

Figure 4-8 shows the author standing next to the transmitting antenna. The reason for the antenna's relatively large size is the ability to handle significant transmitter power. If the coil turn spacing were reduced, there is the possibility of arcing between turns due to the very high voltages on the exposed copper strapping. In



Figure 4-8: The author and the transmitting antenna

a practical, safe system, the transmitting antennas would be installed in a plastic cabinet with a door interlock so that the transmitter would be disabled if the door were opened.

### 4.3 Receiver design

The receiver that has been designed for this thesis is based on a digitized-IF software radio design. The incoming 2.0MHz signal is received by an EMCO 6502 active *H*-field loop antenna, and is downconverted to a 156.8KHz intermediate frequency (IF) by a custom designed filter/downconverter unit. This intermediate frequency is then digitized by a dual channel 12-bit analog to digital converter card made by Measurement Specialties. Subsequent processing, including IF filtering, amplitude estimation, and phase estimation are performed in software written in MATLAB and making use of MATLAB's data acquisition and signal processing toolboxes. This design for the receiver hardware is shown schematically in Figure 4-9.

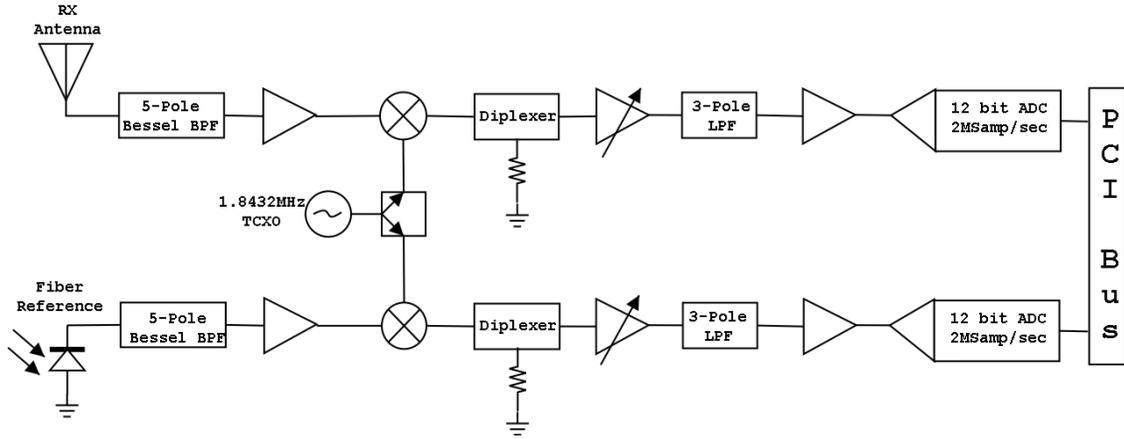


Figure 4-9: Receiver signal flow, hardware

### 4.3.1 Receiving antenna

The receiving antenna is an EMCO 6502 active  $H$ -field loop. This type of antenna was chosen because of its electrostatic shielding properties and its high gain and low noise figure.

Electrostatic shielding is important in this application because it enforces the magnetic field coupling desired between the transmitter and receiver antennae. Additionally, since many types of man made noise (eg motor noise) are primarily electrostatically coupled on electrical wiring, the electrostatic shielding of the EMCO antenna helps to reduce this noise.

The gain, compression point, and dynamic range of this antenna are also good. Gain is important because of the small diameter of the loop. Without a preamplifier at the loop's terminals, the received signal power would be very weak. The gain of a loop like this one is specified in terms of its antenna factor, which is the relationship between the applied magnetic field and the output power across the  $50\Omega$  output port of the device. The EMCO 6502 is specified to have an  $H$ -field antenna factor of  $-41\text{dBA/m}$ .

The 1dB compression point is a measure of the large signal handling capability of the antenna's preamplifier. This antenna exhibits a 1dB compression point of  $5\text{V/m}$  which is a very large field. In practice this translates into the ability to operate close

to the transmitting antenna without experiencing overload. In the experiments of Chapter 6, it was found that the receiving antenna could be located as close as 5m from the transmitting antenna without experiencing preamplifier saturation. This is good performance.

Dynamic range is important to providing good measurement results because the signal from the transmitter varies over a very wide range of power. Close to the antenna, at a distance of 5m, the transmitted signal may be received with a power of 10dBm at the receiver terminals, while at 30m this signal may be only -60dBm. This represents 70dB of received dynamic range. Fortunately the EMCO 6502 is specified to 125dB of dynamic range at these frequencies, so the measurement is not limited by the receiving antenna.

### **4.3.2 Filtering and downconversion**

The downconverter unit is a custom designed unit that is an integral part of the receiver. It is responsible for converting the incoming 2.0MHz signal on two ports to a common 156.8KHz intermediate frequency, which is output on two ports. The reason for this is to process both the incoming radio signal as well as the fiber optic reference signal on equal and identical paths, with the same local oscillator frequency, so that the phase relationship between the two signals is fixed in the receiver at all times. The downconverter is shown schematically in Figure 4-9.

Both the radio signal and the reference signal first pass through a preselection filter, based on a 5-pole Bessel band pass topology. This preselection filter is designed to pass the 2.0MHz incoming transmitter signal while rejecting any noise or interfering signals from an external source. The 3dB bandwidth of this filter is approximately 50KHz, and the phase response is linear through the passband. This is a general characteristic of the Bessel filter and is important in this design not because of its absolute phase shift value but rather because of the smooth nature of the phase variation in the passband. Absolute phase through the filters and other components in the downconverter is double differenced out by the phase measurement algorithm and has no effect on the final phase measurement.

The downconverter's local oscillator is a 1.8432MHz crystal oscillator based on the Butler low-noise topology. This oscillator exhibits excellent phase noise characteristics of -120dBc/Hz at 10KHz offset (measured). This means that there is little phase jitter introduced by the local oscillator. However, even if there were a phase jitter, this phase would be common to both the received radio signal and to the fiber optic reference signal. The error would then cancel out in the differential phase measurement.

The mixers used in the downconverter unit are Mini-Circuits double balanced diode mixers. These mixers have excellent strong signal performance, as evidenced by their third order intercept point (IP3) of +15dBm. Strong signal performance is important because of the very wide variation in received signal strength with distance.

### 4.3.3 IF-DSP: The software radio

The 156.8KHz intermediate frequency is digitized by a PCI card based two channel data acquisition card, model PCI-DAS4012, manufactured by Measurement Specialties Inc. This card contains a precision voltage reference, a crystal oscillator based timebase, and two 12-bit analog to digital converters.

The card is controlled by a MATLAB program written using MATLAB's data acquisition toolbox. The IF signal processing flow is shown in Figure 4-10.

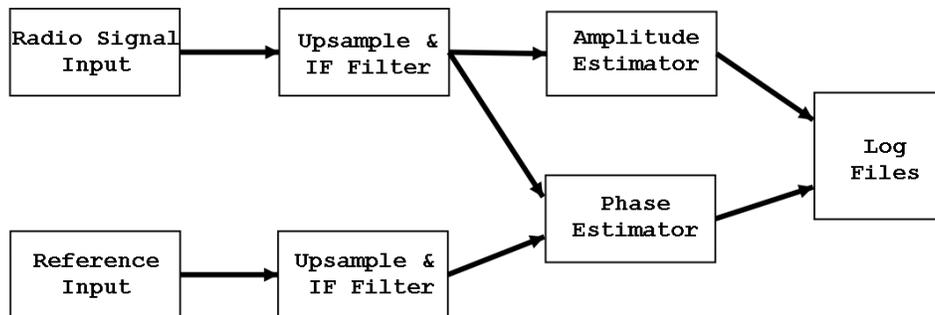


Figure 4-10: Receiver signal flow, software IF processing

The two channels of 156.8KHz IF representing the radio signal and the reference signal are sampled at a 2.0MSPS sampling rate. This rate is limited by the MATLAB toolbox device drivers for the data acquisition card. The IF is well within the 1.0MHz

Nyquist limit, and on average 12.75 samples are taken per cycle of each incoming IF signal.

First, as shown in Figure 4-10, both signals are upsampled with MATLAB's linear interpolation upsampling mechanism, to end up with a reasonable number of samples in the output buffer. Since the sampling interval was 1msec (for a 1KHz update rate), there are 20Ksamples in each upsampled data frame. This data is band pass filtered by identical 40-tap FIR implementations of a Gaussian linear phase filter. A Gaussian shape was chosen because Bessel filters, like those implemented in analog form in the downconverter, are not realizable as an FIR filter [31]. This filter has a bandpass characteristic of 1KHz bandwidth, which was chosen as a compromise between noise bandwidth, filter amplitude response shape, and phase response.

Then, the amplitude and phase estimators described in Chapter 5, Section 5.5 are applied to this data set to produce the final amplitude and phase estimates.

All raw sample data as well as intermediate values are stored to disk for subsequent processing, and to allow for the use of different signal processing strategies in future work.

## 4.4 Conclusion

In this chapter the design and construction of the measurement apparatus are explained. This apparatus is the first known measurement system specifically designed for indoor propagation measurements at a 150m wavelength. It is shown in Chapter 6 that this equipment, when used together to produce amplitude and phase measurements, is capable of better than 1dB of amplitude accuracy and better than 1 degree of phase accuracy, over the more than 70dB of dynamic range that is encountered when making the indoor propagation maps that are the core contribution of this thesis.

# Chapter 5

## Signal processing methods

### 5.1 Introduction

An *estimator* is a function  $\mathcal{F}$  which, given some inaccurate measurements  $\vec{z}$  of an underlying system  $x$ , produces an estimate  $\hat{x} = \mathcal{F}(\vec{z})$  which best represents the state of that underlying system. Estimators can be designed to take advantage of any available information about the system.

Finding one's position in a building is an estimation problem because a set of measurements, each one unreliable when taken on its own, is used in concert with an estimator that is designed to produce the best estimate of the user's position. In this thesis, two distinct types of estimator have been developed to solve the positioning problem.

The first application of estimation theory is the *nonlinear estimation* of each sinusoidal input signal amplitude or phase based on a sampled representation of that signal. The second application is the *linear filtering* of position solutions generated from an application of the hyperbolic position solution.

Both of these estimation problems have well known *maximum likelihood* estimators. In the former nonlinear, sinusoidal estimation case, the maximum likelihood estimator is derived from the periodogram or Fourier transform representation of the sampled input signal. In the latter case, the discrete Kalman filter is used to produce a maximum likelihood estimator for the receiver position at a given point in time.

In Section 5.5 the specialized time domain phase estimator developed for this thesis is presented. This technique produced the phase measurements presented in Chapter 6. Section 5.2 presents the least squares technique used for the amplitude fits presented in Chapter 6.

## 5.2 Linear systems and estimators: the amplitude fit

We will cover the linear estimation problem first as the Kalman filter solution shown at the end of this section can be generalized to cover nonlinear problems.<sup>1</sup> We will start by defining the form of a linear estimator of a random variable, then introduce some various possible forms of an optimal linear estimator. We will then introduce the Kalman filter, which is a linear recursive estimator that is optimal in the least squares sense. The Kalman filter will be used to filter the measurements from the mobile receiver in our positioning system to produce finished position estimates.

Let us begin by assuming that a set of  $l$  measurements  $\vec{z}$  of a variable  $\vec{x} = [x_1, \dots, x_n]$  have been taken. Note that  $l$  is not necessarily equal to  $n$ , for reasons explained shortly. In an experimental scenario,  $\vec{x}$  represents the true value of a particular measurement of a physical system. For the purpose of analysis, we will treat  $x$  as a random variable and analyze it in a statistical manner to arrive at an estimate  $\hat{x}$  which represents our best guess as to the true value of the quantity being measured.

In this linear estimation problem, we will assume that each of the  $l$  measurements represents a linear combination of the (unknown) true value of  $x$ , which may take on  $n$  values, plus a random measurement noise  $v$  that is added to each measurement. Strictly speaking this is an *affine transformation* of  $x$ , but the term linear is applied in the context of estimation theory. Since  $x$  is a random variable,  $z$  is also a random

---

<sup>1</sup>This generalization is called an *extended Kalman filter*. While the extended Kalman filter is based on ad hoc, heuristic methods, EKFs have been shown to work well in many practical problems.

variable, and may be defined as follows:

$$\vec{z} = H\vec{x} + \vec{v} \tag{5.1}$$

where  $\vec{z}$  is an  $l$  element vector of measurements,  $\vec{x}$  is an  $n$  element state vector, and  $\vec{v}$  contains the additive contribution of the random measurement noise.  $H$  is an  $l \times n$  matrix which represents the contribution of each of the  $n$  possible values of  $\vec{x}$  to any given measurement. If  $l = n$  there is exactly enough information to solve the underlying system of equations, but if  $l < n$  or  $l > n$  the problem is underdetermined or overdetermined respectively.

There are many possible ways to recover an estimate of  $\vec{x}$  from  $\vec{z}$ . The method that is chosen to derive this estimate depends on the goals of the person analyzing the system, how many measurements are available, and how much is known about the measurement process. We will first examine the linear least squares estimator and show its many desirable features.

### 5.2.1 The linear least-squares estimator

Let us assume there is a quantity  $J$ , which is an error metric for our estimator. One sensible error metric would be to define  $J$  to represent the matrix of squared deviations between the measured values and the underlying true value, taking in to account the under- or overdetermined cases, as follows:

$$J = (\vec{z} - H\hat{x})'(\vec{z} - H\hat{x}) \tag{5.2}$$

We then set  $\partial J / \partial \hat{x} = 0$  to arrive at the least squares estimator (LSE):

$$\hat{x} = (H'H)^{-1}H'\vec{z} \tag{5.3}$$

Note that  $(H'H)^{-1}H'$  is often called the pseudoinverse of  $H$ .

The least squares estimator has several desirable properties. First, if the problem is exactly determined, the estimator correctly produces the exactly determined solution

to the underlying system of linear equations. If the problem is overdetermined, that is, if there are more measurements  $l$  than there are degrees of freedom  $n$ , the least squares estimator makes use of this additional information to minimize the residual error. If on the other hand the problem is underdetermined, the least squares estimator also produces a good result by effectively ignoring the nullspace of  $H$  and again returning the estimate with least residual error.

Additionally, thinking back to our original statement that  $x$  would be treated as a random variable, the least squares estimator does not require knowledge of the statistical properties of  $x$  or  $z$ ; if there is no further information about the probability density function (PDF) for these variables, the least squares estimator makes use of all available information and is therefore an *optimal* estimator in that case.

## 5.2.2 Bias and error covariance

Now that we have introduced the notion of an estimator, and given the simple example of a linear least-squares estimator, we will proceed to describe methods of evaluating the performance of an estimator that *does* make use of information about the PDFs of  $x$  or  $z$ .

To start intuitively, let us consider the case of a Gaussian distribution for  $x$ . If  $x$  is a Gaussian random variable, it is fully described by its probability density function  $p_x(x)$  as follows:

$$p_x(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left[-\frac{(x-m)^2}{2\sigma^2}\right] \quad (5.4)$$

Here,  $m$  is the mean, which for a Gaussian is equal to the expected value  $E(x)$ , while  $\sigma^2$  is the variance of  $x$ . These two parameters are sufficient to completely describe  $x$ .

Now let us assume that we wish to evaluate the performance of an estimator  $\hat{x}$  where  $x$  is a variable (not necessarily a random variable, nor specifically a Gaussian). First we define the error  $e(\vec{z})$ , which represents the difference between the estimator's output and the true value of  $x$ :

$$e(\vec{z}) = \hat{x}(\vec{z}) - x \quad (5.5)$$

and then define the *bias*  $b_{\hat{x}}(x)$  as

$$b_{\hat{x}}(x) = E[\hat{x}(\vec{z}) - x] \quad (5.6)$$

$$= \left[ \int_{-\infty}^{\infty} \hat{x}(\vec{z}) p_z(\vec{z}; x) dz \right] - x \quad (5.7)$$

Thus, the bias of an estimator is the difference between the expected value of the estimator evaluated at a particular measurement, and the true value of the quantity being measured.

The *error covariance* is a measure of how the bias of an estimator correlates with its error. This metric is useful because its trace is the mean square estimation error.

$$\Lambda_e(x) = E \left[ [e(\vec{z}) - b_{\hat{x}}(x)][e(\vec{z}) - b_{\hat{x}}(x)]' \right] \quad (5.8)$$

Note that in the case of a estimator of a nonrandom process, as many practical estimators are, the error covariance is the same as the covariance of the estimator itself. This result is of importance in the application of the Cramer-Rao bound to practical problems.

