2012-09-13

A modified fingerprinting technique for an indoor, range-free, localization system with dynamic radio map annealing over time

Lesser, Andrew M.

UNIVERSITY OF CALGARY

A Modified Fingerprinting Technique For An Indoor, Range-Free, Localization

System With Dynamic Radio Map Annealing Over Time

by

Andrew M. Lesser

A THESIS

SUBMITTED TO THE FACULTY OF GRADUATE STUDIES

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE

DEGREE OF MASTER OF SCIENCE

DEPARTMENT OF ELECTRICAL AND COMPUTER ENGINEERING

CALGARY, ALBERTA

September, 2012

Abstract

Indoor wireless localization systems have gained considerable interest in the past decade with the wide spread implementation of affordable wireless networks throughout indoor environments. Many organizations have employed these systems to track people, equipment, and merchandise in an effort to reduce operating costs which can include loss or theft, inventory, and efficient utilization of time sensitive assets.

The complex, indoor radio frequency propagation environment introduces many challenges for wireless location systems. In particular, the large and small scale fading of signals introduces uncertainties in the location dependence of radio frequency measurements.

This thesis explores two approaches to mapping the above location dependency of measurements with the primary focus on reducing the time required for extensive environment calibration. The formulation of proposed location estimation algorithms and calibration approaches will be presented. A radio frequency device affixed to a mock hospital asset will be used as a real world example to validate the algorithms.

Acknowledgements

The author would like to thank his supervisors Dr. Michal Okoniewski and Dr. John Neilsen for their insightful ideas and guidance throughout this research.

A special thank you goes to Audrian Sutinjo for his countless discussions, many of which where of approaches utilized within this thesis. The author would like to acknowledge the support of fellow graduate students Trevor, Yonathan, Billy, Andy, Andrea, Marcel, Thomas, Charlotte, and many others.

The author would like to thank the technical staff, most importantly Chris Simon, for his positive attitude and fascinating troubleshooting skills that made the hardware design and fabrication within this research possible.

Table of Contents

List of Tables

List of Figures

List of Abbreviations and Symbols

Abbreviations

ADE	Average Distance Error
ANP	Anchor Point
AP	Access Point
AOA	Angle of Arrival
CDF	Cumulative Distribution Function
GEM	Generalized Expectation-Maximization
GT	Ground Truth
HMM	Hidden Markov Model
IID	Independent and Identically Distributed
IDW	Inverse Distance Weighting
IV	Infusion
KDE	Kernel Density Estimation
LOS	Line-of-Sight
MAP	Maximum A Posteriori
MLE	Maximum Likelihood Estimation
MMSE	Minimum Mean Squared Error
MSE	Mean Squared Error
NNSS	Nearest Neighbors in Signal Space
PCB	Printed Circuit Board
PDF	Probability Density Function
PL	Path Loss
RP	Reference Point
RSS	Received Signal Strength
RTLS	Real-Time Location System
SMD	Surface Mount Device
TDOA	Time Difference Of Arrival
TOA	Time of Arrival
T-R	Transmitter-Receiver
WLAN	Wireless Local Area Network
WNNSS	Weighted Nearest Neighbors in Signal Space

Symbols

ρ	Two-dimensional indoor localization environment
$\mathbf{x}_{ap_i}$	The Cartesian coordinates of the ith anchor point
$\hat{\mathbf{x}}$	The estimated asset location in Cartesian coordinates
$\mathbf{x}_{\text{test}}$	True asset test location
$\mathbf{F}(\mathbf{x}_{ap_i})$	Fingerprint matrix of the ith anchor point
$\mathbf{x}_{rp_i}$	The Cartesian coordinates of the ith reference point
T_{ap_s}	Fingerprint sample time at each anchor point
t	Asset trace sequence time index
T	Length of asset trace time
N_{test}	Number of test locations
N	Number of anchor points used in estimation (location state space)
d	Number of APs used in location estimation (dimensionality)
$\mathbf{r}_t$	Vector of RSS measurements at time t from d APs
$\mathbf{r}_{ap_i}$	Vector of fingerprint RSS at the ith anchor point
n	Number of fingerprint RSS collected at each $\mathbf{x}_i$
K	Number of selected nearest neighbors in signal space
$\mathcal{R}$	Complete radio map: $\mathcal{R} \triangleq \{(\mathbf{x}_{ap_i}, \mathbf{F}(\mathbf{x}_{ap_i})) \,|\, i = 1, \ldots, N\}$
$\mathcal{Q}$	Auxiliary Function
S	The Set of all available states in HMM
Q	Vector of unknown sequence of states at successive times
R	Vector of RSS measurements acquired during Q
$\mathbf{A}$	Location-state transition probability distribution
$\boldsymbol{\pi}$	Initial location-state probability distribution
$\boldsymbol{\lambda}$	Emission or observation probability distribution
m	Location-state transition distance constraint
θ	The tuple of HMM model parameters, $\theta = (\mathbf{A}, \boldsymbol{\lambda}, \boldsymbol{\pi})$
θ^k	The current HMM model parameters at each GEM iteration
θ^{k+1}	The updated HMM model parameters at each GEM iteration
θ^*	The converged HMM model parameters
γ	The Lagrangian multiplier for HMM parameter maximization
$\alpha_i(t)$	Forward partial probability
$\beta_i(t)$	Backward partial probability
$\delta_t(i)$	Highest likelihood along a single path among all the possible paths ending in state s_i at time t
$\psi_t(j)$	Tracking variable of best path ending in state s_i at time t
Q^+	The most probable sequence of states, given R, determined by the Viterbi algorithm: $Q^+ = \left[q_1^+, q_2^+, \ldots, q_T^+\right]$
$\mathcal{R}^+$	Annealed radio map: $\mathcal{R}^+ \triangleq \{(\mathbf{x}_{ap_i}, \mathbf{F}^+(\mathbf{x}_{ap_i})) \,|\, i = 1, \ldots, N\}$ where $\mathbf{F}^+(\mathbf{x}_{ap_i})$ for each selected $ANP = q_t^+, 1 \le t \le T$.

Chapter 1

Introduction

1.1　Motivation

Methods of wireless location estimation have become very popular in Real-Time Location Systems (RTLSs) and are employed by a variety of different commercial products [1, 2, 3, 4]. An important application of such a system is for the accurate localization of equipment and persons within a dense hospital environment. Systems utilizing an existing Wireless Local Area Network (WLAN) 802.11 infrastructure typically, but not always, employ a Received Signal Strength (RSS) based algorithm to estimate a location. This research will focus on the use of RSS measurements which are common in any WLAN. These systems can be classified into two main categories, range-based and range-free location estimation.

In the traditional range-based approach, the location of interest is estimated through the use of geometrical techniques such as Time of Arrival (TOA) or Time Difference Of Arrival (TDOA). All range-based approaches require the determination of the dependence present between any measurements and the distance between the Transmitter-Receiver (T-R). Due to the varying multipath and shadowing effects present in a typical indoor propagation environment, range-based measurements can be shown to be highly variable [5], resulting in significant challenges for accurate lo-

cation estimation.

In the range-free approach, location estimation algorithms are typically split into two stages: an offline and an online phase. In the offline phase, a radio map is constructed from RSS measurements taken within the propagation environment at multiple locations with known coordinates. This radio map is created by sectioning the environment into a dense grid of locations, or Anchor Point (ANP) locations, where sequential RSS calibration measurements are recorded from multiple APs to develop a statistical model of the propagation environment at each location. In literature, the RSS calibration measurements at ANP locations are often referred to as fingerprints of the WLAN radio frequency environment and the creation of a complete radio map is referred to as a site survey. In the online phase, an estimated location is determined from the evaluation, or location-dependence mapping, of multiple online RSS measurements against the radio map.

First, two range-free algorithms that estimate a location by mapping the RSS-location dependency between online RSS measurements and fingerprint RSS measurements will be investigated. These fingerprint RSS measurements are acquired from multiple calibration locations in the offline phase to form a radio map of the environment. The performance of each location estimation algorithm will be analyzed in terms of their error between true and estimated locations. A radio map created from fingerprint RSS measurements at every selected ANP in the environment will be used for the analysis of two AP selection methods.

Second, due to the time consuming nature of fingerprinting every ANP, which is also considered to be too invasive on the normal operations within an environment, an investigation will be conducted into the two components of fingerprinting that

can be adjusted in order to alleviate the problem: a reduction in the time spent collecting fingerprint RSS measurements at each ANP and a reduction in the density of fingerprinted ANPs throughout the environment. An interpolation algorithm will be utilized to construct multiple radio maps from fingerprint RSS measurements acquired by two unique hardware configurations.

Third, a radio map annealing algorithm consisting of a discrete state-space Hidden Markov Model (HMM) will be outlined that will utilize the motion dynamics of an asset throughout the environment to infer its online sequential RSS measurements to the most probable sequence of ANP locations, or update the fingerprints within interpolated radio maps at the ANP locations corresponding to the most probable path of an asset throughout the environment.

1.2   Project Objectives

The goal of this research is to design and improve the calibration techniques required for a range-free indoor WLAN localization system. This research will also test the implementation of manufactured tracking modules that have been specifically design for RSS-based wireless location systems. This thesis differs from the majority of previous research in literature where indoor WLAN localization systems employ laptops to acquire RSS measurements. These do not reflect real life asset localization scenarios and, as such, research in this thesis is based on commercially available modules specifically designed for WLAN localization systems. These modules will be utilized for the acquisition of all RSS measurements. The specific aims of this research are to:

1. Explore deterministic and probabilistic approaches to range-free indoor local-

ization with emphasis on methods for the selection and weighting of RSS measurements from multiple APs;

2. Develop techniques for reducing the radio map calibration effort required for range-free indoor localization, specifically the density and location of ANPs and their corresponding sample time;

3. Develop an interpolation algorithm that populates an initial radio map of dense ANPs from a sparse, incomplete, radio map;

4. Validate the interpolation with varying sparseness of radio map ANPs with fingerprint RSS measurements collected via two hardware configurations: a turntable and multiple stationary reference points;

5. Develop a radio map annealing algorithm that improves the initial radio map over time with multiple unknown asset motion sequences;

6. Validate the above algorithm with varying fingerprint sample times and reductions in the density of ANPs used for interpolation of the radio maps.

1.3 Thesis Outline

The outline of the thesis is as follows:

- In Chapter 2, two indoor WLAN location estimation approaches will be explored that have been used extensively in literature that utilizes RSS measurements: range-based and range-free indoor WLAN location estimation. The high level details of each approach will be explained, each with their own strengths and weaknesses, and will be accompanied by previous research into their specific utilization techniques. The chapter will be concluded with an explanation as to why the range-free approach was chosen for this thesis.

- In Chapter 3, the proposed range-free WLAN localization system will be outlined beginning with the RSS measurement acquisition hardware and setup. The use of a radio map will be introduced which will describe the spatial-RSS dependency mapping throughout different locations within the environment where two different location estimation approaches will also be outlined. An RSS interpolation algorithm will be formulated that will be used to complete an initial radio map collected from sparse locations, as well as a technique that utilizes online RSS measurements from multiple unknown location sequences of an asset to anneal the initial radio map, or infer the online RSS measurements to the most probable ANP location in the initial radio map, over time.

- In Chapter 4, the proposed techniques outlined in Chapter 3 will be validated. Two different AP selection methods will be compared in terms of their performance in each of the location estimation approaches. Details will be provided as to explain the choice of one location estimation algorithm for the remainder of the validation. Both an interpolation and an annealing algorithm will be validated for their performance in improving the accuracy of static location estimation. The annealing algorithm will utilize initial radio maps, interpolated from fingerprint RSS measurements collect by a turntable and multiple stationary reference point devices, and multiple dynamic asset traces with online RSS measurements collected throughout the environment.

- In Chapter 5, the thesis contributions and outline suggestions for future work on this project will be summarized.

Chapter 2

Background

Indoor location estimation systems will be classified as either a range-based or a range-free approach in this thesis. Both require the system designer to determine the dependence between radio frequency measurements and either the distance between a transmitter-receiver or calibration at pre-defined locations throughout an environment. This chapter will explore these two approaches, discuss previous work as reported in literature, and compare each in terms of their strengths and weaknesses. An explanation will be given as to why this research will focus on the range-free approach to location estimation.

2.1 Range-Based Location Estimation

Range-based approaches are generally referred to in literature as *radio propagation modeling* [6, 7, 8, 9, 10, 11, 12, 13, 14]. In TOA, the range is determined from a signals transit time, or delay, between a T-R, where as TDOA uses the difference in transit time between two or more signals. Angle of Arrival (AOA) is an extension of TDOA, where an array of receivers measure a signals transit time individually, and where the difference in range between any two can be used to calculate the AOA. These measurement types, which require precise clocks for both T-R, are considered to be extremely complex and are not available in standard 802.11 WLAN. However, RSS

measurements are readily available in WLAN where devices attempting to establish, or have established, a connection to the WLAN to rank the signal quality, based on RSS, between itself and an associated AP. If a device registers that the signal quality is degrading past its allowable limit for reliable communication, the device will attempt to associate to a different AP with greater signal quality. Most range-based approaches require knowledge of the exact coordinates of every transmitter used for location estimation. The remainder of this thesis will focus exclusively on WLAN where RSS measurements are readily available.

The first step required for range-based approaches is the determination of the dependence present between RSS measurements and the distance between the T-R in WLAN. The attenuation of RSS over distance, referred to as propagation or Path Loss (PL), is modeled to be inversely proportional to the distance, or range, between an AP and a receiver raised to a specific PL exponent. Due to the varying multipath and shadowing effects present in a dense, indoor propagation environment, it can often be quite challenging to model or predict this dependence. Multipath, or small scale fading, is caused by the interference from multiple signal paths which occur around the carrier wavelength[5]. The wavelength for 802.11 WLAN is $\lambda_c = 2.9979 \text{ E } 8/2.4 \text{ E } 9 = 12.5$ cm. Shadowing is a result of reflections, diffractions, and absorptions of a signal off obstacles in an environment.

The second step required for RSS range-based approaches typically involves trilateration, which refers to a location estimate calculated from the intersection region formed by circles of radii equal to the estimated ranges between multiple APs with known coordinates. Non-linear least squares algorithms, which suffer from convergence problems and high computational effort, are used to iteratively solve a system

of equations, formulated by the circles, to attempt to converge on a single estimated location.

2.1.1 Range Determination

The generic formulation used to model the above dependence is known as the Friss equation [5] given as

$$P_r = P_t \frac{G_t G_r g^2 \gamma}{4\pi d^{n_{ex}}} \tag{2.1}$$

where P_t is the transmitted power (W), G_t and G_r are the transmitter and receiver gains, d is the distance between the transmitter and receiver, n_{ex} is the PL exponent, g is the Rayleigh/Rician parameter, and γ is the log-normal parameter.

In order to utilize Eq. 2.1 with RSS measurements, it is often more convenient to take its logarithm, where P_r is the received power (dBm), to form

$$P_r = \alpha - 10 \cdot n_{ex} \cdot \log(d) + \mathcal{X} \tag{2.2}$$

which is utilized extensively in range-based techniques. The parameter $\alpha = P_t + 20\log(g) + 10\log\gamma + 10\log\frac{G_t G_r}{4\pi}$ can be determined *a priori* and will remain constant as long as the transmit power and antenna gains remain unchanged [7, 12]. $\mathcal{X}$ is approximated as a Gaussian random variable with zero mean and represents the shadowing effects within the environment. It can be noted that the free-space forms of Eq. 2.1 - 2.2 occur with a PL exponent of 2. It has been shown [5, 15] that for indoor environments the PL exponent can range from a typical value of 3 to an extreme of > 6. This is attributed to the dramatic increase in the accumulation of non-Line-of-Sight (LOS) components, due to obstructions within the environment,

which arrive at the receiver at different instances in time.

To calculate a distance estimate from Eq. 2.2, an assumption must be stated that over a short period of time the WLAN propagation environment between a T-R does not change and therefore a n_{ex} exists such that Eq.2.2 follows a Gaussian distribution [7].

$$P_{r_t} - (\alpha - 10 \cdot n_{ex} \cdot \log(d_t)) \ \rightarrow \ \mathcal{N}(0, \sigma) \tag{2.3}$$

P_{r_t} is modeled as a Gaussian random variable, at any instant in time t, where a Maximum Likelihood Estimation (MLE) of the distance, or range, between a T-R can be calculated for a single RSS measurement as

$$\hat{d} = 10^{\frac{(\alpha - P_{r_t})}{10 \cdot n_{ex}}} \tag{2.4}$$

If multiple RSS measurements are collected for a short period of time, and each are modeled as independent Gaussian random variables, then a MLE of the this distance becomes

$$\hat{d} = 10^{\frac{(\alpha - \bar{P}_r)}{10 \cdot n_{ex}}} \tag{2.5}$$

where $\bar{P}_r$ is the average of the multiple RSS measurements. During asset motion, it has been shown that RSS measurements averaged over a very short period of time will reduce the effects of the fast fading term [6, 7].

Examples of Eq. 2.4 based on RSS measurements can be seen in Fig. 2.1 for LOS and non-LOS environmental conditions. The PL exponent for the LOS case is slightly higher to that of free space, which is to be expected since there are multiple reflections off surrounding surfaces arriving at the receiver at different times, as well

as a direct LOS component. The non-LOS case does not have a direct component and leads to a further increase in the PL exponent. Both cases demonstrate the severe multipath effects on RSS measurements within an indoor environment where variations by almost 17 dBm can be seen at a single location. A common method used
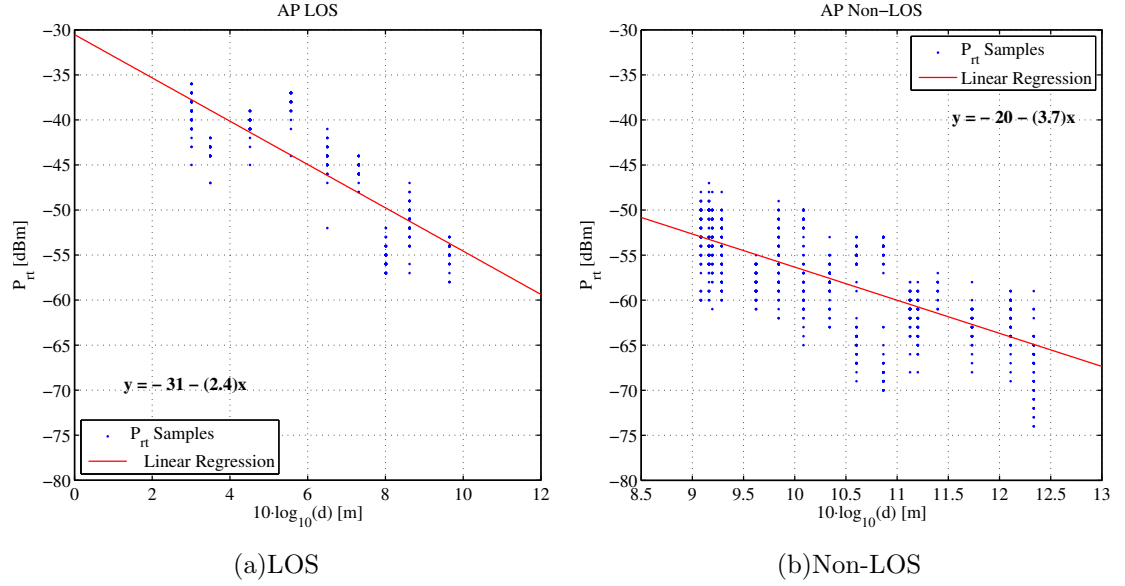
Figure 2.1: Received Signal Strength for Single AP at LOS and non-LOS Locations

to determine the PL exponent for a given range of locations is to use linear regression techniques based on RSS measurements collected a multiple locations within the localization environment. This can be seen as a limited approximation since RSS measurements collected across a large area are used to calculate a single PL exponent for that area. Any estimated range based on Eq. 2.4 that utilizes a single PL exponent will not take into account the large statistical variations in RSS for that particular location. However, while averaging of multiple RSS measurements for the range

calculation of Eq. 2.5 does incorporate statistical variations, the use of a single PL exponent over a large area still remains as a critical influence on inaccurate range estimation.

2.1.2 Trilateration

In a range-based approach, the most common technique used to estimate a location is called *trilateration*. It is based exclusively on T-R geometry, where range estimates based on RSS measurements are used to form radii of intersecting circles with known origin, or in the case of RSS measurements, the origins are APs with known coordinates. The prefix *tri* reflects the requirement of 3 or more estimated ranges, or circles, that are needed to calculate an intersection region. This can be seen in Fig. 2.2 for two different scenarios. Ideally, if the estimated radii equaled the true radii for three or more circles, then the intersection of the circles converges on a single point, the true location. In a severe multipath environment, these radii are over or under estimated yielding an intersection region consisting of multiple solutions. Therefore, an asset can have an estimated location anywhere within this intersection region due to the inaccuracy in range estimation; this directly translates to an inaccuracy in location estimation. To estimate a location region with trilateration, a system of equations formed from the multiple circles must be solved through the iterative minimization of a cost function, generally with a sum of squared distances criterion as shown in Eq. 2.6:

$$\epsilon = \sum_{i=1}^{N_{ap}} \left((\hat{x} - a_i)^2 + (\hat{y} - b_i)^2 - \hat{d}_i^2 \right)^2 \tag{2.6}$$
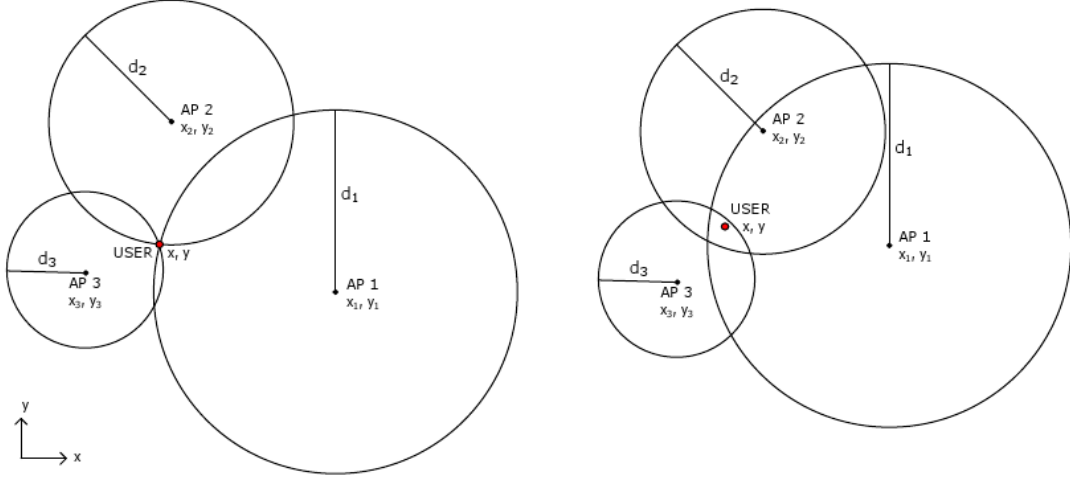
Figure 2.2: Example of True and Estimated Range Distances

where N_{ap} is the number of APs used in localization, (a_i, b_i) are the Cartesian coordinates of the ith AP, $\hat{d}_i$ is the range estimate for the ith AP, and $(\hat{x}, \hat{y})$ is the iteratively estimated location. Non-linear least squares algorithms such as conjugate gradient, steepest decent, Gauss-Newton, Levenberg-Marquardt, and trust region have been investigated for trilateration [6, 7, 10, 11, 16, 17], with each performing an iterative minimization of a cost function based on a overdetermined system of non-linear equations. Some of the main drawbacks are: large convergence time, requirements for accurate initial location estimates, and the possibility of a divergence from any real solutions. The problem can be formulated as a linear set of equations by making use of lines of position, or radial axes, of multiple intersecting circles thus allowing for a linear least squares location estimation. However, this approach is generally disregarded due to the time required to algebraically determine the number of independent lines of position, for different multiples of APs, at each location estimation instance and generally results in multiple solutions.

Some trilateration systems calculate multiple location estimations over a short time period with sets, or windows, of RSS measurements collected at the unknown location where the final estimated location is determined by the average or weighted centroid of the multiple solutions [18].

Notable research conducted in [7, 9] attempted to concurrently estimate the PL exponents and location at any instance in time through the maximization of a compatibility function, which incorporates the PL exponents into Eq. 2.6. The authors reasoned that the estimated ranges required for Eq. 2.6 can be regarded as PL exponent estimations that maximize the specific compatibility function. With the increase in the number of estimated parameters, a non-linear, Levenberg-Marquardt least squares estimation algorithm was formulated that placed a number of heuristic constraints on the values of the estimated PL exponents. The computational effort required to perform such a non-linear least squares approach is high, yet they were able to achieve good location estimation accuracy with an average distance error of 3.97 meters and a 70 percentile of 4.7 meters, without the assistance of tracking techniques. The highest average of online RSS measurements from 4 APs were used for location estimation.

The SELFLOC system proposed by [16] relied on a method of fusing multiple location estimations together through Region of Confidence (ROC) filtering. Prior to estimation, a ROC is constructed from the multiple solutions to the system of equations that form intersection regions of the range estimated circles of multiple APs. Location estimates are either accepted or rejected depending on if they fall within the ROC or not. The filtered location estimates are then sorted into bins where the shortest path connecting at least one estimated location from each bin is found and

averaged together to form the final estimated location. A new ROC is formulated based on previous shortest paths following the expiration of a designated time interval. The system was able to achieve a best mean distance error of 3.8 meters with the use of 4 APs.

2.2 Range-Free Location Estimation

Range-free approaches are generally referred to in literature as *fingerprinting* [10, 19, 20, 21, 22, 23, 24, 25, 26] and consist of two phases: an offline and an online phase.

In the offline phase, a location estimation system is trained or calibrated through a site survey that is performed to collect fingerprint RSS measurements from multiple APs, at N spatially distributed ANP locations, and to store them within a radio map of the environment. Each ANP utilized in the site survey is a known location distributed by a unique grid layout which is typically designed to uniformly cover the localization environment.

This calibration is needed to determine the location dependency of RSS. The site survey is carried out where the fingerprint RSS measurements, as well their Cartesian coordinates, are collected from each visible AP and used to construct a radio map, $\mathcal{R}$, of the environment. The mean and variance of the RSS at each ANP, for each visible AP, gives us insight into the temporal variations throughout the environment caused by time-varying multipath and shadowing.

$$\mathcal{R} \triangleq \{(\mathbf{x}_{ap_i}, \mathbf{F}(\mathbf{x}_{ap_i})) \,|\, i = 1, \ldots, N\} \tag{2.7}$$

where $\mathbf{x}_{ap_i} \triangleq [x\ y]^T$ are the Cartesian coordinates of the ith ANP.

$\mathbf{F}(\mathbf{x}_i) \triangleq [\mathbf{r}_{ap_i}(1), \ldots, \mathbf{r}_{ap_i}(n)]$ is the fingerprinting matrix containing the RSS measurements at each ANP. The fingerprint vector $\mathbf{r}_{ap_i}(\tau) \triangleq [r_i^1(\tau), \ldots, r_i^L(\tau)]$ contains the RSS measurements from each of the L APs at time τ at each spatial point $\mathbf{x}_i$. The time consuming and intrusive nature of fingerprinting an environment at every ANP used in the range-free location estimation system is by far one of its most important limitations. There are two main components that can be adjusted in order to alleviate this problem. The first would be the amount of time spent fingerprinting at each ANP. This directly translates into the number of RSS measurements from each ANP used in the localization algorithm. The second would be the number of fingerprint ANPs used to complete the radio map. This thesis will explore methods to alter these components and attempt to alleviate the time sensitive nature of fingerprinting a radio map.

Another limitation would be the vendor specific nature of acquiring RSS measurements. The IEEE 802.11 standard specifications on RSS, are that they are a relative measure of RSS for internal use by chipsets[27]. Typically, modulated radio frequency signals enter a device through single or multiple antennas, with known gain, and are sampled by analog to digital converters for either the in-phase, quadrature-phase, or a combination of components. An integrator is then used to accumulate samples over a specified time period for a vendor specific calculation of RSS[28, 29]. A more detailed explanation of how vendor specific hardware acquires RSS measurements is beyond the scope of this thesis. Therefore, all fingerprint RSS measurements acquired to construct a radio map will be hardware dependent.

2.2.1 Mapping the RSS-Location Dependency

In the online phase, an asset affixed with a transceiver collects test RSS measurements, $\mathbf{r}_t$, at its unknown location and a location estimate is computed $\hat{\mathbf{x}} = b\left(\mathbf{r}_t, \mathcal{R}\right)$ where $b\left(\ldots\right)$ is a particular mapping between the radio map ANP locations and the RSS measurements. Two approaches to mapping $b\left(\mathbf{r}_t, \mathcal{R}\right)$ will be considered: deterministic and probabilistic.

2.2.1.1 Deterministic

The most well known deterministic approach to indoor range-free location estimation is often referred to as the Nearest Neighbors in Signal Space (NNSS) algorithm and its formulation was first reported in literature [24]. This laid the foundation for subsequent work reported in [30, 31, 32, 33, 34] and has been demonstrated to provide excellent accuracy when compared to more sophisticated techniques.

Let the set $\left\{\mathbf{x}_{ap_{(1)}}, \ldots, \mathbf{x}_{ap_{(N)}}\right\}$ denote the ordering of ANPs with respect to their increasing distance in signal space, or the distance between the sampled RSS measurements and the mean calibration RSS measurement at each ANP.

$$d_{NNSS}\left(\mathbf{r}_t, \mathbf{F}(\mathbf{x}_{ap_{(i)}})\right) = \|\mathbf{r}_t - \bar{\mathbf{r}}_i\|^2 \qquad 1 \leq i \leq N \tag{2.8}$$

where $\bar{\mathbf{r}}_i = \frac{1}{n}\sum_{\tau=1}^{n}\mathbf{r}_i(\tau)$ are vectors of the average fingerprint RSS, for each AP, at $\mathbf{x}_{ap_{(i)}}$. The location estimate is obtained from the average of the K selected ANPs with known locations as

$$\hat{\mathbf{x}} = \frac{1}{K}\sum_{k=1}^{K}\mathbf{x}_{ap_{(k)}} \tag{2.9}$$

The RADAR system described in [24, 30] was the first known deterministic finger-printing approach that utilized the NNSS algorithm for location estimation. The effect of using multiple NNSS for location estimation was analyzed. Their results showed that for a large value of K, a significant degradation in accuracy occurred due to ANPs located far from the true location were being included in the average of Eq. 2.9. The effect of orientation was also investigated with separate radio maps collected for each of the four directions. Location estimations were performed with online RSS measurements collected in an orientation opposite that that of the radio map. Their results showed a decrease in the localization accuracy as significant as 67% for differnt orientations. The number of ANPs, and their corresponding locations, used in a radio map was also investigated. Starting from a randomly located set of $N = 70$ ANPs throughout their environment, a reduction to $N = 20$ and $N = 40$ ANPs saw a median decrease in accuracy of 33% and 10% respectively which led then to invoke a threshold on the density of required ANPs and distributed them uniformly throughout the localization environment. An extension known as Weighted Nearest Neighbors in Signal Space (WNNSS) will be described in greater detail in the following chapter as it forms the basis for a comparison of location estimation approaches explored in this thesis.

A major limitation of the NNSS algorithm is that no statistical properties are utilized in the mapping between online RSS measurements and the fingerprints contained in the radio map, as opposed to the probabilistic approaches outlined below.