### 5.2.3 The Cramer-Rao bound

We stated in Section 5.2.1 that the linear least squares estimator is optimal if we know nothing about the PDF of  $x$  or  $z$ . Let us assume that we do have some statistical information about  $x$ , in the form of the probability density function for  $z$ . Since  $z$  depends linearly on  $x$ , the PDF of  $z$  is obviously related in some way to the PDF of  $x$ . The Cramer-Rao bound is a tool that will allow us to make a useful generalization about the PDF of  $\hat{x}$ , knowing only the PDF of  $z$ . It does this by providing an asymptotic lower limit for the variance of any unbiased estimator of  $x$ .

The Cramer-Rao bound is a result of information theory whose derivation is given in Willsky [51]. The existence of the bound depends on the existence of a quantity

called the *Fisher information* of  $z$ , which is defined by

$$I_z(x) = E \left[ \left( \frac{\partial \ln p_z(z; x)}{\partial x} \right)^2 \right] \quad (5.9)$$

The Cramer-Rao bound is given by the reciprocal of the Fisher information:

$$\lambda_{\hat{x}} \geq \frac{1}{I_z(x)} \quad (5.10)$$

The Fisher information is in a sense a measure of curvature of the natural log of the underlying PDF as  $x$  varies. If the PDF has sharply defined peaks, then the value of  $x$  will be easier to determine from the measurements.

The PDF  $p_z$  must be strictly positive for all  $x$  and  $z$  for the Fisher information to be calculable, so for some problems no Cramer-Rao bound exists. An estimator is said to be *efficient* if it achieves the Cramer-Rao bound given the input conditions.

Let us now determine the Cramer-Rao bound for a problem of practical importance. For example, consider the Cramer-Rao bound for a scalar Gaussian problem  $z = x + v$  where  $v$  is a zero mean Gaussian, representing noise, with variance  $\sigma^2$ . In this case we calculate the Fisher information as follows:

$$\ln p_z(z; x) = -\frac{1}{2\sigma^2}(x - z)^2 - \frac{\ln 2\pi\sigma^2}{2} \quad (5.11)$$

Here the Fisher information is

$$I_z(x) = \frac{1}{\sigma^2} \quad (5.12)$$

and the Cramer-Rao bound for the variance of any unbiased estimator for  $x$  is

$$\lambda_{\hat{x}} \geq \sigma^2 \quad (5.13)$$

This is a useful result as many problems can be simplified to this sort of scalar Gaussian subproblem, for example in a large data set where the Central Limit Theorem is applicable.

## 5.2.4 Estimation, filtering, and smoothing

As we stated in Section 5.2, estimation is defined as generating a guess as to the state of a system given one or more unreliable measurements of that system. In many practical cases, one is concerned with the time evolution of the state of a system. Three related concepts for applying estimation theory to time-varying systems are *filtering*, *smoothing*, and *prediction* [18].

In a time varying system, filtering refers to using a combination of current and past measurements to estimate the current state of a system. Smoothing refers to using a set of measurements up to the present time to estimate the state of a system at some time in the past. Prediction is using past and present measurements to estimate the system state in the future.

In this thesis we are concerned with filtering and smoothing of time varying data sets, which include raw sampled signals from a radio receiver, as well as state estimation to determine a position solution.

## 5.2.5 Recursive filters

In the 1940s Norbert Wiener developed the method of spectral filtering and showed its many benefits for filtering and smoothing linear systems. However, because the Wiener filter is noncausal, requiring data from both  $t_-$  and  $t_+$  to form its estimate at a time  $t_0$ , its use is limited in many practical problems for which a causal response is desired. Additionally, the amount of state storage required to form an estimate at  $t_0$  grows with time, which is a significant barrier to practical utility. Fortunately, in 1960 Kalman published his solution [26] to the Wiener problem, an optimal recursive filtering algorithm based on the Wiener filter.

A recursive filter is one for which the current state estimate  $\hat{x}_{t_0}$  is based on the current measurement  $\bar{z}_t$  and the immediate past state estimate  $\hat{x}_{t-1}$ . The amount of state storage required to implement a recursive filter is therefore fixed and does not grow with time. This is of great advantage in nearly every practical problem.

Since its introduction, the Kalman filter has been applied to nearly every type of

control problem ranging from missile guidance systems [27] to terrestrial navigation and generalized control problems. It is “optimal” in almost any way that one might want a filter to be optimal. The Kalman filter takes advantage of *a priori* state information including initial conditions, knowledge of the system and measurement process, and first- and second-order noise statistics. Since the first and second moments (mean and standard deviation) are often the only characterizations of noise available to the engineer, and therefore the Kalman filter incorporates all available information.

### 5.2.6 Discrete Kalman filters

Assume we are trying to estimate the state  $\vec{x}$  of a discrete time process which, at a certain time  $t$ , can be described by the linear stochastic difference equation

$$\vec{x}_t = A\vec{x}_{t-1} + B\vec{u}_t + \vec{w}_{t-1} \quad (5.14)$$

In this difference equation the matrix  $A$  is  $n \times n$  and gives the state transition from  $x_{t-1}$  to  $x_t$ , neglecting both noise and a driving function. The matrix  $B$  is  $x \times l$  and gives the contribution from the driving function, if there is one. In this system,  $w$  represents noise arising from the process itself; it is assumed to be a Gaussian with zero mean and variance  $Q$ .

We will now (as in Eq 5.1) take a series of measurements  $z_t$ , related to this system by:

$$\vec{z}_t = H\vec{x}_t + \vec{v}_t \quad (5.15)$$

Here,  $v$  represents measurement noise. We assume that  $v$  is Gaussian with zero mean and variance  $R$ . The matrix  $H$  encapsulates the relationship between the system state and the measurement vector.

Now we define the two steps of the discrete Kalman filter, which we will present from an algorithmic point of view. Good mathematical treatments of the derivation of the Kalman filter appeared first in [26], although more a more modern references

is found in [18]. These references also provide proof of the various ways in which the Kalman filter is optimal. We will concentrate here on the application of the Kalman filtering algorithm to a practical engineering problem rather than on the underlying theory.

At its heart the discrete Kalman filter algorithm operates a feedback mechanism. The first step, called the *prediction* step, takes the current system state estimate and projects it forward in time, to the time that the next measurement is expected. Then the second step, the *correction* step, adjusts that state prediction by including measured data to form the final estimate at that time. The benefits of this sort of recursive algorithm are numerous, but chief among them is that only the immediate prior state estimate need be retained, so the storage memory for the algorithm does not grow with time. Similarly, computational complexity is constant with time so position solutions do not need to take longer to compute as time passes.

### The prediction step

The first step of the discrete Kalman filter algorithm begins by predicting the future state estimate of the system forward in time. To do this, we first project the state estimate forward in an *a priori* sense; that is, first without the inclusion of measurement data. We will denote the predicted value of the state estimate from this first phase with a superscript *prj* to indicate that it is based on projection from *a priori* data; eg  $\hat{x}_t^{prj}$ .

$$\hat{x}_t^{prj} = A\hat{x}_{t-1} + B\hat{u}_t \quad (5.16)$$

Here  $A$  and  $B$  are the state transition matrix and driving function matrices, respectively. Then we project the covariance estimates forward as well:

$$P_t^{prj} = AP_{t-1}A' + Q \quad (5.17)$$

The state prediction and covariance predictions are then corrected by application of measurement data in the second step of the algorithm.

## The correction step

Now the estimate formed by the projection of the first step is corrected by the application of measured data. The Kalman gain matrix  $K_t$  is used to add a certain amount of information from the measurements to the state estimate, in an amount dependent on the measurement's error covariance. By this method, unreliable measurements have less weight than reliable ones (and vice versa). This covariance is assumed to be time varying and is continuously updated. First the Kalman gain matrix is computed:

$$K_t = P_t^{prj} H' (H P_t^{prj} H' + R)^{-1} \quad (5.18)$$

This optimal form of the gain matrix  $K_t$  is given by an optimization problem on the sum of the diagonal elements of the covariance matrix  $P_t$ . The derivation of  $K_t$  in the form shown here is given in [18]. Then the finished state estimate at time  $t$  is computed:

$$\hat{x}_t = \hat{x}_t^{prj} + K_t (z_t - H \hat{x}_t^{prj}) \quad (5.19)$$

and the error covariance matrix is updated as well, where  $I$  represents the identity matrix of appropriate dimensions:

$$P_t = (I - K_t H) P_k^{pr} \quad (5.20)$$

Note that the covariance matrix is also updated at each time, providing a convenient estimate of error alongside the state estimate. This is a very important feature of the Kalman filter.

## System identification and filter parameters

The two step time update approach for the discrete Kalman filter implementation given in this section requires the modeling of the physical process being measured (in order to determine  $H$ ) as well as the tuning of the parameters  $A$ ,  $B$ ,  $Q$  and  $R$  to give the “best results” from measured data. This is the *system identification* problem. There are a wide variety of approaches to the system identification problem,

ranging from building and testing physical models (eg in the case of systems that can be successfully modelled with the classical equations of motion, which is a common case) to more complex systems that do not have simple models.

In general, if one has a plausible candidate model for a process with unknown parameters, simply selecting an appropriate value of the process noise covariance  $Q$  will yield acceptable results [49]. That is the approach taken in this thesis.

### 5.3 Nonlinear systems and estimators: Generalizations

Many practical systems are nonlinear; measurements of their state have the following general form, where  $h$  is a nonlinear function of  $x$ :

$$\vec{z} = h(\vec{x}) + \vec{v} \tag{5.21}$$

We will now show that there is no generalized efficient estimator  $\hat{x}$  for this system by attempting to find its Cramer-Rao bound. This argument parallels Willsky [51].

If  $v$  is Gaussian with zero mean and variance  $\sigma^2$ , we can find the PDF for  $z$ :

$$p_z(\vec{z}; \vec{x}) = N(\vec{z}; h(\vec{x}), \sigma^2) \tag{5.22}$$

and attempt to find the Fisher information, proceeding as in Section 5.2.3. We begin by calculating the partial derivative:

$$\frac{\partial \ln p_z(\vec{z}, \vec{x})}{\partial \vec{x}} = \left( \frac{\vec{z} - h(\vec{x})}{\sigma^2} \right) \frac{dh(\vec{x})}{dx} \tag{5.23}$$

and then compute the Fisher information itself:

$$I_z(\vec{x}) = E \left[ \left[ \left( \frac{\vec{z} - h(\vec{x})}{\sigma^2} \right) \frac{dh(\vec{x})}{dx} \right]^2 \right] \tag{5.24}$$

$$\begin{aligned}
&= \left[ \frac{dh(\vec{x})}{dx} \right]^2 E \left[ \left( \frac{\vec{v}}{\sigma^2} \right)^2 \right] \\
&= \frac{1}{\sigma^2} \left[ \frac{dh(x)}{dx} \right]^2
\end{aligned}$$

and then the Cramer-Rao bound:

$$\lambda_{\hat{x}}(\vec{x}) \geq \frac{\sigma^2}{(dh(\vec{x})/dx)^2} \quad (5.25)$$

For equation 5.25 to represent an efficient estimator, it must also be a *valid* estimator; that is, if

$$\vec{x} + \frac{1}{I_z(\hat{x})} \frac{\partial}{\partial x} \ln p_z(\vec{z}; \vec{x}) = \left( \vec{x} - \frac{h(\vec{x})}{dh(\vec{x})/dx} \right) + \frac{\vec{z}}{dh(\vec{x})/dx} \quad (5.26)$$

is a function only of  $z$ . Note that the right hand term of eq. 5.26 is the only one with a dependence on  $z$ . Since  $z$  can be arbitrary, no efficient estimate  $\hat{x}$  can exist unless  $dh(\vec{x})/dx$  does not depend on  $x$ . Only the linear case for  $h(\vec{x})$  has this property- all nonlinear cases have an  $x$  dependence. Therefore no efficient estimator  $\hat{x}$  can exist for a nonlinear system.

## 5.4 The nonlinear estimator: phase estimation

We have now shown that, in general, efficient estimators for nonlinear systems do not exist. However, for certain practical cases there are nonlinear estimators that are useful and practical to implement. One of these is the *sinusoidal estimation* case, which figures prominently in the analysis for this thesis.

### 5.4.1 The sinusoidal estimation problem

The sinusoidal estimation problem, central to the analysis of the data collected for this thesis, concerns the estimation of the amplitude, frequency, and phase of a noisy sinusoidal signal based on distinct observations  $n$ :

$$z[n] = A \cos(\omega_0 n + \phi) + v(n) \quad (5.27)$$

where  $A > 0$  and  $0 \leq \omega_0 < \pi$ . We assume that the noise contributions  $v$  are independent, identically distributed (iid) Gaussian random variables with zero mean and variance  $\sigma^2$ .

### 5.4.2 Cramer-Rao bounds for sinusoidal estimation

We will first calculate the Cramer-Rao bounds for each of the parameters  $A$ ,  $\omega_0$  and  $\phi$ , given  $N$  observations. Let us collect these parameters into a single state vector  $x$  as follows:

$$\vec{x} = \begin{bmatrix} A \\ \omega_0 \\ \phi \end{bmatrix} \quad (5.28)$$

Similarly, we will gather the observations  $z[n]$  as well:

$$\vec{z} = \begin{bmatrix} z[0] \\ \vdots \\ z[N-1] \end{bmatrix} \quad (5.29)$$

We will then define the *signal to noise ratio* or SNR for these measurements<sup>2</sup>, which is:

$$\gamma = \frac{A^2}{2\sigma^2} \quad (5.30)$$

and calculate the elements of the Fisher information matrix:

$$[I_z(\vec{x})]_{ij} = -E \left[ \frac{\partial^2}{\partial x_i \partial x_j} l(\vec{z}; \vec{x}) \right] \quad (5.31)$$

where  $l$  is defined as follows:

$$l(\vec{z}; \vec{x}) = \ln p_z(\vec{z}; \vec{x}) = -\frac{N \ln(2\pi\sigma^2)}{2} - \frac{1}{2\sigma^2} \sum_{n=0}^{N-1} [z[n] - A \cos(\omega_0 n + \phi)]^2 \quad (5.32)$$

---

<sup>2</sup>In a radio frequency engineering sense, SNR is the ratio of signal power to noise power in a given receiver bandwidth. In this thesis the accuracy of the sinusoidal estimators used depend critically on SNR and great effort has been made to ensure the maximum SNR possible.

which is too cumbersome to be useful as an engineering rule of thumb. Therefore Willsky[51] uses “order of dependence” notation to denote the Cramer-Rao bounds for amplitude, frequency, and phase for a large number of measurements  $N$  as follows:

$$\text{var} \left( \frac{\hat{A}(\vec{z})}{A} \right) \geq \frac{1}{A^2} [I_z^{-1}(\vec{x})]_{11} \approx \frac{2\sigma^2}{A^2 N} = \frac{1}{\gamma N} \approx O \left( \frac{1}{\gamma N} \right) \quad (5.33)$$

$$\text{var} (\hat{\omega}_0(\vec{z})) \geq [I_z^{-1}(\vec{x})]_{22} \approx \frac{12}{\gamma N(N^2 - 1)} \approx O \left( \frac{1}{\gamma N^3} \right) \quad (5.34)$$

$$\text{var} (\hat{\phi}_z) \geq [I_z^{-1}(\vec{x})]_{33} \approx \frac{2(2N - 1)}{\gamma N(N + 1)} \approx O \left( \frac{1}{\gamma N} \right) \quad (5.35)$$

### 5.4.3 A maximum-likelihood sinusoid estimator

Given the constraints on nonlinear estimators put forth in Section 5.3, one might expect to be limited to inconvenient or *ad hoc* estimation solutions. But as it turns out, the maximum-likelihood estimator for the sinusoidal estimation problem can be built from the the very versatile discrete Fourier transform (DFT) when the number of samples  $N$  is large. We can therefore say that the DFT is the *asymptotic maximum likelihood estimator* for amplitude, frequency, and phase of a sinusoid.