2.2.1.2 Probabilistic

Probabilistic range-free location estimation is formulated based on Bayesian statistics with the most common approaches including MLE and Maximum A Posteriori (MAP)

estimation which modeled the fingerprint RSS measurements, $\mathbf{r}_{ap_i}$, conditioned on each ANP location, $\mathbf{x}_{ap_i}$, contained in the radio map each as conditional probability densities. The MLE of a location is approximated as

$$\hat{\mathbf{x}}_{\mathrm{MLE}} \approx \arg \max_{\mathbf{x}_{ap_i}} p\left(\mathbf{r}_{ap_i}|\mathbf{x}_{ap_i}\right) \qquad i = 1 \ldots N \tag{2.10}$$

with the MAP estimate of a location approximated as

$$\hat{\mathbf{x}}_{\mathrm{MAP}} \approx \arg \max_{\mathbf{x}_{ap_i}} p\left(\mathbf{x}_{ap_i}|\mathbf{r}_{ap_i}\right) = \frac{p\left(\mathbf{r}_{ap_i}|\mathbf{x}_{ap_i}\right) p\left(\mathbf{x}_{ap_i}\right)}{\int p\left(\mathbf{r}_{ap_i}|\mathbf{x}_{ap_i}\right) p\left(\mathbf{x}_{ap_i}\right) \mathrm{d}\mathbf{x}_{ap_i}} \qquad i = 1 \ldots N \tag{2.11}$$

where $p\left(\mathbf{r}_{ap_i}|\mathbf{x}_{ap_i}\right)$ and $p\left(\mathbf{x}_{ap_i}|\mathbf{r}_{ap_i}\right)$ are conditional probability densities referred to as the *likelihood* and *posterior* densities. $p\left(\mathbf{x}_{ap_i}\right)$ is the *prior* knowledge of an assets location before any observed RSS measurements and is typically assumed to be uniform in density as no location knowledge is available before acquiring RSS measurements. Such an assumption results in a correspondence between the MLE and MAP formulations and, since the approximated likelihood density is determined solely from the fingerprint RSS measurements, the MLE of location is bound to the discrete, uniformly placed ANPs within the environment.

A common method to approximate the likelihood densities, $p\left(\mathbf{r}_{ap_i}|\mathbf{x}_{ap_i}\right)$, $i = 1 \ldots N$, is to represent the fingerprint RSS measurements for each available AP as a *histogram* with defined origin and bin width, with reported use in literature by [26, 35, 36, 37]. The resulting histogram density estimate is constructed by counting the number of RSS measurements that fall within each bin defined as $\{w(b),\ b_{min} \le b \le b_{max}\}$. This process is repeated for each AP with the overall likelihood density calculated as the dimensional product of these individual densities. The

overall histogram estimate of the approximated likelihood density at a single ANP is given as [22]

$$\hat{p}\left(\mathbf{r}_{ap_i}|\mathbf{x}_{ap_i}\right) = \prod_{l=1}^{d} \hat{p}\left(r_{ap_i}^{l}|\mathbf{x}_{ap_i}\right) = \prod_{l=1}^{d} \frac{1}{nh} \sum_{\tau=1}^{n} \delta\left(r_{ap_i}^{l}\left(\tau\right)\right) \qquad (2.12)$$

where $h = w\left(b+1\right) - w\left(b\right)$ is the bin width and

$$\delta\left(r_{ap_i}^{l}\left(\tau\right)\right) = \begin{cases} 1 & \text{if } w\left(b\right) \leq r_{ap_i}^{l}\left(\tau\right) < w\left(b+1\right) \\ 0 & \text{otherwise} \end{cases} \qquad (2.13)$$

The histogram density estimate requires the assumption that the fingerprint RSS measurements are Independent and Identically Distributed (IID). This assumption is violated where correlations can exist for RSS measurements acquired over time [22]. The histogram estimate of approximated likelihood densities is strongly dependent on the choice of origin and bin width which, even with optimal parameters, has a relatively slow convergence to the true density due to the large number of required samples [38, 39].

A third approach to probabilistic location estimation known as the Mean Squared Error (MSE) estimator, explored in literature by [19, 20, 21, 22, 23, 40], will utilize a technique referred to as multivariate Kernel Density Estimation (KDE) to approximate the required posterior densities and will be explored in greater detail in Chapter 3 as it forms the basis of the indoor range-free localization system detailed in this thesis. The work in [22] reported superior Minimum Mean Squared Error (MMSE) location estimation performance of a multivariate KDE when compared to a histogram density estimate. Therefore, the use of a kernel density estimate for the calculation

of posterior densities will not be explored in this thesis.

2.2.2 Radio Map Interpolation

A variety of radio map interpolation approaches have been investigated in the literature [25, 26, 41, 42] with each employing either deterministic or probabilistic location estimation algorithms. To validate performance, various reductions to a radio map were explored including either the density of ANP locations, the sample time at each ANP location, or both.

The work in [41] detailed the formulation of a cubic spline interpolation algorithm to interpolate a complete radio map from fingerprint RSS measurements at sparse ANP locations. The cubic spline formulation required the computation of polynomial coefficients based on either one of four different boundary conditions. The fingerprinting sample time at each ANP was fixed with the analysis of multiple radio maps interpolated using each of the four boundary conditions through reductions in the density of ANPs. A single boundary condition was then selected through a comparison of the multiple interpolated radio maps versus a complete radio map without any reduction in the number of ANPs. Location estimation was calculated with the K NNSS algorithm and the optimal radio map. The authors provide no indication as to the selected value of K or the number of utilized APs. Without interpolation, the average distance error was reported as 2.77 meters using a complete radio map of fingerprint RSS measurements. The best average distance error was reported as 2.92 meters and occurred with a 38% reduction in the total number of ANPs used for cubic spline radio map interpolation. The worst average distance error was reported as 3.52 meters and occured with a 13% reduction in the total number of ANPs.

The work in [42] formulated a localization estimation approach that utilized a set of radial basis functions to interpolate between a reduced radio map and online RSS measurements and produce a estimated location. Each radial basis function was dependent on the choice of kernel function, which mapped location-RSS dependency between the radio map and online RSS measurements, which was selected as Gaussian. A location offset, determined as the centroid of all ANP coordinates, was also required along with an optimal kernel bandwidth. This complex approach required a least squares estimation of radial basis function parameters for each variation of the kernel bandwidth, with the optimal value selected as the one resulting in the smallest Root Mean Squared (RMS) distance error using a separate test radio map. This approach did not attempt to interpolate fingerprint RSS measurements from a reduced radio map to a complete one. The density of ANPs within a radio map and their corresponding sample time were both investigated with results showing only a 21% increase in the RMS distance error for a reduction in sample time of 1/2 and a reduction to only 40% of the initial ANP in the complete radio map.

The work in [26] outlined a probabilistic location estimation system which utilized a histogram estimate of the likelihood densities, at each ANP, stored in a radio map. Therefore, histogram estimates of the likelihood densities at the reduced radio map ANP locations were used to linearly interpolate the missing likelihood densities to a complete radio map. The density of a skipped ANP in the complete radio map was interpolated from its two neighboring ANP likelihood densities based on the linear distance between them. Intuitively, an ANP closer in distance to the skipped ANP would have a greater contribution then the ANP further away. This relatively efficient interpolation approach preformed well with the overall accuracy of localization

increasing from 49% to 59%, resulting from a 2/3 reduction in the number of original ANPs and 20 fingerprint RRS measurements used in the histogram estimate of the likelihood densities. With an increase to 60 fingerprint RSS measurements, the overall accuracy increased further from 56% to 68%.

The work in [25] focused solely on incomplete outdoor radio maps which utilized an Inverse Distance Weighting (IDW) algorithm for the interpolation of a complete radio map with fingerprint RSS measurements from sparse, scattered ANP locations. This technique will be explored further in Chapter 3 as it forms the basis for the interpolation approach used in this thesis.

2.2.3   Radio Map Annealing

The radio map annealing algorithm outlined in [26, 40] models each dynamic unknown user trace sequence as a Markov process governed by a stochastic finite state machine known as a Hidden Markov Model (HMM). The model parameters a transition matrix which governs how a user may move through the available states, an initial state distribution that accounts for prior knowledge about were a user may have started its sequence, and an emission distribution which describes the likelihood of a particular observation given that the user is within a certain state. The emission distribution can be formulated using any probabilistic approach to mapping the RSS-location dependency using the fingerprint RSS measurements in an initial radio map. The number of states in the model is identical to the number of ANPs in the initial radio map. The term *hidden* is used to emphasize that the sequence of states corresponding to each user trace is unknown and that only observations are emitted from each state in the unknown sequence, or in this case online RSS measurements. In the first step

to improve the initial radio map over time, multiple unknown user traces as used to train the HMM model parameters via a Generalized Expectation-Maximization (GEM) algorithm. This iterative technique adjusts the HMM model parameters in an attempt to maximize the probability of observing each sequence of online RSS measurements emitted from the unknown user traces. The second step to improve the initial radio map over time involves the use of a Viterbi algorithm, along with the adjusted HMM parameters, to find the most probable sequence of states, or path of the user trace, that could have produced the observed online RSS measurements. The initial radio map can now be annealed with the state, or ANP, labeled online RSS measurements from each user trace. The radio map annealing algorithm will be detailed in Chapter 3 as it forms the basis for the approach used in this thesis and was not explained in great detail in the literature.

2.3 Discussion

A detailed background of range-based and range-free location estimation approaches has been given with a high level comparison of their individual strengths and weaknesses shown in Table 2.1. Based on this comparison, along with literature presented in the previous sections, the probabilistic range-free location estimation approach was chosen for for remainder of the thesis. The proposed range-free localization system, outlined in Chapter 3, will implement a MSE estimator which utilizes a probabilistic technique referred to as multivariate KDE to approximate the required posterior densities. To reduce the exhaustive effort of acquiring fingerprint RSS measurements at each ANP, radio map interpolation and annealing algorithms will be presented based on IDW and HMM parameter training via GEM with multiple unknown asset traces.

Range-based Approaches	
Strengths	Weaknesses
• no extensive radio map database required • no exhaustive offline fingerprinting phase	• precise AP coordinates required • determination of multiple location dependent PL exponents for range calculation • iterative non-linear solvers required for estimation location

Range-free Approaches	
Strengths	Weaknesses
• AP coordinates are not required • higher accuracy reported in literature • probabilistic approaches can be used to anneal an incomplete radio map over time • measurements at specific orientations can be incorporated into radio map • radio map interpolation can reduce site survey	• Extensive fingerprinting required at each ANP • location estimation limited to scope of fingerprint environment • radio map is hardware dependent • large database to store radio map

Table 2.1: High Level Comparison of Strengths and Weaknesses

Chapter 3

Proposed System and Concepts

As stated in chapter 2, the range-free localization technique will be used in this thesis. The initial location estimation system concept is illustrated at a high level in Fig. 3.1. This chapter will outline and discuss each of the sections required for the operation of the overall system and will begin with the measurement system that utilized specialized hardware to acquire fingerprint RSS measurements that will be used to construct a radio map of the localization environment. As an example of an asset, we used a mock intravenous pump equipped with the specialized hardware to acquire RSS measurements at multiple online locations. To investigate the effects that an asset has on RSS attenuation, asset radiation patterns were collected in an
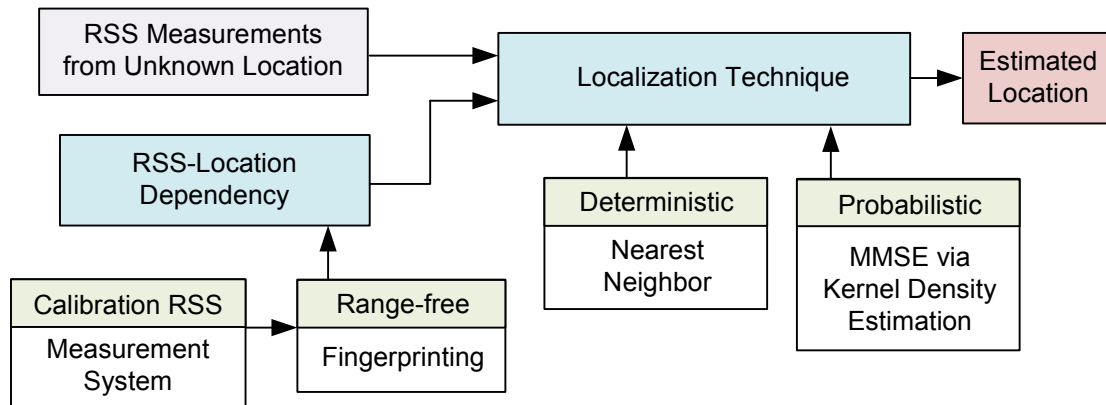
Figure 3.1: Diagram of Proposed Range-Free Localization System

anechoic chamber for the mock Infusion (IV) pump as well as a human subject.

Two range-free location estimation methods will be explored that take different approaches to the mapping of the dependence between the radio map fingerprint and the RSS measurements collected at test locations. To try and alleviate the time consuming and intrusive nature of fingerprinting, the reduction in sampling time at each ANP and the total number of ANP required for an initial radio map will be explored. This will be demonstrated later in the chapter with a modified location estimation system shown in Fig. 3.16. Finally, an annealing algorithm will be outlined that utilizes multiple sequences of RSS measurements with unknown locations to improve the radio map over time.

Tracking algorithms, which utilize trajectories based on a combination of location and velocity estimations to help improve an overall solution will not be applied in this thesis. Concepts applied in this thesis can be extended to easily include tracking methodologies and will be discussed in the conclusion along with potential future work.

3.1 Measurement System

Throughout the thesis, all RSS measurements were acquired using WiFly® RN-134G 802.11b modules, a product family manufactured by Roving Networks Inc.[43]. A standalone module can be seen in Fig. 3.2(a) with an attached omnidirectional antenna. Fig. 3.2(b) displays a Reference Point (RP) which houses one of the standalone modules and was designed to safely secure the device within the environment and to protect the hardware from tampering. These modules are considered ideal
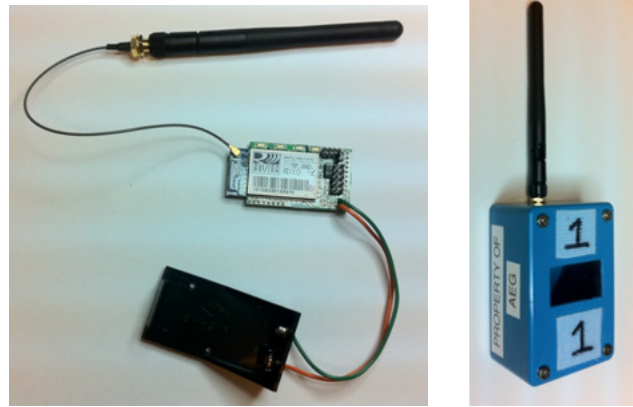
(a)Module with Omnidirectional An-
tenna

(b)Reference
Point

Figure 3.2: WiFly RN-131G 802.11b Modules

for this research because of their ability to perform real-time RSS measurement ac-

quisition. Some of the features include: selective channel scanning, variable channel

scan time, wakeup/sleep timers, on-chip antenna, and ultra low power consumption

(sleep/awake). The WiFly® modules run specialized firmware that allows specific

events to wake up the module, scan the environment, and transmit the data for col-

lection. Online RSS measurements from test locations were collected using the on

board Surface Mount Device (SMD) antenna. Radiation patterns for each configura-

tion were measured in the anechoic chamber at the University of Calgary to determine

the best orientation of the module, on a human subject and on a mock IV pump, and

determine effects cause by RF attenuation.
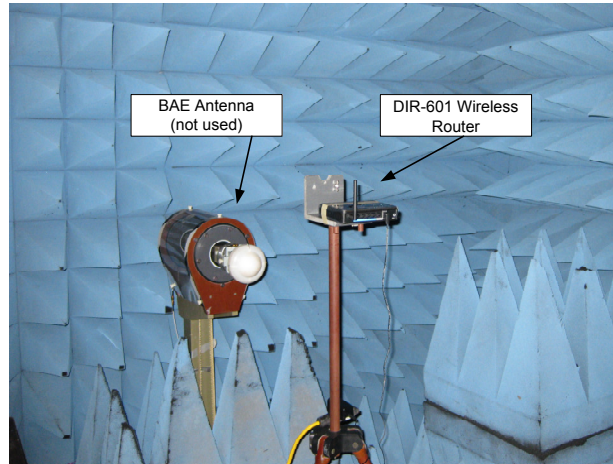
Figure 3.3: Access Point Located in Anechoic Chamber

3.2 Radiation Patterns

The anechoic chamber at the University of Calgary was used to characterize the radiation patterns of a module with the omnidirectional and the chip antennas. An investigation into these radiation patterns was required to characterize the RSS attenuation introduced by affixing a module to an asset as different materials vary widely in their electromagnetic characteristics [5]. Significant RSS attenuation, throughout any 360 degree orientation, can dramatically reduce the accuracy of any WLAN localization system.

To simplify the problem, a DIR-601 wireless router was placed and aligned atop a tripod and used as the transmitter as shown in Fig. 3.3. This was the logical choice since the experimental site uses these same wireless routers as APs. To limit the addition of extra RF interference between the AP and the module, C++ code routines
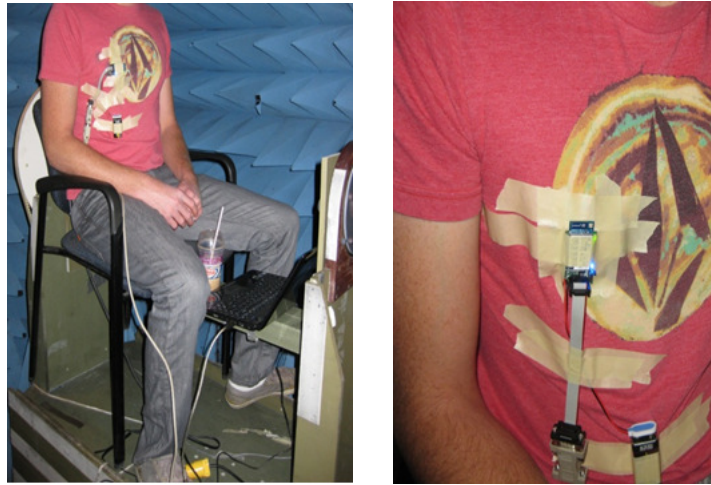
Figure 3.4: Vertically Orientated Module on Human Subject in Anechoic Chamber

were written to scan, collect, and transmit the RSS measurements via RS-232 outside the chamber. If the specialized firmware were used, the module would attempt to transmit the data to a web server following each RSS scan. The radiation patterns were constructed at 1.5 degree increments from the average of 5 RSS measurements. Each radiation pattern plot is orientated so that the transmitting AP is in direct line with the module at +90 degrees (right side of each plot). To measure the radiation pattern of the omnidirectional antenna, a chair was placed on the chamber turntable and a user held the antenna base at arm's length while focusing on the antenna place-ment over the turntable pivot point. For the chip antenna, radiation patterns were collected for the module aligned in a vertical and horizontal orientation. A human test asset was placed in a chair on the chamber turntable with a module fixed to the left chest in each orientation, as shown in Fig. 3.4, with the resulting radiation patterns shown in Fig. 3.5. Nulls in the radiation pattern can be seen at angles of −62 and 100 degrees for the horizontal orientation of the module, and are absent
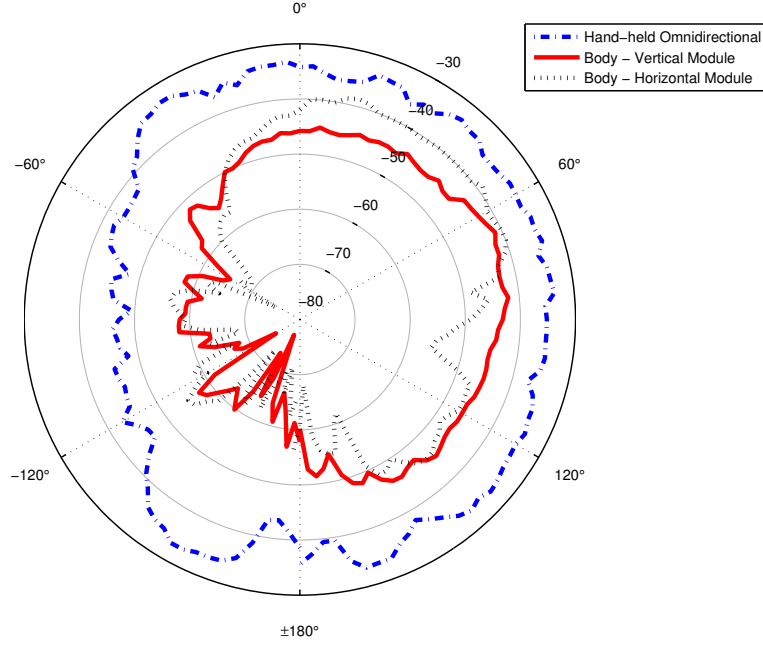
Figure 3.5: Radiation Pattern for Vertically and Horizontally Orientated Module on Human Subject [dBm]

in the vertical orientation. These nulls are a characteristic design in the SMD chip antenna radiation patterns. The results further indicate, for angles between −50 and 170 degrees, that the RSS is severely degraded and is attributed to the human body's ability to act as a large RF attenuator [44]. These large variations in the reported RSS, from either a single or multiple APs, would inflict additional location estimation errors due to inaccuracies in the mapping of the RSS-location dependency. Methods to alleviate these effects will not be explored in this thesis. The radiation pattern of a module attached to the mock IV pump was also conducted for the said orientations as shown in Fig. 3.6. A reference radiation pattern was also constructed by suspending a module above the chamber turntable at the exact height of the mock IV pump using polystyrene, a material with the electromagnetic characteristics close to air [45], as
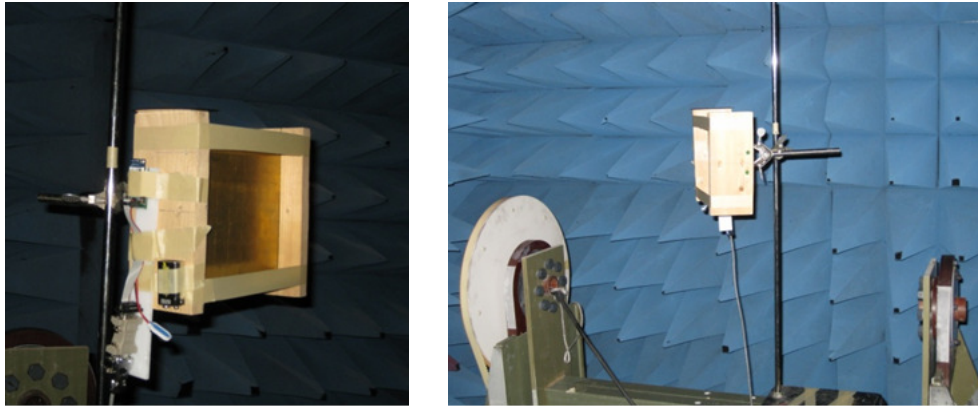
Figure 3.6: Vertically Orientated Module on Mock IV Pump
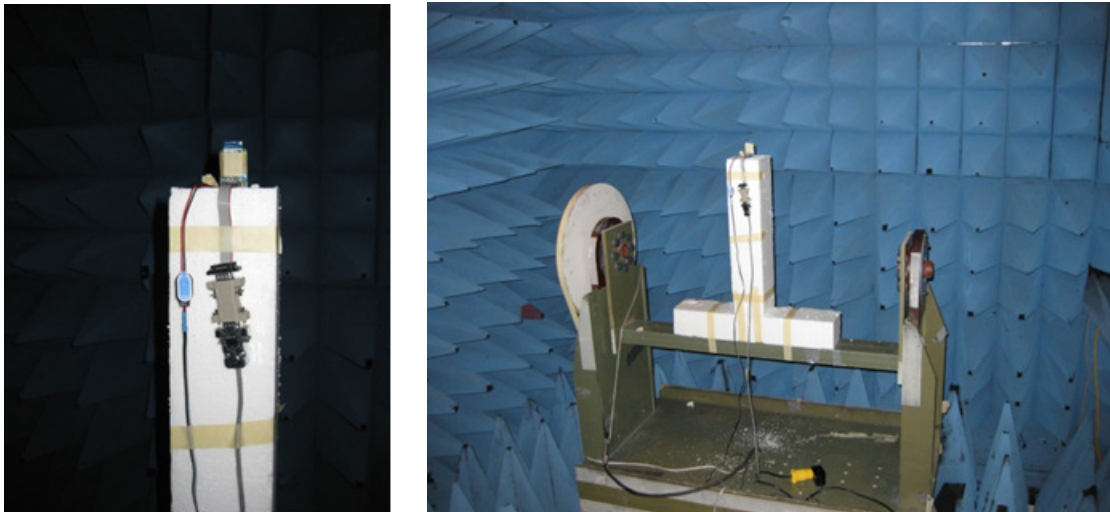
Figure 3.7: Vertically Orientated Module on Styrofoam
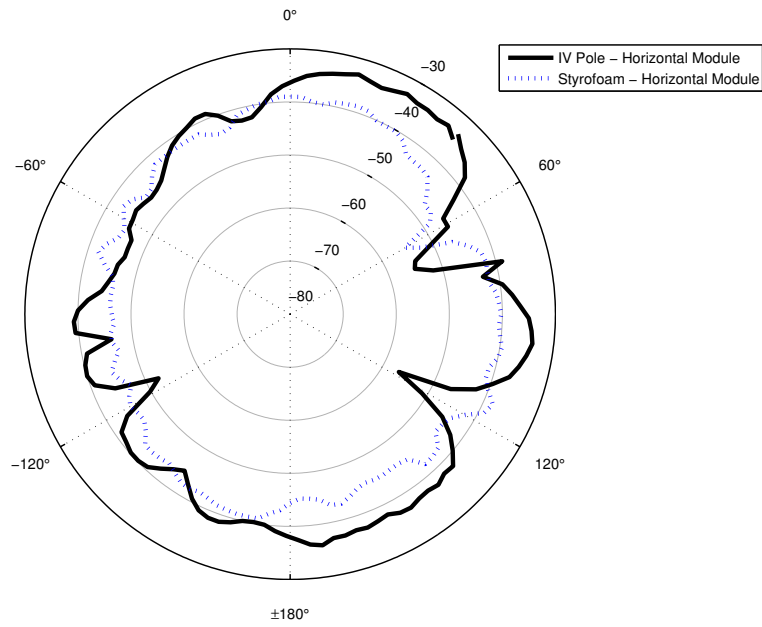
Figure 3.8: Radiation Patterns for Mock IV Pump and Styrofoam - Horizontal [dBm]
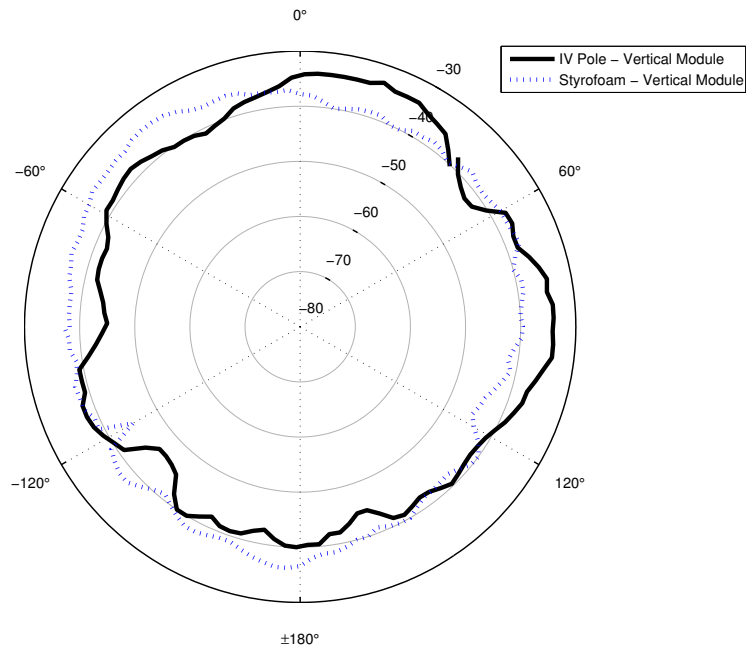
Figure 3.9: Radiation Patterns for Mock IV Pump and Styrofoam - Vertical [dBm]

shown in Fig. 3.7. From this we can determine if the turntable itself is affecting the radiation pattern. The specifications of the Rufa 2.4 GHz SMD Chip antenna [46], manufactured by Antenova, details the radiation patterns for these two alignments but not the radiation patterns produced with the addition of a human or equipment asset. The radiation patterns for the horizontal and vertical module orientations are shown in Fig. 3.8 - 3.9. The horizontal module configuration of Fig. 3.8 illustrates nulls in the radiation pattern as described for the human body case. The radiation patterns for the vertical module orientation, shown in Fig. 3.9, indicate small variations in the RSS over 360 degrees, for both the mock IV pump and the polystyrene, which indicates that the placement of a vertical module on the mock IV pump would not greatly effect the reported RSS from APs. Therefore, the mock IV pump, with a vertically orientated module, will be used for acquiring RSS measurements at all online locations and is shown in Fig. 3.10.

Figure 3.10: Mock Infusion Pump Located in Hallway

3.3 Localization Site and 802.11 WLAN Infrastructure

The existing wireless infrastructure, located on the third floor of the University of Calgary's Information and Communication Technology (ICT) building, consists of three sporadically located Aruba AP-70 access points. After further investigation, it was determined that these access points would not be suitable for RSS-based location systems due to their inherent ability to vary their transmit power throughout the day as shown in Fig. 3.11. The analysis of RSS measurements taken from multiple days lead to the conclusion that the transmit power is not varied at predictable time intervals and therefore cannot be adapted for. Although this setting could probably be altered, access to these parameters was restricted to the University of Calgary IT department. To alleviate this problem, 6 D-Link DIR-601 wireless routers, or APs, were positioned throughout the localization environment. These APs allow a system

(a)Varying Transmit Power   (b)Static Transmit Power
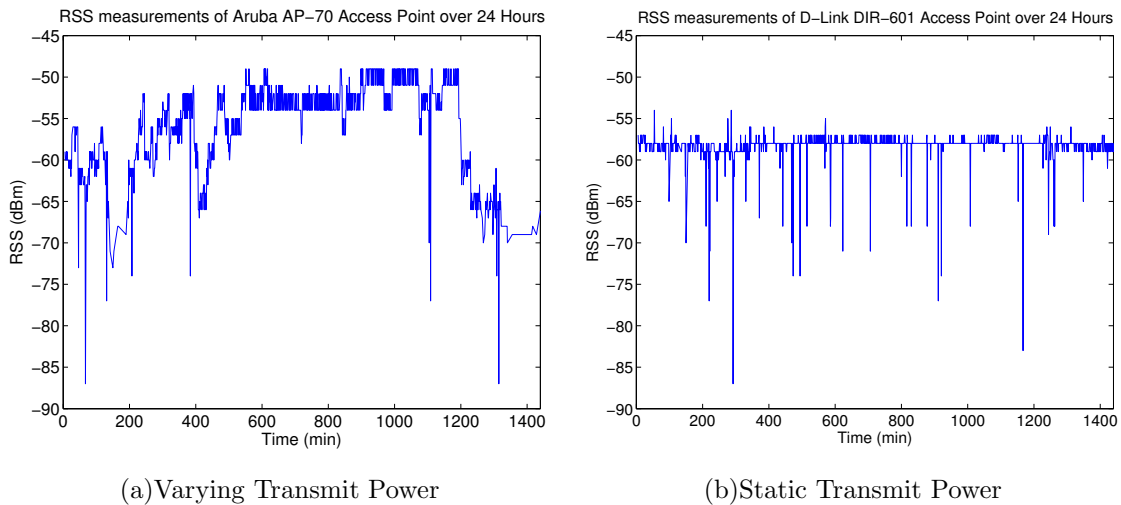
Figure 3.11: RSS measurements From a Varying and a Static Transmit Power Access Point

designer to specify a static transmit power level for which the RSS measurements, even with presence of outliers which are inherent to any indoor multipath environment, can be seen to follow a stationary level. Since the WiFly modules were set to scan each channel for 100 ms, the access points were configured to have a beacon rate of 20ms. This would increase the probability that a WiFly module would hear multiple APs during each scan operation. The third floor of the University of Calgary's ICT building is constructed from a wide variety of materials with each having their own unique electromagnetic characteristics such as reflection and attenuation factors. The main support beams, and floors, are constructed from reinforced concrete. The majority of the interior walls consist of aluminum frames layered with drywall. The building layout was surveyed into Matlab® for plotting as shown in Fig. 3.12.

3.4   Hardware for Ground Truth Radio Map Acquisition

Each ANP used in the system is at a known location spatially distributed by a unique grid layout which is typically designed to uniformly cover the localization environment. This thesis will investigate a uniform grid of ANPs separated by 1 meter throughout three corridors as illustrated in Fig. 3.12. In order to collect reliable fingerprint RSS measurements at ANP locations, a variable-speed, TCP/IP controlled turntable was constructed and tested (Fig. 3.13). Two WiFly modules were connected to opposite sides of the turntable with their omnidirectional antennas positioned vertically and were set to scan the environment on the desired channels at $1Hz$. A total of 200 scans were obtained at each ANP location translating to a fingerprint sample time of 100 seconds. Since RSS measurements a known to be dependent on
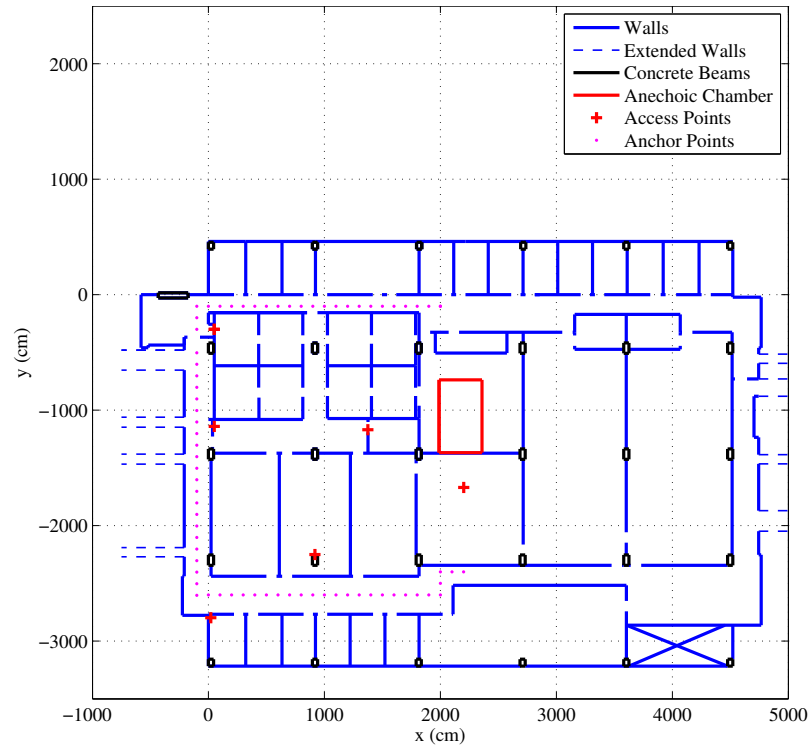
Figure 3.12: Third Floor ICT Floor Plan with Anchor Point Layout
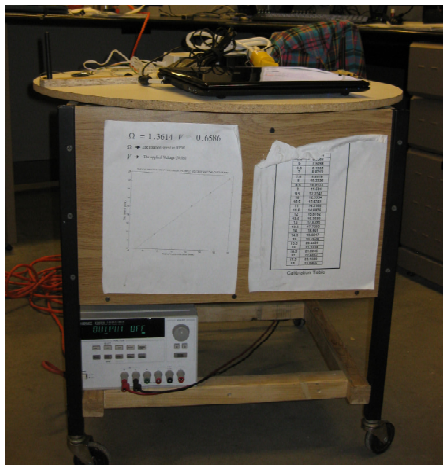
(a)Side
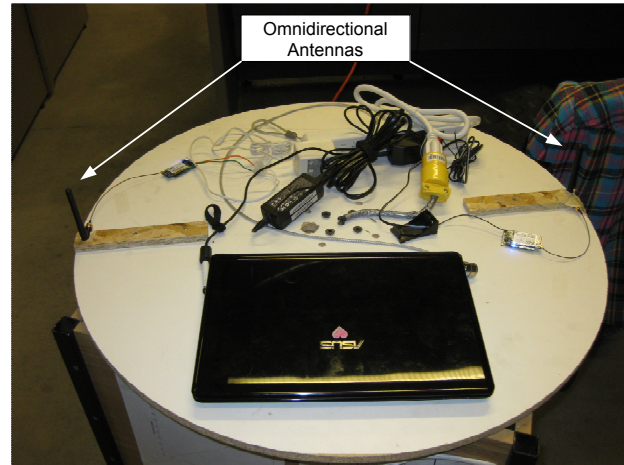
(b)Top

Figure 3.13: Turntable System Used for Site Survey

device orientation[31], the rotation of the WiFly modules throughout the data collection process provided a means as to capture more statistical variation in the RSS measurements throughout a planer 360 degree range in orientation. This was accomplished by rotating at a rate of approximately one half of a wavelength per RSS measurement at the operating frequency calculated as

$$\frac{\lambda}{2} = \frac{f}{2 \cdot c} = \frac{2.452 \text{ E } 9}{2 \cdot 2.9979 \text{ E } 8} = 0.0611[\text{m}] \tag{3.1}$$

A site survey was carried out where the fingerprint RSS measurements, as well their Cartesian coordinates, were collected from each visible AP and used to construct a radio map, $\mathcal{R}$, of the environment. The mean and variance of the fingerprint RSS measurements at each ANP, for each visible AP, gave insight into the temporal variations throughout the environment caused by time-varying multipath and shadowing.

$$\mathcal{R} \triangleq \{(\mathbf{x}_i, \mathbf{F}(\mathbf{x}_i)) \,|\, i = 1, \ldots, N\} \tag{3.2}$$

where $\mathbf{x}_i \triangleq [x \, y]^T$ are the Cartesian coordinates of the ith ANP.

$\mathbf{F}(\mathbf{x}_i) \triangleq [\mathbf{r}_i(1), \ldots, \mathbf{r}_i(n)]$ is the fingerprinting matrix containing the RSS measurements at each ANP. The fingerprint vector $\mathbf{r}_i(\tau) \triangleq [r_i^1(\tau), \ldots, r_i^L(\tau)]$ contains the RSS measurements from each of the L APs at time τ at each spatial point $\mathbf{x}_i$. The complete turntable radio map will be assumed as the Ground Truth (GT) for accuracy comparison. An example of fingerprint RSS measurements for one random ANP location are shown in Fig. 3.14.
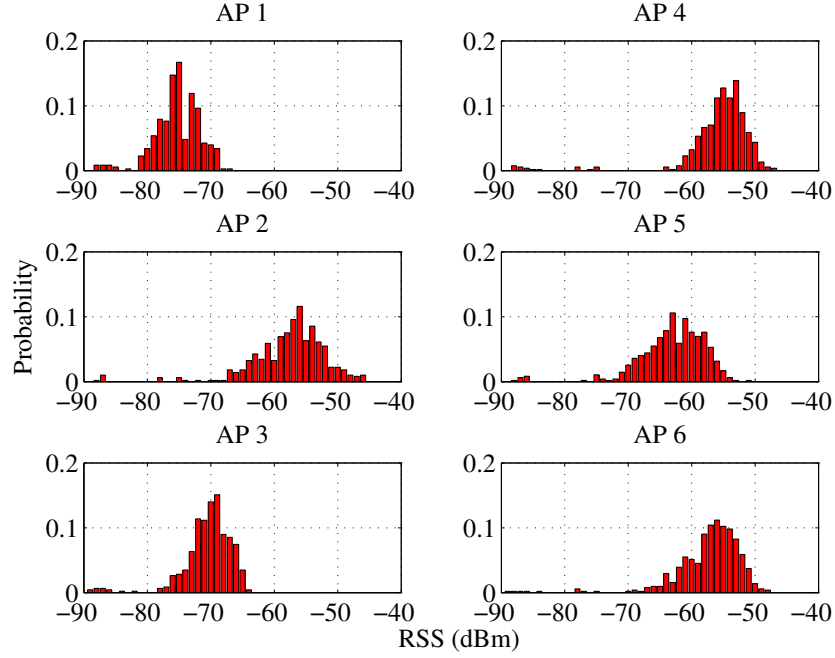
Figure 3.14: Fingerprint at Random Grid Point for Six Access Points

3.4.1 Sources of Turntable RSS Measurement Errors

For each ANP location, the turntable was positioned center with an ANP marker on the ground which was used to help with alignment. The main source of measurement error in this case would be the misalignment of the turntable with the floor markers. These alignment errors are empirically estimated to be smaller than 5 cm.

3.5 Location Estimation Methods

In this section, zero memory location estimation techniques will be explored. The term zero memory refers to an estimator that does not require knowledge of a previous estimate to compute its present estimate. Utilizing the radio map given by Eq. 3.2

and observed RSS measurements $\mathbf{r}_t$, a location estimate is computed $\hat{\mathbf{x}} = b\left(\mathbf{r}_t, \mathcal{R}\right)$ where $b\left(\ldots\right)$ is a particular mapping between the radio map ANP locations and the RSS measurements. Two approaches to mapping $b\left(\mathbf{r}_t, \mathcal{R}\right)$ will be considered. First, a deterministic NNSS approach will be formulated and discussed, followed by a probabilistic approach based on a MMSE criterion with nonparametric KDE. Specific variations of the above techniques will also be explored to show their improvement to the overall location accuracy when compared to their traditional methods. Furthermore, these methods do not rely on approximated T-R distances, the so called PL models, for localization of an asset as discussed in Chapter 2.

3.5.1 Deterministic Range-free Location Estimation

3.5.1.1 K-Nearest Neighbors in Signal Space

Introduced in Section 2.2.1 and reviewed here for convenience, let $\left\{\mathbf{x}_{ap_{(1)}}, \ldots, \mathbf{x}_{ap_{(N)}}\right\}$ denote the ordering of ANPs with respect to their increasing distance in signal space, or the distance between the online RSS measurements and the mean fingerprinted RSS measurement at each ANP.

$$d_{NNSS}\left(\mathbf{r}_t, \mathbf{F}(\mathbf{x}_{ap_{(i)}})\right) = \|\mathbf{r}_t - \bar{\mathbf{r}}_i\|^2 \qquad 1 \leq i \leq N \tag{3.3}$$

where $\bar{\mathbf{r}}_i = \frac{1}{n}\sum_{\tau=1}^{n}\mathbf{r}_i(\tau)$ are the vectors of the average of fingerprint RSS measurements, for each AP, at $\mathbf{x}_{ap_{(i)}}$. The location estimate is obtained from the average of the K selected ANPs with known locations as

$$\hat{\mathbf{x}} = \frac{1}{K}\sum_{k=1}^{K}\mathbf{x}_{ap_{(k)}} \tag{3.4}$$

The number of nearest-neighbours K chosen for location estimation must be investigated. If $K = 1$, complete trust is given to a single anchor point within the radio map that has the smallest Euclidean distance in signal space. This can lead to low location accuracy if, in fact, multiple anchor points throughout the environment have comparable Euclidean distances in signal space and are not contributing to the overall weighted average. If $K \gg 1$, the weighted average of K ANP coordinates could contain ANPs with potentially low Euclidean distance in signal space, thus lowering location accuracy. The effect of high $K \gg 1$ is visualized in Fig. 3.12, where if an asset were located in the top right ANP and several K are used within the average, the estimated location would be skewed to an ANP many meters away due to the averaging of the nearest-neighbours locations even if the RSS measurements were accurate to the top right ANP.

3.5.1.2   K-Weighted Nearest Neighbors in Signal Space

An extension to the NNSS technique outlined in Section 2.2.1, known as WNNSS, assigns a weight to each ANP ordered with respect to increasing distance in signal space.

$$\hat{\mathbf{x}} = \sum_{k=1}^{K} \frac{\beta_j}{\sum_{j=1}^{K} \beta_j} \mathbf{x}_{ap_{(k)}} \tag{3.5}$$

where

$$\beta_j = \frac{1}{\epsilon + d_{NNSS}\left(\mathbf{r}_t, \mathbf{F}(\mathbf{x}_{ap_{(j)}})\right)} \tag{3.6}$$

and ϵ is a small number used to avoid division by zero. This approach will elevate some of the inaccuracy in location estimate due to the number of nearest-neighbours chosen since it is a weighted average of the nearest-neighbour locations. WNNSS has

been reported to increase accuracy in positioning error by 2% over NNSS [47].

3.5.2 Probabilistic Range-free Location Estimation

This approach to location estimation will be formulated based on a MSE criterion utilizing KDE techniques as presented in [23, 19, 20, 22]. The MSE location estimator is the minimization of the l_2 norm of the location error stated as

$$\hat{\mathbf{x}} = \arg \min_{\hat{\mathbf{x}}} \mathbb{E} \left\{ (\mathbf{x} - \hat{\mathbf{x}})^T (\mathbf{x} - \hat{\mathbf{x}}) \right\} \tag{3.7}$$

where the expected value of the MMSE location estimate is the conditional mean of $\mathbf{x}$ given the RSS measurements [48] such that

$$\hat{\mathbf{x}} = \mathbb{E} \left\{ \mathbf{x} | \mathbf{r}_t \right\} = \int_\rho \mathbf{x} f \left(\mathbf{x} | \mathbf{r}_t \right) d\mathbf{x} \tag{3.8}$$

The location estimation problem can be reduced to the challenge of determining the posterior conditional Probability Density Function (PDF) $f \left(\mathbf{x} | \mathbf{r}_t \right)$ where prior information about the asset location can be extracted from the radio map. With our focus on indoor location estimation in the presence of a combination of LOS and non-LOS conditions, there is no simple parametric model for approximating the density $f \left(\mathbf{x} | \mathbf{r}_t \right)$ and as such, nonparametric techniques must be adopted for the density estimation of $f \left(\mathbf{x} | \mathbf{r}_t \right)$ from the underlying structure of the radio map RSS measurements without any prior assumptions.

3.5.2.1 Multivariate Joint Kernel Density Estimation

From Bayes theorem,

$$f\left(\mathbf{x}|\mathbf{r}_t\right) = \frac{f\left(\mathbf{r}_t|\mathbf{x}\right) f\left(\mathbf{x}\right)}{\int_\rho f\left(\mathbf{r}_t|\mathbf{x}\right) f\left(\mathbf{x}\right) \mathrm{d}\mathbf{x}} \tag{3.9}$$

the likelihood $f\left(\mathbf{r}_t|\mathbf{x}\right)$ provides relevant information about the asset location based on the observed RSS measurements $\mathbf{r}_t$. The prior $f\left(\mathbf{x}\right)$ is the belief in the asset location before any observed RSS measurements. Direct approximations of the prior and likelihood densities are needed to estimate the posterior density $f(\mathbf{x}|\mathbf{r}_t)$ and since no explicit equation is available to determine the prior, this density is estimated from the underlying structure of the radio map RSS measurements associated with each selected ANP through the use of the joint density $f(\mathbf{r}_t, \mathbf{x})$ [48]. The posterior density estimate can be formulated using the definition of conditional density

$$f(\mathbf{x}|\mathbf{r}_t) = \frac{f(\mathbf{r}_t, \mathbf{x})}{\int_\rho f(\mathbf{r}_t, \mathbf{x})\mathrm{d}\mathbf{x}} \tag{3.10}$$

Given the fingerprint training data, the joint density $f(\mathbf{r}_t, \mathbf{x})$ can be estimated through the multiplication of weighted kernel functions, or marginal densities, summed over each ANP used in positioning. The product kernel density estimator [38] is given as

$$f(\mathbf{r}_t, \mathbf{x}) \approx \frac{1}{N\sigma_{\mathbf{x}_{ap_i}}\sigma_{\mathbf{r}_{ap_i}}} \sum_{i=1}^{N} K\left(\frac{\mathbf{r}_t - \bar{\mathbf{r}}_{ap_i}}{\sigma_{\mathbf{r}_{ap_i}}}\right) K\left(\frac{\mathbf{x} - \mathbf{x}_{ap_i}}{\sigma_{\mathbf{x}_{ap_i}}}\right) \tag{3.11}$$

where $\{(\mathbf{x}_{ap_i}, \bar{\mathbf{r}}_{ap_i})|i = 1, \ldots, N\}$ is the mean of the fingerprint RSS measurements at each selected ANP and $\left\{\sigma_{\mathbf{x}_{ap_i}}, \sigma_{\mathbf{r}_{ap_i}}\right\}$ are the kernel bandwidths. The choice of kernel function was investigated in[23, 19, 20] that showed that the shape of the kernel did not lead to drastic variations in location estimation. The consensus was to implement

the Gaussian kernel due to its ease of mathematical manipulation. In general, for KDE a recommended multivariate, or d-dimensional, kernel function should be unimodal, smooth, and adhere to the following conditions [38, 22]

- $K\left(\mathbf{r}_t\right) \geq 0, \forall\ \mathbf{r}_t \in\ \mathbb{R}^d$
- $\int_{\mathbb{R}^d} K\left(\mathbf{r}_t\right) = 1$
- $\int_{\mathbb{R}^d} \mathbf{r}_t K\left(\mathbf{r}_t\right) = 0$
- $\int_{\mathbb{R}^d} \mathbf{r}_t \mathbf{r}_t^{\mathrm{T}} K\left(\mathbf{r}_t\right) = \mathbf{I}_d$

This thesis will focus on the Gaussian kernel, therefore the multivariate product KDE becomes

$$f(\mathbf{r}_t, \mathbf{x}) \approx \frac{1}{N} \sum_{i=1}^{N} \mathcal{N}\left(\mathbf{r}_t; \bar{\mathbf{r}}_{ap_i}, \boldsymbol{\sigma}_{\mathbf{r}_{ap_i}}\right) \mathcal{N}\left(\mathbf{x}; \mathbf{x}_{ap_i}, \boldsymbol{\sigma}_{\mathbf{x}_{ap_i}}\right) \tag{3.12}$$

where $\left\{\boldsymbol{\sigma}_{\mathbf{x}_{ap_i}}, \boldsymbol{\sigma}_{\mathbf{r}_{ap_i}}\right\} = \left\{\sigma^*_{\mathbf{x}_{ap_i}} \mathbf{I}_d, \sigma^*_{\mathbf{r}_{ap_i}} \mathbf{I}_d\right\}$ and $\mathcal{N}$ is a Gaussian kernel defined as

$$\mathcal{N}\left(\mathbf{r}_t; \bar{\mathbf{r}}_{ap_i}, \boldsymbol{\sigma}_{\mathbf{r}_{ap_i}}\right) \equiv \frac{1}{(2\pi)^{d/2} \left|\boldsymbol{\sigma}_{\mathbf{r}_{ap_i}}\right|^{1/2}} \exp\left(-\frac{1}{2}\left(\mathbf{r}_t - \bar{\mathbf{r}}_{ap_i}\right)^{\mathrm{T}} \boldsymbol{\sigma}_{\mathbf{r}}^{-1}\left(\mathbf{r}_t - \bar{\mathbf{r}}_{ap_i}\right)\right) \tag{3.13}$$

Given that RSS measurements have been observed, the denominator, or normalizing factor of the posterior density can be rewritten as

$$f(\mathbf{x}|\mathbf{r}_t) = \frac{f(\mathbf{r}_t, \mathbf{x})}{\int_\rho f(\mathbf{r}_t, \mathbf{x})\mathrm{d}\mathbf{x}} = \frac{f(\mathbf{r}_t, \mathbf{x})}{f(\mathbf{r}_t)} \tag{3.14}$$

and substituting the multivariate product kernel density estimate into the above posterior density leads to

$$f(\mathbf{x}|\mathbf{r}_t) \approx \frac{\sum_{i=1}^{N} \mathcal{N}\left(\mathbf{r}_t; \bar{\mathbf{r}}_{ap_i}, \boldsymbol{\sigma}_{\mathbf{r}_{ap_i}}\right) \mathcal{N}\left(\mathbf{x}; \mathbf{x}_{ap_i}, \boldsymbol{\sigma}_{\mathbf{x}_{ap_i}}\right)}{\sum_{i=1}^{N} \mathcal{N}\left(\mathbf{r}_t; \bar{\mathbf{r}}_{ap_i}, \boldsymbol{\sigma}_{\mathbf{r}_{ap_i}}\right)} \tag{3.15}$$

$$= \sum_{i=1}^{N} \omega_i(\mathbf{r}_t) \mathcal{N}\left(\mathbf{x}; \mathbf{x}_{ap_i}, \boldsymbol{\sigma}_{\mathbf{x}_{ap_i}}\right) \tag{3.16}$$

where

$$\omega_i(\mathbf{r}_t) = \frac{\mathcal{N}\left(\mathbf{r}_t; \bar{\mathbf{r}}_{ap_i}, \boldsymbol{\sigma}_{\mathbf{r}_{ap_i}}\right)}{\sum_{i=1}^{N} \mathcal{N}\left(\mathbf{r}_t; \bar{\mathbf{r}}_{ap_i}, \boldsymbol{\sigma}_{\mathbf{r}_{ap_i}}\right)} \tag{3.17}$$

The posterior density shown in Eq. 3.16 follows the form of a Gaussian mixture and as such, the MMSE location estimate and its covariance can be approximated by a single Gaussian density with the mean and covariance. The estimated location and its covariance are often referred to as the first and second moments[48]. Referring back to the MMSE in Eq. 3.8

$$\hat{\mathbf{x}} = \mathbb{E}\left\{\mathbf{x}|\mathbf{r}_t\right\} = \int_{\rho} \mathbf{x} f(\mathbf{x}|\mathbf{r}_t) \mathrm{d}\mathbf{x} \tag{3.18}$$

and substituting in the posterior density estimate of Eq. 3.14,

$$\hat{\mathbf{x}} = \int_{\rho} \mathbf{x} \frac{f(\mathbf{r}_t, \mathbf{x})}{f(\mathbf{r}_t)} \mathrm{d}\mathbf{x} = \frac{\int_{\rho} \mathbf{x} f(\mathbf{r}_t, \mathbf{x}) \mathrm{d}\mathbf{x}}{f(\mathbf{r}_t)} \tag{3.19}$$

where the denominator becomes

$$f(\mathbf{r}_t) = \int_{\rho} f(\mathbf{r}_t, \mathbf{x}) \mathrm{d}\mathbf{x} \tag{3.20}$$

$$= \int_{\rho} \frac{1}{N} \sum_{i=1}^{N} \mathcal{N}\left(\mathbf{r}_t; \bar{\mathbf{r}}_{ap_i}, \boldsymbol{\sigma}_{\mathbf{r}_{ap_i}}\right) \mathcal{N}\left(\mathbf{x}; \mathbf{x}_{ap_i}, \boldsymbol{\sigma}_{\mathbf{x}_{ap_i}}\right) \mathrm{d}\mathbf{x} \tag{3.21}$$

$$= \frac{1}{N} \sum_{i=1}^{N} \mathcal{N}\left(\mathbf{r}_t; \bar{\mathbf{r}}_{ap_i}, \boldsymbol{\sigma}_{\mathbf{r}_{ap_i}}\right) \left(\int_{\rho} \mathcal{N}\left(\mathbf{x}; \mathbf{x}_{ap_i}, \boldsymbol{\sigma}_{\mathbf{x}_{ap_i}}\right) \mathrm{d}\mathbf{x}\right) \tag{3.22}$$

$$= \frac{1}{N} \sum_{i=1}^{N} \mathcal{N}\left(\mathbf{r}_t; \bar{\mathbf{r}}_{ap_i}, \boldsymbol{\sigma}_{\mathbf{r}_{ap_i}}\right) \tag{3.23}$$

since $\int K(\mathbf{z})\mathrm{d}\mathbf{z} = 1$ assuming a zero-mean, non-negative kernel with unit area. The final position estimate becomes

$$\hat{\mathbf{x}} = \mathbb{E}\left\{\mathbf{x}|\mathbf{r}_t\right\} = \sum_{i=1}^{N} \omega_i(\mathbf{r}_t, \mathbf{F}(\mathbf{x}_{ap_i}))\mathbf{x}_{ap_i} \tag{3.24}$$

where $\omega_i(\mathbf{r}_t, \mathbf{F}(\mathbf{x}_{ap_i}))$ is defined by Eq. 3.17 using the information provided from each radio map fingerprint $\mathbf{F}(\mathbf{x}_{ap_i})$. A similar derivation [48] can be done to formulate the associated covariance

$$\mathbf{P} = \mathbb{E}\left\{(\mathbf{x} - \hat{\mathbf{x}})(\mathbf{x} - \hat{\mathbf{x}})^{\mathrm{T}}|\mathbf{r}_t\right\} \tag{3.25}$$

$$= \sum_{i=1}^{N} \mathbb{E}\left\{(\mathbf{x} - \hat{\mathbf{x}})(\mathbf{x} - \hat{\mathbf{x}})^{\mathrm{T}}|\mathbf{r}_t\right\} \omega_i(\mathbf{r}_t, \mathbf{F}(\mathbf{x}_{ap_i})) \tag{3.26}$$

$$= \sum_{i=1}^{N} \mathbb{E}\left\{(\mathbf{x} - \mathbf{x}_i + \mathbf{x}_i - \hat{\mathbf{x}})(\mathbf{x} - \mathbf{x}_i + \mathbf{x}_i - \hat{\mathbf{x}})^{\mathrm{T}}|\mathbf{r}_t\right\} \omega_i(\mathbf{r}_t, \mathbf{F}(\mathbf{x}_{ap_i})) \tag{3.27}$$

$$= \sum_{i=1}^{N} \omega_i(\mathbf{r}_t, \mathbf{F}(\mathbf{x}_{ap_i}))\mathbf{C}_i + \sum_{i=1}^{N} \omega_i(\mathbf{r}_t, \mathbf{F}(\mathbf{x}_{ap_i}))(\mathbf{x}_i - \hat{\mathbf{x}})(\mathbf{x}_i - \hat{\mathbf{x}})^{\mathrm{T}} \tag{3.28}$$

$$= \sum_{i=1}^{N} \omega_i(\mathbf{r}_t, \mathbf{F}(\mathbf{x}_{ap_i})) \left[\mathbf{C}_i + (\mathbf{x}_i - \hat{\mathbf{x}})^{\mathrm{T}}\right] \tag{3.29}$$

$$= \left[\sum_{i=1}^{N} \omega_i(\mathbf{r}_t, \mathbf{F}(\mathbf{x}_{ap_i})) \left(\mathbf{C}_i + \mathbf{x}_i\mathbf{x}_i^{\mathrm{T}}\right)\right] + \hat{\mathbf{x}}\hat{\mathbf{x}}^{\mathrm{T}} \tag{3.30}$$

where $\mathbf{C}_i = \mathbb{E}\left\{(\mathbf{x} - \mathbf{x}_i)(\mathbf{x} - \mathbf{x}_i)^{\mathrm{T}}|\mathbf{r}_t\right\}$ is the variance of the ANP positions determined during the fingerprint site survey.

It is important to note, as previously mentioned for the histogram density estimate outline in Section 2.2.1.2, that the multivariate product kernel density estimate requires the assumption that the fingerprint RSS measurements are IID. This assumption is violated due to correlations in RSS measurements collected over time,

however, the density estimate has been shown to maintain high performance with the proper kernel bandwidth tuning [22].

3.5.2.2   Kernel Bandwidths

Also known as the window width, these bandwidths control the width of the kernel or the region of influence of each training sample[38, 39]. Let's consider the two extreme cases:

- $\sigma_{\mathbf{r}_{ap_i}} \to 0$

  The KDE becomes a train of delta functions centered at each fingerprint RSS measurement and risks significantly over-fitting the data.

- $\sigma_{\mathbf{r}_{ap_i}} \to \infty$

  The KDE approaches a Uniform distribution that captures the global structure of the data and risks significantly under-fitting the data.

Therefore, it is important to determine the dependence of this parameter on each set of fingerprint RSS measurements. It has been shown in the literature [38, 39] that for a Gaussian kernel the optimal bandwidth in terms of minimizing the asymptotic mean integrated error between the estimated and true densities is given by

$$\sigma_{\mathbf{r}_{ap_i}}^* = \left(\frac{4}{d+2}\right)^{\frac{1}{d+4}} \hat{\sigma}_{\mathbf{r}_{ap_i}} n^{\frac{-1}{d+4}} \tag{3.31}$$

where d is the number of APs used, n is the number of fingerprint RSS measurements, and $\hat{\sigma}_{\mathbf{r}_{ap_i}}^2 = 1/d \sum_{l=1}^{d} \left(\hat{\sigma}_{\mathbf{r}_{ap_i}}^2\right)^2$ is the average of the estimates of the marginal variances for each AP at $\mathbf{x}_i$. Therefore $\boldsymbol{\sigma}_{\mathbf{r}_{ap_i}} = \sigma_{\mathbf{r}_{ap_i}}^* \mathbf{I}_d$. The primary disadvantage of this optimal bandwidth selector is that a single value is used to apply the same bandwidth to each set of fingerprint RSS measurements, for each AP, at each ANP.

To account for the variation of the fingerprint RSS measurements from a Gaussian distribution, a smoothing factor, s, is introduced to scale the optimal bandwidth. The *smoothed* optimal kernel bandwidth is then used for location estimation

$$s = \frac{\tilde{\sigma}^*_{\mathbf{r}_{ap_i}}}{\sigma^*_{\mathbf{r}_{ap_i}}} \tag{3.32}$$

where $\tilde{\sigma}^*_{\mathbf{r}_{ap_i}}$ is smoothed optimal kernel bandwidth used in estimation. A basic example of how the smoothing factor selection, applied to the optimal kernel bandwidth, affects KDE for a single AP ($d = 1$) is shown below in Fig. 3.15.
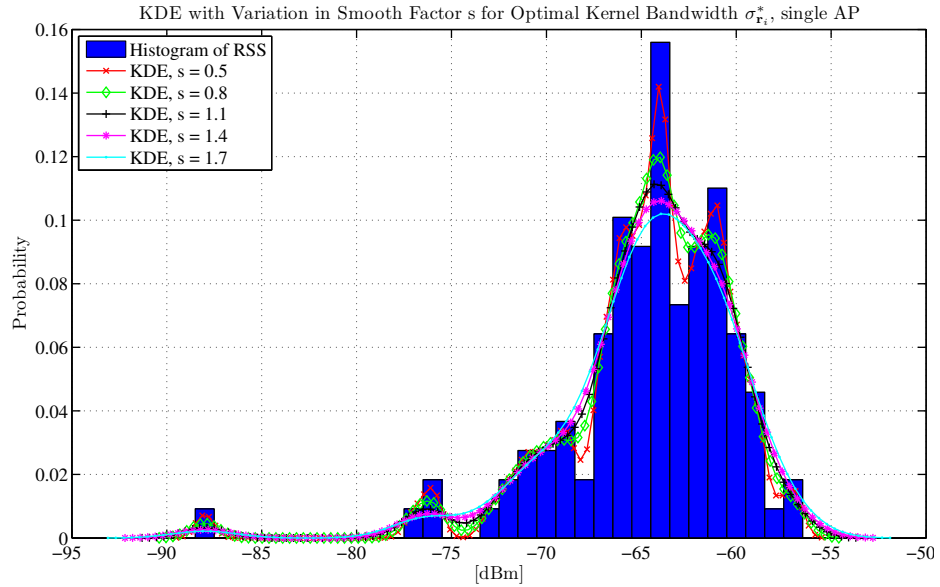
Figure 3.15: KDE of RSS Measurements from Single AP with Varying Smooth factor

3.6 Radio Map Annealing Via Dynamic Asset Traces

The time consuming and intrusive nature of location fingerprinting an environment is one of the most important disadvantages of range-free indoor location estimation techniques. There are two main components of a range-free system that can be adjusted in order to alleviate this problem. The first is the amount of time spent collecting data at each ANP. This directly translates into the number of fingerprinted RSS measurements, n, from each ANP used in the localization algorithm. The second is a reduction in the number of fingerprinted ANPs which can lead to an incomplete radio map.