Recall from our discussion of the least-squares estimator that the MLE produces the following estimator  $\hat{x}$ :

$$\hat{x}(\vec{z}) = \arg \max_x l(\vec{z}; \vec{x}) \quad (5.36)$$

which is the solution to the nonlinear least squares optimization problem:

$$\begin{bmatrix} \hat{A} \\ \hat{\omega}_0 \\ \hat{\phi} \end{bmatrix} = \arg \min_{(A, \omega_0, \phi)} J(A, \omega_0, \phi) \quad (5.37)$$

where  $J$  is defined as

$$J(A, \omega_0, \phi) = \sum_{n=0}^{N-1} [z[n] - A \cos(\omega_0 n + \phi)]^2 \quad (5.38)$$

This least squares optimization problem has as an asymptotic solution the periodogram, or discrete Fourier transform  $Y_N(e^{j\omega})$ .

$$Y_N(e^{j\omega}) = \frac{1}{\sqrt{N}} \sum_{n=0}^{N-1} z[n]e^{j\omega n} \quad (5.39)$$

For example, the asymptotic MLE for frequency is

$$\hat{\omega}_0 = \arg \max_{\omega} |Y_N(e^{j\omega})|^2 \quad (5.40)$$

while the asymptotic MLE for amplitude is found by first determining the frequency of the peak, and using its amplitude:

$$\hat{A}^2 = \frac{4}{N} |Y_N(e^{j\omega_0})|^2 \quad (5.41)$$

and the asymptotic MLE for phase is the negated phase of  $Y_N(e^{j\omega})$  at its peak:

$$\hat{\phi} = -\tan^{-1} \left( \frac{\text{Im} (Y_N(e^{j\omega_0}))}{\text{Re} (Y_N(e^{j\omega_0}))} \right) \quad (5.42)$$

The implementation of the asymptotic MLE for amplitude, frequency, and phase can therefore be simplified to one pass through a discrete Fourier transform algorithm. From a computational point of view, the Fast Fourier Transform algorithm (FFT) has  $O(N \log N)$  run time and is therefore an excellent candidate for a practical implementation. In this thesis work the native MATLAB FFT implementation was used successfully.

Figure 5-1 shows the performance of this periodogram estimator in terms of amplitude estimation variance with respect to SNR, for varying  $N$ . It is clear that this estimator asymptotically reaches the CRB for amplitude estimation, as  $N$  increases.

Figure 5-2 shows the performance of this periodogram estimator in terms of frequency estimation variance with respect to SNR, for varying  $N$ . It is clear that this estimator also asymptotically reaches the CRB for frequency estimation, as  $N$  increases.

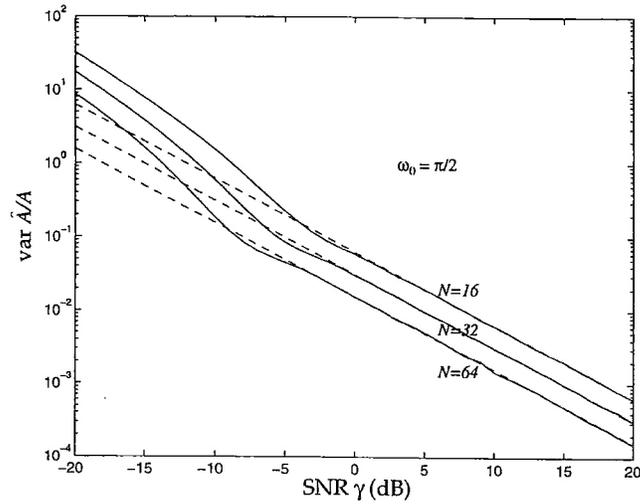


Figure 5-1: FFT amplitude estimation performance with respect to the CRB for varying  $N$ . Note that since amplitude and phase estimation are both  $O(\frac{1}{\gamma N})$  the phase estimation plot is identical.

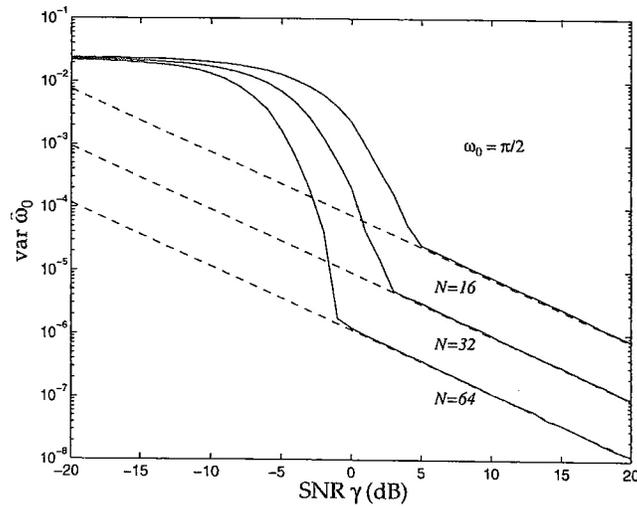


Figure 5-2: FFT frequency estimation performance with respect to the CRB for varying  $N$

## 5.5 A time domain variant of the periodogram estimator

Due to the occasionally large data sets gathered in the course of this thesis research, it was desirable to find a time domain estimator for phase that is computationally efficient and does not have a large memory requirement. The typical FFT algorithm (as used by MATLAB) requires that all samples in the input dataset be available simultaneously because of the butterfly operation at the heart of the FFT [31]. Therefore an *ad hoc* time domain estimator was devised, with the intent of preserving the spirit of the FFT-based maximum likelihood phase estimator. It is not claimed that this estimator is a maximum likelihood estimator for an arbitrary signal, but it is postulated that in the limiting case where the bandpass filter used is infinitely narrow and correctly positioned it will yield an equivalent result.

This estimator works because the input signal is narrowband, at a known frequency, and the desired phase output is differential with respect to the reference signal. The input dataset is first filtered with a narrowband FIR bandpass filter, an operation which is  $O(N)$ . If the filter's center frequency is equal to the (known) frequency of maximum energy, and the bandwidth of the bandpass filter is sufficiently narrow, this operation returns a sampled representation of the input signal equivalent to the frequency peak of the FFT that is the first step of the periodogram based MLE.

Note that in the present experimental case, because the incoming radio signal and the reference signal have the same exact frequency (being direct copies of the same internal oscillator signal) this approach also yields the sinusoids which have the same phase relationship to each other. If the bandwidth of the filter is sufficiently narrow, the noise rejection property of the FFT peak finding method is also preserved.

Since the bandpass of the filter is assumed not to include DC, the resulting two sinusoids are by definition zero mean. The signum function is then applied to convert the sinusoidal input to a square wave output. This is equivalent to the limiting amplifier used in an FM demodulator.

The intersection area of the two square waves is then directly proportional to

phase. The square waves are multiplied, and their resultant (-1 to +1) is mapped to 0 to 180 degrees of relative phase. The measured performance of the entire measurement system, including this estimator, is documented in Chapter 6. In practice, this estimator performed to expectations, yielding measured phases within 1 degree of the test signal input.

## 5.6 Conclusions

In this chapter the basics of linear and nonlinear estimation have been presented and the specific cases applicable to this thesis have been covered in detail. In particular, the techniques of linear least-squares estimation, Kalman filtering, and FFT-based sinusoidal estimation find application in the analysis of the data gathered for this thesis. The *ad hoc* time domain based estimator used in the real time data collection part of the thesis is also presented.

# Chapter 6

## Experimental Results and Analysis

### 6.1 Introduction

In this chapter we present the results of the experiments carried out using the equipment described in Chapter 4. These experiments were designed to characterize the amplitude and phase response of a variety of indoor propagation paths at the test frequency of 2.0MHz. The author's initial working hypothesis was that at these frequencies, where the wavelength is on the order of the building's dimensions, the internal structure of the building would not have a great deal of effect on signal propagation, and therefore a simple model based on the free space propagation model would be appropriate to describe the observed propagation behavior.

The experiments were carried out in three phases. First, several experiments were carried out using NIST-traceable HP test equipment to simulate the expected over-the-air signals, and to characterize the amplitude and phase accuracy of the measurement apparatus itself. These experiments, described in Section 6.2, show that the measurement apparatus provides amplitude measurements accurate to within 1dB over a dynamic range of over 70dB. Phase accuracy was shown to be better than 1 degree over the same dynamic range. These results are essentially as good as the HP test equipment used to produce this characterization.

Secondly, to confirm that the custom-built propagation measurement apparatus was functioning properly, a series of outdoor measurements was carried out. The

transmitter was placed outdoors in an open area adjacent to the Wiesner Building. After initial calibration as described in Section 6.2, the receiver was then moved away from the transmitter in a straight line path, radially away from the transmitter. Measurements of signal amplitude and phase were taken at intervals of one foot (30.48cm) over a total distance of 25m. In these experiments, results agreed well with the free space model; signal phase was shown to be linear with distance, and to exhibit the expected slope of 2.4 degrees/meter (0.0418 radians/meter). Signal amplitude exhibited the expected power law dependence on distance, but the observed attenuation was significantly greater ( $p \propto d^{-7.6}$ ) than that expected for the free space dipole-dipole model ( $p \propto d^{-6.0}$ ) presented in Chapter 3. This is likely due to the presence of a poorly conductive ground beneath the measurement apparatus.

With these results showing that the apparatus was working properly, indoor measurements were then carried out in the corridors of the Wiesner Building. The surprising results of these experiments show that, even at the relatively long wavelength of 150m, the structure of the building has a very significant impact on indoor radio wave propagation. At very close ranges (transmitter-receiver separations of approximately 10m to 15m) received signal amplitude and phase agree reasonably well with the free space model. However, at transmitter-receiver separations exceeding approximately 15m, very pronounced deviations from free space behavior are observed. Signal amplitude and phase are shown to vary widely, exhibiting abrupt changes in slope, often including a large “bump” in phase whose peak occurred typically at a transmitter-receiver separation of approximately 10m-25m. The position of the bump varied from experiment to experiment, but the bump was nevertheless present in every data set.

Indoor measurements were carried out with six different transmitter locations and over 4000 different receiver locations. In each case, significant deviations from free space propagation were observed. These results are partially explained by the parallel plate waveguide model presented in Chapter 7, which suggests that these deviations may be partially caused by the floor and ceiling of the building acting as an unterminated parallel plate waveguide. However it is likely that more than one propagation mode is being observed; without the use of modulation or interferometric

techniques it is not possible to separate the contributions from various sources of in-channel interference, as discussed in Section 3.7.4.

Because of the aforementioned “bump” in phase, distance solutions using phase were not successful at distances exceeding approximately 15m; distance errors were in the range of tens of meters. This simple approach is obviously useless for use in a practical system because of these intolerable errors.

It was then decided to examine the possibility of using the measured amplitude information alone to provide distance estimates. Received signal amplitude in dBm, which are logarithmic units, was plotted against the logarithm of distance. This would yield a straight line fit in log-log space if the amplitude followed a power law, as would be expected from the free space model. This approach proved to be more successful. Applying a least-squares fit to the measurements yielded position estimates with approximately zero mean and standard deviations of approximately 1m to 2m in typical cases. This approach, while still requiring some location specific calibration, is an encouraging result suggesting that further investigation of indoor mid-field propagation is worthwhile. Chapter 8 suggests further work that would be useful to further understand indoor propagation in this frequency regime.

## **6.2 Calibration and performance of experimental apparatus**

The experimental apparatus described in Chapter 4 has been carefully characterized to ensure that the measurements presented in this thesis are accurate. Two types of calibration were performed, which calibrate amplitude and phase independently.

Since each experiment took 4 to 6 hours to complete, measurements were usually taken on different days. It was found that receiver gain exhibited an experiment-to-experiment variation, on the order of several dB from day to day. Amplifier gain variation with thermal changes or small power supply voltage changes are the best explanation for this finding. Since it was desired to provide amplitude results accurate

to better than 1dB, a pre-measurement calibration procedure was created.

### 6.2.1 Amplitude calibration and results

Before each measurement run, an amplitude calibration was performed. This was performed with an HP E4421 signal generator and a calibrated step attenuator, both of which were calibrated in a NIST-traceable facility. The generator was adjusted to 2.00000 MHz, and its output signal was attached via the attenuator to the RF signal input of the receiving downconverter unit. The attenuator was set to 0dB for this calibration. A special MATLAB calibration script was then run, which directed the operator to set the generator to various output levels. The calibration was then performed automatically by MATLAB and the calibration factors were saved for use in subsequent measurements. The amplitude uncertainty of the HP E4421 is specified by HP to  $\pm 0.5\text{dB}$ , while the step attenuator has similar uncertainty. Therefore the calibration apparatus has a worst case uncertainty of 1dB.

Figure 6-1 shows the results of a typical amplitude calibration, performed prior to measurements taken on October 26, 2002. The y-axis for this figure is the signal

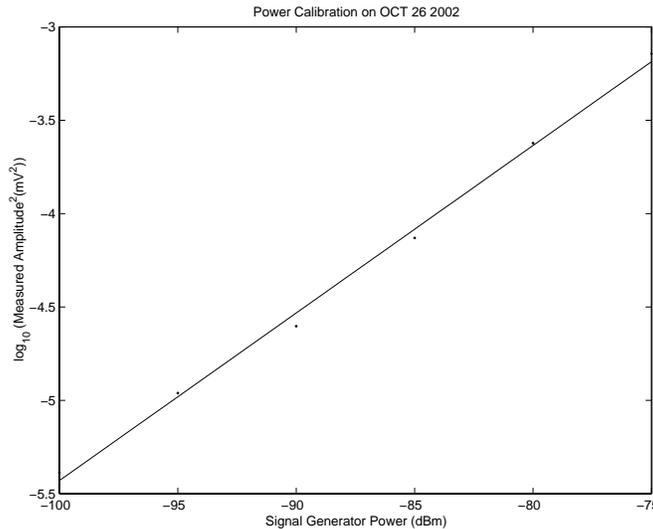


Figure 6-1: Typical amplitude calibration results

processing code's internal measure of amplitude, which is the logarithmic power relationship between the applied voltage at the analog to digital converter, referenced to

50 ohms. These amplitude units (AU) represent ten dB each. This amplitude measurement includes all gains of all components in the RF signal chain except for the receiving antenna, which was separately calibrated. Signal powers referred to in this thesis are with reference to the input power at the antenna port of the receiver chain. From Figure 6-1 it is apparent that the worst case deviation from the least squares fit occurs at an amplitude of -90dBm, which exhibited a deviation of approximately -0.075 amplitude units, corresponding to 0.75dB of variation at that point, since the y axis has units of dB/10.

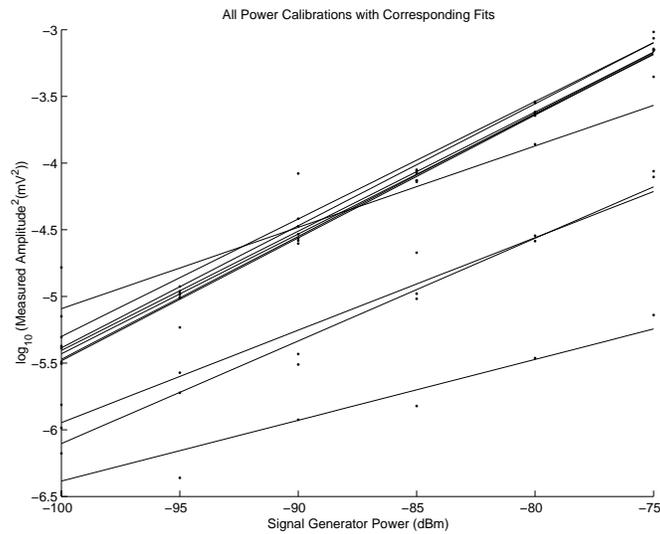


Figure 6-2: Results of ten amplitude calibrations, taken over two months, showing apparatus gain drift

Figure 6-2 shows the results of ten amplitude calibrations taken over the course of two months of experimental work. The gain of the receiver chain exhibited a day to day drift of about one amplitude unit (AU) corresponding to a worst case total change of 10dB over all measurements. The standard deviation of these values, prior to calibration, was 0.39 AU corresponding to a standard deviation of 3.9dB. After calibration the standard deviation of amplitude errors was 0.78dB, well within the 1dB target.