In order to dynamically anneal an initial radio map with unknown asset traces,
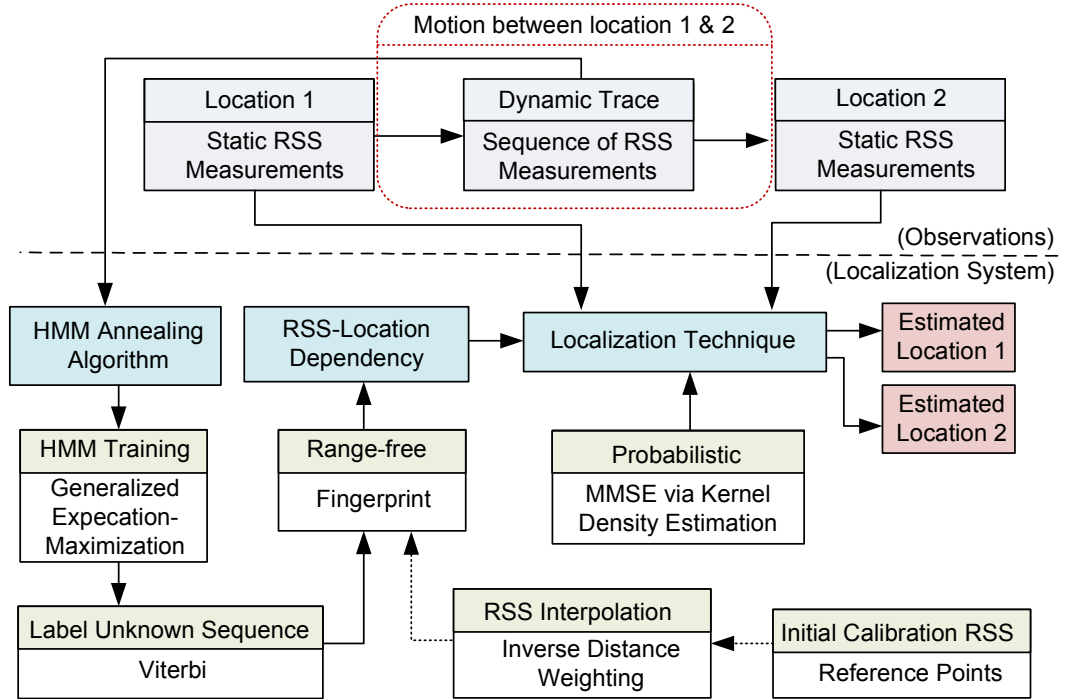
Figure 3.16: Localization System with Dynamic Radio Map Annealing

one must first interpolate the skipped ANPs to complete the radio map. A HMM finite state machine, combined with a GEM algorithm, trains the HMM model parameters with multiple unknown asset traces to anneal the initial radio map over time. A block diagram of the overall system, with the implementation of radio map annealing for a single unknown asset trace, can be seen in Fig. 3.16. The system is divided into two sections: observations and localization system. First, initial calibration RSS measurements are collected at sparse locations throughout the environment where an interpolation algorithm is used to complete the radio map for each ANP in the system. This portion of the localization system is indicated as a dashed line, in the lower right hand section of Fig. 3.16, since it will only be utilized once prior to location estimation. Second, the static RSS measurements from the first unknown asset location are used for MMSE location estimation. Third, an unknown asset trace, used by the HMM annealing algorithm, is shown as the sequence of online RSS measurements collected during the dynamic motion of the asset between static locations 1 and 2. The sequence of RSS measurements is used to train the HMM model parameters through the use of a GEM. Third, the most probable sequence of states, based on the observed sequence of RSS measurements, is determined via a Viterbi algorithm. The newly labeled sequence of RSS measurements are added into the radio map at their respective ANP locations. Finally, the annealed radio map and the static RSS measurements from the second unknown asset location are used for MMSE location estimation.

The stock WiFly® RN-134 802.11b modules did not support a feature for the detection of dynamic motion, therefore it was necessary to develop additional circuitry, the accelerometer daughter board outline in Appendix A, to allow for the

determination between static and dynamic online RSS measurements.

3.6.1 Interpolation of Skipped Anchor Points

To reduce the number of fingerprinted locations and ensure the system still has a complete radio map, we interpolate the fingerprint RSS measurements to each skipped ANP while keeping the measurement acquisition as translucent as possible to the normal operations within an environment. To try to alleviate the intrusiveness of using a turntable to calibrate ANPs, a second method was developed that utilized the designed RPs placed along the length of each hallway wall. Pairs of RPs located directly across from one another were used to interpolate RSS to each ANP location. The reduction in the total number of RPs used for interpolation was also investigated. In the first approach, the number of fingerprinted ANPs has been reduced to every third ANP as shown in Fig. 3.17. To interpolate the RSS of the skipped ANPs, the two closest fingerprinted ANPs are used. It can be seen that the contribution from each fingerprinted ANP will depend solely on the distance between it and each of the skipped ANPs. In the second approach, the number of RPs has also been reduced to every third ANP spacing as shown in Fig. 3.18. It can be seen that the RPs spaced directly in line with an ANP are used solely to interpolate that specific ANP. The interpolation of the off-spaced ANPs will require the combined contributions from each of its four closest RPs.

3.6.1.1 Inverse Distance Weighting for RSS

An appropriate algorithm is needed in order to interpolate the RSS measurements of the skipped ANPs. We have chosen IDW as our method of interpolation which is
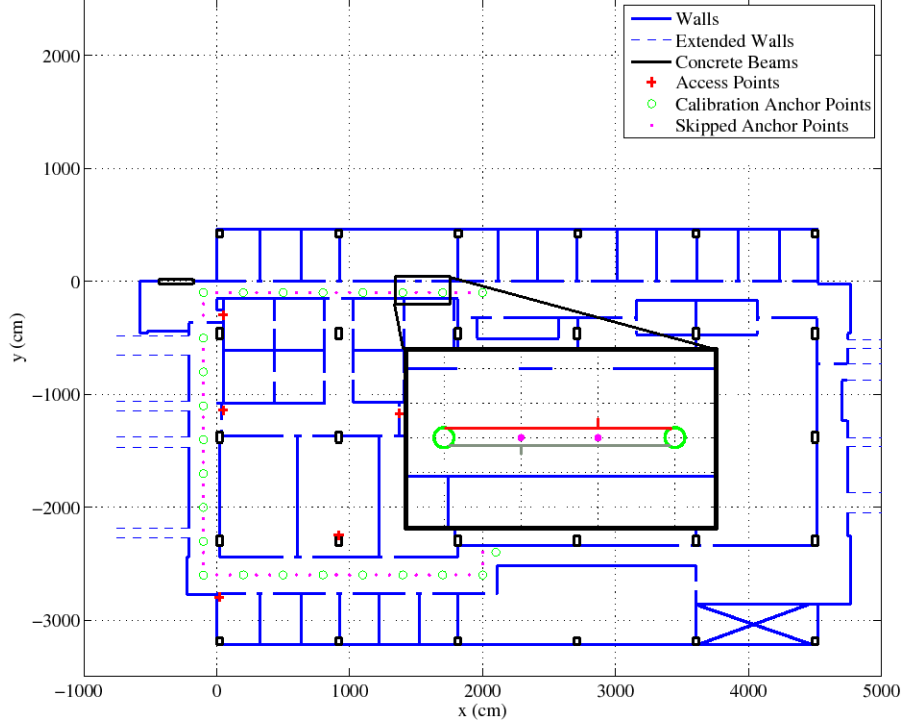
Figure 3.17: Example Interpolation from Calibration Anchor Points

typically used to interpolate data at sparse locations within dense grids[49]. It has been previously implemented for the interpolation of outdoor radio maps[25]. This elegant, yet simple approach weights the RSS measurements from the RPs, the known scattered locations, for each AP by the inverse of its difference in Euclidean distance to each skipped ANP. To find a interpolated RSS value, $\mu_{ap\{RSS\},l}$, at a skipped ANP, $\mathbf{x}_{ap}$, based on the RSS samples $\mu\left(\mathbf{x}_{rp_i}\right)_{\{RSS\},l}$ for $i = 1, \ldots, N$ is given as

$$\mu_{ap}(\mathbf{x}_{ap})_{\{RSS\},l} = \sum_{i=1}^{N} \frac{\alpha_i(\mathbf{x}_{ap})}{\sum_{i=1}^{N} \alpha_i(\mathbf{x}_{ap})} \mu\left(\mathbf{x}_{rp_i}\right)_{\{RSS\},l} \tag{3.33}$$

where $l = 1, \ldots, d$ is the number of selected APs and
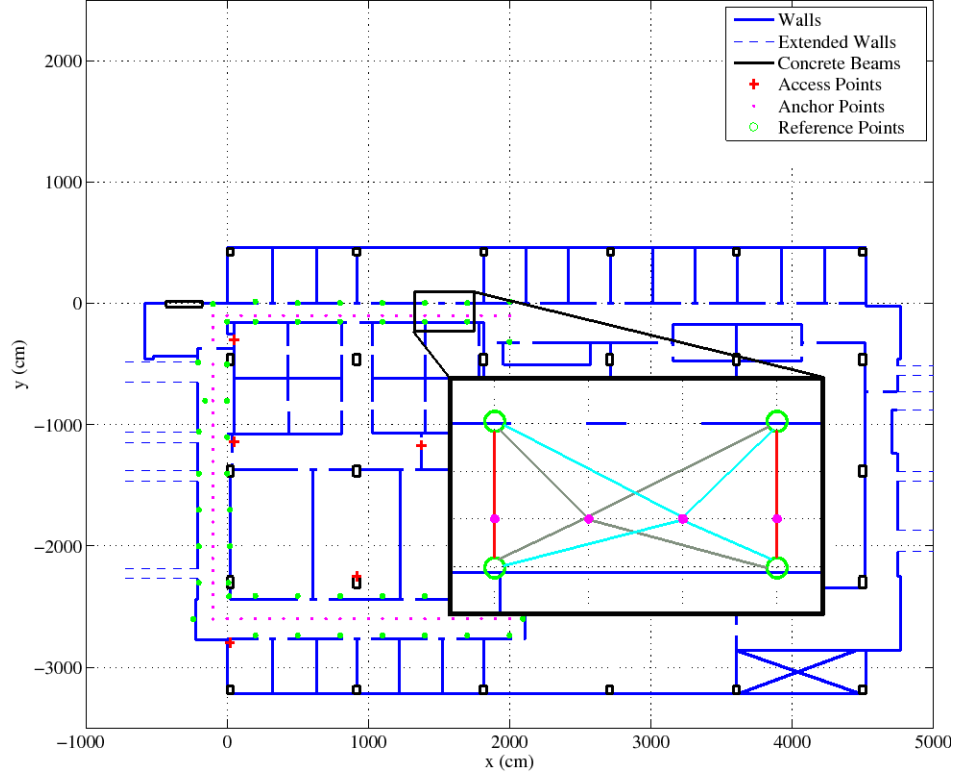
Figure 3.18: Example Interpolation from Calibration Reference Points

$$\alpha_i(\mathbf{x}_{ap}) = \frac{1}{d(\mathbf{x}_{ap}, \mathbf{x}_{rp_i})^\rho} = \frac{1}{\|\mathbf{x}_{ap} - \mathbf{x}_{rp_i}\|^\rho} \tag{3.34}$$

The exponent ρ is termed the power parameter and controls the weight of each RP on the current ANP in the interpolation. It follows that for RSS measurement IDW, the RPs closest to the interpolated ANP should be influenced most, with distant RPs contributing less, hence $\rho \geq 2$ for 2-d interpolation. Ideally, setting $\rho = 2$ will result in Eq. 3.34 becoming the inverse of the Euclidean distance between each RP and the interpolated ANP and has been shown to lead to satisfactory empirical results with low computational complexity [49].

3.6.2 Hidden Markov Model for Unknown Asset Traces

A HMM is a stochastic finite state machine that can model discrete Markov processes with a set of parameters. The term hidden refers to the fact that the underlying stochastic process is not observable and can only be observed through a second set of stochastic processes that produce observation sequences [50, 51]. The HMM had its earliest known success in the field of speech recognition [51]. The total number of available states within the HMM, N, also equals the number of ANP used in the location estimation system. This set of states will be referred to as $S = \{s_1, \ s_2, \dots \ s_N\}$. We will denote the hidden state at any online time t as q_t with a particular asset trace sequence of hidden states of length T denoted by $Q = [q_1, \ q_2, \dots \ q_T]$ where $q_t \in \{s_1, \ s_2, \dots \ s_N\}$. The collection of observed RSS measurements from d APs corresponding to the sequence of hidden states will be denoted as $R = [\mathbf{r}_1, \ \mathbf{r}_2, \dots \ \mathbf{r}_T]$. There are two conditional independence assumptions[52] that need to be made in order to move forward with any HMM algorithm and are outlined as follows.

1. For any asset trace sequence of hidden states, Q, the probability of a certain hidden state q_t at time t only depends on the hidden state q_{t-1} at time $t - 1$. This is known as the first-order Markov assumption.

$$P\left(q_t|q_{t-1}, q_{t-2}, \dots, q_1\right) = P\left(q_t|q_{t-1}\right) \tag{3.35}$$

2. Any RSS measurement, $\mathbf{r}_t$, observed from a hidden state, q_t, is independent of any other state.

$$P\left(\mathbf{r}_t | q_T, \mathbf{r}_T, q_{T-1}, \mathbf{r}_{T-1}, \ldots, q_{t+1}, \mathbf{r}_{t+1}, q_t, \mathbf{r}_t, q_{t-1}, \mathbf{r}_{t-1}, \ldots, q_1, \mathbf{r}_1\right) = P\left(\mathbf{r}_t | q_1\right)$$

$$(3.36)$$

3.6.2.1 Parameter Overview

The production of any hidden asset trace sequence, Q, is described by a time independent stochastic transition matrix that governs how an asset may move throughout the environment subject to a certain set of constraints. In general, the transition probabilities do not necessarily have to be symmetric ($a_{i,j} \neq a_{j,i}$

$$\mathbf{A} = \{a_{i,j}\} = \{p\left(q_t = s_j | q_{t-1} = s_i\right)\} \qquad 1 \leq i \leq N$$
$$1 \leq j \leq N$$

$$(3.37)$$

An example of the transition probabilities for one state, or one column of $\mathbf{A}$, is given in Fig. 3.19 for two scenarios. A maximum distance constraint of 2 meters is set governing how far the asset can jump forward or backward in state space, and the associated probability of each throughout the state sequence. In Fig. 3.19(a), the probabilities follow a triangular distribution where an asset is more likely to transition to its closest neighboring states either forwards or backwards, with less probability that the asset will transition further. In Fig. 3.19(b), the probabilities follow a uniform distribution where an asset is equally likely of transitioning forwards or backwards to its closest neighboring states, or further states depending on the maximum distance constraint. An example of $\mathbf{A}$ representing only the state transition matrix distance
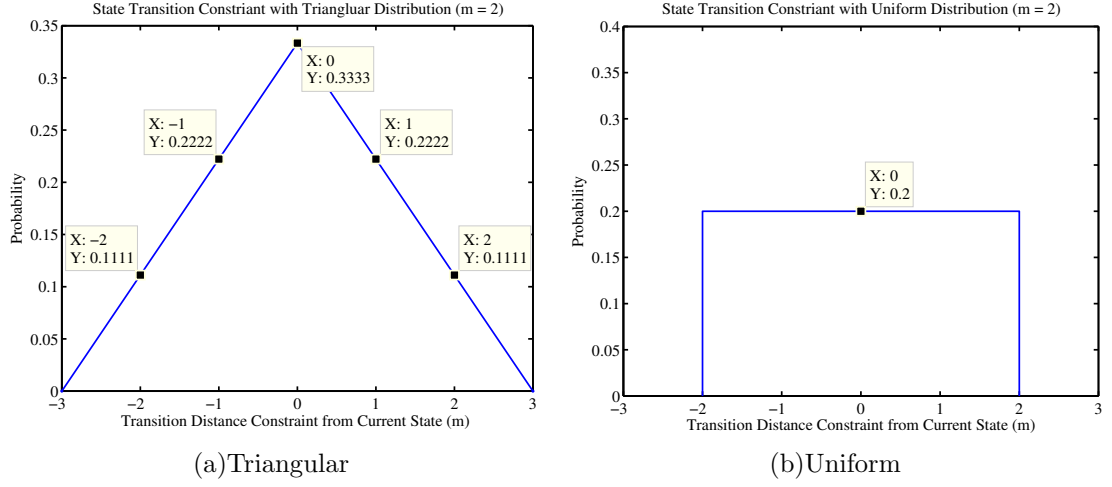
Figure 3.19: State Transition Matrix Distance Constraints

constraint shown in Fig. 3.19(a) can be seen in Eq. 3.38.

$$
\mathbf{A} = \begin{bmatrix}
\cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\
\vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\
\vdots & 0 & 0.11 & 0.22 & 0.33 & 0.22 & 0.11 & 0 & \cdots & \cdots & \vdots \\
\vdots & \cdots & 0 & 0.11 & 0.22 & 0.33 & 0.22 & 0.11 & 0 & \cdots & \vdots \\
\vdots & \cdots & \cdots & 0 & 0.11 & 0.22 & 0.33 & 0.22 & 0.11 & 0 & \vdots \\
\vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\
\cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots
\end{bmatrix}_{(N \times N)}
\tag{3.38}
$$

For time $t = 1$, the initial state probabilities are formulated to account for any prior knowledge about where an asset may start their trace sequence.

$$
\boldsymbol{\pi} = \{\pi_i\} = \{p(q_1 = s_i)\} \qquad 1 \leq i \leq N
\tag{3.39}
$$

The emission, or conditional, probabilities describe the likelihood of a certain observation at a particular time, $\mathbf{r}_t$, given that the HMM is in state s_i. For discrete observations, where $\mathbf{r}_t \in \{v_1,\ v_2, \ldots\ v_H\}$, the matrix form

$$\boldsymbol{\lambda} = \{\lambda_{i,h}\} = \{p\left(\mathbf{r}_t = v_h | q_t = s_i\right)\} \qquad 1 \le i \le N$$
$$1 \le h \le H \tag{3.40}$$

This is the discrete observation assumption that an observation is any one of H possible values, v_h, where $V = \{v_1,\ v_2, \ldots\ v_H\}$. For continuous observations, the probability densities are described using a set of PDFs for the observation space.

$$\boldsymbol{\lambda} = \{\lambda_i(\mathbf{r}_t)\} = \{p\left(\mathbf{r}_t | q_t = s_i\right)\} \qquad 1 \le i \le N \tag{3.41}$$

In this research, the conditional probabilities are calculated using the technique outlined in Section 3.5.2.1 where each $\mathbf{r}_t$ in the observation sequence R is used within the likelihood kernel density estimate to map its RSS-dependency against the fingerprint RSS measurements corresponding to each interpolated ANP in the set of all available states S. The conditional probabilities are also normalized to result in a total probability of 1 over all states. Assuming that any observed RSS measurement is an allowable value, $r_t = v_t$, then

$$\boldsymbol{\lambda} = \{p\left(\mathbf{r}_t = v_t | q_t = s_i\right)\} = \{p\left(\mathbf{r}_t | \mathbf{x}_i\right)\} \tag{3.42}$$
$$= \{\omega_i(\mathbf{r}_t)\} = \left\{ \frac{\mathcal{N}\left(\mathbf{r}_t; \bar{\mathbf{r}}_i, \boldsymbol{\sigma}_{\mathbf{r}}\right)}{\sum_{i=1}^{N} \mathcal{N}\left(\mathbf{r}_t; \bar{\mathbf{r}}_i, \boldsymbol{\sigma}_{\mathbf{r}}\right)} \right\} \qquad 1 \le i \le N \tag{3.43}$$

As a standard stochastic constraint, we must guarantee that a state transition occurs from $t \rightarrow t + 1$, therefore normalization conditions are enforced on each model parameter to ensure that the appropriate vector dimension has a probability summation equal to 1.

$$\sum_{j=1}^{N} a_{i,j} = 1 \qquad \sum_{h=1}^{H} \lambda_{i,h} = 1 \qquad \sum_{i=1}^{N} \pi_i = 1 \qquad (3.44)$$

An example of how the above assumptions are invoked on a HMM can be seen in Fig. 3.20. The white circles represent the hidden asset trace states as they progress through time with interconnecting location state transition probabilities. Their associated observations, RSS measurements, are denoted as grey circles with emission or likelihood probabilities. The initial location state probability at time $t = 1$ is also shown. The complete set of HMM model parameters outlined in Eq. 3.37 - 3.44 will be denoted as the tuple $\theta = (\mathbf{A}, \boldsymbol{\lambda}, \boldsymbol{\pi})$. The distributions of the above parameters will
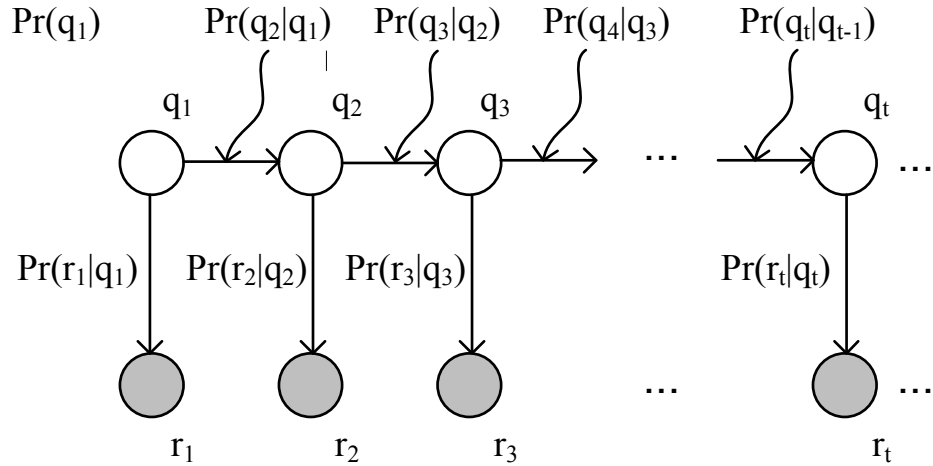
Figure 3.20: Example of First Order HMM State Transition Through Time

follow as

$$P\left(q_1|\boldsymbol{\pi}\right) = \prod_{i=1}^{N} \pi_i \tag{3.45}$$

$$P\left(q_t|q_{t-1}, \mathbf{A}\right) = \prod_{i=1}^{N}\prod_{j=1}^{N} a_{i,j} \tag{3.46}$$

$$P\left(\mathbf{r}_t|q_t, \boldsymbol{\lambda}\right) = \prod_{i=1}^{N}\prod_{h=1}^{H} \lambda_{i,h} \tag{3.47}$$

with the complete-data likelihood, or the joint distribution of R and Q given θ for a sequence of length T, given by

$$P\left(R, Q|\theta\right) = P\left(q_1|\boldsymbol{\pi}\right) \left[\prod_{t=2}^{T} P\left(q_t|q_{t-1}, \mathbf{A}\right)\right] \prod_{m=1}^{T} P\left(\mathbf{r}_m|q_m, \boldsymbol{\lambda}\right) \tag{3.48}$$

$$= \prod_{i=1}^{N} \pi_i \left[\prod_{t=2}^{T}\prod_{i=1}^{N}\prod_{j=1}^{N} a_{i,j}\right] \prod_{t=1}^{T}\prod_{i=1}^{N}\prod_{h=1}^{H} \lambda_{i,h} \tag{3.49}$$

3.6.2.2 Generalized Expectation-Maximization for Model Training

In the first setup to improve the initial interpolated radio map over time, unknown asset traces will be used to train the HMM model parameters via a GEM algorithm. Also known as Baum-Welch parameter estimation[52, 53, 54, 51, 55, 50, 56], this iterative technique improves an initial estimate of the model parameters, denoted as $\theta^1 = \left(\mathbf{A}^1, \boldsymbol{\lambda}^1, \boldsymbol{\pi}^1\right)$,to a set of model parameters θ^* that best explains $P(R|\theta)$, which is the likelihood of the observed RSS measurement sequence R given θ. The maximum likelihood problem becomes

$$\theta^* = \arg\max_{\theta} \log P(R|\theta) \tag{3.50}$$

Although the maximization should be over $P(R|\theta)$, it is equivalent to maximizing $\log P(R|\theta)$ since log is monotonic and it makes the GEM algorithm mathematically easier to handle. Furthermore, with a HMM the state sequence Q that would have produced R is unknown and must be incorporated into the maximization through $P(R, Q|\theta)$, which is the joint distribution of R and Q summed over all possible hidden state sequences.

$$\theta^* = \arg\max_\theta \log \sum_{Q \in S} P(R, Q|\theta) \tag{3.51}$$

This solidifies our understanding as to why $P(R|\theta)$ is often referred to as the incomplete data likelihood and $P(R, Q|\theta)$ the complete data likelihood.

Maximizing Eq. 3.51 would be difficult since it involves taking the log of a sum of variables. Therefore, we turn to GEM where the goal is to iteratively maximize an auxiliary function, $\mathcal{Q}$, through the expectation(E) and maximization(M) steps until convergence is reached at some local optimum θ^*. Only an improved $\theta^{(k+1)}$ is required to be set at each iteration k of the M step, not the optimal as required by standard EM.

$$\theta^{(k+1)} = \arg\max_\theta \mathcal{Q}\left(\theta, \theta^{(k)}\right) \tag{3.52}$$

where θ are the HMM parameters to be maximized, $\theta^{(k)}$ are the maximized parameters from the previous iteration, and the $\mathcal{Q}$ function is

$$\mathcal{Q}\left(\theta, \theta^{(k)}\right) = \mathbb{E}\left[\log P(R, Q|\theta) \,|\, R, \theta^{(k)}\right] \tag{3.53}$$

$$= \sum_{Q \in S} \log P(R, Q|\theta) \, P\left(Q|R, \theta^{(k)}\right) \tag{3.54}$$

$$= \sum_{Q \in S} \log P\left(R, Q | \theta\right) \frac{P\left(R, Q | \theta^{(k)}\right)}{P\left(R | \theta^{(k)}\right)} \tag{3.55}$$

and the term $P\left(R | Q, \theta^{(k)}\right)$ is the posterior distribution of the latent state sequence Q, given R and $\theta^{(k)}$, and is evaluated during the E-step which will be discussed later. The $\mathcal{Q}$ function can be reformulated in terms of the joint probability, $P\left(R, Q | \theta\right)$, using Bayes' theorem as shown in Eq. 3.55. An explanation as to how GEM maximizes the likelihood $P\left(R | \theta\right)$ through the iterative maximization of the $\mathcal{Q}$ function is arduous and beyond the scope of this thesis. Interested readers are encouraged to read the detailed formulation provided by [57]. It can be guaranteed that if we choose $\theta^{(k+1)}$ so that $\mathcal{Q}\left(\theta, \theta^{(k+1)}\right) \geq \mathcal{Q}\left(\theta, \theta^{(k)}\right)$, then $\log P\left(R | \theta^{(k+1)}\right) \geq \log P\left(R | \theta^{(k)}\right)$.

M-Step

To make the HMM parameter maximization manageable, the $\mathcal{Q}$ function can be rewritten as the sum of each parameter allowing for the maximization to be performed individually. With knowledge of the joint distribution for the current iteration, as given in Eq. 3.49, the $\mathcal{Q}$ function expands to

$$\mathcal{Q}\left(\theta, \theta^{(k)}\right) = \sum_{Q \in S} \log P\left(q_1 | \boldsymbol{\pi}\right) \frac{P\left(R, Q | \theta^{(k)}\right)}{P\left(R | \theta^{(k)}\right)}$$

$$+ \sum_{Q \in S} \left(\sum_{t=2}^{T} \log P\left(q_t | q_{t-1}, \mathbf{A}\right)\right) \frac{P\left(R, Q | \theta^{(k)}\right)}{P\left(R | \theta^{(k)}\right)}$$

$$+ \sum_{Q \in S} \left(\sum_{t=1}^{T} \log P\left(\mathbf{r}_t | q_t, \boldsymbol{\lambda}\right)\right) \frac{P\left(R, Q | \theta^{(k)}\right)}{P\left(R | \theta^{(k)}\right)} \tag{3.56}$$

The first term in Eq. 3.56 can be evaluated further since by selecting all $Q \in S$, the values of q_0 are just repeatedly selected, or the marginal probabilities for $t = 1$.

$$\sum_{Q \in S} \log P(q_1|\boldsymbol{\pi}) \frac{P\left(R, Q|\theta^{(k)}\right)}{P\left(R|\theta^{(k)}\right)} = \sum_{i=1}^{N} \log \pi_i^{q_{1,i}} \frac{P\left(R, q_1 = s_i|\theta^{(k)}\right)}{P\left(R|\theta^{(k)}\right)} \qquad (3.57)$$

The second term in Eq. 3.56 will become the sum of the joint-marginal probabilities for $t-1$ and t since for each t we are progressing over all state transitions from i to j and weighting by the corresponding probability.

$$\sum_{Q \in S} \left(\sum_{t=2}^{T} \log P(q_t|q_{t-1}, \mathbf{A}) \right) \frac{P\left(R, Q|\theta^{(k)}\right)}{P\left(R|\theta^{(k)}\right)} =$$
$$\sum_{i=1}^{N} \sum_{j=1}^{N} \sum_{t=2}^{T} \log a_{i,j}^{q_{t,j} q_{t-1,i}} \frac{P\left(R, q_{t-1} = s_i, q_t = s_j|\theta^{(k)}\right)}{P\left(R|\theta^{(k)}\right)} \qquad (3.58)$$

The third term in Eq. 3.56 will become the sum of the marginal probabilities for t since for each t we are progressing over the emission probabilities for all states and weighting each possible emission by the corresponding probability

$$\sum_{Q \in S} \left(\sum_{t=1}^{T} \log P(\mathbf{r}_t|q_t, \boldsymbol{\lambda}) \right) \frac{P\left(R, Q|\theta^{(k)}\right)}{P\left(R|\theta^{(k)}\right)} = \sum_{i=1}^{N} \sum_{t=1}^{T} \sum_{h=1}^{H} \log \lambda_{i,h}^{q_{t,i} \mathbf{r}_{t,h}} \frac{P\left(R, q_t = s_i|\theta^{(k)}\right)}{P\left(R|\theta^{(k)}\right)}$$
$$(3.59)$$

Following the procedure detailed in Appendix B, the technique of Lagrange multipliers can be used to define the model parameters that will iteratively maximize $P(R, Q|\theta^*)$.

$$\pi_i^{(k+1)} = \frac{P\left(R, q_1 = s_i|\theta^{(k)}\right)}{P\left(R|\theta^{(k)}\right)} = P\left(q_1 = s_i|R, \theta^{(k)}\right) \qquad (3.60)$$

$$a_{i,j}^{(k+1)} = \frac{\sum_{t=2}^{T} P\left(R, q_{t-1} = s_i, q_t = s_j | \theta^{(k)}\right)}{\sum_{t=2}^{T} P\left(R, q_{t-1} = s_i | \theta^{(k)}\right)} = \frac{\sum_{t=2}^{T} P\left(q_{t-1} = s_i, q_t = s_j | R, \theta^{(k)}\right)}{\sum_{t=2}^{T} P\left(q_t = s_i | R, \theta^{(k)}\right)}$$

$$\tag{3.61}$$

$$\lambda_{i,h}^{(k+1)} = \frac{\sum_{t=1}^{T} P\left(R, q_t = s_i | \theta^{(k)}\right) \delta_{\mathbf{r}_t, v_h}}{\sum_{t=1}^{T} P\left(R, q_t = s_i | \theta^{(k)}\right)} = \frac{\sum_{t=1}^{T} P\left(q_t = s_i | R, \theta^{(k)}\right) \delta_{\mathbf{r}_t, v_h}}{\sum_{t=1}^{T} P\left(q_t = s_i | R, \theta^{(k)}\right)} \tag{3.62}$$

E-Step

At each iteration in the maximization of $\mathcal{Q}$, the posterior and joint-posterior probabilities

$$P\left(q_t = s_i | R, \theta^{(k)}\right) \tag{3.63}$$

$$P\left(q_{t-1} = s_i, q_t = s_j | R, \theta^{(k)}\right) \tag{3.64}$$

required for the M-Step are evaluated efficiently using a two-stage technique known as the forward-backward, or Baum-Welch, algorithm. The forward partial probability, denoted as $\alpha_i(t)$, is the joint probability of observing the first $\mathbf{r}_\tau$, where $\tau = 1, 2, \ldots, t$ and being in state q_i at time t.

$$\alpha_i(t) = P\left(\mathbf{r}_1, \mathbf{r}_2, \ldots, \mathbf{r}_t, q_t = s_i | \theta^{(k)}\right) \tag{3.65}$$

Following the derivation detailed in Appendix B, $\alpha_i(t)$ can be evaluated recursively as

$$\alpha_i(1) = \pi_i^{(k)} \lambda_i^{(k)}(\mathbf{r}_t) \qquad 1 \le i \le N \tag{3.66}$$

$$\alpha_i(t) = \lambda_i^{(k)}(\mathbf{r}_t) \sum_{j=1}^{N} \alpha_j(t-1) a_{j,i}^{(k)} \qquad 1 \le i \le N \tag{3.67}$$

$$1 \le t \le T$$

The backward partial probability, denoted as $\beta_i(t)$, is the conditional probability of observing the remaining $\mathbf{r}_\tau$, where $\tau = t+1, t+2, \ldots, T$ given the state at t is q_i.

$$\beta_i(T) = 1 \qquad \forall i \tag{3.68}$$

$$\beta_i(t) = P\left(\mathbf{r}_{t+1}, \mathbf{r}_{t+2}, \ldots, \mathbf{r}_T | q_t = s_i, \theta^{(k)}\right) \qquad 1 \le t \le T-1 \tag{3.69}$$

Following the derivation detailed in Appendix B, $\beta_i(t)$ can be evaluated recursively as

$$\beta_i(T) = 1 \tag{3.70}$$

$$\beta_i(t) = \sum_{j=1}^{N} a_{i,j}^{(k)} \lambda_i^{(k)}(\mathbf{r}_{t+1}) \beta_j(t+1) \qquad 1 \le t \le T \tag{3.71}$$

The posterior probability detailed in Eq. 3.63 can be now be formulated in terms of $\alpha_i(t)$ and $\beta_i(t)$ since

$$P\left(q_t = s_i | R, \theta^{(k)}\right) = \frac{P\left(R, q_t = s_i | \theta^{(k)}\right)}{P\left(R | \theta^{(k)}\right)} \tag{3.72}$$

$$P\left(R, q_t = s_i | \theta^{(k)}\right) = P\left(R | q_t = s_i, \theta^{(k)}\right) P\left(q_t = s_i | \theta^{(k)}\right) \tag{3.73}$$

$$= P\left(\mathbf{r}_{1:t}, \mathbf{r}_{t+1:T} | q_t = s_i, \theta^{(k)}\right) P\left(q_t = s_i | \theta^{(k)}\right) \tag{3.74}$$

$$= P\left(\mathbf{r}_{1:t} | q_t = s_i, \theta^{(k)}\right) P\left(\mathbf{r}_{t+1:T} | q_t = s_i, \theta^{(k)}\right) P\left(q_t = s_i | \theta^{(k)}\right) \tag{3.75}$$

$$= P\left(\mathbf{r}_{1:t}, q_t = s_i | \theta^{(k)}\right) P\left(\mathbf{r}_{t+1:T} | q_t = s_i, \theta^{(k)}\right) \tag{3.76}$$

$$= \alpha_i(t) \beta_i(t) \tag{3.77}$$

$$P\left(R|\theta^{(k)}\right) = \sum_{i=1}^{N} P\left(R, q_t = s_i|\theta^{(k)}\right) = \sum_{i=1}^{N} \alpha_i\left(t\right)\beta_i\left(t\right) \tag{3.78}$$

resulting in

$$P\left(R, q_t = s_i|\theta^{(k)}\right) = \frac{\alpha_i\left(t\right)\beta_i\left(t\right)}{\sum_{i=1}^{N}\alpha_i\left(t\right)\beta_i\left(t\right)} \tag{3.79}$$

The joint posterior probability detailed in Eq. 3.64 can also be formulated in terms of $\alpha_i\left(t\right)$ and $\beta_i\left(t\right)$ since

$$P\left(q_{t-1} = s_i, q_t = s_j|R, \theta^{(k)}\right) = \frac{P\left(R, q_{t-1} = s_i, q_t = s_j|\theta^{(k)}\right)}{P\left(R|\theta^{(k)}\right)} \tag{3.80}$$

$$P\left(R, q_{t-1} = s_i, q_t = s_j|\theta^{(k)}\right) = P\left(\mathbf{r}_{1:t-1}, q_{t-1} = s_i, \mathbf{r}_t, q_t = s_j, \mathbf{r}_{t+1:T}|\theta^{(k)}\right) \tag{3.81}$$

$$= P\left(\mathbf{r}_{t+1:T}|q_t = s_j, \theta^{(k)}\right)\dots$$

$$\times P\left(q_t = s_j|q_{t-1} = s_i, \theta^{(k)}\right)\dots$$

$$\times P\left(\mathbf{r}_t|q_t = s_j|\theta^{(k)}\right)\dots$$

$$\times P\left(\mathbf{r}_{1:t-1}, q_{t-1} = s_i|\theta^{(k)}\right) \tag{3.82}$$

$$= \beta_i\left(t\right)a_{i,j}\lambda_j\left(\mathbf{r}_t\right)\alpha_i\left(t-1\right) \tag{3.83}$$

$$P\left(R|\theta^{(k)}\right) = \sum_{i=1}^{N} P\left(R, q_t = s_i|\theta^{(k)}\right) = \sum_{i=1}^{N} \alpha_i\left(t\right)\beta_i\left(t\right) \tag{3.84}$$

resulting in

$$P\left(q_{t-1} = s_i, q_t = s_j|R, \theta^{(k)}\right) = \frac{\beta_i\left(t\right)a_{i,j}\lambda_j\left(\mathbf{r}_t\right)\alpha_i\left(t-1\right)}{\sum_{i=1}^{N}\alpha_i\left(t\right)\beta_i\left(t\right)} \tag{3.85}$$

In summary, the GEM algorithm for HMM parameter training based on unknown asset trace sequences maximizes the likelihood of $P(R|\theta)$ to some local optimum θ^* through the iterative maximization of the HMM parameters. This is accomplished through the construction and evaluation of the $\mathcal{Q}$ function at each iteration until a convergence criterion is reached, typically when the change in likelihood does not increase between successive iterations. The GEM algorithm can be visualized in Fig. 3.21 as 'climbing' the surface of the likelihood function to arrive at local optimum θ^* [57].
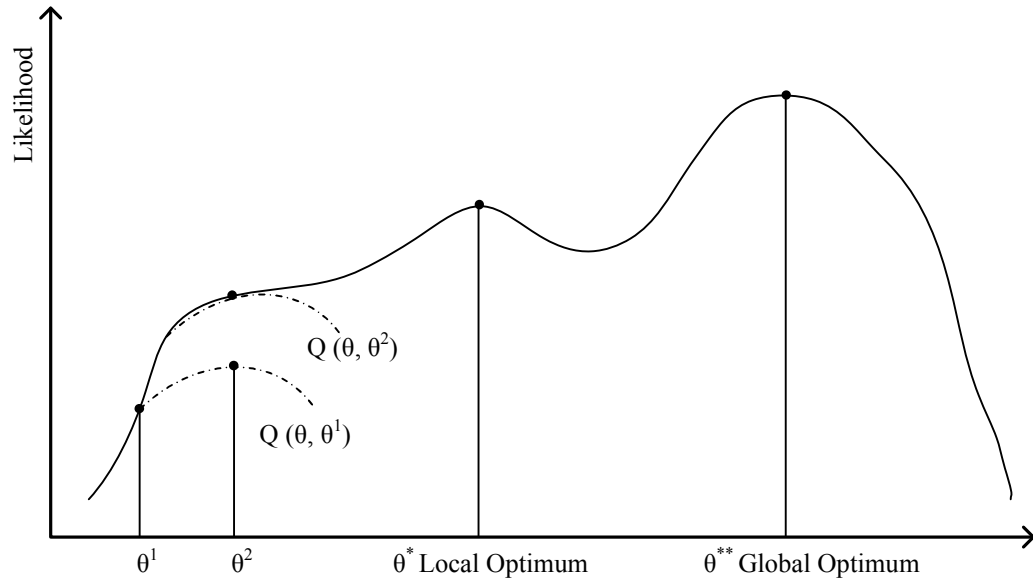
Figure 3.21: Iterative Maximization of the Observation Sequence Likelihood via GEM

3.6.2.3 Probability Lattice Tracing via Viterbi Algorithm for Most Probable State Sequence

Now that the HMM model parameters have been trained through GEM, the second setup to improve the initial fingerprinted radio map may be explained. The sequence of observed RSS measurements, $R = [\mathbf{r}_1, \mathbf{r}_2, \ldots \mathbf{r}_T]$, collected during the unknown asset trace sequence, $Q = [q_1, q_2, \ldots q_T]$, can be used to anneal the radio map if the locations of where the observations were collected can be determined. The Viterbi algorithm can be used to find the most probable asset trace sequence, or path, of states that could have produced their associated observations[58, 51, 55] and is defined as the maximization of $P(Q|R, \theta^*)$. In fact, the Viterbi algorithm can also be used to find the most probable asset trace sequence of states using the observed RSS and the set of untrained HMM parameters. This would not be advisable since the optimized HMM parameters resulting from the maximized likelihood of $P(R|\theta^*)$ through GEM best represents the sequence of observed measurements R.

There are two variables that are required for the Viterbi algorithm and are defined as

$$\delta_t(i) = \max_{q_1, q_2, \ldots, q_{t-1}} P(q_1, q_2, \ldots, q_t = s_i, \mathbf{r}_1, \mathbf{r}_2, \ldots, \mathbf{r}_t | \theta_*) \tag{3.86}$$

which is the best score, or highest likelihood, along a single path among all the possible paths ending in state s_i at time t and

$$\psi_t(i) = \arg \max_{q_1, q_2, \ldots, q_{t-1}} P(q_1, q_2, \ldots, q_t = s_i, \mathbf{r}_1, \mathbf{r}_2, \ldots, \mathbf{r}_t | \theta_*) \tag{3.87}$$

which is used to keep track of the best path ending in state s_i at time t. The full procedure of the Viterbi algorithm is outline as

Initialization

$$\delta_t(i) = \pi_i \lambda_i(\mathbf{r}_1) \qquad 1 \le i \le N \tag{3.88}$$

$$\psi_t(i) = 0 \tag{3.89}$$

Recursion

$$\delta_t(j) = \max_{1 \le i \le N} \left(\delta_{t-1}(i)\, a_{i,j} \lambda_j(\mathbf{r}_t) \right) \qquad 2 \le t \le T$$
$$1 \le j \le N \tag{3.90}$$

$$\psi_t(j) = \arg \max_{1 \le i \le N} \left(\delta_{t-1}(i)\, a_{i,j} \right) \qquad 2 \le t \le T$$
$$1 \le j \le N \tag{3.91}$$

Termination

$$P^+(Q|\theta^*) = \max_{1 \le i \le N} \delta_T(i) \tag{3.92}$$

$$q_T^+ = \arg \max_{1 \le i \le N} \delta_T(i) \tag{3.93}$$

Backtracking

$$q_t^+ = \psi_{t+1}\left(q_{t+1}^+\right) \qquad T - 1 \ge t \ge 1 \tag{3.94}$$

$$Q^+ = \left[q_1^+, q_2^+, \ldots, q_T^+ \right] \tag{3.95}$$

Following the above calculation of the most probable sequence of states, Q^+, and remembering that each HMM state corresponds to a unique ANP in the system, the radio map is annealed by updating the fingerprint matrix with the observed RSS

measurements for each AP resulting in

$$\mathcal{R}^+ \triangleq \left\{ \left(\mathbf{x}_{ap_i}, \mathbf{F}^+(\mathbf{x}_{ap_i})\right) | i = 1, \ldots, N \right\} \tag{3.96}$$

where $\mathbf{F}^+(\mathbf{x}_{ap_i})$ is the annealed fingerprint matrix at each ANP$= q_t^+$, $1 \leq t \leq T$.

3.7 Discussion

This chapter discussed each component required for the proposed indoor range-free localization system that will be experimentally validated in Chapter 4. It is important to note that this thesis will not explore the implementation of tracking algorithms and their associated benefits to location estimation accuracy. However, the Bayesian statistical development of the proposed probabilistic location estimation and radio map annealing algorithms will lead to an interesting discussion as to the feasibility of tracking adaptation into the proposed system.

Chapter 4

Experimental Validation

This chapter explores the validation of proposed concepts outlined in Chapter 3. First, the experiment setup used to perform the site survey environment, placement of ANPs and RPs, and the acquisition of online RSS measurements at static test locations and during dynamic test traces will be described.

The GT radio map constructed via the turntable setup of Section 3.4 was used to analyze the number and selection of d APs required within the deterministic and probabilistic location estimation techniques of WNNSS and MMSE. The selectable parameters of both techniques, K for WNNSS and s for optimal kernel bandwidth smoothing of the KDE, were examined to determine their most appropriate selection to produce the best distance error results possible for this system.

The first component which addresses the laborious fingerprinting effort will be a reduction in the total number of calibration ANP locations. Fingerprint RSS measurements acquired from sparse calibration ANP locations will be used to interpolate fingerprints to a complete radio map of all available ANPs. Validation will be performed with different degrees of sparseness for both the turntable fingerprints collected at each ANP as well as fingerprints acquired from stationary RPs attached to walls throughout the environment. The second component, or the n number of initial fingerprint RSS measurements at each of the sparse ANPs, will also be explored.

4.1 Setup and RSS Measurement Acquisition Procedures

This section details the offline production of the complete GT radio map with the turntable setup, the placement of RPs throughout the environment, the acquisition of online RSS measurements for each available AP at each online location, the acquisition of online RSS measurements during each dynamic trace of the asset, and the figure of merit used in this thesis. The total number of available APs within the localization environment, as outlined in Section 3.3, is fixed at 6.

4.1.1 Ground Truth Radio Map

The localization environment was divided into 71 ANPs, equally spaced by 1m as shown in Fig. 3.12. The two modules attached to the turntable setup, shown in Fig. 3.13, were programmed to collect a single fingerprint RSS measurement vector every 1 second. Since some scans do not sense every AP available to the localization system, some ANP locations vary in the number of fingerprint RSS measurements for each available AP. Our tests will evaluate the proposed system with a total $N = 200$ fingerprint RSS measurement vectors for each ANP location. The number of RSS measurement vectors, from each AP, did vary in length due to the amount of time spent acquiring at each ANP, therefore only the first 200 RSS measurement vectors can be guaranteed and will be used for this validation. Therefore, $T_{ap_s} = 200$ seconds were spent at each ANP location. Several days were required to collect all GT fingerprint RSS measurements.

4.1.2 Interpolated Radio Map from Reference Points

As shown in Fig 3.18, two RPs were affixed to the opposing walls, in direct line with each ANP, at a height of 180 cm along the corridors forming a direct line between the ANPs and the two RPs as shown in Fig 3.18. RSS measurements were collected for approximately 10 minutes with the scan period set to 2 seconds. This resulted in approximately $220 - 240$ RSS measurement vectors used to interpolate each ANP in the complete radio map. Once again, only the first 200 RSS measurement vectors will be used for this validation. A second set of several days was required to collect all RP RSS measurements and did not coincide with the collection of the GT fingerprint RSS measurements. With no changes to the propagation environment, this is assumed to not effect the validation of proposed techniques.

4.1.3 Static Online Locations

The mock IV pump setup, detailed in Section 3.2, was used to collect online RSS measurements at each of 101 static online locations with its Cartesian coordinates, $\mathbf{x}_{\text{test}}$, recorded for accuracy calculations. The static locations were randomly selected with each of the 71 ANPs included at least once in the set of 101 online locations. Each online location was observed for approximately 180 seconds with the module set to scan for available APs every 5 seconds. This yielded approximately 36 online RSS measurement vectors for each static location. The orientation of the mock IV pump was selected randomly at location. This thesis will utilize all acquired RSS measurements at each online location. The effects of varying the number of online RSS measurements used for location estimation will not investigated.

4.1.4   Dynamic Asset Traces

In addition to the static online locations, the mock IV pump setup was used to collect online RSS measurements every second during dynamic motion between each of the static online locations, resulting in 100 dynamic asset traces of varying length from $2 - 60$ meters. The speed at which the asset was moved throughout each trace was approximately between $2 - 3$ meters per second. A volunteer was used to move the mock IV pump through each trace as to mimic an actual patient in a hospital. The accelerometer daughter board, detailed in Appendix A, was configured to wake the module for scanning every 1 second during dynamic motion. Since the dynamic traces were of varying length, it follows that each of fingerprint vectors would also vary in length. A third set of several days was required to collect online RSS measurements at all static location and dynamic traces. As mentioned in Section 4.1.2, this is assumed to not effect the validation.

4.1.5   Figure of Merit

Distance error is calculated as the Euclidean distance between the estimated, $\hat{\mathbf{x}}$, and the true static online location, $\mathbf{x}_{\text{true}}$. This thesis will report the Average Distance Error (ADE), which is defined as the average of l_2 norms between N_{true} estimated and true static online locations as

$$d_{\text{ADE}} = \frac{1}{N_{\text{true}}} \sum_{i=1}^{N_{\text{true}}} \left\| \hat{\mathbf{x}} - \mathbf{x}_{\text{true}}\left(i\right) \right\|^2 \tag{4.1}$$

4.2   Localization Algorithm Comparison with Ground Truth

The two range-free localization algorithms discussed in Section 3.5, the deterministic WNNSS and probabilistic MMSE, were each evaluated utilizing all 101 static online locations with the complete radio map comprised of a total of 200 fingerprint RSS measurements vectors. The selection of APs for each algorithm will be determined based on the sample RSS measurements taken at each of the static online locations. The d number of selected APs will be evaluated using two different methodologies: the d highest average of online RSS measurement vectors and the d highest frequency of online RSS measurement vectors for each static online location. For the first method, each algorithm calculates the average value of online RSS measurements from each AP, sorts the averages from highest to lowest, and selects the d APs from the sorted list. For the second method, each algorithm counts the number of RSS measurements from each AP, sorts the counts from highest to lowest, and selects the d APs from the sorted list. These will be referred to as *method 1* and *method 2* for the sake of convenience.

4.2.1   K Weighted Nearest Neighbors in Signal Space

Along with the determination of the d number of APs for location estimation, it was also necessary to vary the selectable parameter K, or the number of WNNSS, to see its effect on the ADE. As can be seen in Table 4.1, the d selected number of APs used in the location estimation was varied from 3 to 5 with the K selected number of WNNSS varied from 1 to 7 for the complete radio map of 200 fingerprint RSS measurement vectors. For clarification, method 1 and 2 for AP selection will be referred to as *Avg. RSS* and *Freq. RSS*, or *High Avg. RSS* and *High Freq. RSS* in

various figures and tables throughout the validation.

The value of K corresponding to the smallest ADE for each set of selected APs in Table 4.1 was used to further investigate the effects of varying the n number of initial fingerprint RSS measurements on the ADE as shown in Fig. 4.1. It can be seen that the ADE is highly dependent on the number of selected APs, regardless of which selection method is used, with an approximate decrease in the ADE of 0.5 meters, per AP, with an increase from 3 to 4 to 5 APs. The selection of APs with method 1, which is the most common technique found in literature, outperforms method 2 for the selection of 3 and 4 APs. In contrast, for the selection of 5 APs method 2 results in the smallest ADE. The results in Fig. 4.1 further show that a fairly consistent ADE is maintained with the increase in n used for WNNSS location estimation. This could be attributed to the turntable configuration, which was described in Section 3.4, where the rotation attempts to average out the spatial fast fading effects on RSS

K WNNSS - ADE (m) - Ground Truth Radio Map - 200 Fingerprint RSS Measurements - APs Selected by Highest Avg. RSS or Highest Freq. RSS Measurements at Online Locations					
	3 APs		4 APs		5 APs	
K	Freq. RSS	Avg. RSS	Freq. RSS	Avg. RSS	Freq. RSS	Avg. RSS
1	4.16	4.02	3.12	3.28	2.46	2.51
2	3.99	3.84	2.98	3.01	2.34	2.43
3	3.75	3.57	3.03	2.99	2.33	2.40
4	3.65	3.59	2.99	2.94	2.31	2.39
5	3.65	3.50	2.99	2.95	2.35	2.41
6	3.57	3.40	2.96	2.91	2.39	2.47
7	3.62	3.42	3.00	2.93	2.40	2.53

Table 4.1: K WNNSS - Ground Truth Radio Map - 200 RSS Measurements
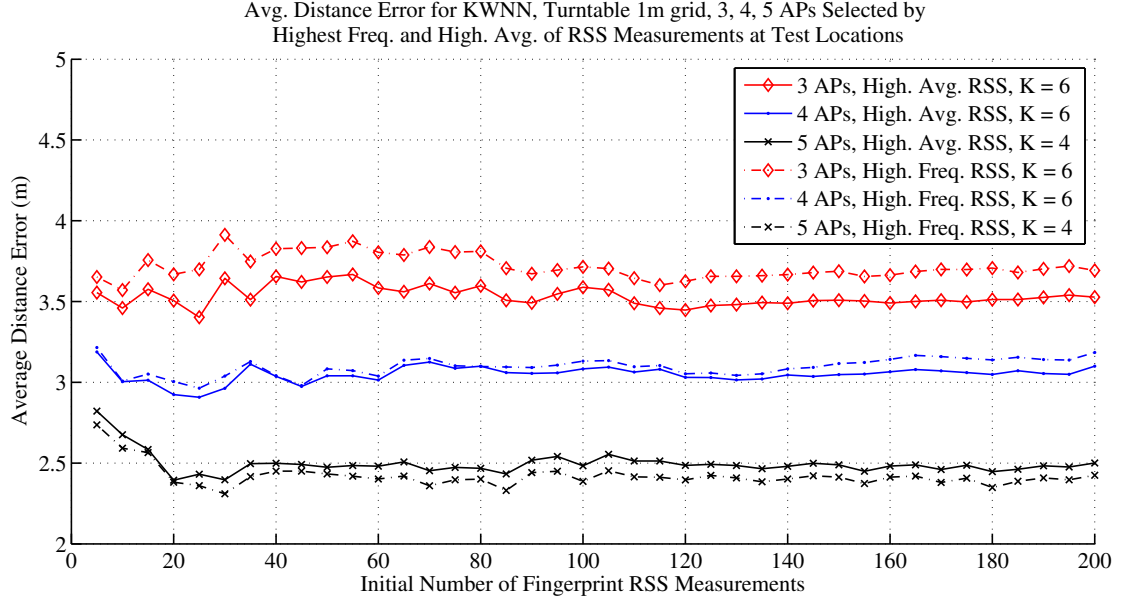
Figure 4.1: WNNSS for Minimum Distance Error with Ground Truth - 3, 4, 5 APs

measurements. The turntable does not improve the residual effects caused by long term shadowing of walls and obstacles.

With a fixed number of 5 selected APs, method 1 and 2 were compared with variations in K and n as shown in Fig. 4.2. The variation of K can be seen to directly influence the ADE with $K = 1$ producing the greatest variation in the ADE with a further increase in K resulting in a more uniform ADE as n is increased. As $K > 5$, the ADE increases due to WNNSS from more distant ANPs contributing to the location estimation. This effect can be best seen through the contour plots in Fig. 4.3 with the selection of 5 APs by method 1 and 2 shown in Fig. 4.3(a) and Fig. 4.3(b) respectively. The ADE resulting from 5 APs selected via method 1 is more consistent with an increase in n.
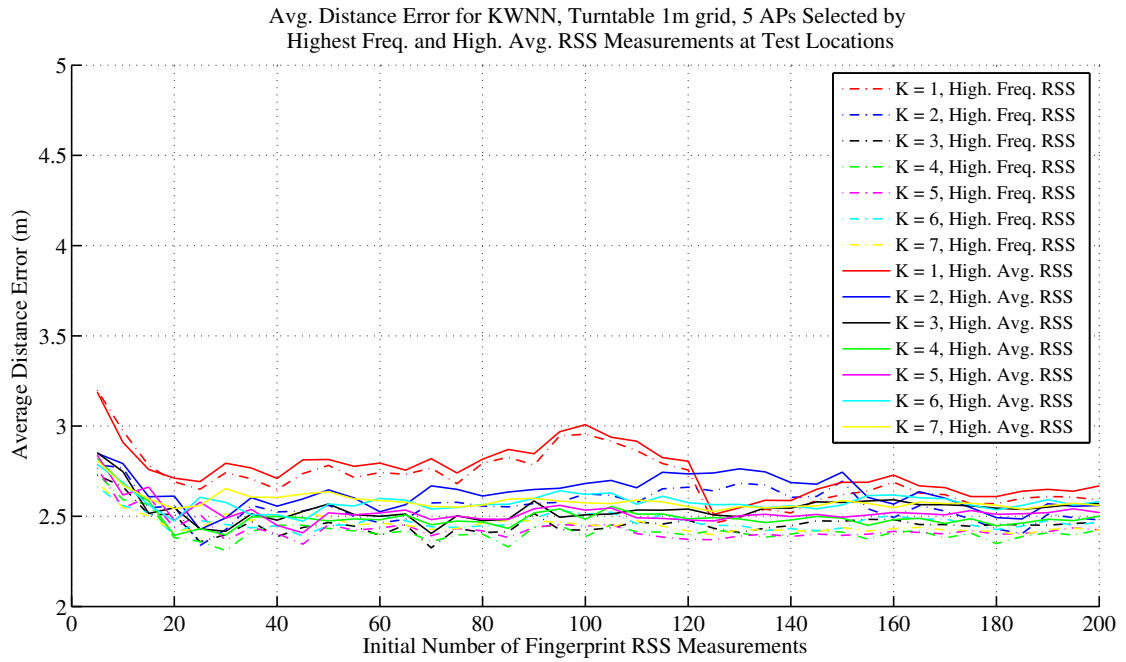
Figure 4.2: Vary K with WNNSS for Average Distance Error with Ground Truth - 5 APs

(a)APs Selected by Highest Freq. RSS

(b)APs Selected by Highest Avg. RSS
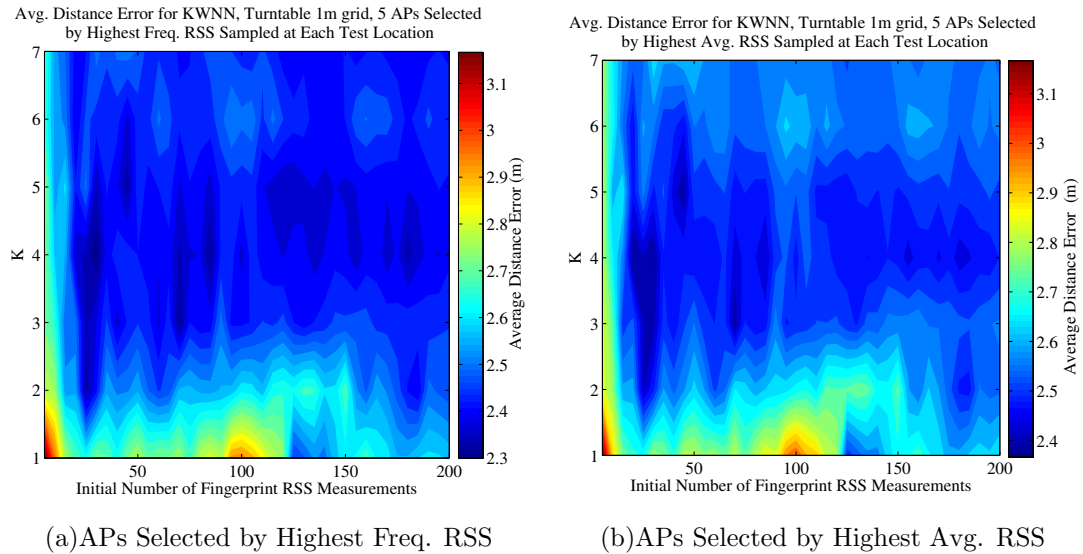
Figure 4.3: WNNSS - Average Distance Error for Vary K with 5 APs

A Cumulative Distribution Function (CDF) is used extensively in literature to conveniently show the probabilities resulting from the accumulation of all distance errors for a given localization technique. This can be visualized in Fig. 4.4 for WNNSS with fixed values of K, variations in n, and APs selected via method 2. For example,

(a)3 APs

(b)4 APs

(c)5 APs
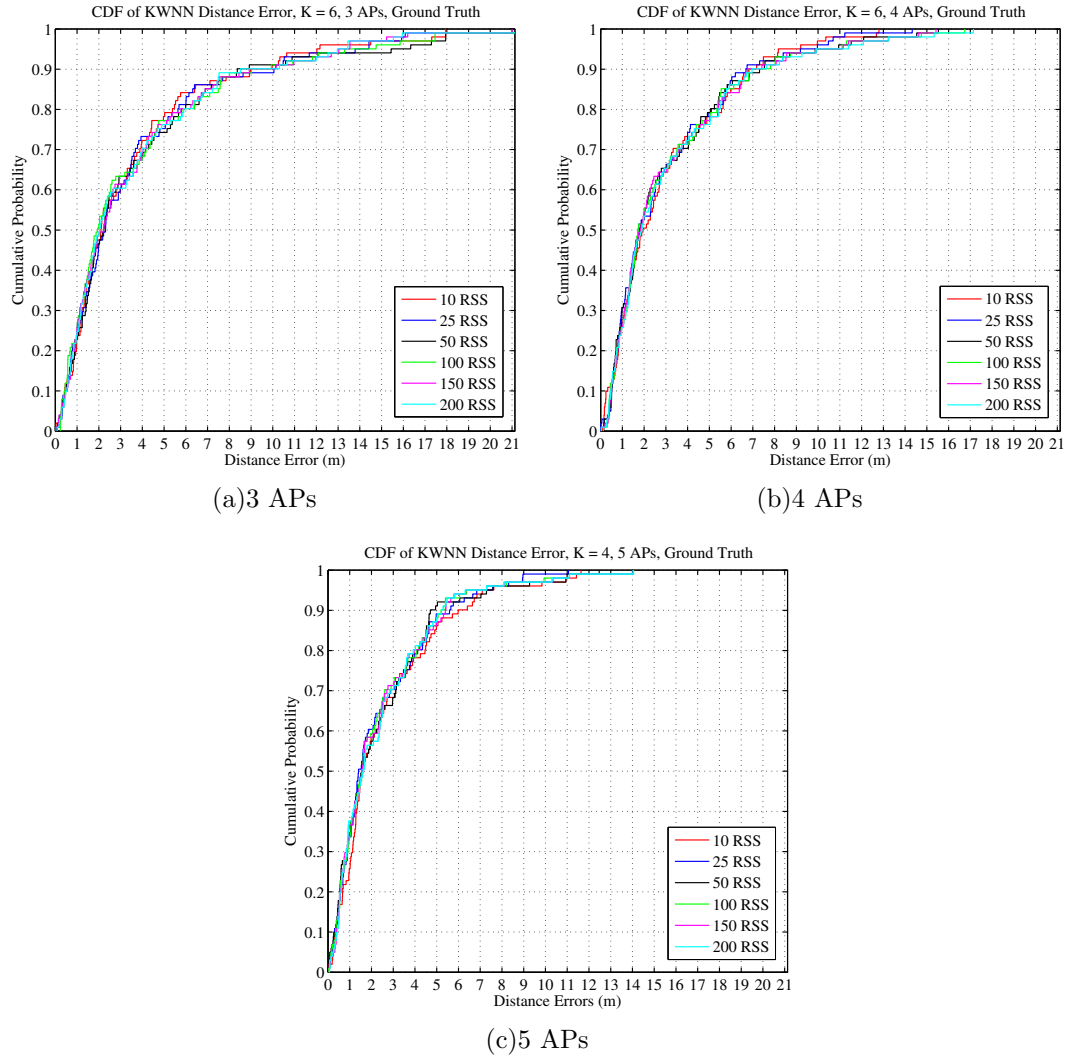
Figure 4.4: CDF of K WNNSS for Varying Number of RSS Measurements

K WNNSS - Distance Errors (m) - APs Selected by Highest Freq. RSS Measurements at Online Locations								
	3 APs, K = 6			4 APs, K = 6			5 APs, K = 4		
# RSS	AVG	50 %	70 %	AVG	50 %	70 %	AVG	50 %	70 %
10	3.57	2.10	3.92	3.01	1.95	3.34	2.59	1.54	3.10
25	3.70	2.31	3.73	2.96	1.73	3.73	2.36	1.40	2.88
50	3.84	2.23	4.22	3.08	1.79	3.65	2.43	1.59	3.12
100	3.71	1.94	4.27	3.13	1.75	3.51	2.39	1.55	2.61
150	3.69	2.22	4.12	3.12	1.85	3.58	2.41	1.58	2.70
200	3.69	2.03	4.07	3.19	1.90	3.54	2.42	1.61	2.85

Table 4.2: K WNNSS - Distance Errors - APs Selected by Highest Freq. RSS at Online Locations

to determine the cumulative probability of a distance error less than or equal to 3 meters, one would locate this value on the x axis and trace the figure vertically to the desired CDF plot. As in Fig. 4.4(a), the resulting cumulative probability for a distance error of less than or equal to 3 meters corresponds to approximately 63% for $n = 100$ initial fingerprint RSS measurements. The distance error results from the WNNSS CDF plots presented in Fig. 4.4 are shown below in Table 4.2. These include the ADE, the 50th (median) and 70th percentile of distance error, for fixed values of K, variations in n, and APs selected via method 2. Even with a small value of n, WNNSS shows superior location estimation performance with $d = 5$ selected APs and $K = 4$.