## 6.2.2 Phase measurement characterization

The phase measurement mechanism used in this thesis was fully differential, meaning that both the direct signal and the phase reference signal were simultaneously applied to the receiver. The receiver phase measurement algorithm was therefore inherently self-calibrating, since any oscillator drift in the transmitter, the receiver's local oscillator, or the analog to digital converter's sampling clock was guaranteed to apply equally to both the incoming radio signal as well as the fiber optic cable-supplied reference signal. The phase measurement mechanism was therefore treated as a system to be characterized, but not individually calibrated before each measurement. Figure

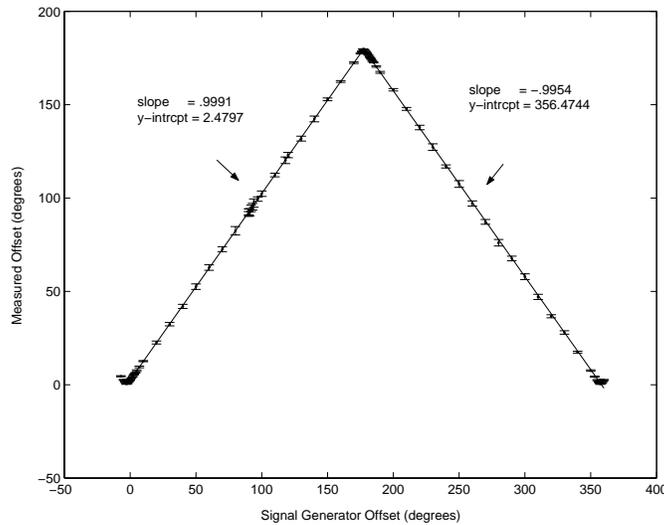


Figure 6-3: Result of phase characterization showing conformance to DDS synthesizer phase

6-3 shows the result of this characterization. To produce this plot, two HP 33250A signal generators with the phase lock option were set to 2.0MHz, and they were connected through equal-length cables to the antenna and reference signal inputs of the measurement apparatus. In order to eliminate the phase offset caused by the cables interconnecting the 10MHz internal reference signal used by the generators' phase lock options, a Tektronix TDS 3052 500MHz oscilloscope set to 10nsec per division was used to simultaneously view the two generators' outputs, and to set the zero phase as accurately as possible. This is a limitation of the way the HP generators are

designed.

In practice there was an uncertainty in the absolute phase within about  $\pm 2$  degrees (0.034 radians) caused by the visual nature of the zero phase set procedure. This accounts for the non-zero y-intercept shown in Figure 6-3. The more important measure of the phase measurement accuracy is the slope of this measurement. From zero to 180 degrees, the slope is 0.9991 while from 180 degrees to 360 degrees the slope is -0.9954. Because the phase measurement mechanism used in this apparatus has an ambiguity at 180 degrees, the transfer function has to be treated in a piecewise fashion. The mean of the absolute value of the slope over the entire cycle was therefore 0.9973, which means that the measurement receiver tracks the signal generator phase to approximately 2.7 parts per thousand, which is 0.972 degrees or 0.0170 radians. This is well within the specified phase tracking accuracy of the two HP 33250A generators, so this accuracy is essentially limited by the calibration of the generators used for calibration.

### **6.2.3 Calibration and characterization summary**

It has been shown in this section that with calibration, amplitude may be measured to an accuracy of better than 1dB with this apparatus, and phase may be measured to better than 1 degree. These results are nearly as good as the HP equipment used to calibrate the apparatus.

As is shown in the following sections, the experimental results presented in this thesis are not limited by equipment calibration accuracy. The behavior of the radio channel itself included spatial variations in amplitude and phase whose effects were far greater than any caused by inaccuracies in the measurement apparatus itself.

## **6.3 Channel noise measurement methodology**

Once accurate calibration of the apparatus was reliably achieved, attention was turned to the question of the noise profile of the 2.0MHz channel. This was necessary to establish the necessary transmitted signal power, and to establish the system signal

to noise ratio (SNR) and the attendant theoretical bounds on estimator performance, as discussed in Chapter 5.

Two different locations were chosen to reflect the extremes in the in-building environment. The first location was deep in the windowless basement of the Wiesner Building, adjacent to E15-042. The second location was in the office part of the Wiesner Building, on the 4th floor (E15-415). The 4th floor location was open with glass walls on all sides, and glass windows open to the outdoors.

It was expected that the two locations would favor different dominant sources of noise. The basement was expected to exhibit primarily man-made noise caused by electrical equipment or machinery. The 4th floor office location was expected to exhibit noise from office equipment (eg spurious radiation from computer equipment) as well as externally generated noise from atmospheric sources or other processes that would be received indoors due to the many windows surrounding the office.

The measurement apparatus used was the standard receiver and antenna hardware combined with amplitude estimation code. This code was run with a 1Hz receiver bandwidth. The transmitter unit was not used in these tests. A wrapper script was written in MATLAB causing the apparatus to sample at intervals of 1 minute. The script was left to run for 24 hours, and the sampling times were labelled in local time. The total sample length was therefore 1440 points.

### **6.3.1 Spatial and temporal analysis of noise data**

Figure 6-4 shows the observed noise power versus time in the basement test location. The noise is relatively constant during the entire 24 hour period, and the average noise power is approximately -80dBm. The ITU372 man-made noise estimate (see the discussion in Section 3.5.2) as well as the thermal noise limit are marked in red on this graph. The gray dots represent individual sub-interval samples, while the bars represent the standard deviation of the noise level. The dark black line is the interval average of the noise level, as a moving average. It is apparent that the noise does not vary substantially during the day, and the noise power at any given time did not exceed about -75dBm. Figure 6-5 shows the noise received in the 4th floor office

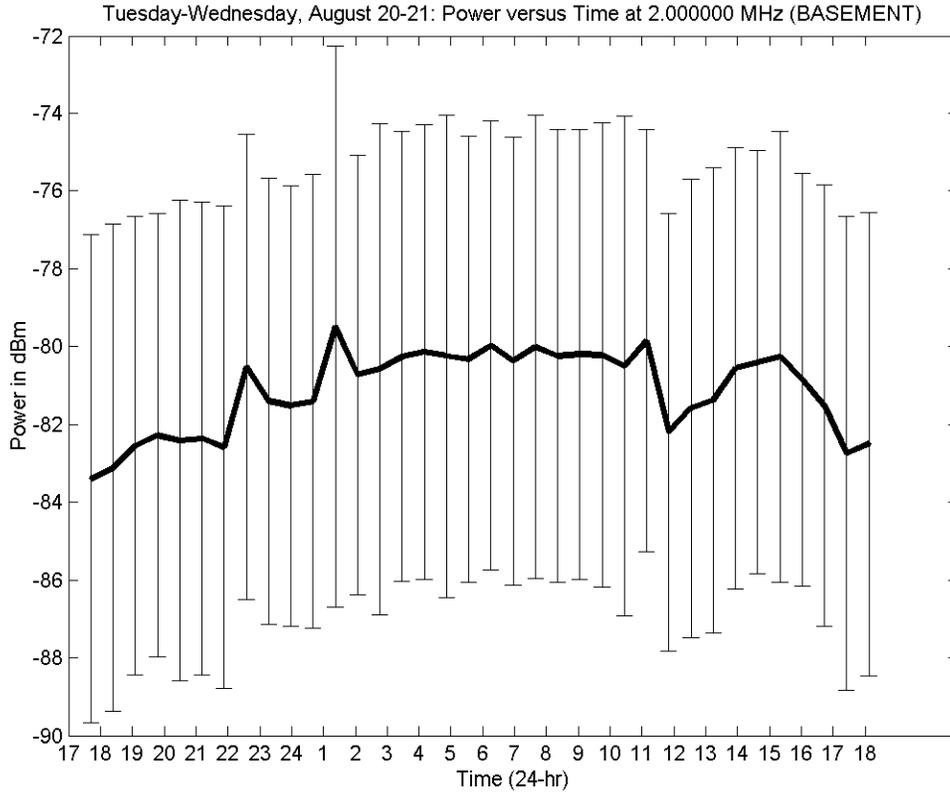


Figure 6-4: Noise power, August 20, 2002, basement

test location. These measurements show an interesting effect; there is a slow rise in the received noise power overnight, from about 10pm local time until approximately 2am, until the peak value was observed. The peak noise power received in the 4th floor test location was about -78dBm. The noise level at 2am is almost 20dB higher than it is during the day. Then, abruptly around 8:30am, the noise level drops to its daytime average of about -95dBm.

The noise power received in the basement was generally larger than that received in the 4th floor, probably due to the large amount of machinery nearby, for example, the elevator machine room and the machine shop adjacent to the measurement location. This noise was relatively constant during the day. Upstairs on the 4th floor, however, a very pronounced day-night variation was observed. It is surmised from these two very different results that an ionospheric process is manifest in locations where outdoor (ambient) signals can be received. As is well known (eg [41], [42], [12]) at night the

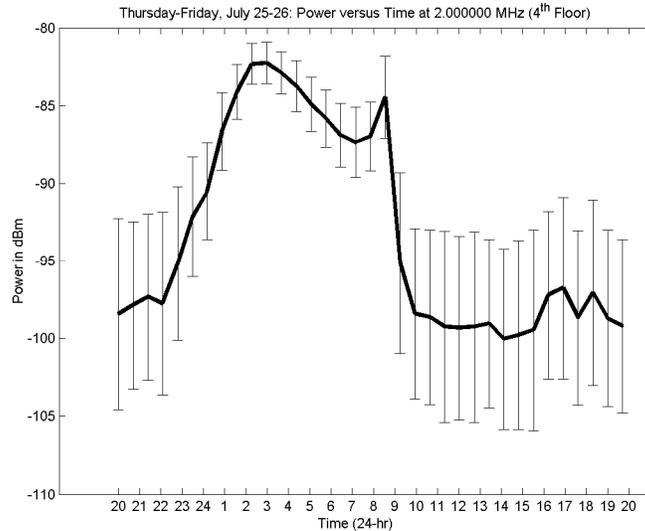


Figure 6-5: Noise power, July 25, 2002, fourth floor

upper, ionized layers of the ionosphere rise dramatically in the absence of incoming solar radiation. This dramatic increase in reflective layer height leads to “ionospheric skip” propagation, where signals from both natural and man-made sources can travel great distances by bouncing between the the surface of the Earth and the conductive layers of the ionosphere. Because the 4th floor test location was largely open to the outdoors, this effect was observable indoors. It is strongly suspected (though not tested in the context of this thesis) that similar results would be observed if the measurement apparatus were placed just outside the building.

These noise measurements were used to determine the transmitted signal power necessary to achieve the target signal to noise ratio (SNR) of at least 15dB (see Section 3.7.2 for a discussion of this issue). Because the worst case measured noise power in any location did not exceed -75dBm, the transmitter was adjusted to yield a signal strength of at least -60dBm in the worst case, which occurred when the receiver was as far as possible from the transmitter, typically at or around 35m apart. Early in the preliminary experiments it was found that a transmitter output power of 15.1W (+41.8dBm) was sufficient to achieve these signal levels in all measurements. Therefore the transmitter output was maintained at that level. Excessive transmitter power would have made it impossible to make measurements close to the transmitter

due to the likelihood of receiver overload problems.

## 6.4 Amplitude and phase measurements

With the preparatory work completed, attention was then focused on the amplitude and phase measurements that are at the heart of this thesis. The experiments were designed by first choosing an experiment site, including transmitter location (which was always fixed in these experiments) and receiver path, which began as close to the transmitter as was possible, and moved in a straight line away from the transmitter, stopping at regular intervals to take amplitude and phase measurements. Given the tendency of the receive preamplifier to go into saturation in very close proximity to the transmitter, which was indicated by an “overload” LED, the receiver’s starting location was between 6m and 10m from the transmitting antenna.

In these experiments, the phase center of the transmitting antenna was assumed to be its physical center. As the transmitting antenna was a small loop, fed from a balanced feedline, it was expected to produce a radially symmetric field. This was confirmed by rotating the transmitter in place around a point on the floor. Since no change was observed at a receiver in a fixed location, the phase center was assumed to lie in the center of the antenna coil. The height of the receiving loop antenna was measured and used to determine the z-axis (in the direction of the Earth’s radius) offset.

### 6.4.1 Transmitter and receiver “truth” data

Indoors, the transmitter location was surveyed with respect to the Wiesner Building blueprints, which were supplied by MIT’s Physical Plant department. These blueprints were found to be inaccurate in places where recent construction had taken place. Therefore positions inside the building were surveyed with respect to the building’s columns, which were taken as the “truth” for these measurements. This was found to be much more self-consistent than using other features found on the blueprints. The surveying equipment used was a steel tape measure calibrated in

inches and feet. These measurements were input as English units, then converted to metric units automatically by a MATLAB script. The accuracy of these surveyed locations is estimated to be approximately 5cm to 10cm in x, y, and z. As with the amplitude and phase calibrations mentioned in Section 6.2, this was not found to be a dominant error source for these measurements. In future work, if more precision were expected from the radio signal based position estimator, a more accurate means of surveying within the building would be necessary to reduce position accuracies below that possible with the building blueprints. Additionally an automated mechanism to survey the receiver location while it is moved around the building would greatly reduce the time needed to manually survey receiver locations.

Receiver location was measured in a 1-D sense, using the tape measure to determine the straight line distance between the physical center of the transmitting antenna. The tape measure was used to mark the measurement intervals, which were typically 1ft or 30.48cm. If high spatial frequency changes in either amplitude or phase were observed, the measurement was done at finer intervals in an attempt to capture the shape of the curves more precisely. Because the presence of the steel tape measure was found to influence the measurements, as described in Section ??, the measured locations were marked in advance of measurement using small pieces of masking tape, and the tape measure was then removed.

These measurements were carried out using seven different transmitter locations, six indoors and one outdoors. The outdoor location was used as a control experiment to confirm the proper operation of the apparatus before beginning indoor measurements.

## 6.5 Outdoor control experiment

The outdoor control experiment was performed in an outdoor courtyard adjacent to the Wiesner Building, between buildings E15 and E23. The courtyard is paved with granite blocks and has a triangular open area of dimensions exceeding 100m by 100m. The experimental setup is shown on a map of this area in Figure 6-6, while

a photograph of this test in progress is shown as Figure 6-7. The transmitter was located in the middle of the open area, and the receiver unit was placed on a cart and moved away from the transmitter in a straight line.

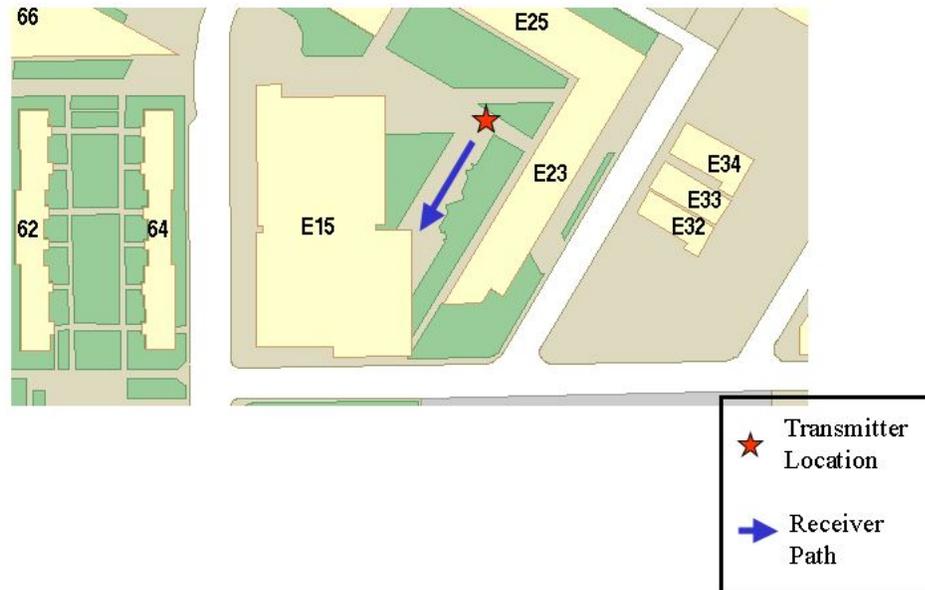


Figure 6-6: Outdoor measurement configuration

### 6.5.1 Raw amplitude and phase data

Figure 6-8 shows the amplitude and phase data collected in this experiment. The measured amplitude and phase data are shown as a composite of ten submeasurements. Each individual measurement is shown with bars marking the standard deviation of the ten submeasurements. Each submeasurement is an individual output of the amplitude or phase estimator described in Chapter 5, made from a 1msec acquisition of the radio frequency signal. The interval between submeasurements was 1 second, resulting in a 10 second total acquisition time. This was done in order to get a sense of the short term temporal variation in the received signal. This information, presented



Figure 6-7: A photograph of the outdoor test configuration

on all subsequent amplitude and phase data plots shown in this thesis, shows that in general, short term averaging (that is, averaging on a time scale over which no major physical changes occur to the indoor environment) will not significantly improve the accuracy of position estimates, because the major source of error is spatially varying bias in the radio propagation itself.