4.2.2 Minimum Mean Squared Error Estimator via KDE

A comparison of method 1 and 2 for the selection of d APs must be made against the probabilistic MMSE localization approach. The selectable parameter s, or smoothing factor outlined in Section 3.5.2.2, adjusts the optimal kernel bandwidth used with

the multivariate KDE to account for RSS measurement variations from the Gaussian distribution and directly effects the ADE. Therefore, it was necessary to calculate the parameter s with a variation in n, or the number of initial fingerprint RSS measurements, for the selection of APs based on method 1 and 2. An example for $d = 5$ APs can be in Fig. 4.5, where the value of s resulting in a minimum ADE was determined for each increment of n from the GT radio map. Further visualization can be seen in Fig. 4.6, where contours of the ADE are shown with variations in s and n. Optimum values of the smoothing factor, for $d = 5$, were calculated as the average of s across the variation in n for both method 1 and 2 as $s_{avg} = 1.0175$ and $s_{freq} = 1.0125$ respectively. The results of MMSE with optimum smoothing factors can be seen in Table 4.3, which reports the ADE resulting from the selection of 3, 4, and 5 APs based on methods 1 and 2 utilizing the entire radio map of 200 fingerprint RSS measurement vectors.

The optimum values of s, for each set of selected APs, were used to investigate the effects of varying n on the ADE as shown in Fig. 4.7. Similar to the results of the WNNSS location estimation shown in Fig. 4.1, the MMSE ADE is

MMSE via KDE - Avg. Distance Error (m) - 200 Fingerprint RSS Measurements - APs Selected by Highest Avg. RSS or Highest Freq. RSS Measurements at Online Locations, Optimum Smooth Factor					
3 APs		4 APs		5 APs	
Freq. RSS	Avg. RSS	Freq. RSS	Avg. RSS	Freq. RSS	Avg. RSS
3.39	3.22	2.98	2.97	2.41	2.48

Table 4.3: MMSE via KDE - Average Distance Error - 200 RSS Measurements
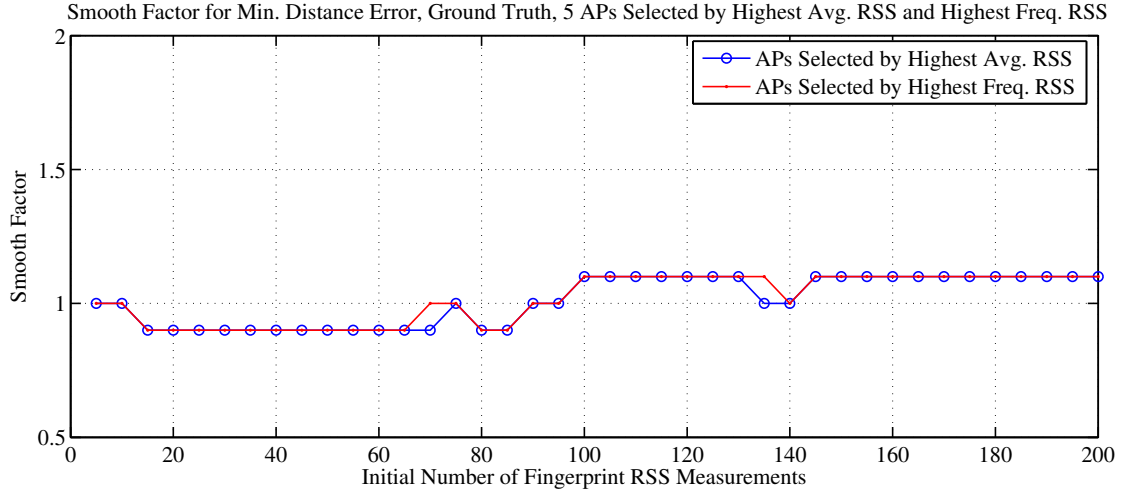
Figure 4.5: MMSE via KDE- Smoothing Factor for Min. Distance Error, Ground Truth

(a)APs Selected by Highest Freq. RSS

(b)APs Selected by Highest Avg. RSS
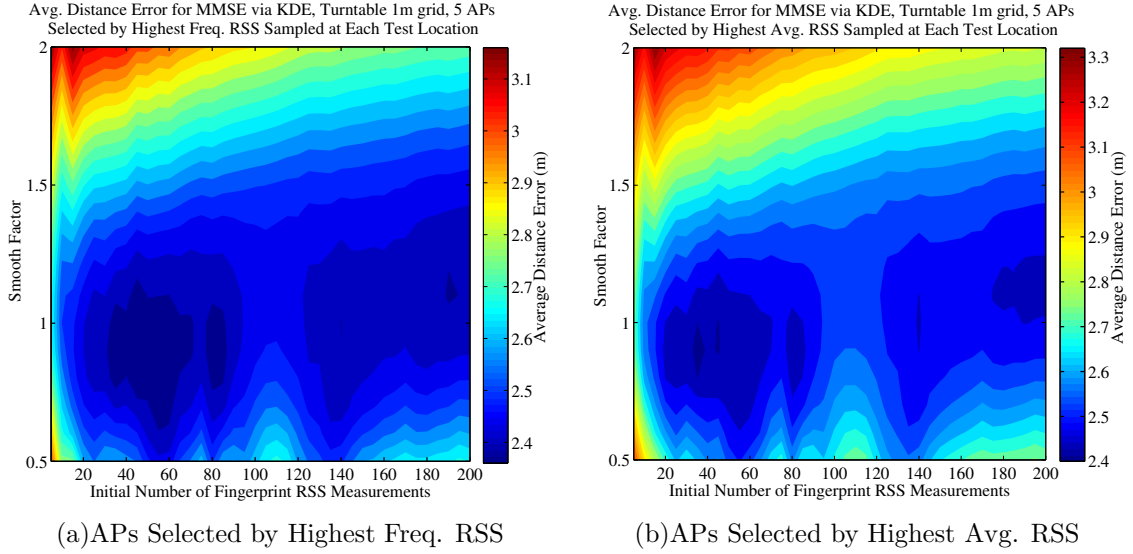
Figure 4.6: MMSE via KDE - Average Distance Error for Variations in Smoothing Factor with 5 APs Selected by Highest Freq. RSS at Test Locations
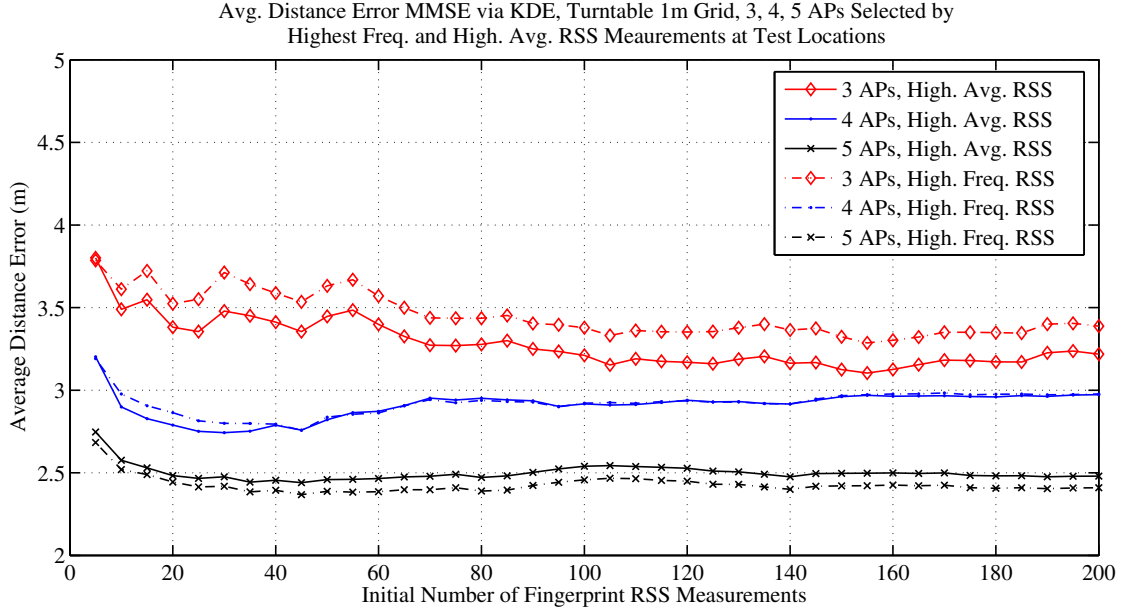
Figure 4.7: MMSE via KDE for Average Distance Error

highly dependent on the number of selected APs regardless of which selection method is used with an approximate decrease in the ADE of 0.3 meters with an increase from 3 to 4 and an approximate decrease in the ADE of 0.5 meters from 4 to 5 APs. The selection of APs with method 1 again outperforms method 2 for the selection of 3 and 4 APs. In contrast, for the selection of 5 APs method 2 results in the smallest MMSE ADE.

A CDF is used to visualize the cumulative probabilities of all distance errors and is shown in Fig. 4.8 for MMSE location estimation with optimal values of s, variations in n, and APs selected via method 2. The distance error results from the MMSE location estimation CDF plots are shown in Table 4.4.

(a)3 APs
(b)4 APs
(c)5 APs
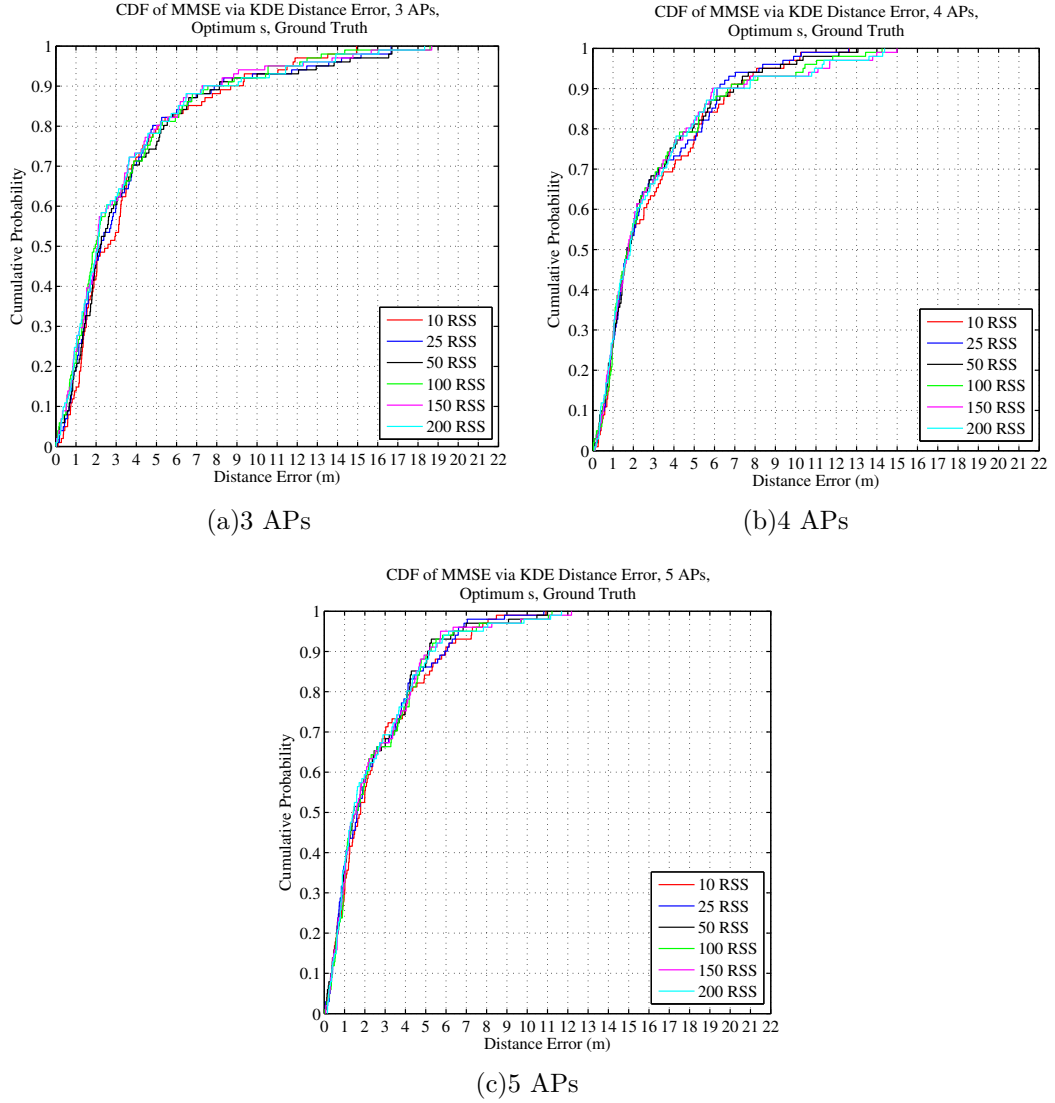
Figure 4.8: CDF of MMSE via KDE for Varying Number of RSS Measurements, APs Selected by Highest Freq. RSS at Online Locations

4.2.3 Remarks

Two different methods were investigated for the selection of APs using both WNNSS and MMSE location estimation algorithms with optimal values of parameters K and s. The results showed that both location estimation algorithms achieved comparably

MMSE via KDE - Distance Errors (m) - APs Selected by Highest Freq. RSS Measurements at Online Locations								
	3 APs			4 APs			5 APs		
# RSS	AVG	50 %	70 %	AVG	50 %	70 %	AVG	50 %	70 %
10	3.61	2.59	3.77	2.98	1.81	3.94	2.52	1.79	2.99
25	3.55	2.24	3.83	2.82	1.70	3.27	2.41	1.61	3.46
50	3.63	2.21	3.82	2.84	1.69	3.24	2.39	1.44	3.39
100	3.38	1.97	3.80	2.92	1.76	3.19	2.46	1.57	3.46
150	3.32	2.08	3.56	2.97	1.75	3.30	2.42	1.57	3.38
200	3.39	2.11	3.59	2.98	1.86	3.47	2.41	1.45	3.31

Table 4.4: MMSE via KDE - Distance Errors - APs Selected by Highest Freq. RSS at Online Locations

low ADE of approximately 2.5 meters when $d = 5$ APs were selected with method 2, regardless of the n number of initial fingerprint RSS measurements from the GT radio map. As a result of these findings, the number of selected APs by method 2, calculated by the highest frequency of online RSS measurements for each online location, with be fixed at $d = 5$ for the remainder of the validation. It is important to note that the optimum values of K and s were determined from a single data set specific to the location estimation environment used for this validation. Changes to the locations of APs, building modifications, or the density of personal would alter the influences of shadowing on the fingerprint RSS measurements, thus resulting in inaccurate optimum values.

Although comparable ADE was achieved for both the deterministic WNNSS and probabilistic MMSE location estimation algorithms, the Bayesian statistics supplied by the KDE utilized within the MMSE location estimation are advantageous as they can be used to calculate the likelihood, or emission, distributions required by the GEM algorithm used to train the HMM model parameters of the dynamic radio

map annealing algorithm outlined in Section 3.6.2. Therefore, the MMSE location estimation algorithm will be used throughout the remainder of this validation. It is also interesting to note that for MMSE location estimation, the average value of the smooth factor s was approximately a value of 1 which indicates that the turntable GT radio map of 200 fingerprint RSS measurement vectors approximated the Gaussian kernel with optimum selected bandwidths as outlined in Section 3.5.2.2. The use of a complete radio map, consisting of fingerprint RSS measurements from dense ANP locations, would be ideal for range-free location estimation, but not very practical. Therefore, two scenarios were introduced, as outlined in Section 3.6.1, to geometrically reduce the density of fingerprint ANPs.

4.3   Radio Map Annealing via Dynamic Asset Traces

The proposed radio map interpolation and annealing algorithms, as outlined in Section 3.6, were analyzed for two different scenarios. First, a set of 3 complete radio maps were constructed with the IDW interpolation algorithm, as outlined in Section 3.6.1.1, and reductions in the density of fingerprinted locations from the original GT turntable radio map of 71 ANPs. These will be referred to as interpolated turntable radio maps. Second, a set of 3 complete radio maps were interpolated with reductions in the density of RPs located along opposing walls to the ANP locations. These will be referred to as interpolated RP radio maps. It is important to recall that the ANPs, or states, which form a complete radio map are located uniformly at 1 meter increments throughout the localization environment. The reduction in the density of locations used for interpolation will be referred to as an increase in the *sparseness*. The removal of every 2nd, 3rd, and 4th location results in an increase in

sparseness to 1/2, 1/3, and 1/4 of the original locations, be it ANP locations from the original GT radio map or locations of the RPs. Both scenarios were visualized in Fig. 3.17 - 3.18 for a sparseness of 1/3. The multiple unknown asset traces were utilized within the annealing algorithm to study its effects on MMSE location estimation for increases in sparseness of both the interpolated turntable and interpolated RP radio maps. The maximum number of iterations and convergence criterion of the GEM used for the annealing algorithm were fixed at 100 and 0.001 respectively. The number of APs used for this validation was fixed at $d = 5$ and were selected by method 2, as described Section 4.2.3.

4.3.1 Smoothing Factors of Interpolated Turntable Radio Maps

To implement MMSE location estimation on each of the interpolated turntable radio maps, an investigation into the smoothing factor that adjusts the optimal kernel bandwidth, required by the multivariate KDE, was conducted with variations in both s and n. As in previous section, the smoothing factors for each interpolated turntable radio map was selected as the value of s resulting in the smallest location estimation error at each increment of n. The interpolated radio maps resulting from increases in the sparseness of fingerprint locations are referred to as 2, 3, and 4 meter for convenience as shown in Fig. 4.9. The original GT turntable radio map, referred to as 1 meter, was included for as comparison of s without interpolation. Further visualization can be seen in Fig. 4.10, where contours of the ADE are shown with variations in s and n. The results illustrate relatively small variations in both the smoothing factor and the ADE with increases in n, regardless of the sparseness of
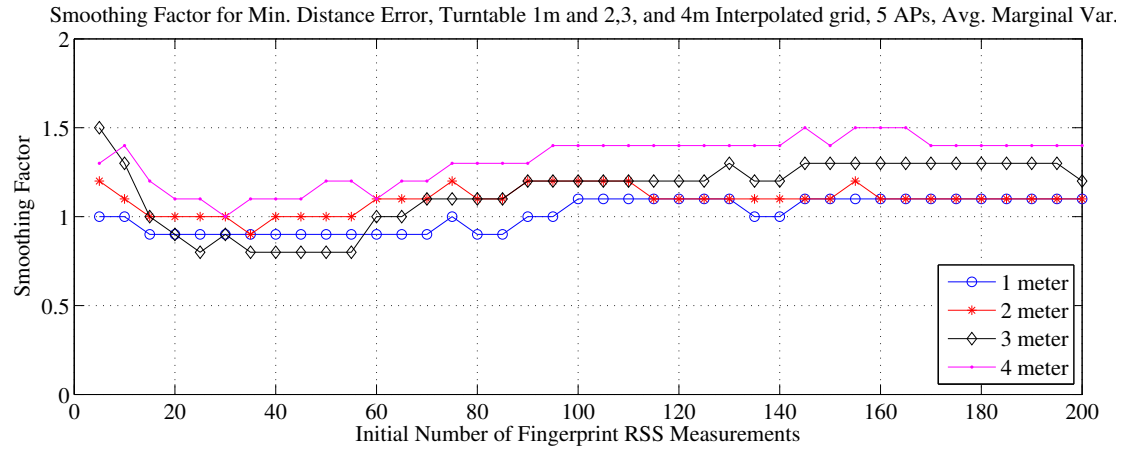
Figure 4.9: Interpolated Turntable Radio Maps - Variations in Smoothing Factor for Min. Distance Error
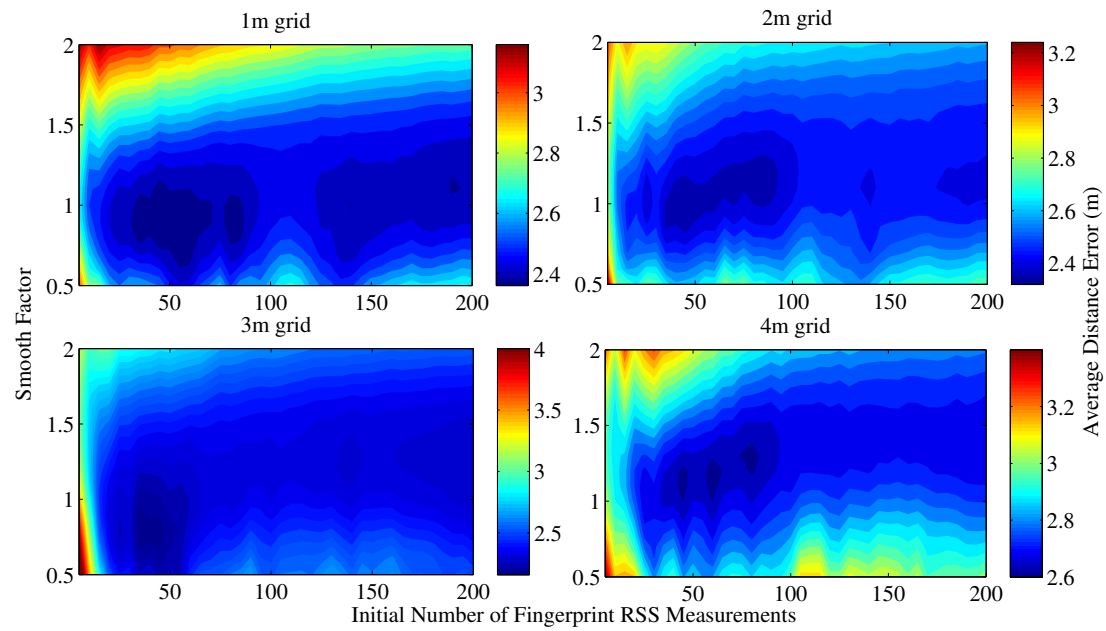
Figure 4.10: Interpolated Turntable Radio Maps - Variations in Smoothing Factor with Contour of Avg. Distance Error

the interpolated turntable radio map. This indicates that the use of an optimum smoothing factor, calculated as the average of s across the variation in n, can be utilized with the MMSE location estimation algorithm regardless of the selected value of n. Optimum values were calculated as the average of s across the variation in n for each radio map and are shown in Table 4.5. It can be seen that the optimum values increase as the sparseness of interpolated ANPs is increased. These optimum smoothing values were fixed within the MMSE location estimation algorithm during the investigation of annealing the interpolated turntable radio maps. It is important to again note that the optimum values of s, as shown in Table 4.5, were determined from a single data set specific to the location estimation environment used for this validation.

Interpolated Turntable Radio Maps Optimum Values of Smoothing Factor				
	1 meter	2 meter	3 meter	4 meter
s	1.013	1.095	1.139	1.317

Table 4.5: Interpolated Turntable Radio Maps - Values of Smoothing Factor

4.3.2 Dynamic Annealing of Interpolated Turntable Radio Maps

The annealing algorithm outlined in Section 3.6.2 required the configuration of constraints on the HMM parameter **A**, the state transition probability distribution. These constraints govern how an asset may move throughout the environment every 1 second, as modeled by the HMM state machine. The probabilities of the maximum distance constraint were selected based on a triangular distribution for the validation

as shown in Fig. 3.19(a). For a distance constraint of $m = 2$, the probabilities depict that an asset is most likely to transition to its closest neighboring states with less probability that the user will transition to further states. The HMM parameter $\boldsymbol{\pi}$, the initial state distribution, was selected to be uniform across all states since no prior knowledge is available as to the starting state of an asset. The likelihood, or emission, distribution $\boldsymbol{\lambda}$ has probabilities describing the likelihood of observing an online RSS measurement given that the asset is in a particular state within the HMM. Recall that the states within the HMM correspond to the ANPs within a complete radio map. These likelihood probabilities are calculated using the technique outlined in Section 3.6.2.1 where each online RSS measurement within the sequence R, collected during the unknown asset trace, are used within the KDE to map its RSS-dependency against the selected interpolated radio map.

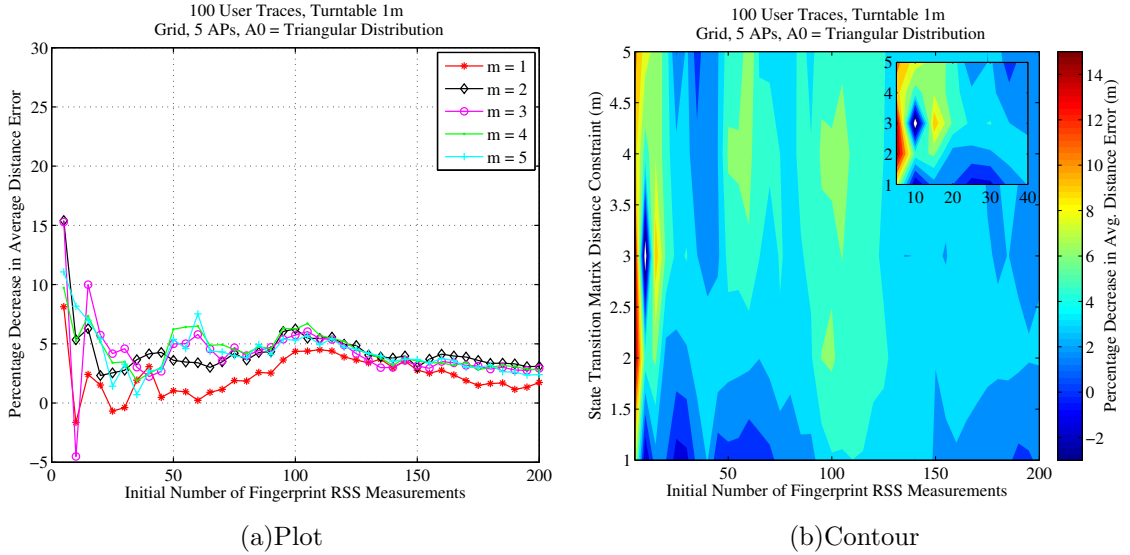The annealing of each interpolated turntable radio map was performed with

(a)Plot   (b)Contour

Figure 4.11: Annealing of 1 meter Turntable Radio Map

variations in the maximum distance constraint, for $m = 1 \ldots 4$, to analyze how well the HMM state machine captured the actual speed of the asset throughout each trace. Variations were also conducted on the initial number of fingerprint RSS measurements, n, used for the calculations of λ and MMSE location estimation. The annealing algorithm was first utilized on the original GT turntable radio map to investigate its effects on the ADE without any interpolation. This is shown in Fig. 4.11. The improvement to location estimation from annealing a radio map is reported as the *percentage decrease in ADE*. The ADE was calculated through the use of MMSE localization of all static asset locations with and without the annealing of each radio map. The use of the same online RSS measurements from the static asset locations was needed to offer a fair comparison of only the improvements to the ADE with and without annealing.

The results in Fig. 4.11(a) indicate that annealing of the original GT turntable radio map can provide some improvement to the ADE with variations in n. The percentage decrease in ADE was consistent with state transition distance constraints of $m = 2 \ldots 5$, regardless of the initial value of n. The constraint $m = 1$ performing lower which was to be expected since the the HMM state machine could not reflect transitions greater than 1 meter per second. The speed of the user during each of the dynamic asset traces was maintained at between $2 - 3$ meters per second. The greatest percentage decrease was between $8 - 16\%$ and was accomplished with an initial value of $n = 10$. As n was increased, the percentage decrease in ADE became relatively consistent for particular values of m, with a maximum and minimum of 8% and 3% respectively. Improved visualization can be seen in Fig. 4.11(b) were the contour of the percentage decrease in ADE is zoomed for small values of n.