### 6.5.2 Analysis of outdoor data

As shown in Figure 6-8, measured phase outdoors is linear and generally exhibits the expected slope of 2.4 deg/meter, which is found by dividing the 360 degrees in a full cycle by the wavelength of 150m. For transmitter-receiver separations of less than 22m, the measured slope was 2.452 deg/meter. At 22m, however, a small “bump” is observed in the data, which returns to the expected phase at 24m. The cause of the “bump” observed in this data is not known. Potential causes include

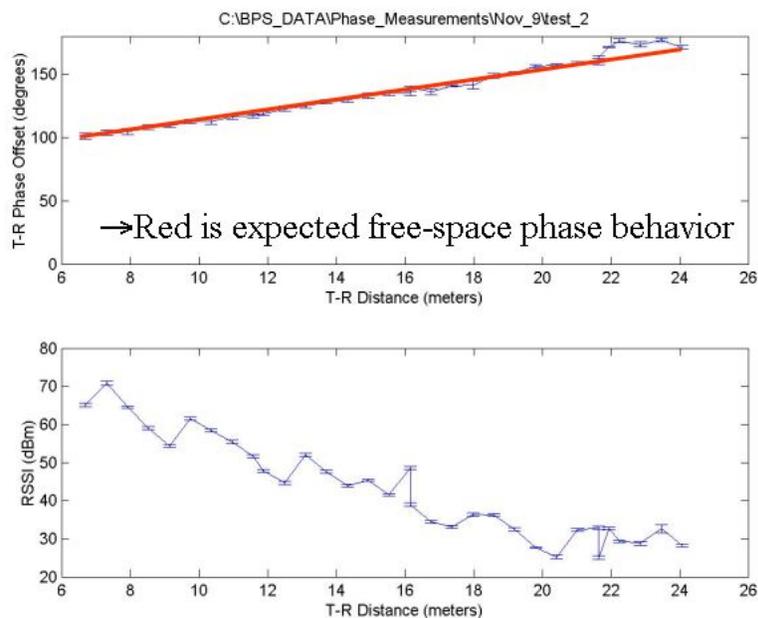
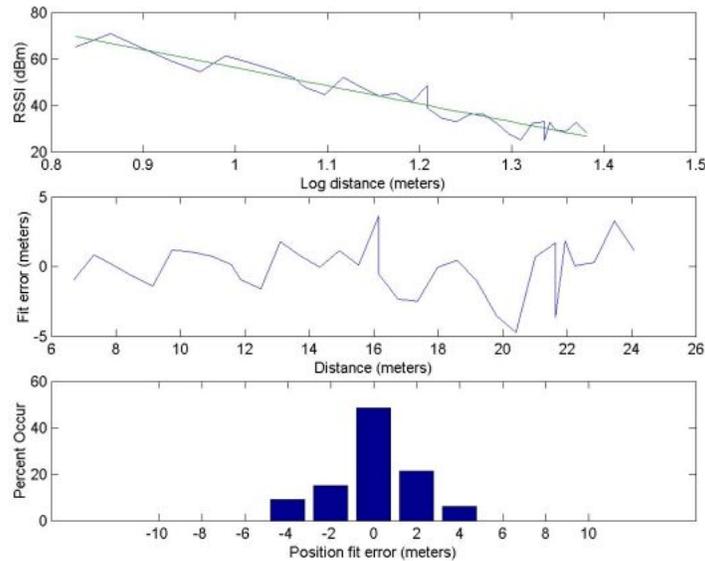


Figure 6-8: Amplitude and phase versus transmitter-receiver distance, outdoors

reflections from the buildings adjacent to the courtyard, temperature related changes in the calibration of the test apparatus, or some type of human error. The latter two potential causes seem less likely to the author at this time, because the temperature of the apparatus would be unlikely to change dramatically over the span of a few minutes' time, and great pains were taken to ensure the accuracy of the measured positions of the transmitter and receiver. To distinguish among these potential causes, one could repeat these measurements with the transmitter-receiver axis in a different orientation with respect to the adjacent buildings, or better, to find a much larger open area. Typical antenna test ranges are specified to be clear of any obstruction for several wavelengths in all directions.

As mentioned earlier, the amplitude was expected to behave according to a power law, so amplitude versus distance would fit a line in log-log space. Figure 6-9 shows this fit. The least-squares best fit line fit the expression  $p \propto d^{-7.6}$ , with  $p$  having

units of power. In terms of either  $E$  or  $H$  fields, this would be  $\{E, H\} \propto d^{-3.8}$ . Theoretically, assuming strictly near-field magnetic field coupling, the expected exponent would be  $d^{-3}$  (see Section ??). Power loss therefore exceeds the theoretical expectation by a significant amount. One potential explanation for this observation is the non-isotropic behavior of the transmitter and receiver antenna radiation patterns, leading to a loss in magnetic field coupling due to field lines not arriving in the preferred direction, perpendicular to the plane of the loop.



→ Distance error mean=0.09m,  $\sigma$ =1.84m

Figure 6-9: Amplitude fit to distance, outdoors

Using the straight line fit to the amplitude data, a simple estimator for transmitter-receiver distance was constructed. The measured amplitude was treated as the input to the fit line, and an estimate of the transmitter-receiver distance was then recovered by treating the distance as the unknown. Figure 6-9 shows the result of this position fit. The mean of the position error was 0.09m, while the standard deviation was 1.84m. This result was generally typical of the indoor amplitude-based position fits

as well.

Another potential explanation would be losses caused by eddy currents induced in the ground beneath the antennas. Because the transmitting and receiving antennae were oriented with their major axis in the direction of the Earth's radius, it would be expected that some of the field lines returning from the antenna would have to meet the boundary conditions imposed by a conductive Earth. If the soil beneath the antennae were not particularly conductive, the skin depth could be appreciable. Skin depth in a conductor is given by the following formula [12]:

$$S = \frac{1}{\sqrt{\pi f \mu \sigma}} \quad (6.1)$$

Here  $f = 2.0\text{MHz}$  and  $\mu = 4\pi \times 10^{-7}\text{H/m}$  because the soil is assumed to have the susceptibility of free space. Typical soil conductivity ranges from 0.001 siemen/meter for dry soil to 0.01 siemen/meter for wet, salty soil [12]. At a frequency of 2.0MHz, this corresponds to a skin depth of 19.9m for dry soil or 3.56m for wet soil. Therefore some of the magnetic field lines will pass through appreciable amounts of soil, with attendant attenuation due to the lossy medium.

## 6.6 Indoor experiments

Next we will turn our attention to the indoor experiments that are the heart of this thesis. The indoor experiments were not intended to be an exhaustive and detailed survey of propagation at every point in the Wiesner Building. Rather, a set of six transmitter locations and accompanying receiver paths were chosen that are likely to be representative of the various different environments that can be found in the building. Figure 6-10 shows the areas of the Wiesner Building that have been covered by experimental measurements. Figure 6-11 is a photograph of the experimental setup. The transmitting antenna is visible at the end of the corridor in the background. The receiving loop antenna was placed on a nonconductive, plastic sawhorse and kept in the middle of the corridor. The receiving equipment is visible

on the plastic cart near the data collection operator, who in this case was the author's undergraduate assistant, Mike Krypel.

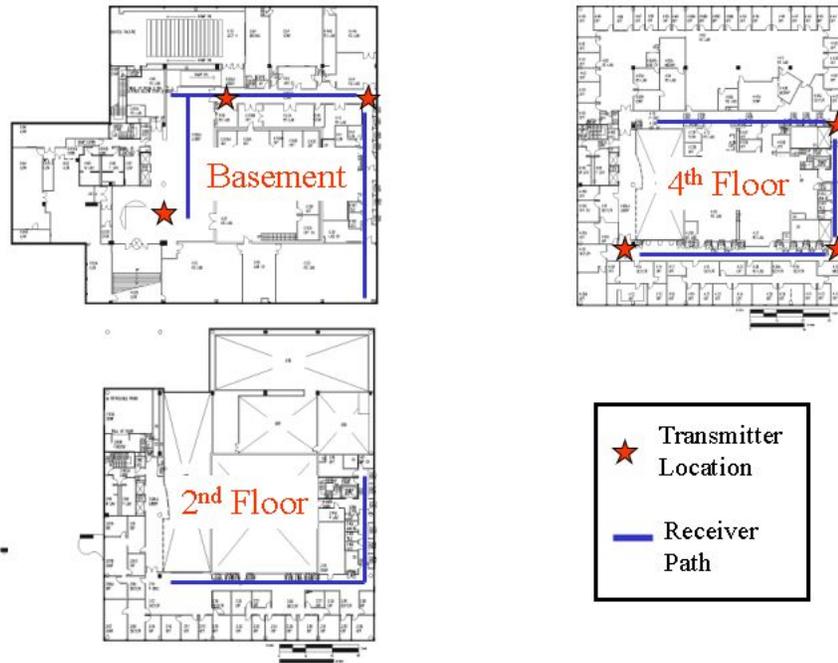


Figure 6-10: Measurement coverage of the Wiesner Building

### 6.6.1 Limitations of these experiments

It is impractical to present all of the individual data sets gathered from the 6 different transmitter locations and approximately 4000 different receiver locations<sup>1</sup>. Two representative data sets are presented and explained in this section. They are typical of the results obtained in the course of this thesis work.

It should also be noted that these data were taken in the corridors of the Wiesner Building after business hours. In general, very few people were present in the building during this time. No special preparation was performed. The corridors and adjacent

---

<sup>1</sup>The data gathered in the course of these observations may be obtained electronically from the author by emailing [matt.reynolds@alum.mit.edu](mailto:matt.reynolds@alum.mit.edu).



Figure 6-11: A photograph of an indoor experiment in progress

rooms on the 4th floor contained an assortment of normal office equipment and furniture. The basement data was taken in laboratory space that is light-industrial in character. Large metal objects, including machine tools, metal stock, metal bookcases, and other laboratory supplies, were present in and adjacent to the corridors in which testing was done.

The author's access to the building was limited to the corridors for the usual security reasons, as well as the prohibitive amount of effort that would have been required to survey non-straight-line paths through the building. Also, the accuracy of the building blueprints decreased markedly where the interior walls are concerned. They are usually made of glass, or gypsum wallboard with metal framing studs, and are frequently moved around to accommodate the occupant's wishes.

It was the author's intent to provide proof-of-concept level data that could be used to plan further experimentation. Limiting the data collection to the corridors



Figure 6-13 shows the measured amplitude and phase as a function of distance. The expected phase slope of  $-2.4$  degrees/meter is plotted in red<sup>2</sup>. The measured phase agrees reasonably well with the expected behavior until a distance of approximately 17m is reached. At that point a prominent bump in the measured phase is observed. There are some smaller subfeatures of this bump, including a local minimum (trough) at 21m that has a total change of over 30 degrees in one meter of increasing distance.

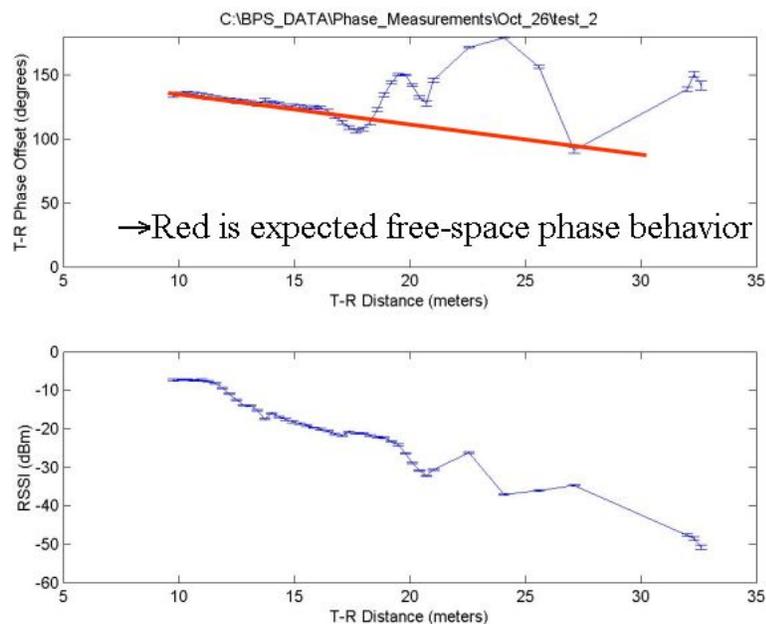


Figure 6-13: Amplitude and phase versus transmitter-receiver distance, basement

This was a very discouraging result when first observed. This behavior was found in every data set, with a prominent bump present between 10m and 20m of transmitter-receiver separation. After exhaustively recharacterizing the experimental apparatus, it was found that this is, in fact, the behavior of the radio channel itself and is not a measurement artifact. Further discussion on this feature is presented in

<sup>2</sup>Note that it is not possible to control the quadrant of the total phase shift through the transmitter due to the 180 degree ambiguity of the receiver's phase estimator. Therefore this experiment has a negative phase slope rather than a positive slope. This is immaterial to the results presented here.

Section 6.7. Because of this enormous deviation from the expected linear change of phase with distance, position solutions based on phase were not attempted.

Features are also apparent in the amplitude data beginning around 17m and continuing through the remainder of the data set at 32m. Looking at the amplitude fit presented in Figure 6-14, a much better result than was expected given the phase data was obtained. In this case, using a linear fit to amplitude, which treated the amplitude as known and the position as unknown, yielded an error mean of 0.10m and a standard deviation of 1.68m. This was a surprisingly good result, given the enormous standard deviation that would have been observed if the phase data were used.

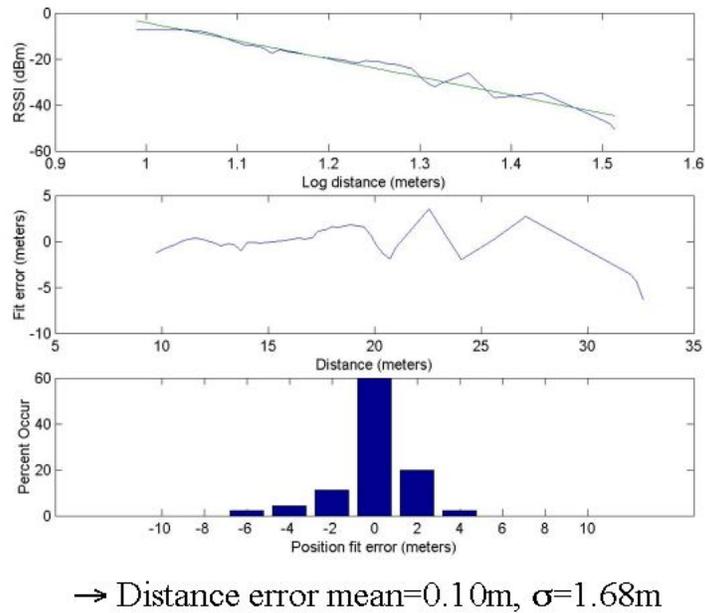


Figure 6-14: Amplitude fit to distance, basement

In a similar fashion to the outdoor results, the exponent of the signal power law is found to be  $p \propto d^{-7.8}$ . This shows slightly more attenuation per unit distance than the outdoor result, which in turn exceeded that expected from the free space

magnetic field result. It seems likely that a similar phenomenon to the conductive ground effect proposed as an explanation to the outdoor result is at work. Both the floor and the ceiling are made of reinforced concrete, which consists of steel reinforcing rods that are wired together and then cast in concrete. The conductivity of concrete is low, approximating the 0.001 siemen/meter of dry soil. The conductivity of the steel reinforcing rods is much higher. It is plausible that eddy currents induced in the steel reinforcing rods could have the effect of increasing the attenuation beyond that observed outdoors because of the close proximity of the metal structural materials to both the transmitting and receiving antennae.

### **6.6.3 Typical experiment II: 4th floor, south hallway**

The Wiesner Building houses many different research activities and therefore is broken in to many different sub-spaces. Therefore there are many different environments inside. The measurement presented in the prior section is representative of the results obtained in the basement of the building, which is light industrial in character. In the basement there are machine shops and laboratories that contain a lot of large metal machinery. Measurements were also taken in the upper floors of the building, where the typical environment is of an office character. The results shown in this section were taken on the 4th floor, with the transmitter placed adjacent to E15-434. The receiver was then moved in a straight line away from the transmitter toward the east side of the building.

As shown in Figure 6-15, in this experiment the transmitter placement was approximately 8m from the west wall of the building, and approximately 2m from the south wall of the building. The measured amplitude and phase as a function of distance are presented in Figure 6-16. The expected phase slope of -2.4 degrees/meter is plotted in red. As described in Section 6.6.2, the measured phase slope agrees well with the expected slope until a distance of approximately 8m is reached. Again, in a manner similar to the previously described experiment, an abrupt change in phase is observed with a nearly vertical slope occurring at approximately 9.5m. Then, at 10m the abrupt change in slope ends and a region of linearity begins, extending to a

distance of 15m. This region has a slope of approximately -12.2 degrees/meter, which is approximately 5 times the expected slope.

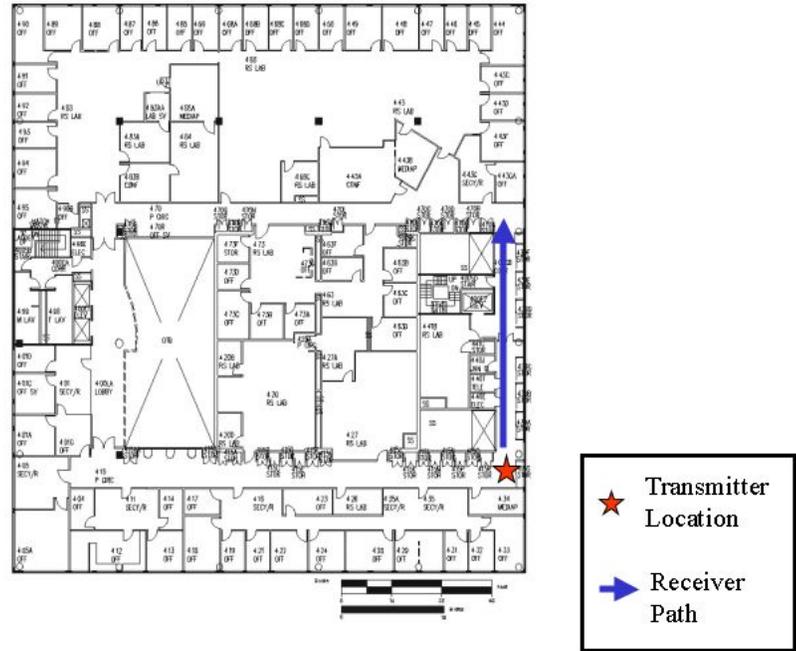


Figure 6-15: Fourth floor measurement configuration

At a distance of approximately 15.5m, time varying behavior is observed, along with an abrupt change in phase slope and a corresponding variation in the behavior of amplitude with respect to increasing distance. The presence of this time varying behavior is indicated by the very large standard deviation among the ten sub-measurements making up one data point. Until this observation, very small standard deviations are found in this dataset. Typical standard deviations for phase values among sub-measurements in the prior 15m of measurement were  $\pm 1$  degree or less. Standard deviations in the measurements made from 15.5m to 21m are as much as twenty degrees or more.

This unexpected result could be explained with three different hypotheses. One hypothesis is that there was an external source of interference, for example a nearby

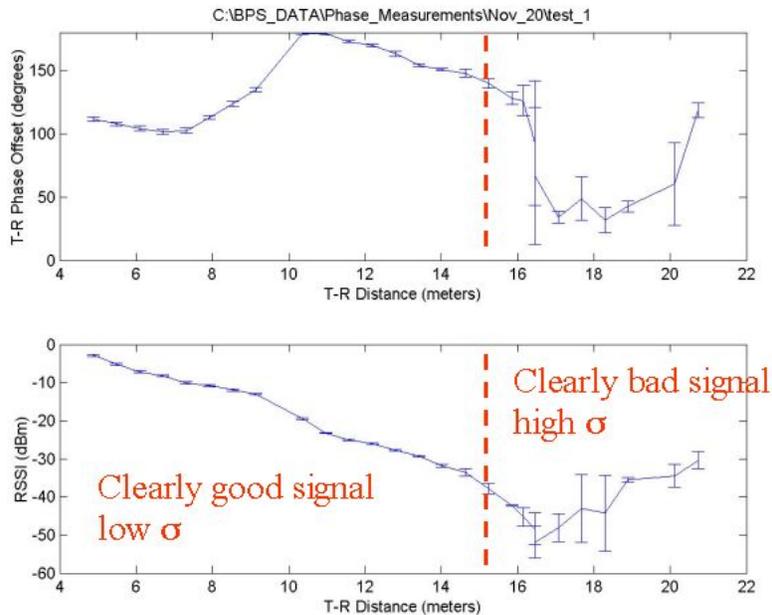
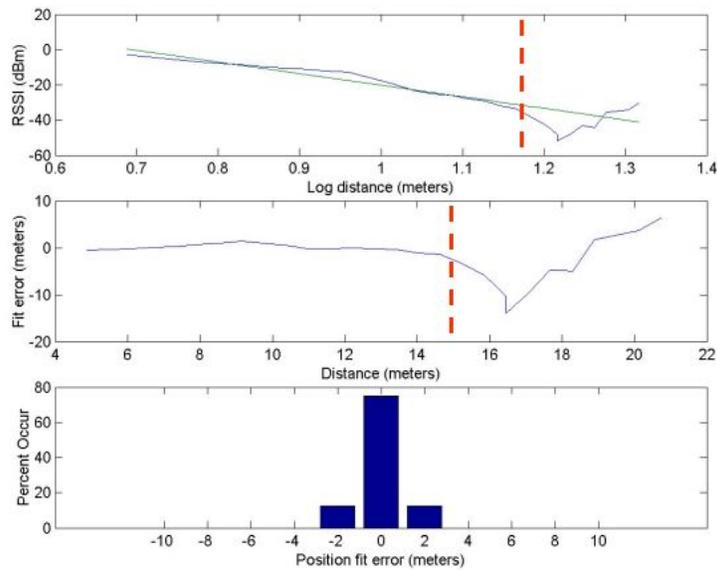


Figure 6-16: Amplitude and phase versus transmitter-receiver distance, fourth floor

radio transmitter with significant energy at 2.0MHz that was normally inactive, but happened to become active at the time the 15.5m measurement was made, and which remained active for the remainder of the experiment. Since the large jump in standard deviation was flagged as an exception by the MATLAB data collection software, the operator was able to repeat several earlier measurements, and obtained results which agreed to within one standard deviation. This coincidentally timed interference hypothesis is therefore unlikely to be true.

The second hypothesis is that some type of propagation anomaly was observed in which the transmitter's signal strength and phase were not well behaved functions of distance, perhaps due to reflections or conduction in some structure. In this case, the received signal strength would exhibit spatially varying behavior, such as that caused by diffraction or other wave-interference effects. Knife edge diffraction at door openings and corridor intersections has been experimentally observed at mi-



→ Good signal: distance error mean=-0.05m,  $\sigma=0.71$ m

Figure 6-17: Amplitude fit to distance, fourth floor

crowave frequencies, but in general diffractive or interference effects are manifest with periodicities on the order of a significant fraction of the signal's wavelength (typically  $0.25\lambda$ ). In this experiment the first observable deviation from the expected results was manifest with an abrupt change in phase slope at only 8m, which is  $0.053\lambda$ . The next change in slope occurred at 9.5m, which is  $0.063\lambda$ , a difference of only  $0.01\lambda$ . This very abrupt change with respect to a wavelength, at transmitter-receiver separations that are far less than  $0.25\lambda$ , suggests that diffractive effects are less likely to be responsible for this behavior. A different reflection-based hypothesis is presented in Chapter 7; this hypothesis suggests that the floor and ceiling of a particular floor act as an unterminated parallel plate waveguide, excited by the transmitter. It is shown from this simple model that spatially varying phase data can be obtained with realistic parameters, even taking in to account the long wavelenth of 150m, but an entirely different experimental campaign would be needed to confirm or refute this

hypothesis.

The third hypothesis is that an interfering signal was present at all times, but its effects were masked by the much stronger signal from the measurement transmitter, when the receiver was close to the measurement transmitter. If so, this masking effect would become less significant at increasing transmitter-receiver distances. Since in this experiment the amplitude was found to fit a power law with an exponent of  $p \propto d^{-7.4}$ , the decrease in transmitted signal strength with distance is substantial. At a certain point the majority of the 2.0MHz signal sampled by the receiving antenna would therefore consist of the interfering signal, and the phase measurement would be incorrect because of the strong interference. If this were the case the measured signal amplitude would be expected either to remain constant with increasing distance, in the unlikely event that the amplitude of the interfering signal were to exactly make up for the path loss in the signal from the measurement transmitter, or more likely, the interfering signal would become dominant and would lead to an increase in the received signal strength. The latter was observed to be the case in these measurements; an increasing trend in signal amplitude was found at approximately 15.5m to 16m. But this does not explain the time varying behavior of the received amplitude and phase, except as a “beat note” between the desired signal and the interfering one if the two signals were close enough in frequency to appear in the receiver bandwidth simultaneously.

The obvious test to separate the second and third hypotheses would be to turn the measurement transmitter off to separate signals resulting from the measurement transmitter from those resulting from an external source. This test was performed at a transmitter-receiver separation of 17m; unfortunately the results were inconclusive. With the transmitter off, the phase measurements are meaningless due to the lack of a reference signal. The amplitude measurements were not time varying and were lower than those received when the transmitter was on, indicating that the presence of the transmitter signal was in some way contributing to the observed results.

It was found that with the transmitter off, the overload indicator on the receiving antenna would illuminate if the receiving antenna were brought within 1m of a certain

electrical outlet located on the wall at the 17m point. This suggests that a source of an interfering signal was present, and that this signal was being carried by the electrical wiring in the building. Due to the very narrow 2.0MHz bandpass filter at the input of the measurement receiver, this interfering signal was rejected, accounting for the decrease in the received signal strength. However the receiving antenna might have been close to its saturation limit and nonlinear effects (eg intermodulation distortion) might have been present.

It was not possible to make useful measurements from 15.5m to 21m, but data gathered from 4m to 15.5m yielded good results using a least squares amplitude fit that excluded the high- $\sigma$  data points. Figure 6-17 shows the result of the least squares amplitude fit, which treated the amplitude as known and the distance as unknown. Including data only from this region, the power law exponent was found to be  $p \propto d^{-7.4}$ , which corresponds to an attenuation between that observed outdoors, which was found to be  $d^{-7.6}$  and in the basement, which was found to be  $d^{-7.8}$ . The error mean was -0.05m while the standard deviation was 0.71m. These are very good results, in fact better than those achieved in the outdoor control case (outdoors, the error mean was 0.09m with a standard deviation of 1.84m). This result could be attributed to a combination of numerical coincidence and selective use of the data set. Selective use of the data set is not unfair in this case, because an actual system built around this measurement methodology would definitely include a signal quality metric similar in principle to the RAIM (“receiver automated integrity monitoring”) mechanism used by modern GPS receivers [1]. Such a system, like the RAIM algorithms, would consider several signal quality metrics (eg time variation of amplitude and phase submeasurements, or local slope compared to expected values) in concert when deciding whether to accept or reject a particular measurement as part of the overall position solution. The use of the Kalman filter to integrate these results into a navigation solution is discussed in Chapter 5.

## 6.7 Conclusions

Many different approaches to indoor navigation have been tried, both by this author and by many others. No single system has yet been demonstrated that meets the simultaneous criteria of high accuracy, fast update rate, inexpensive receivers, and low infrastructure requirements. Given these requirements, which past results have shown are difficult to achieve individually, much less simultaneously, a new approach was conceived involving measuring the amplitude and phase of a relatively low frequency unmodulated signal. Indoor radio propagation in this spectral window has not been previously studied, so this thesis presents the first known investigatory experiments intended as a proof-of-concept and to gather enough data to be able to bound the performance of a system using these principles.

Prior art indoor radiopositioning systems have operated at very short wavelengths, where quasi-optical propagation is observed. At the very short wavelengths, multipath interference, diffractive effects, and other anomalous propagation effects have been observed, making it very difficult to separate the direct path signal from multipath. It was the author's initial hypothesis that, at wavelengths comparable to the dimensions of a building, these anomalous effects would not be seen because the entire system, including the building in which it is installed, would fit well within one wavelength. Thus all measurements would take place within a single cycle of phase. Fine spatial structure, meaning variations in the signal observables on very short distance scales, would not be apparent, and propagation characteristics would mimic those of free space to a great extent.

As shown conclusively in this chapter, this initial hypothesis is not correct. Indoor propagation at a wavelength of 150m exhibits complex behavior that would not be expected given this initial hypothesis. In the two indoor experiments presented in detail in this chapter it has been shown that signal phase exhibits abrupt changes in slope, exceeding 5 times the expected value, with very small changes ( $0.01\lambda$ ) in transmitter-receiver distance. These changes are not trivially explained by any single feature of the measurement environment. A parallel plate waveguide model that

produces features of similar appearance is given in Chapter 7. Given these abrupt changes (almost discontinuities in some datasets) the simple linear phase-distance model could not be applied successfully. These effects are shown to be the result of the indoor propagation environment; the outdoor experiments and careful characterizations that were performed indicate that the measurement apparatus properly measures amplitude and phase, and verifies the free space expectations.

Received signal amplitude is found to be a better predictor of transmitter-receiver distance than phase. A power law relationship was expected and found between received signal amplitude and distance. Typical values for the exponent range from  $d^{-7.4}$  to  $d^{-7.8}$ . Using a least squares fit of the logarithm of distance to the received power in dB, transmitter-receiver distance indoors could be determined to meter-class accuracy ( $\sigma \approx 1m$ ). Although the exponent varies from location to location, it was relatively simple to perform these measurements, and a very simple map of these scalar values could be made. Operation of a system built on this amplitude measurement method can be expected to yield a useful indoor positioning system with 1KHz update rate and meter-class accuracy.

# Chapter 7

## A simplified indoor mid-field propagation model

### 7.1 Introduction

This chapter presents a simplified indoor mid-field propagation model based on the hypothesis that the floor and ceiling of the Wiesner Building, which are made of reinforced concrete, guide waves in a manner similar to a parallel plate waveguide. This model is shown to provide qualitatively suggestive results, accounting for the observed “bump” in phase mentioned in Chapter 6. We start by introducing the basics of electromagnetic wave propagation and then address the parallel plate waveguide model specifically.

In 1864, James Clerk Maxwell gave a series of lectures on the subject of electricity and magnetism before the Royal Society. In these lectures, Maxwell unified everything that was then known about electricity and magnetism into a coherent mathematical framework. In 1873, he published his *Treatise on Electricity and Magnetism*, a work still well worth reading today. Maxwell’s equations are one of the first examples of a modern physical law. These equations correctly described all phenomena of electricity and magnetism known at the time, and contained a hint of a then-unseen phenomenon in the form of a wave solution to those equations. In 1887, Heinrich Hertz proved the physical reality of this wave solution when he used a

spark gap transmitter to create an electromagnetic wave that could be detected some distance away with a primitive spark gap receiver.

The substance of Maxwell's equations have not changed since they were published in 1873 as a group of 20 separate equations. Considerable progress has been made since then in terms of mathematical sophistication and methods of application of Maxwell's equations to practical problems. For example, ten years after Maxwell's Treatise was published, Oliver Heaviside's innovative methods for vector calculus allowed the simplification of Maxwell's equations from the original 20 equations down to the now well-known four equation differential form:

$$\begin{aligned}\nabla \times \vec{E} &= -\mu \frac{\partial \vec{H}}{\partial t} \\ \nabla \times \vec{H} &= \epsilon \frac{\partial \vec{E}}{\partial t} \\ \nabla \cdot \vec{H} &= 0 \\ \nabla \cdot \vec{E} &= 0\end{aligned}$$

In this section we apply Maxwell's equations to the problem of low frequency indoor radio propagation. We show that a parallel-plate waveguide model can be used to approximate the propagation characteristics that were experimentally found in the 2.0MHz propagation measurements conducted for this thesis.