The percentage decrease in ADE for the annealing of the 2, 3, and 4 meter interpolated turntable radio maps can be seen in Fig. 4.12 - 4.14. The results follow similar patterns with variations in the initial value of n. Interestingly, with small values of n, the percentage decrease in ADE is dramatically high for all values of m. This indicates that the annealing algorithm is able to dynamically improve the initial radio maps with a high degree of accuracy given relatively inaccurate initial radio maps. With increases in n, the annealing algorithm results in a relatively consistent percentage decrease in ADE. The percentage decrease in ADE remains lower for $m = 1$ as was previously shown for the annealing of the original GT turntable radio map.
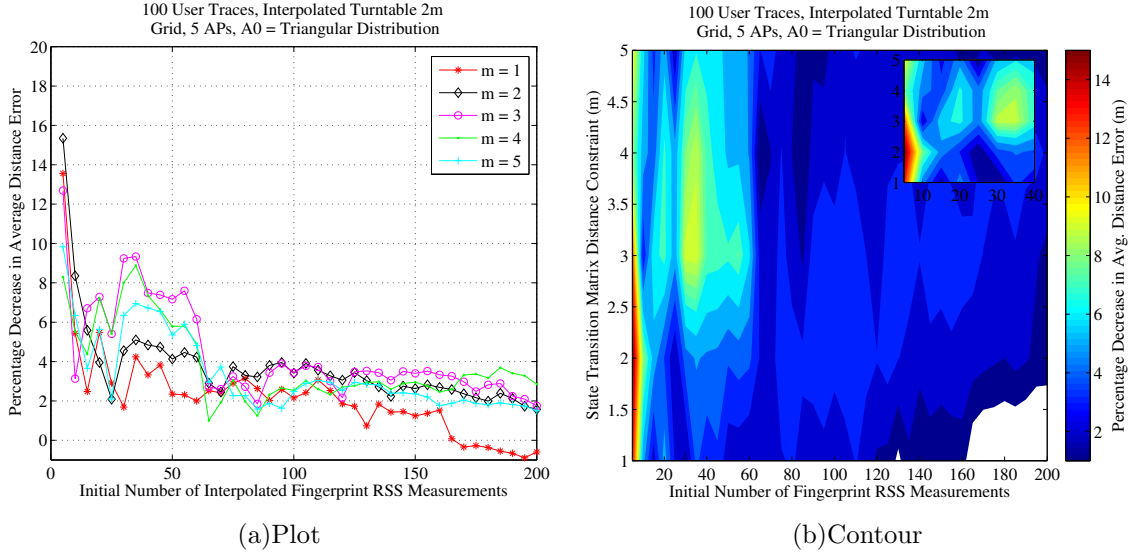

(a)Plot   (b)Contour

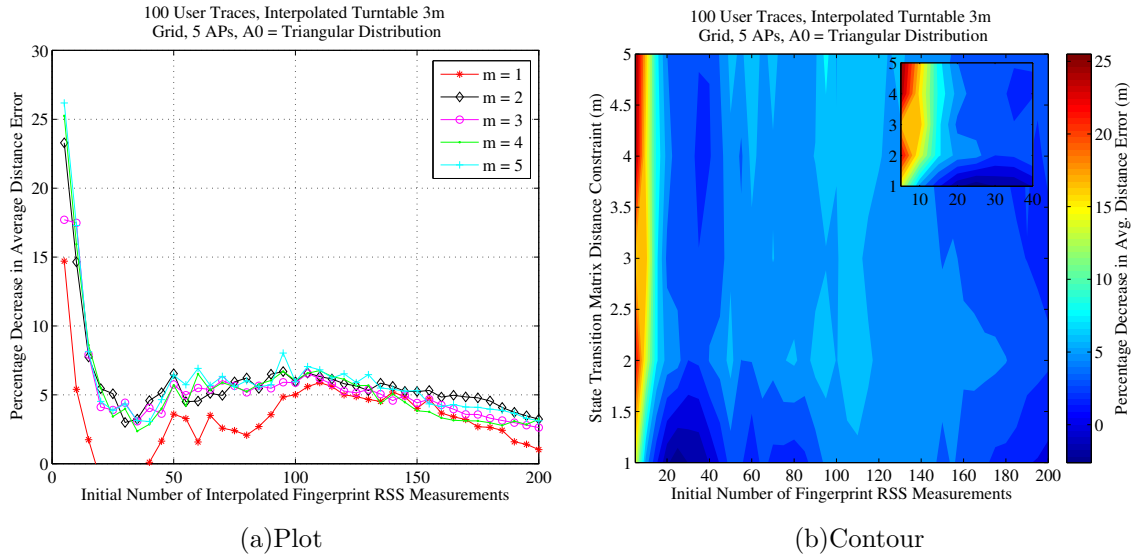Figure 4.12: Annealing of 2 meter Interpolated Turntable Radio Map

Figure 4.13: Annealing of 3 meter Interpolated Turntable Radio Map
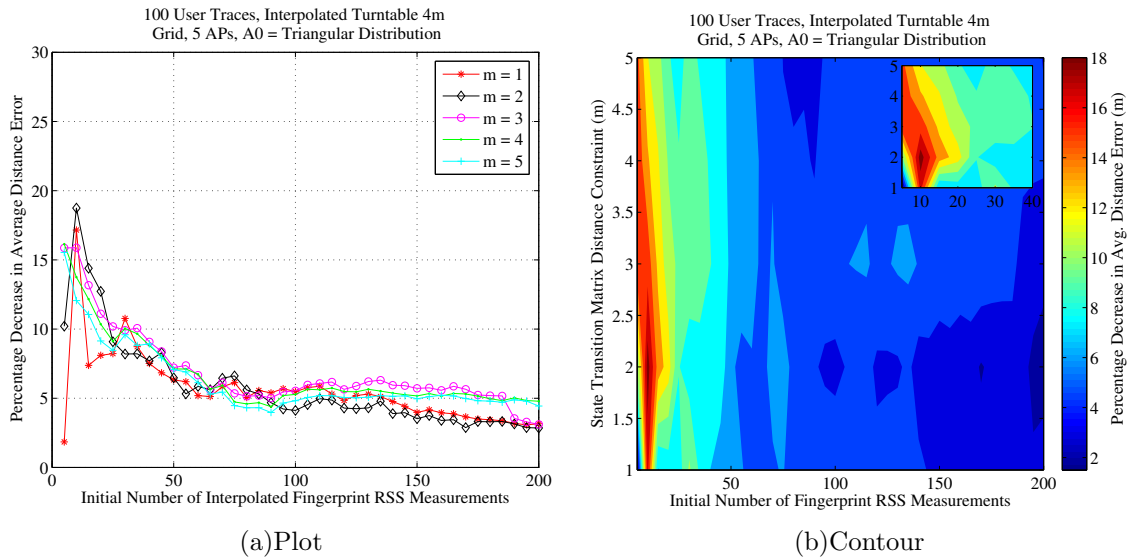
Figure 4.14: Annealing of 4 meter Interpolated Turntable Radio Map

MMSE Localization Distance Errors (m) - Interpolated
Turntable Radio Maps - With and Without Annealing - 5 APs Selected
by Highest Freq. RSS Measurements at Online Locations, m = 3

No Traces

# RSS	1m			2m			3m			4m		
	AVG	50 %	70 %	AVG	50 %	70 %	AVG	50 %	70 %	AVG	50 %	70 %
10	2.52	1.79	2.99	2.54	1.59	3.21	2.79	1.81	3.45	2.89	1.87	3.77
25	2.41	1.61	3.46	2.41	1.67	3.15	2.33	1.68	2.77	2.76	1.90	3.46
50	2.39	1.44	3.39	2.40	1.60	3.45	2.30	1.63	2.79	2.68	1.82	3.55
100	2.46	1.57	3.46	2.43	1.61	3.31	2.34	1.52	2.83	2.68	1.69	3.87
150	2.42	1.57	3.38	2.45	1.60	3.36	2.33	1.53	2.66	2.71	1.70	3.90
200	2.41	1.45	3.31	2.43	1.55	2.90	2.28	1.54	2.76	2.69	1.85	4.16

100 Traces

# RSS	1m			2m			3m			4m		
	AVG	50 %	70 %	AVG	50 %	70 %	AVG	50 %	70 %	AVG	50 %	70 %
10	2.63	1.57	3.60	2.46	1.62	3.42	2.30	1.68	2.77	2.43	1.73	3.28
25	2.31	1.59	2.80	2.28	1.50	2.86	2.24	1.47	2.54	2.48	1.79	3.12
50	2.27	1.42	2.73	2.22	1.29	3.06	2.17	1.44	2.43	2.49	1.67	3.10
100	2.32	1.45	3.36	2.34	1.46	3.04	2.20	1.52	2.62	2.54	1.73	3.52
150	2.35	1.45	2.87	2.37	1.44	2.73	2.23	1.48	2.63	2.56	1.63	3.59
200	2.34	1.35	2.84	2.39	1.54	2.69	2.22	1.49	2.62	2.60	1.77	3.50

Table 4.6: MMSE Distance Errors - With and Without Annealing - Interpolated Turntable Radio Maps

MMSE Localization Standard Deviations (m) Interpolated Turntable Radio Maps With and Without Annealing - 5 APs, m = 3				
	No Traces			
	1m	2m	3m	4m
# RSS	AVG	AVG	AVG	AVG
10	2.29	2.35	2.62	2.68
25	2.20	2.22	2.02	2.53
50	2.25	2.29	2.16	2.39
100	2.32	2.32	2.18	2.38
150	2.34	2.35	2.23	2.45
200	2.37	2.38	2.22	2.38
	100 Traces			
	1m	2m	3m	4m
# RSS	AVG	AVG	AVG	AVG
10	2.54	2.14	1.95	2.04
25	2.18	2.09	1.98	2.12
50	2.09	2.13	2.01	2.21
100	2.14	2.23	2.01	2.19
150	2.28	2.29	2.09	2.28
200	2.30	2.35	2.16	2.27

Table 4.7: MMSE Standard Deviations - With and Without Annealing - Interpolated Turntable Radio Maps

The overall changes in the MMSE location estimation distance errors and standard deviations can be seen in Table 4.6 - 4.7 for each of the interpolated turntable radio maps. These include the ADE, the 50th (median) and 70th percentile of distance error, for variations in n with a fixed value of m. The results are tabulated for a state transition distance constraint of $m = 3$ as to conform with the approximate speed of the asset during each dynamic trace. With $n = 10$, which is approximately a

20 fold decrease in the original number of fingerprint RSS measurements collected at each ANP, the ADE decreases with the use of each interpolated turntable radio map. Annealing the 2, 3, and 4 meter interpolated radio maps results in a decrease in the ADE of MMSE location estimation from $2.54 \rightarrow 2.46$, $2.79 \rightarrow 2.30$, and $2.89 \rightarrow 2.43$ meters or 3.2%, 17.6%, and 15.9% resectively. With $n = 100$, which is approximately a 2 fold decrease in the original number of fingerprint RSS measurements collected

Figure 4.15: CDF for Annealed 1 meter Turntable Radio Map
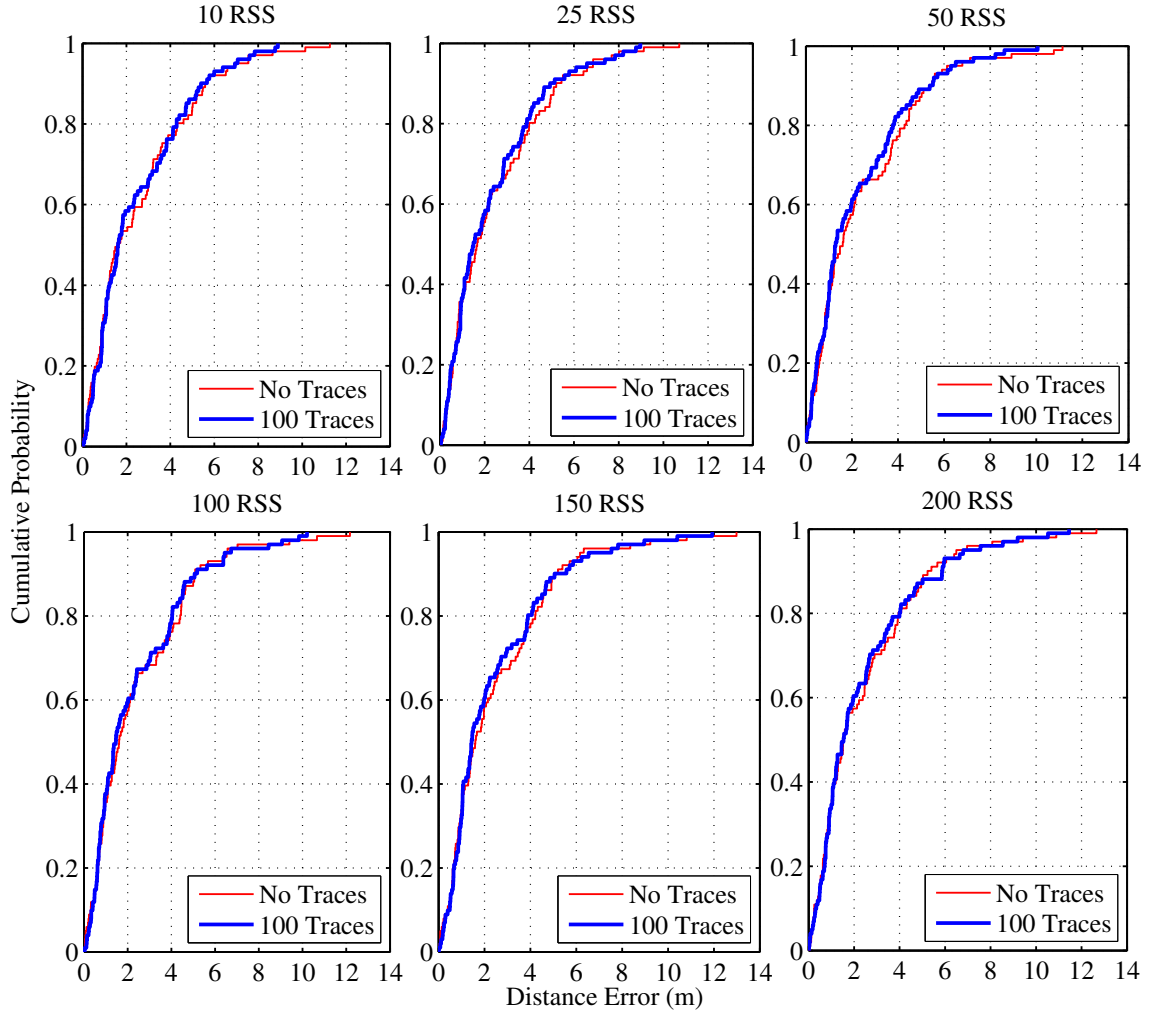
Figure 4.16: CDF for Annealed 2 meter Interpolated Turntable Radio Map

at each ANP, annealing the 2, 3, and 4 meter interpolated radio maps results in a decrease in the ADE from 2.43 → 2.34, 2.34 → 2.20, and 2.83 → 2.54 meters or 3.7%, 6.0%, and 10.2% respectively. It is interesting to note that the original GT radio map results in higher ADE compared to the sparse interpolated radio maps. This can be attributed to higher correlation between turntable figureprints at neighbouring 1 meter ANP locations. The distance errors presented in Table 4.6 can be

seen in Fig. 4.15 - 4.18 where a CDF was used to visualize the changes in distance error, with and without radio map annealing, for increments in n.

An interesting fact is observed that with an increase in the sparseness of finger-printed ANPs, the annealing of the interpolated radio maps results in a general increase in the accuracy of the MMSE location estimation. The effect is most prominent with $n = 10$ and can be attributed to what is called deficient radio map interpolation
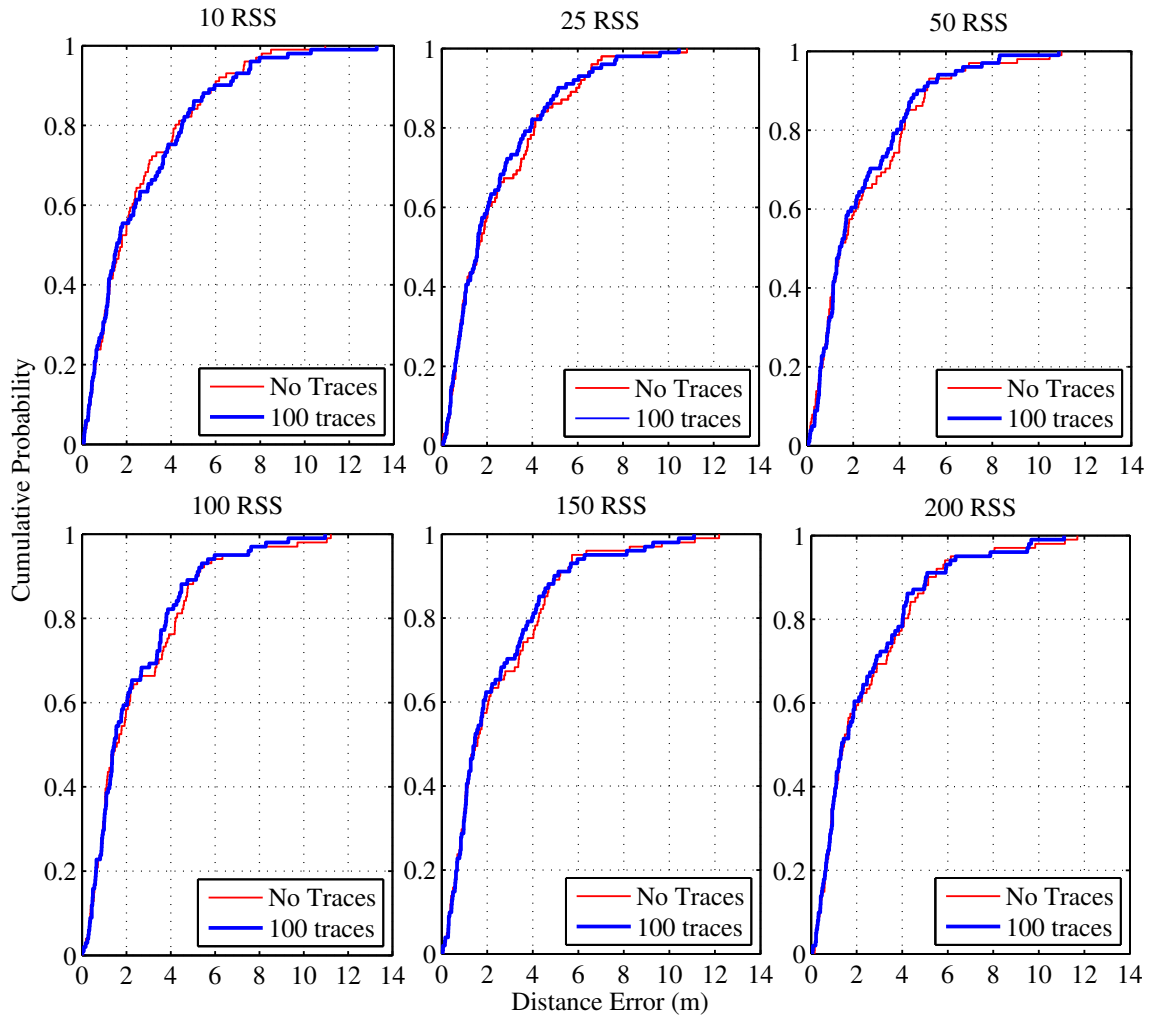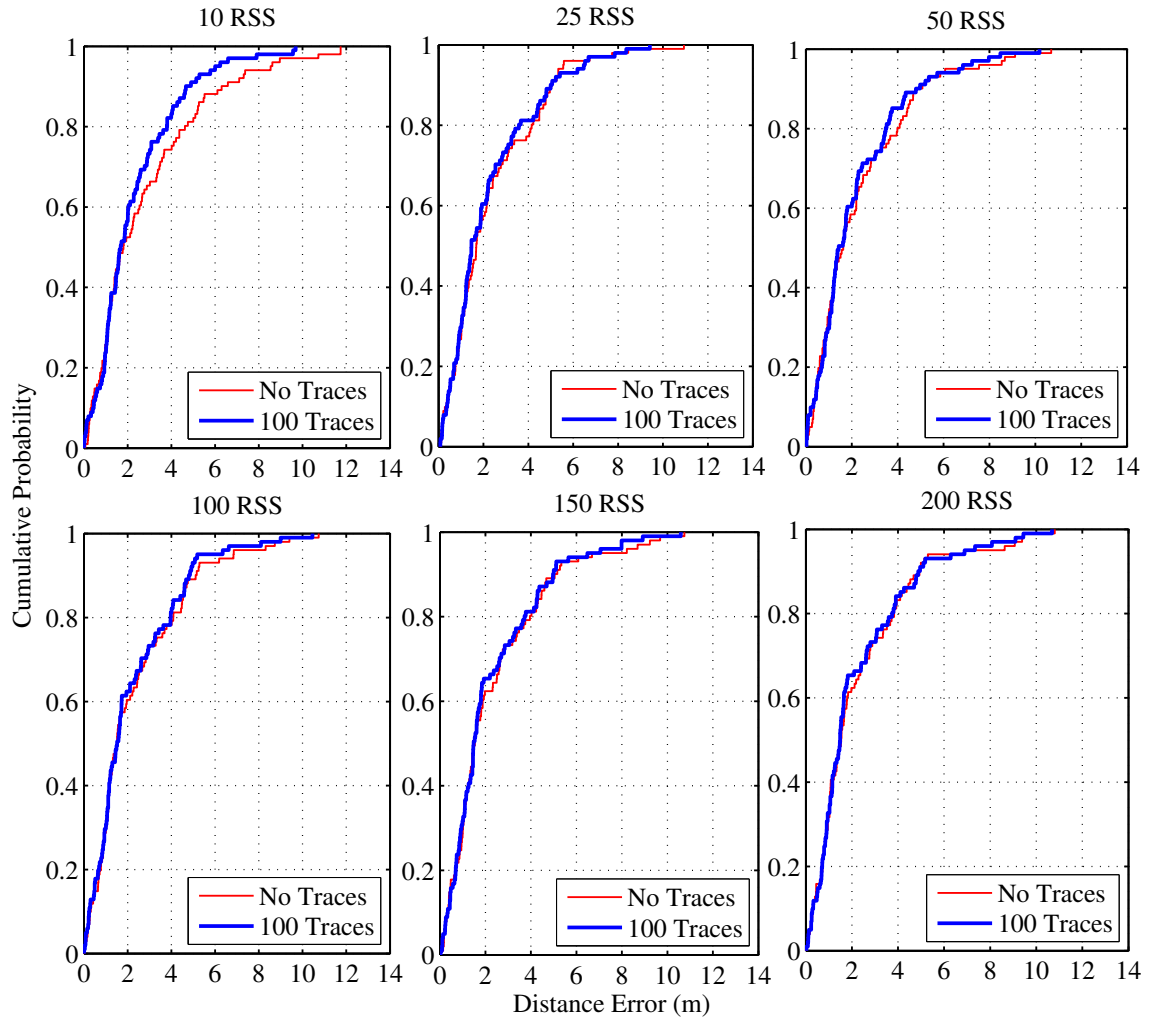

Figure 4.17: CDF for Annealed 3 meter Interpolated Turntable Radio Map
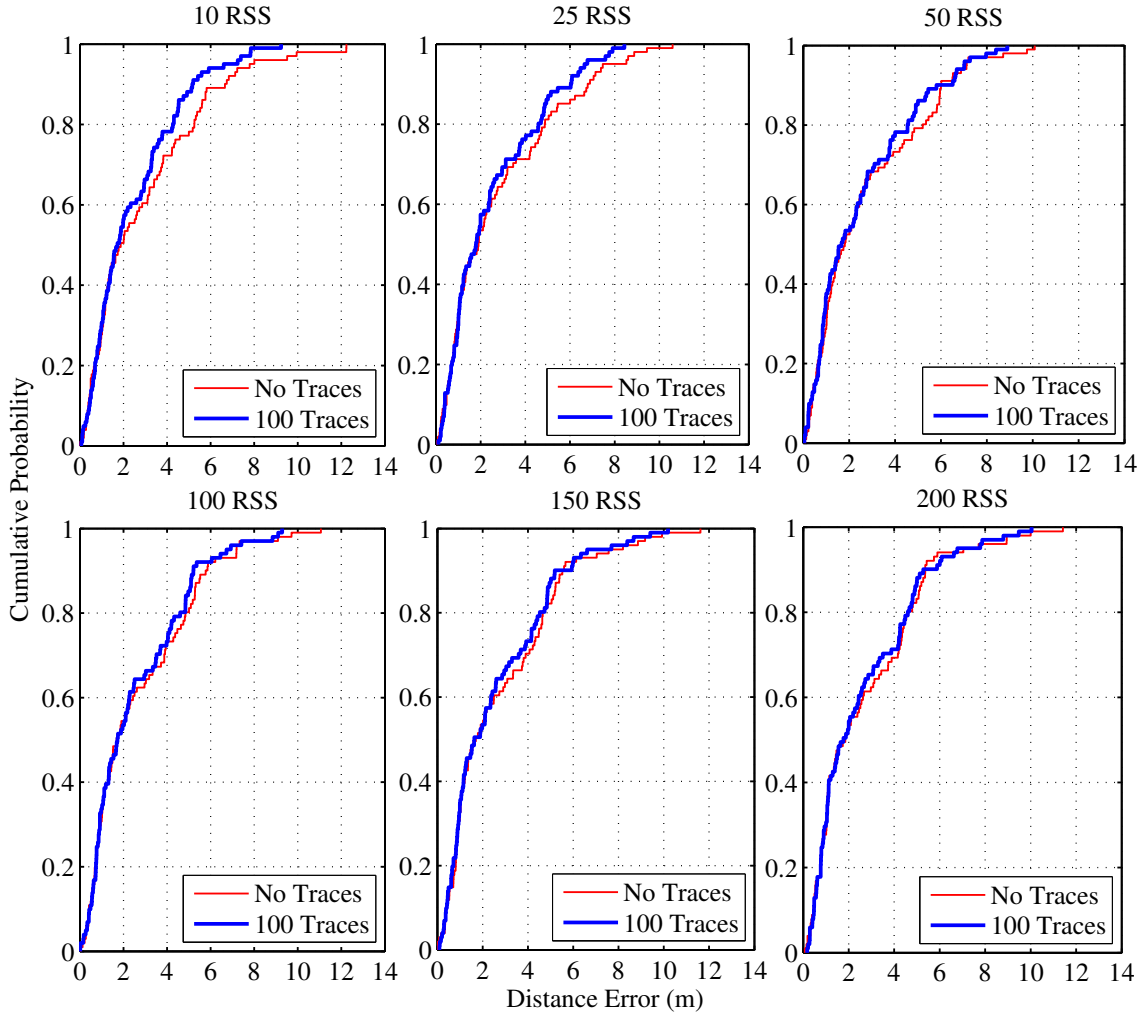
Figure 4.18: CDF for Annealed 4 meter Interpolated Turntable Radio Map

and is detailed as follows. The small number of initial fingerprint RSS measurements from the ANPs used for the interpolation of a complete turntable radio map capture a marginally small statistical view of the RSS-location dependency. This limited view was then transferred to the skipped ANP locations through the interpolation. However, the performance of the annealing algorithm in labeling the most probable sequence of states for each unknown asset trace was maintained regardless of the de-

ficiency in the interpolated radio map. The online RSS measurements are added into the initial interpolated radio map at each ANP, or state, corresponding to the now labeled trace sequences, thus increasing the number of fingerprint RSS measurements. Therefore, the increase in the accuracy of the MMSE location estimation corresponds to an improvement in the statistical view of the RSS-location dependency within the annealed radio map. As n is increased, the fingerprint RSS measurements used for interpolation capture more statistical variations in the RSS-location dependency, attributed to the rotation of the turntable setup during acquisition, and the addition of labeled RSS measurements have less influence. Ultimately, the annealing algorithm provided relatively small improvements to the interpolated GT radio maps.

4.3.3 Smoothing Factors of Interpolated Reference Point Radio Maps

The interpolation of radio maps in the previous section involved skipping ANPs in the original GT turntable radio map. The skipped ANPs were then interpolated from fingerprint RSS measurements of the two closest ANPs that were not skipped, resulting in a set of complete, interpolated turntable radio maps with the same number of ANPs, or states, as the original GT radio map. The significant difference in the RP configuration, as outlined in Fig. 3.18, is that there is no original radio map consisting of fingerprint RSS measurements at each ANP. Therefore, the RPs will be used to interpolate every ANP in each of the interpolated RP radio maps which results in the addition of a 1 meter interpolated RP radio map.