## 7.2 The wave solutions to Maxwell's equations

Maxwell's equations have wave solutions which can be found by first taking the curl of the first equation and substituting the second, applying the vector identity  $\nabla \times (\nabla \times \vec{E}) = \nabla(\nabla \cdot \vec{E}) - \nabla^2 \vec{E}$  :

$$\nabla^2 \vec{E} - \mu\epsilon \frac{\partial^2 \vec{E}}{\partial t^2} = 0 \tag{7.1}$$

This is a second order partial differential equation with plane wave solutions of the form  $e^{i(kx-\omega t)}$ . We define the *wavenumber*  $k$  as:

$$k = \sqrt{\mu\epsilon}\omega \quad (7.2)$$

and further the *phase velocity*  $\nu$ :

$$\nu = \frac{\omega}{k} = \frac{1}{\sqrt{\mu\epsilon}} = \frac{c}{n} \quad (7.3)$$

where  $n$ , the *index of refraction*, relates the permittivity and permeability in a particular medium to that of free space:

$$n = \sqrt{\frac{\mu\epsilon}{\mu_0\epsilon_0}} \quad (7.4)$$

We then define the *wave vector*  $\vec{k}$  which is the direction of propagation for a general electromagnetic wave:

$$\vec{k} = \hat{x}k_x + \hat{y}k_y + \hat{z}k_z \quad (7.5)$$

and rewrite the wave solution in terms of  $\vec{k}$  and a position vector  $\vec{r}$ . Since the dot product  $\vec{k} \cdot \vec{r}$  defines a plane, this is a plane wave :

$$\vec{E}(\vec{r}, t) = \vec{E}_0 \cos(\vec{k} \cdot \vec{r} - \omega t) \quad (7.6)$$

This wave transfers energy in the direction of  $\vec{k}$ ; the magnitude of this energy is given by the real part of the Poynting vector  $\vec{S} = \frac{1}{2}(\vec{E} \times \vec{H})$ .

### 7.2.1 Reflection and refraction of plane waves

Consider a plane wave, which has been traveling in one medium, incident upon a surface of a different medium with surface normal vector  $\vec{x}$ . The original wave is

described as follows:

$$\vec{E} = \vec{E}_0 e^{i\vec{k}\cdot\vec{x} - i\omega t} \quad (7.7)$$

$$\vec{B} = \sqrt{\mu\epsilon} \frac{\vec{k} \times \vec{E}}{k} \quad (7.8)$$

The wave's direction of travel, phase, and amplitude are forced to change in accordance with the properties of the new medium and to satisfy the relevant boundary conditions. In general, a wave incident on a new medium produces two output waves—one *reflected* wave which remains in the old medium, and one *refracted* wave which is transmitted into the new medium.

### Reflected wave

A reflected wave's angle of reflection is equal to the angle of incidence; therefore the reflected wave, referred to by  $\vec{E}''$  and  $\vec{H}''$ , has the following form:

$$\vec{E}'' = \vec{E}_0'' e^{i\vec{k}''\cdot\vec{x} - i\omega t} \quad (7.9)$$

$$\vec{B}'' = \sqrt{\mu\epsilon} \frac{\vec{k}'' \times \vec{E}''}{k} \quad (7.10)$$

where the new wavenumber  $k'' = k = \omega\sqrt{\mu\epsilon}$  because the medium in which the wave is traveling is the same.

### Refracted wave

Any portion of the incident wave that is transmitted into the new medium is refracted according to *Snell's law*:

$$\frac{\sin \theta_i}{\sin \theta_r} = \frac{k'}{k} \quad (7.11)$$

The refracted wave has the form:

$$\vec{E}' = \vec{E}_0' e^{i\vec{k}'\cdot\vec{x} - i\omega t} \quad (7.12)$$

$$\vec{B}' = \sqrt{\mu\epsilon} \frac{\vec{k}' \times \vec{E}'}{k} \quad (7.13)$$

and has a different wavenumber  $k' = \omega\sqrt{\mu'\epsilon'}$  because the refracted wave is traveling in the new medium.

### Magnitude of reflected and refracted waves

The magnitude of the reflected and refracted waves depends on the ratio of the new medium's index of refraction to the prior medium's index of refraction.

$$E_0 + E_0'' - E_0' = 0 \quad (7.14)$$

$$\sqrt{\frac{\epsilon}{\mu}} (E_0 - E_0'') \cos \theta_i - \sqrt{\frac{\epsilon'}{\mu'}} E_0' \cos \theta_r = 0 \quad (7.15)$$

Energy is conserved in that the total power incident on the surface is either reflected or transmitted. The amount of power transmitted from the wave in the first medium in to the second medium depends on the ratio of permittivities and permeabilities, as well as the angle of incidence of the incoming plane wave.

## 7.3 Waves in structures

In addition to propagating in free space, electromagnetic waves can also propagate in a wide variety of structures. Waves in structures must obey the boundary conditions enforced by the electrically conductive or magnetically permeable materials composing those structures. These boundary conditions give rise to *modes* which describe the spatial distribution of  $E$  and  $B$  fields.

### 7.3.1 TE and TM modes

Transverse electric (TE) and transverse magnetic (TM) modes arise in hollow conductive pipes, such as a rectangular waveguide. They are differentiated by the boundary conditions that they meet. TM modes have  $B_z = 0$ , and  $E_z|_S = 0$ , while TE modes

have  $E_z = 0$ , and  $\frac{\partial B_z}{\partial n} |_S = 0$ . The electric and magnetic fields for TE and TM modes are given by:

$$\vec{E} = \frac{i}{\mu\epsilon\omega^2 - k^2} [k\nabla E_z - \omega\hat{z} \times \nabla B_z] \quad (7.16)$$

for TE modes, and

$$\vec{B} = \frac{i}{\mu\epsilon\omega^2 - k^2} [k\nabla B_z + \mu\epsilon\omega\hat{z} \times \nabla E_z] \quad (7.17)$$

for TM modes. Both TE and TM modes require closed conductive surfaces on all sides of the propagation region; in the context of a hallway, for example, the floor, ceiling, and corridor walls must all be conductors. TE and TM modes in structures exhibit a *cutoff* behavior; there is some minimum frequency below which the modes are degenerate and propagation does not occur. Instead an *evanescent* wave is present, which rapidly decays along the waveguide. The cutoff frequency for the lowest order TE<sub>01</sub> mode in such a waveguide is:

$$f_c = \frac{c}{2a} \quad (7.18)$$

where  $c$  is the speed of light in a vacuum, and  $a$  is the longest cross-sectional dimension of the waveguide. Since in a given corridor the width and height are approximately 4m,  $a$  is estimated at that value. The cutoff frequency is therefore 37.5MHz for TE<sub>01</sub>, and any other TE or TM mode would have a higher cutoff frequency. Since this is an order of magnitude higher than the 2.0MHz measurement frequency, any TE or TM wave propagating in a corridor sized waveguide would be evanescent.

### 7.3.2 TEM modes and the parallel plate model

The possibility of a TEM mode being responsible for the observed behavior is interesting, because TEM modes have no low frequency cutoff, which is suggestive of their applicability to this problem. TEM modes can only exist in a parallel plate waveguide

without continuous conducting walls because of their boundary conditions:

$$\nabla \times \vec{E}_{\text{TEM}} = 0 \quad (7.19)$$

$$\nabla \cdot \vec{E}_{\text{TEM}} = 0 \quad (7.20)$$

Their wavenumber  $k$  is given by

$$k = k_0 = \omega \sqrt{\mu\epsilon} \quad (7.21)$$

where  $\mu$  and  $\epsilon$  are the effective permeability and permittivity, which are typically geometry dependent in these problems.

## 7.4 The parallel plate model

The geometry of such a waveguide is shown in Figure 7-1, along with the real-world dimensions of an entire floor of a building. In this simplified model, the internal walls, which are often made of glass or wallboard, are neglected, and the reinforced concrete floor and ceiling of an entire building floor are treated as the conductive surfaces for the waveguide.

As shown in Figure 7-1, the impedance of the waveguide is approximately  $28.7\Omega$ , far less than the  $377\Omega$  of free space. This suggests that we consider a floor of the building as an unterminated waveguide, where the sides of the building (which are typically glass walled) are considered to be an open circuit for the purposes of a transmission line analysis.

Figure 7-2 shows a schematic representation of this transmission line model. The transmitter acts as a source driving one end of the transmission line. The other end of the transmission line is terminated only in the impedance of free space; since it is an order of magnitude higher in impedance than the transmission line's characteristic impedance, it acts as an open circuit. The receiver, moving from the transmitter end of the transmission line toward the open circuit end, is in effect measuring the standing wave arising in this configuration.

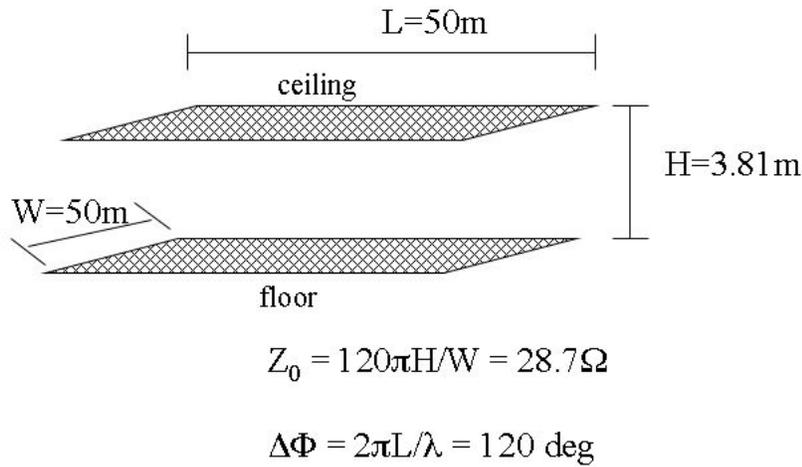


Figure 7-1: Parallel plate waveguide properties

#### 7.4.1 Qualitative fit to data

A simple MATLAB analysis using a sum of two phase delayed complex exponentials, representing the direct path and the standing wave, was prepared for this model. The resultant phase versus distance plot is shown in Figure 7-3 below the actual measured data from one of the experiments in Chapter 6.

As is shown in the figure, the standing wave in the transmission line caused by the reflection from the unterminated end produces a phase “bump” that is somewhat similar to that observed in the accompanying experimental data. The “bump” occurs at a larger transmitter-receiver separation, and it does not have the fine structure of the experimental data, but nevertheless it reproduces the key behavior seen in the experiment.

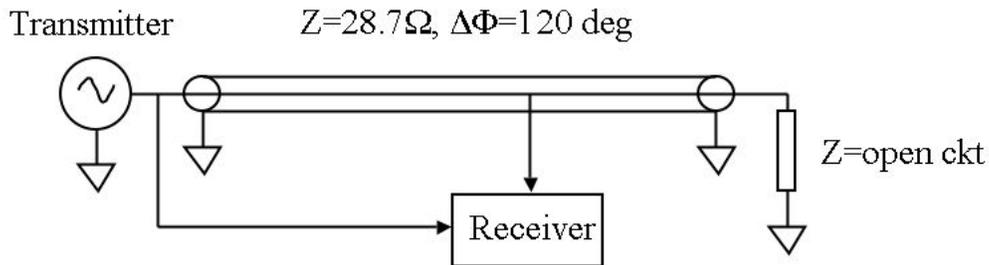


Figure 7-2: Transmission line model

## 7.5 Conclusions, and questions for the future

It is not claimed that the transmission line model presented in this chapter is descriptive in a strict sense; it does not numerically agree with the experimental data, and it does not reproduce the fine structure seen therein. However, it is suggestive of the behavior that can arise from a single reflection in the transmitter-receiver environment, and leads to several interesting questions about the nature of the indoor environment:

- Presumably the indoor radio channel is strictly linear. Therefore any behavior that is observed can be explained in terms of a linear superposition of sources. What is the minimum number of reflections necessary to explain a given observed data set?
- Given this minimum number of reflections, how are they placed in a 3D config-

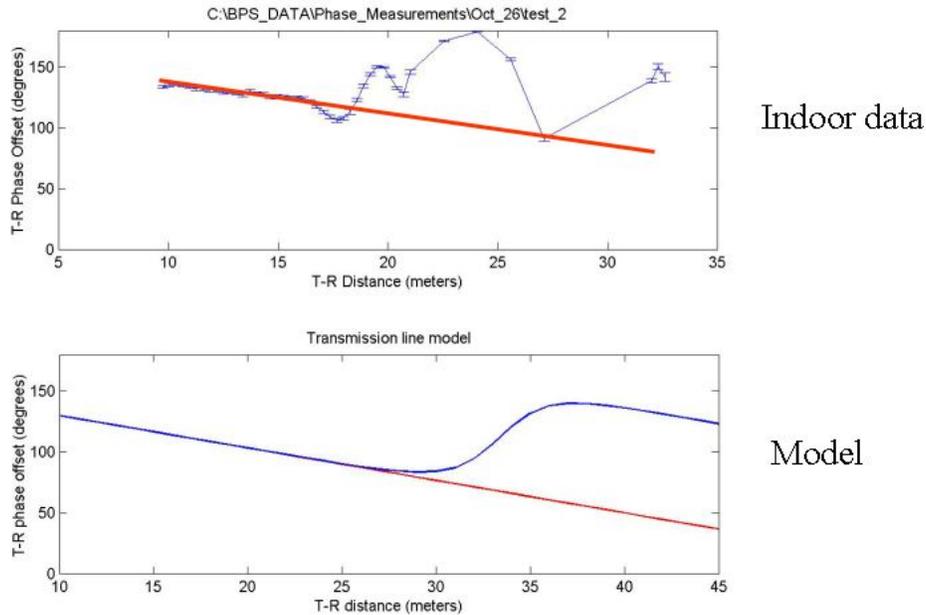


Figure 7-3: Transmission line model results vs experiment

uration to yield the observed results? Does this configuration bear any resemblance to the actual construction of the building under test?

- From a set of building blueprints, is it possible to extract the features of a certain building that are most likely to cause reflections and apply these to a useful predictive model?
- If such a model could be made, what is the minimum amount of data required to encode it? What are its computational requirements?
- If a new building were being built, is there a way to slightly change its design to yield smooth indoor propagation, without compromising its structure or form?

These questions, and many others, have been raised in the author's mind by the results of the experiments presented in this thesis. It is clear that a great deal of additional experimentation will be required to uncover the answers.

# Chapter 8

## Conclusions

This thesis marks the end of a long period of speculation about the behavior of “mid-field” wavelengths indoors. It was the author’s initial hypothesis, based on a suggestion from Neil Gershenfeld, that at wavelengths comparable to the dimensions of a building, the internal structure of that building would not affect signal propagation, and that amplitude and phase would be well-behaved functions of transmitter-receiver distance.

After an exhaustive literature search, a great deal of prior art was found on the propagation of UHF and microwave signals in buildings, primarily related to cellular telephony and GPS applications, but there was no previous study of indoor propagation at longer wavelengths. Therefore the author undertook to design a set of experiments that would represent a “proof of concept” for an indoor low frequency radiolocation system. It was decided to perform the measurements described in this thesis at a frequency of 2.0MHz, corresponding to a wavelength of 150m, approximately three times the length of the nearly-cubic Wiesner Building in which the measurements were to be performed.

The desired properties for such a system were determined after a detailed survey of position tracking systems for human-computer interfaces was performed. The results of this survey are presented in Chapter 2. Position aware ubiquitous computing was the primary application envisioned for this system, but many others are possible, including security or logistics management, or optimization of services such as health

care delivery. It was determined by studying prior art that most existing systems require too much infrastructure, exhibit short useful range indoors, require extensive training or map-building, and are neither fast enough nor accurate enough. These guiding principles were integral to the author's experimental plan and design process.

As described in Chapter 3, a great deal of system engineering work was performed to determine the required performance of the measurement equipment, and to synthesize what could be learned from the available radiolocation literature into a coherent engineering plan. This plan accounted for the best available estimates of channel noise, required signal to noise ratio, and transmitter, receiver, and antenna choice.

Once this plan was determined, the detailed design of the experimental apparatus was begun. Designing and building the measurement apparatus turned out to be a long and challenging process because of the need to custom build almost all parts of the measurement equipment. Because of the unavailability of commercially made low frequency radio equipment suitable for indoor use, with the requisite accuracy and stability needed to make useful amplitude and phase measurements, a variety of different designs were prototyped. After careful characterization and a great deal of omphaloskepsis, most designs were ultimately rejected due to a variety of technical issues. In the end, the design documented in Chapter 4 was arrived at after a lot of research and experimentation. This measurement apparatus has been shown to provide amplitude and phase measurements accurate to within 1 dB and 1 degree, respectively; this performance is as good as the NIST-traceable HP test equipment used to characterize it.

After the equipment was designed, built, and characterized, a series of outdoor measurements were made to ensure that the apparatus was functioning properly and to confirm the understanding of idealized free space behavior developed in Chapter 3. These measurements are documented in Chapter 6.

It was found that outdoors, phase is indeed a well behaved function of distance. The expected rate of change of phase with distance is 2.4 degrees/meter at a wavelength of 150m; outdoor measurements yielded a value of 2.452 degrees/meter, indicating agreement with theory to within 2.2%. Amplitude change with distance was

found, as expected, to conform to a power law. In contrast to the free space dipole-dipole model presented in Chapter 3, which suggests that  $p \propto d^{-6.0}$ , a measured power law of  $p \propto d^{-7.6}$  was observed. This increased attenuation seems most likely to be due to the presence of a lossy ground medium (soil) directly beneath the transmitting and receiving antennae, leading to skin effect loss and eddy current damping.

Indoor measurements were then performed in the Wiesner Building. Six different transmitter locations were selected, corresponding to a variety of indoor environments, including light industrial laboratory spaces, relatively open spaces, and a typical office environment. These environments were expected to be typical of the application environments in which one might want to use an indoor radiolocation system. From the six different transmitter locations, over 4000 different receiver locations were surveyed, each on a straight line path to a particular transmitter location. The transmitter was placed at its surveyed location, and the receiver was moved along a straight line path away from the transmitter, stopping at regular intervals to sample the amplitude and phase of the received signal.

These experiments show that the “good behavior hypothesis” for low frequency indoor propagation is definitely untrue, at least at the 2.0MHz frequency used for these measurements. In every data set, anomalous behavior of phase versus distance is observed. A large “bump” in phase was observed in every dataset, occurring at transmitter-receiver separations of between 10m and 25m. This behavior is location-specific and is related to the interaction of the transmitted signal with the internal structure of the building. Given the existence of this phenomenon, position solutions were not performed using a linear phase-distance model due to the extraordinary inaccuracy that this solution method would have produced.

It is the surprising conclusion of this thesis that amplitude change with distance was much more useful for determining user position indoors. A power law model was fit to the amplitude measurements gathered in each experiment. Typical results for this position estimation method are mean position errors of 0.1m with a standard deviation of between 1m and 2m. The update rate for these measurements is 1KHz, far exceeding the requirements for tracking human motion. These are good results

for wide coverage area indoor radiolocation systems and compare favorably with the prior art systems described in Chapter 2.

A model was sought that would explain the anomalous behavior found in the measurements described in Chapter 6. This work is summarized in Chapter 7. A candidate explanation for the phase behavior was found by noting that the floor and ceiling of each floor of the Wiesner Building are made of reinforced concrete, which is made by casting a grid of steel reinforcing rods in concrete. For the purpose of the model, the floor and ceiling were treated as perfect conductors. It is then shown in Chapter 7 that this parallel plate model supports TE, TM, and TEM modes, the latter with no low frequency cutoff. When unterminated, as would be the case if the building were acting as such a waveguide, reflections from the unterminated ends produce amplitude and phase behavior qualitatively similar to that observed in the experiments shown in Chapter 6. Future work is suggested in the form of experiments that could be carried out to support or undermine this model.

## 8.1 Contributions of this thesis

This thesis contributes to the field of indoor navigation and indoor electromagnetic propagation in three main areas.

1. A portable, precise amplitude and phase measurement apparatus has been designed, constructed, and characterized. The transmitter is based on a novel high efficiency, phase stable class-E amplifier with integrated antenna. The receiver unit is based on a digitized-IF software radio design. The apparatus has been shown to be accurate to within 1 dB in amplitude measurement, and 1 degree in phase.
2. Empirical propagation maps have been produced from six different transmitter locations and over 4000 different receiver locations in the Wiesner Building. These maps, the first known survey of indoor radio propagation at “mid-field” wavelengths comparable to the size of the building, show surprisingly interesting

amplitude and phase behavior, including abrupt changes in phase given receiver movement of only  $0.01\lambda$  or less.

3. A candidate for a theoretical model has been produced which partially explains the measured amplitude and phase data with respect to certain features of the Wiesner Building. This parallel-plate waveguide model is neither a first principles finite element model nor a data driven empirical model, but is instead intended to be suggestive of future experiments in  $E$  and  $H$  field distribution in the building's structure.

## 8.2 Further work

In many ways, this thesis represents more of a beginning than an end. As it turns out, indoor radio propagation exhibits much more complex behavior than was initially expected. Judging by the difficulty and time consuming nature of the measurements performed for this thesis, further work in the field of long wavelength, mid-field indoor radio positioning will require a lot more theoretical modeling and time consuming data gathering than was initially expected.

### 8.2.1 Experiment design and operating frequency

For a variety of reasons, including time, logistics, and available engineering labor, the measurements made in this thesis were made at a single frequency of 2.0MHz. This frequency was chosen somewhat arbitrarily because it is conveniently located in an ITU sanctioned radionavigation band, and because its 150m wavelength fit the initial definition of a mid-field system. The results achieved at this frequency are surprising enough that the evaluation of other frequencies in different orders of magnitude of wavelength, for example 1km and 10m wavelengths, would probably be illuminating as far as determining which possible explanatory models fit a particular building configurations. Perhaps a much longer wavelength of 1km will not exhibit the anomalous propagation effects discovered in the course of this thesis. Alternatively,

the use of a smaller wavelength, such as 10m, might allow one to build a scale model of a building and therefore gain insight about transmitter and receiver placement, signal processing and modulation strategies, and other related issues.

Additionally it makes sense to consider locating the transmitter outside of the building being examined, to see what the outdoor to indoor coupling parameters are, and how signals either leak out of the building, or are confined within the building. This experiment has not been carried out in the course of this thesis, and its results will definitely bear on the feasibility of a practical system because of the possibility of building-to-building interference, and possible oddities of propagation at the building to outdoors boundary.

### **8.2.2 Testing in multiple buildings**

Different buildings have different dimensions and construction techniques. Variation in results given these different environments was not tested in this thesis. It is the author's hypothesis that a wood frame building would have yielded very different results than the reinforced concrete Wiesner Building produced. For a practical system to be built, it would have to be shown that the results are largely building-invariant, or that a simple calibration procedure could be used to guarantee good results.

### **8.2.3 Propagation modeling**

Further work is definitely needed in the area of propagation modeling at long wavelengths, where the transmitter and receiver are very small compared to the operating wavelength. This is the exact opposite of the typical microwave modeling problem. The existing electromagnetic tools do not handle this situation very well, because any "probe" that one might use to determine field coupling or geometry will undoubtedly affect the measurement being gathered.

Some way to digitize the geometry of the building in which testing is being carried out would also be helpful to generate more accurate results that are more representative of observed behavior. However it may prove impossible to include sufficient

detail about the internal metalwork of the building and the connectivity or lack thereof between that metalwork. These construction methods do not seem to be perfectly consistent from building to building, and blueprints are often wrong.

#### **8.2.4 System design and development**

The measurement apparatus designed in the course of this thesis has evolved considerably since it was initially designed. Many issues, such as making every phase measurement differential, and simplifying signal processing software to enable easy testability, seem obvious in retrospect but were not obvious at the beginning of the project.

It is very hard to make careful measurements in a building that is in active use. The constant flow of people in the Wiesner Building, even late at night, is very frustrating as people tend to move apparatus. Also human beings do have complex dielectric properties that affect the measurements being made. One way to mitigate these effects would be to design an automated measurement apparatus to replace the cart-and-assistant method used in this thesis. Perhaps a robotic cart that could be trained to follow a tape line on the floor, or a semi-robotic cart equipped with dead reckoning sensors, would improve the speed of the measurement and would allow much more data to be gathered in a shorter amount of time. This would allow a more complete survey of a given building and would make the process much less frustrating for the experimenter.

### **8.3 Caution and Benediction**

The author humbly offers the following:

Caution: Do not underestimate the difficulty of obtaining good data, nor overestimate your ability to understand the data you have gathered.

Benediction: Good fortune to any who start off in my footprints. May you rapidly eclipse my progress!

# Bibliography

- [1] Global Positioning System Standard Positioning Service Signal Specification, 2nd Ed. June 1995, <http://www.navcen.uscg.gov/pubs/gps/sigspec/gpssps1.pdf>.
- [2] *Fcc rules, part 15.223*, Tech. report, 1999, <http://www.fcc.gov/oet/info/rules/>.
- [3] Jonathan S. Abel and Julius O. Smith, *Source range and depth estimation from multipath range difference measurements*, IEEE Transactions on Acoustics, Speech, and Signal Processing **37** (1989), no. 8, 1157–1165.
- [4] J.B. Andersen, T.S. Rappaport, and S. Yoshida, *"propagation measurements and models for wireless communication channels"*, IEEE Communications **33** (1995), no. 1, 42–49.
- [5] Ascension Corporation, *Technical description of dc magnetic trackers*, 2001.
- [6] R. Azuma, *A survey of augmented reality*, Presence **6** (1997), no. 4, 355–385.
- [7] C. Balanis, *Antenna theory: Analysis and design*, 2 ed., John Wiley and Sons, 1996.
- [8] S. Bryson, *Measurement and calibration of static distortion of position data from 3d trackers*, SPIE Conf. Stereoscopic Displays and Applications III, February 1992, pp. 244–255.
- [9] James Chaffee and Jonathan Abel, *Gdop and the cramer-rao bound*, Proceedings IEEE Position Location and Navigation Symposium '94 (1994), 663–668.

- [10] ———, *On the exact solutions of pseudorange equations*, IEEE Transactions on Aerospace and Electronic Systems **30** (1994), no. 4, 1021–1030.
- [11] Y.T. Chan and K.C. Ho, *A simple and efficient estimator for hyperbolic location*, IEEE Transactions on Signal Processing **42** (1994), no. 8, 1905–1915.
- [12] ITT Corporation, *Reference data for radio engineers*, 6 ed., ITT Corporation, 1975.
- [13] Creamer, *Omega navigation system position fix accuracy*, NAVIGATION **31** (1984-85), no. 4, 251–274.
- [14] ETS-Lindgren, *Model 6502 technical manual*, <http://www.ets-lindgren.com/pdf/Ant6.pdf>.
- [15] Bertrand T. Fang, *Simple solutions for hyperbolic and related position fixes*, IEEE Transactions on Aerospace and Electronic Systems **26** (1990), no. 5, 748–753.
- [16] Borje Forssell, *Radionavigation systems*, Prentice Hall International, 1991.
- [17] Benjamin Friedlander, *Accuracy of source localization using multipath delays*, IEEE Transactions on Aerospace and Electronic Systems **24** (1988), no. 4, 346–359.
- [18] A. Gelb, *Applied optimal estimation*, MIT Press, 1974.
- [19] Ralf Gelhaar, *Sound=space: an interactive musical environment*, Contemporary Music Review **6** (1991), no. 1, 59–72.
- [20] Michael Geyer and Anastasios Daskalakis, *Solving passive multilateration equations using bancroft's algorithm*, Proceedings IEEE Digital Avionics Systems Conference '98, 1998, pp. f41.1–f41.8.
- [21] M. Ghazisaedy, D. Adamczyk, D. Sandin, R. Kenyon, and T. DeFanti, *Ultrasonic calibration of a magnetic tracker in a virtual reality space*, Proceedings of Virtual Reality Annual International Symposium, 1995, pp. 179–188.

- [22] Andy Harter and Andy Hopper, *A distributed location system for the active office*, IEEE Network **8** (1994), no. 1.
- [23] H. Hashemi, *The indoor radio propagation channel*, Proceedings of the IEEE **81** (1993), no. 7, 943–968.
- [24] Jeffrey Hightower and Gaetano Borriello, *Location systems for ubiquitous computing*, IEEE Computer **34** (2001), no. 8, 57–66.
- [25] Hiroshi Ishii and Brygg Ullmer, *Tangible bits: Towards seamless interfaces between people, bits and atoms*, CHI, 1997, pp. 234–241.
- [26] R.E. Kalman, *A new approach to linear filtering and prediction problems*, Transactions of the ASME- Journal of Basic Engineering **82** (1960), 35–45, Series D.
- [27] D. Mackenzie, *Inventing accuracy: A historical sociology of nuclear missile guidance*, MIT Press, 1990.
- [28] F.G. Major, *The quantum beat: The physical principles of atomic clocks*, Springer-Verlag, 1998.
- [29] Mark Moeglein and Norman Krasner, *An introduction to snaptrack server-aided gps technology*, <http://www.snaptrack.com/AtWork/ion.pdf>, 2001.
- [30] H. Nikookar and H. Hashemi, *Phase modeling of indoor radio propagation channels*, Proc IEEE Vehicular Technology **3** (1994), 1759–1763.
- [31] A.V. Oppenheim, R.W. Schaffer, and J.R. Buck, *Discrete time signal processing*, 2 ed., Prentice Hall, 1999.
- [32] J. Paradiso, K. Hsiao, and A. Benbassat, *Interfacing the foot: Apparatus and applications*, Proceedings of the ACM CHI Conference 2000: Extended Abstracts, 2000, pp. 175–176.
- [33] Parkinson, *A history of satellite navigation*, NAVIGATION **42** (1995), no. 1, 109–164.

- [34] J.A. Pierce, *Loran: Long range navigation*, MIT Radiation Laboratory Series, McGraw-Hill Book Co., New York, 1948.
- [35] Matthew Reynolds, Bernd Schoner, Joey Richards, Kelly Dobson, and Neil Gershenfeld, *An immersive, multi-user musical stage environment*, Proceedings of ACM Siggraph (2001).
- [36] R. Schmidt, *Least squares range difference location*, IEEE Transactions on Aerospace and Electronic Systems **32** (1996), no. 1, 234–242.
- [37] N.O. Sokal and A.D. Sokal, "class e - a new class of high-efficiency tuned single-ended switching power amplifiers", "IEEE Journal of Solid-State Circuits" "SC-10" (1975), 167–176.
- [38] T. Starner, D. Kirsch, and S. Assefa, *The locust swarm: an environmentally-powered, networkless location and messaging system*, IEEE International Symposium on Wearable Computing, October 1997.
- [39] W. Stutzman and G. Thiele, *Antenna theory and design*, 2 ed., John Wiley and Sons, 1998.
- [40] John Underkoffler, Brygg Ullmer, and Hiroshi Ishii, *Emancipated pixels: Real-world graphics in the luminous room*, Siggraph 1999, Computer Graphics Proceedings (Los Angeles) (Alyn Rockwood, ed.), Addison Wesley Longman, 1999, pp. 385–392.
- [41] International Telecommunications Union, Tech. report, Report RPI.372-6.
- [42] ———, *World distribution and characteristics of atmospheric radio noise*, Tech. report, Comite Consultatif International Des Radiocommunications, 1963, Documents of Xth Plenary Assembly, Geneva, Switzerland.
- [43] Roy Want, Andy Hopper, Veronica Falcao, and Jon Gibbons, *The active badge location system*, Tech. Report 92.1, ORL, 24a Trumpington Street, Cambridge CB2 1QA, 1992.

- [44] Roy Want, Bill N. Schilit, Norman I. Adams, Rich Gold, Karin Petersen, David Goldberg, John R. Ellis, and Mark Weiser, *The ParcTab ubiquitous computing experiment*, Tech. Report CSL-95-1, Xerox PARC, 1995.
- [45] A. Ward, A. Jones, and A. Hopper, *A new location technique for the active office*, IEEE Personal Communications **4** (1997), no. 5.
- [46] Andrew Martin Robert Ward, *Sensor-driven computing*, Ph.D. thesis, University of Cambridge, Corpus Christi College, 1998.
- [47] A.D. Watt, *Vlf radio engineering*, Pergamon Press, 1967.
- [48] M. Weiser, *The computer for the twenty-first century*, Scientific American **265** (1991), no. 3, 94–104.
- [49] G. Welch and G. Bishop, *An introduction to the kalman filter*, (2002), <http://www.cs.unc.edu/~welch/kalman/kalmanIntro.html>.
- [50] Greg Welch, Gary Bishop, Leandra Vicci, Stephen Brumback, and Kurtis Keller, *The hiball tracker: High performance wide-area tracking for virtual and augmented environments*, Proceedings of the ACM Symposium on Virtual Reality Software and Technology, December 1999.
- [51] A. Willsky, *Stochastic processes and estimation*, 2001, Notes for MIT Course 6.433, to be published.