As was performed in Section 4.3.1 for MMSE location estimation, an investigation into the smoothing factor was conducted with variations in both s and n for each
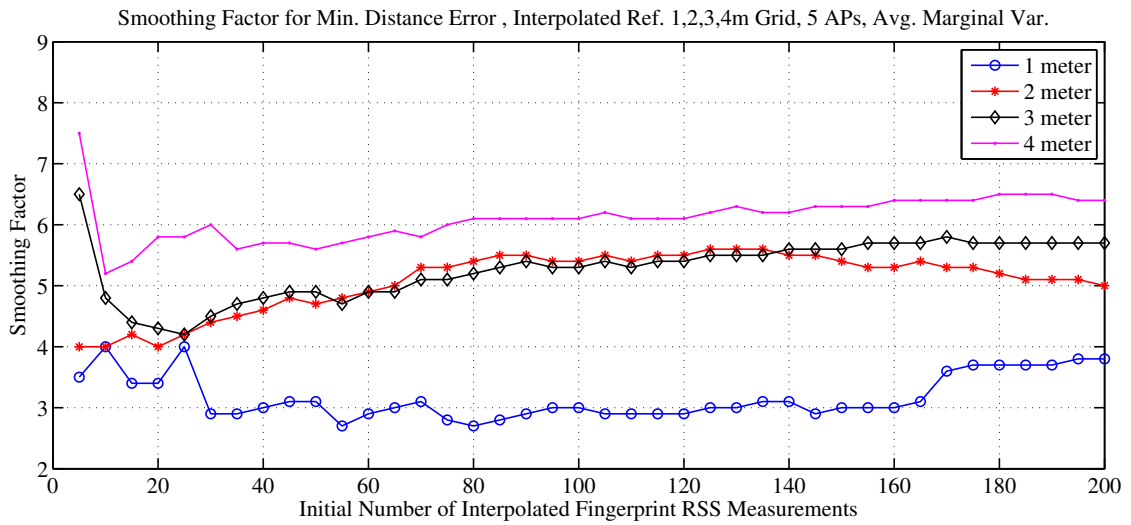
Figure 4.19: Variations in Smoothing Factor with Interpolated Reference Point Radio Map for Min. Distance Error
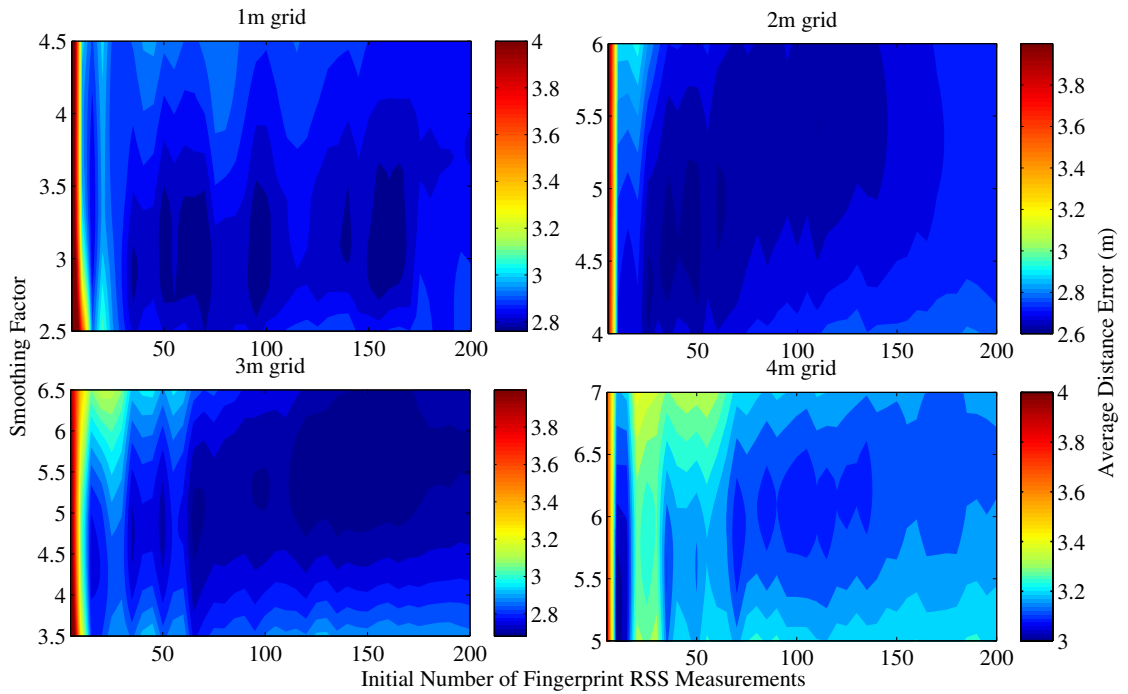
Figure 4.20: Variations in Smoothing Factor with Interpolated Reference Point Radio Map for Avg. Distance Error Contour

of the interpolated RP radio maps. The smoothing factor for each interpolated RP radio map was selected as the value of s resulting in the smallest location estimation distance error at each increment of n. The interpolated radio maps resulting from increases in the sparseness of RP locations are referred to as 1, 2, 3, and 4 meter for convenience as shown in Fig. 4.19. The results show that as the sparseness of RPs was increased, the smoothing factor also increased. When compared to the values of s calculated for the interpolated turntable radio maps, $s \approx 1$, these results seem higher than expected. This was attributed to the stationarity of the fingerprint RSS measurements acquired by the RPs, or their lack of statistical variation over time. This is most evident in a comparison of the 1 and 2 meter interpolated RP radio maps. The 1 meter radio map required two RPs with stationary RSS measurements for interpolation, with the RPs attached to opposing walls directly in line with each of the ANPs. The 2 meter radio map required the combination of 4 RPs with stationary RSS measurements for interpolation and thus encountered a higher smoothing factor than the 1 meter radio map. The 3 and 4 meter interpolated RP radio maps saw a slight increase in the smoothing factor attributed to the increase in sparseness of the RPs. The effects of deficient radio map interpolation can be seen for $n = 5$, where a small statistical view of the RSS-location dependency leads to generally higher values of s required to smooth the optimal Gaussian kernel bandwidth in the KDE required in MMSE location estimation.

With acceptance of these relatively high values of s, there is reassurance in the use of RPs for interpolation by noting that the values remain fairly consistent with increases in n. The ADE contours of the variation in s and n, shown in Fig. 4.20, illustrate relatively small variations in both the smoothing factor and ADE with increases

in n for each of the interpolated RP radio maps. Therefore, the use of an optimum smoothing factor, calculated as the average of s across the variation in n, can be utilized with the MMSE location estimation algorithm regardless of the selected value of n. Optimum values were calculated as the average of s across the variation in n for each radio map and are shown in Table 4.8. These optimum smoothing values were fixed within the MMSE location estimation algorithm during the investigation of annealing the interpolated RP radio maps.

4.3.4 Annealing of Interpolated Reference Point Radio Maps

The annealing of each interpolated RP radio map was performed with the identical HMM parameters used for the annealing of the interpolated turntable radio maps outlined in Section 4.3.2. Variations were performed on the state transition distance constraint, for $m = 1 \ldots 4$, as well as the initial number of fingerprint RSS measurements n. The percentage decrease in ADE from the annealing of each of the interpolated RP radio maps can be seen in Fig. 4.21 - 4.24.

The results of the 1 meter interpolated RF radio map, shown in Fig. 4.21, conform to the results presented for the interpolated turntable radio map, for values of

Interpolated Reference Point Radio Maps Optimum Values of Smoothing Factor				
	1 meter	2 meter	3 meter	4 meter
s	3.175	5.078	5.278	6.105

Table 4.8: Interpolated Reference Point Radio Maps - Values of Smoothing Factor

$n > 10$. The percentage decrease in ADE remained relatively consistent with values approximated between $5 - 9\%$. This was attributed to the use of only 2 RPs, with relatively stationary fingerprint RSS measurements, in the interpolation of the 1 meter radio map. Small values of n resulted in a dramatic precentage decrease in the ADE of $> 25\%$. This indicated that the performance of the annealing algorithm in labeling the most probable sequence of states for each unknown asset trace was maintained regardless of the deficiency in the interpolated radio map resulting from the small statistic view of the RSS-location dependency. With a value of $m = 1$, the annealing algorithm performed lower, for all interpolated RP radio maps, which was to be expected and was explained in the previous section for the annealing of the interplated turntable radio maps where the HMM state machine could not reflect transitions greater than 1 meter per second. The precentage decrease in ADE for the annealing of the 2, 3, and 4 meter interpolated RP radio maps followed a similar

(a)Plot   (b)Contour

Figure 4.21: Annealing of 1 meter Interpolated Reference Point Radio Map
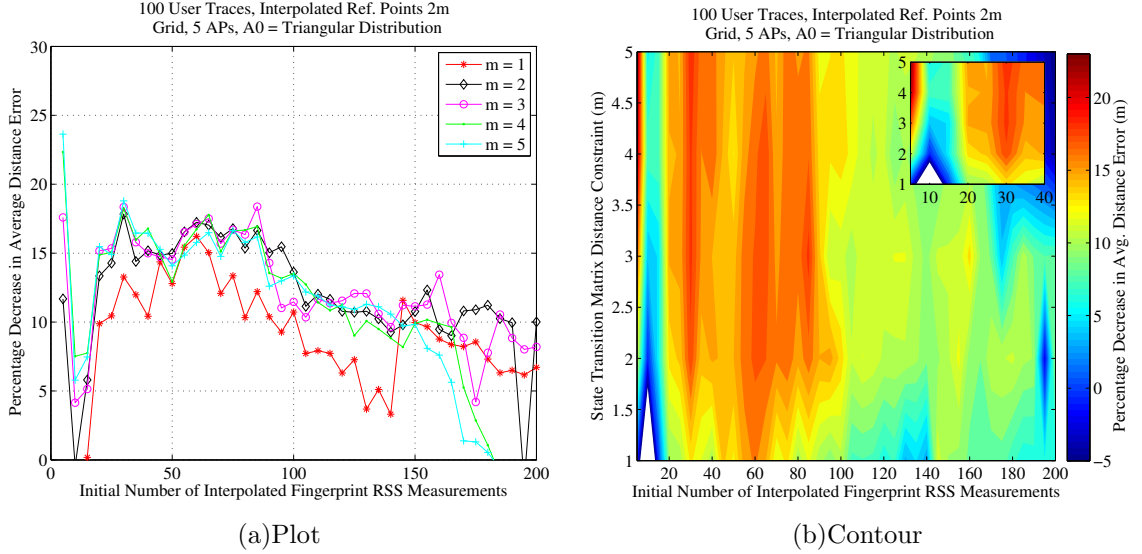
Figure 4.22: Annealing of 2 meter Interpolated Reference Point Radio Map

pattern for values $n \leq 10$. With greater increases in n, the precentage decrease in ADE remained remarkably high. This reiterated the fact that the annealing algorithm maintained high performance in labeling the most probable sequence of states for each unknown asset trace. This dramatic result is attributed to the use of 4 RPs, each with relatively stationary fingerprint RSS measurements, in the interpolation of the 2, 3, and 4 meter radio maps. These deficient radio maps maintained relatively small statistical views of the RSS-location dependency, regardless of the value of n, throughout the interpolation. Again, this indicated that the annealing alogrithm is able to dynamically improve the initial radio maps with a high degree of accuracy given relatively inaccurate radio maps.

As was presented for the annealing of the interpolated turntable radio maps, the

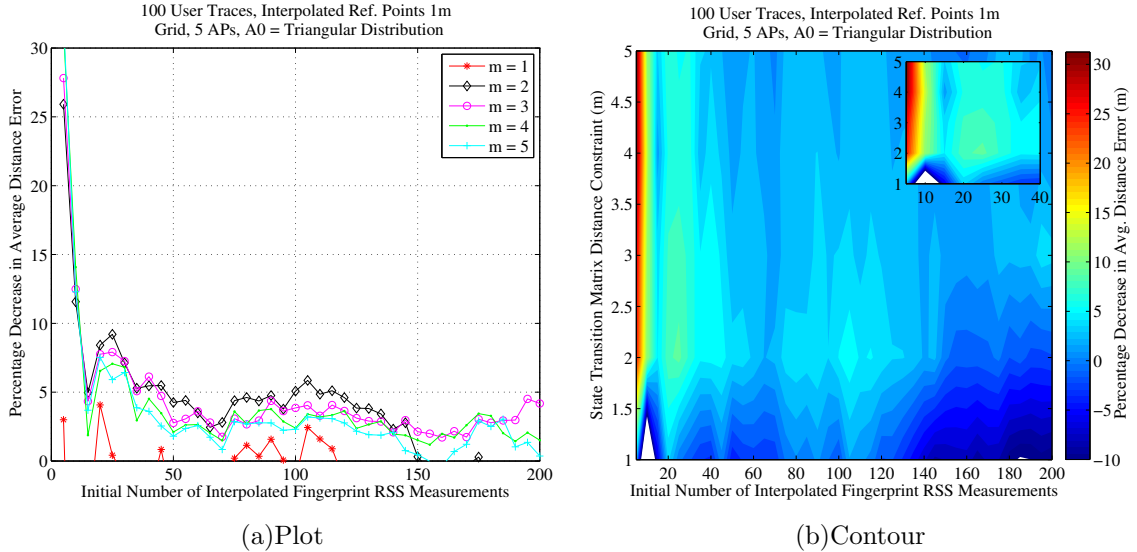(a)Plot

(b)Contour
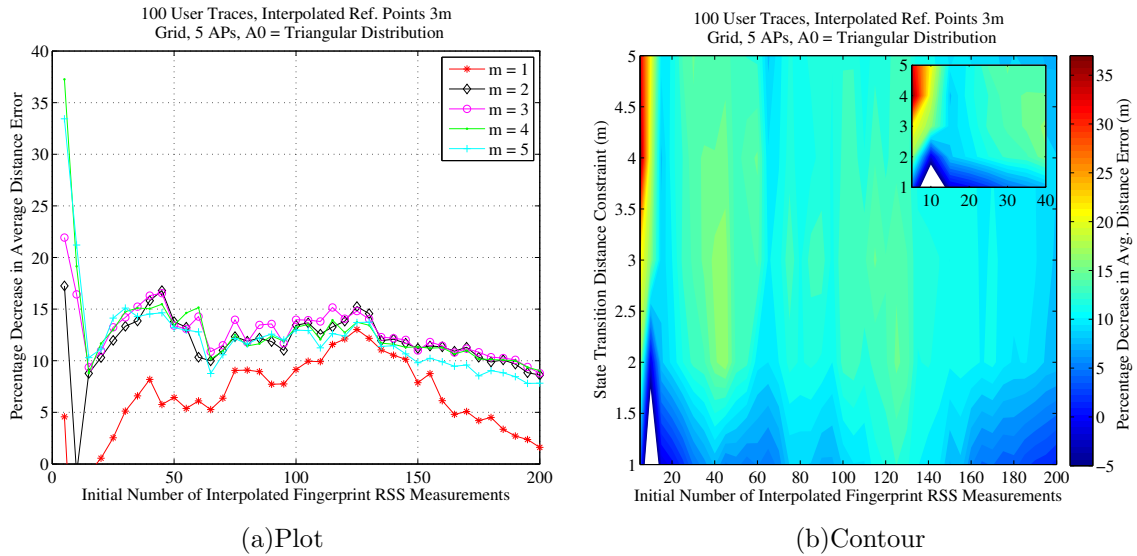
Figure 4.23: Annealing of 3 meter Interpolated Reference Point Radio Map

(a)Plot

(b)Contour
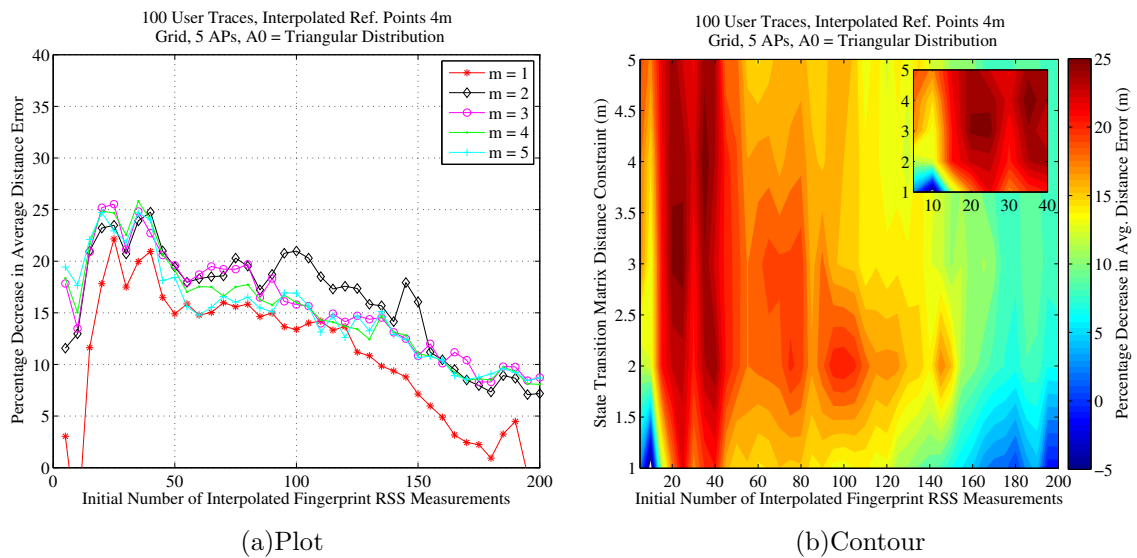
Figure 4.24: Annealing of 4 meter Interpolated Reference Point Radio Map

MMSE Localization Distance Errors (m) - Interpolated
Reference Point Radio Maps - With and Without Annealing - 5 APs Selected
by Highest Freq. RSS Measurements at Online Locations, m = 3

No Traces

# RSS	1m			2m			3m			4m		
	AVG	50 %	70 %	AVG	50 %	70 %	AVG	50 %	70 %	AVG	50 %	70 %
10	3.17	2.66	3.92	2.87	2.45	3.47	3.56	2.87	4.47	3.40	3.04	4.50
25	2.93	2.40	3.80	3.02	2.43	3.59	3.01	2.38	3.66	3.89	2.71	4.21
50	2.77	2.19	3.34	2.98	2.40	3.66	2.98	2.13	3.46	3.58	2.35	4.08
100	2.78	2.20	3.10	3.05	2.41	3.56	3.01	2.17	3.61	3.74	2.33	4.35
150	2.80	2.33	3.18	3.06	2.52	3.62	3.02	2.23	3.50	3.64	2.36	4.18
200	2.88	2.26	3.41	3.04	2.58	3.70	2.99	2.15	3.60	3.61	2.56	4.24

100 Traces

# RSS	1m			2m			3m			4m		
	AVG	50 %	70 %	AVG	50 %	70 %	AVG	50 %	70 %	AVG	50 %	70 %
10	2.78	2.29	3.09	2.75	2.59	3.31	2.97	2.62	3.70	2.94	2.60	3.32
25	2.70	2.19	3.17	2.55	2.14	3.12	2.61	2.18	3.29	2.90	2.35	3.57
50	2.70	2.22	3.07	2.55	2.07	2.90	2.58	2.13	2.90	2.88	2.31	3.54
100	2.67	2.16	3.23	2.70	2.04	3.12	2.59	2.01	3.29	3.15	2.19	3.43
150	2.74	2.16	3.40	2.72	2.15	3.06	2.69	1.99	3.35	3.24	2.24	3.55
200	2.76	2.21	3.35	2.79	2.19	3.34	2.73	2.02	3.23	3.29	2.34	3.54

Table 4.9: MMSE Distance Errors - With and Without Annealing - Interpolated Reference Point Radio Maps

MMSE Localization Standard Deviations (m) Interpolated Reference Point Radio Maps With and Without Annealing - 5 APs, m = 3				
	No Traces			
	1m	2m	3m	4m
# RSS	AVG	AVG	AVG	AVG
10	2.40	2.34	3.23	2.51
25	2.31	2.55	2.75	3.64
50	2.27	2.48	2.85	3.50
100	2.29	2.58	2.70	3.66
150	2.25	2.43	2.74	3.46
200	2.29	2.41	2.74	3.47
	100 Traces			
	1m	2m	3m	4m
# RSS	AVG	AVG	AVG	AVG
10	1.87	1.81	2.04	2.01
25	2.00	1.92	1.93	2.15
50	2.07	2.07	2.07	2.28
100	2.22	2.37	2.37	3.06
150	2.20	2.44	2.44	3.21
200	2.24	2.49	2.49	3.15

Table 4.10: MMSE Standard Deviations - With and Without Annealing - Interpolated Reference Point Radio Maps

overall changes in the MMSE location estimation distance errors and standard deviations can be visualized in Table 4.9 - 4.10 for each of the interpolated RP radio maps and $m = 3$. With $n = 10$, which is approximately a 20 fold decrease in the number of interpolated RP fingerprint RSS measurements at each ANP, the ADE decreases with the use of each interpolated RP radio map. Annealing the 1, 2, 3, and 4 meter interpolated radio maps results in a decrease in the ADE of MMSE location estimation from $3.17 \rightarrow 2.78$, $2.78 \rightarrow 2.75$, $3.56 \rightarrow 2.62$, and $3.40 \rightarrow 2.94$ meters or

12.3%, 1.1%, 26.4%, and 13.5% resectively. With $n = 100$, which is approximately a 2 fold decrease in the number of interpolated RP fingerprint RSS measurements at each ANP, annealing the 1, 2, 3, and 4 meter interpolated radio maps results in a decrease in the ADE from $2.78 \to 2.67$, $3.05 \to 2.70$, $3.01 \to 2.59$, and $3.74 \to 3.15$ meters or 4.0%, 12.5%, 14.0%, and 15.8% respectively.

Figure 4.25: CDF for Annealed 1 meter Interpolated Reference Point Radio Map
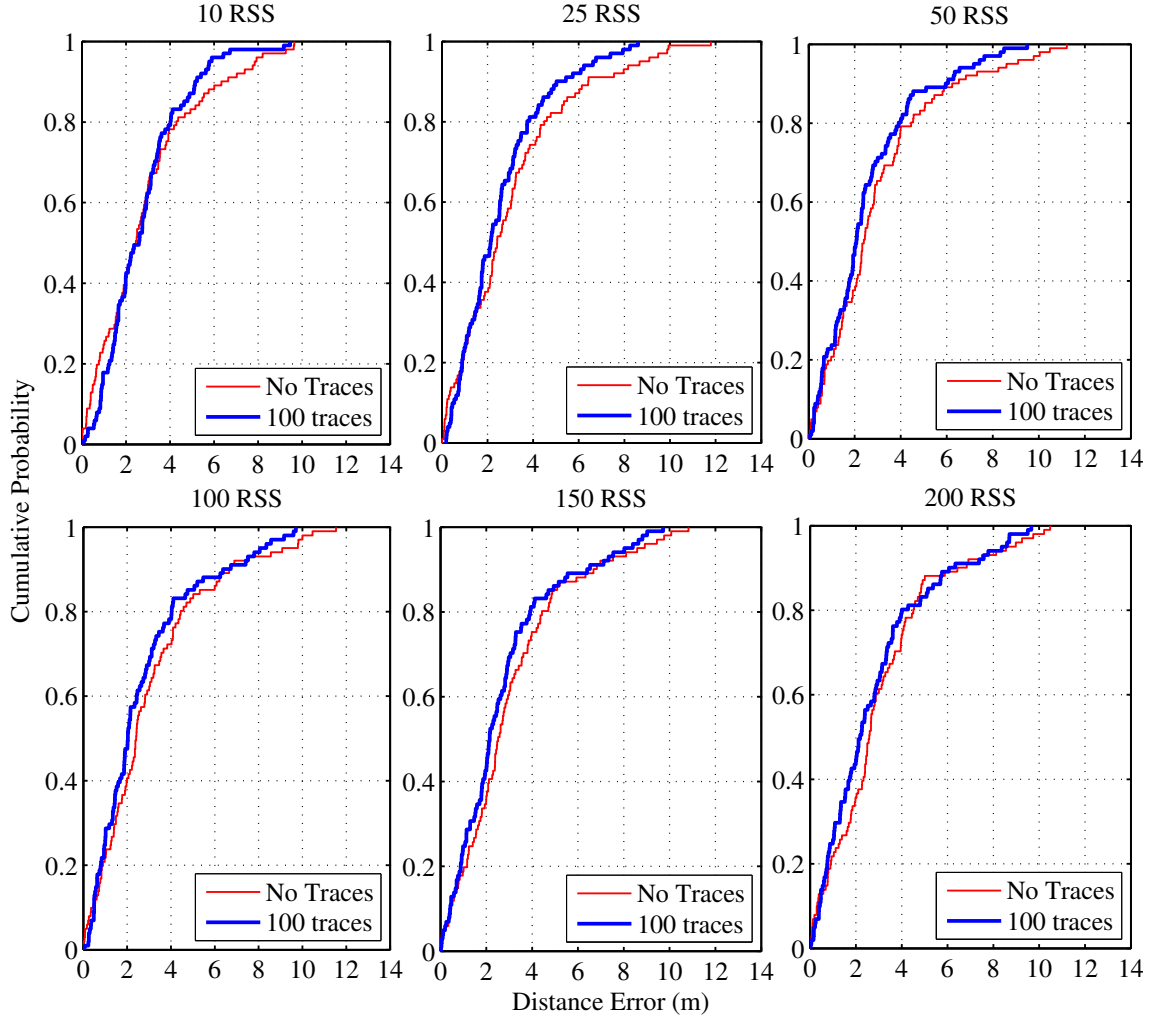
Figure 4.26: CDF for Annealed 2 meter Interpolated Reference Point Radio Map

The distance errors presented in Table 4.9 can be seen in Fig. 4.25 - 4.28 where a CDF was used to visualize the changes in distance error, with and without radio map annealing, for increments in n. An interesting fact is that with an increase in the sparseness of RPs, the annealing of the interpolated radio maps resulted in a continued increase in the accuracy of the MMSE location estimation. The effect is most

prominent with $n = 10$ and can be attributed to deficient radio map interpolation. The small number of initial fingerprint RSS measurements, from the RPs used for the interpolation of a complete RP radio map, captured a marginally small statistical view of the RSS-location dependency. However, the performance of the annealing algorithm in labeling the most probable sequence of states for each unknown asset trace was maintained regardless of the deficiency in the interpolated radio map. The
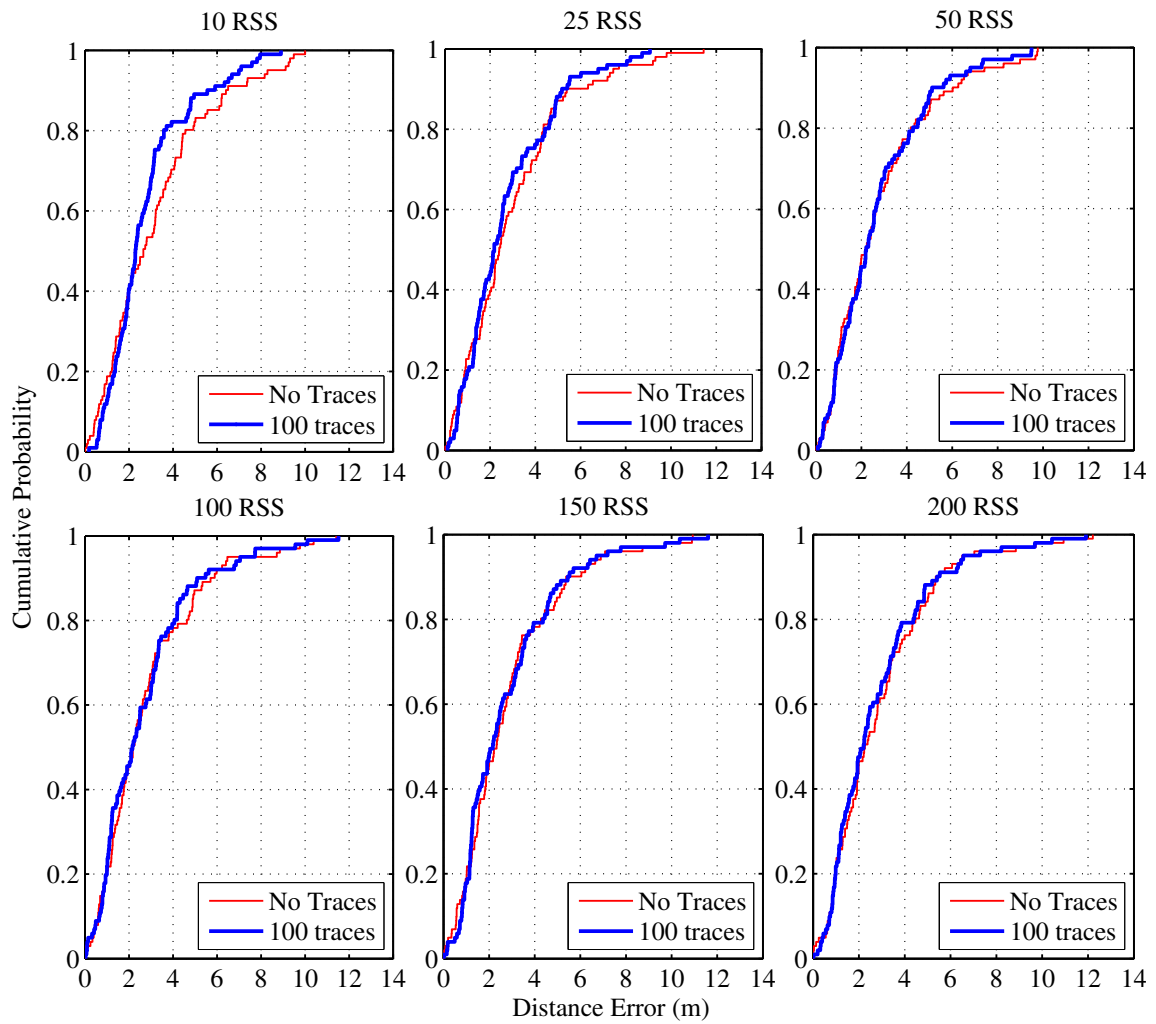

Figure 4.27: CDF for Annealed 3 meter Interpolated Reference Point Radio Map
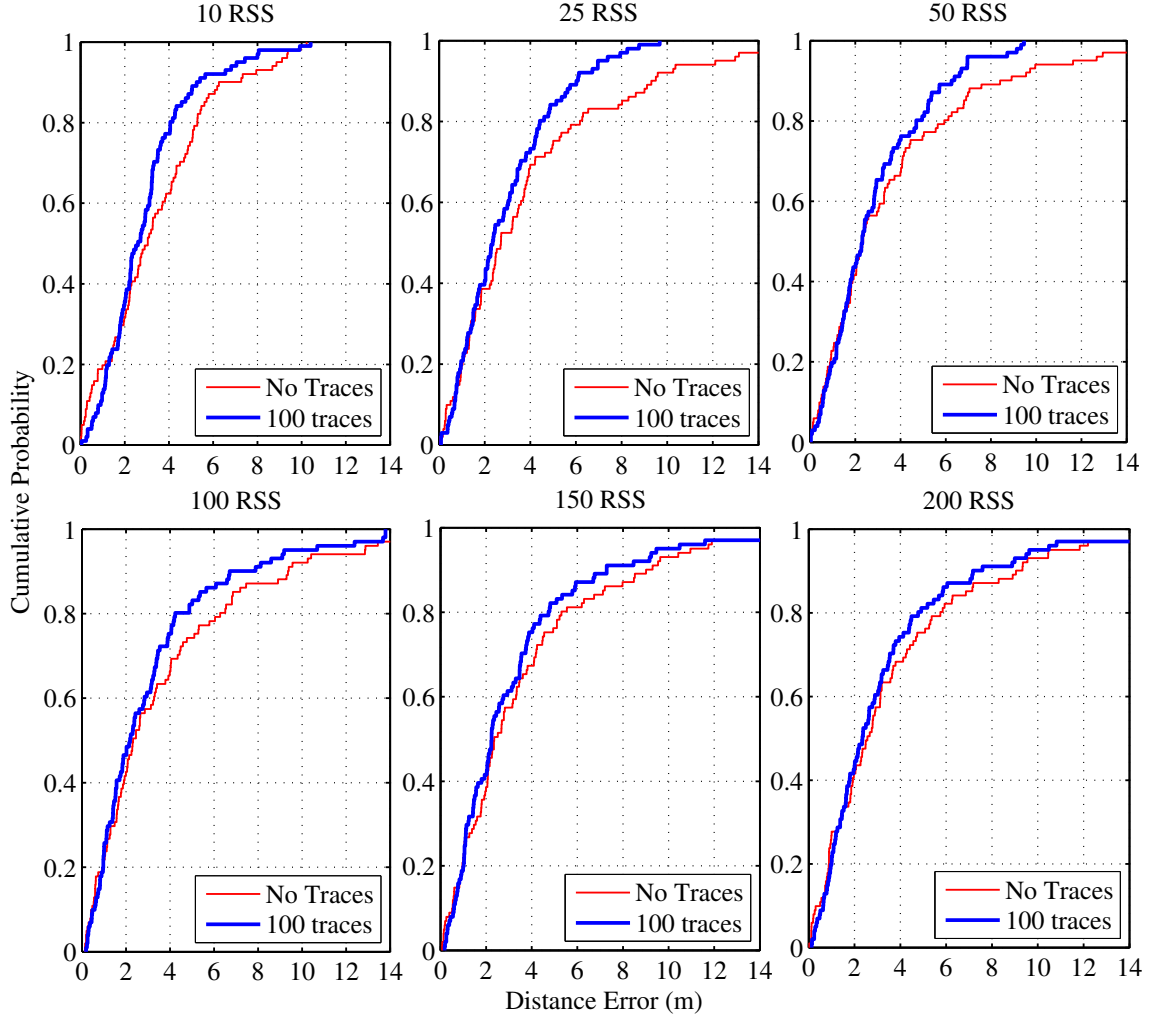
Figure 4.28: CDF for Annealed 4 meter Interpolated Reference Point Radio Map

online RSS measurements are added into the initial interpolated radio map at each ANP, or state, corresponding to the now labeled trace sequences, thus increasing the number of fingerprint RSS measurements. Therefore, the increase in the accuracy of the MMSE location estimation corresponds to an improvement in the statistical view of the RSS-location dependency within the annealed radio map. As the sparseness of RPs and n were increased, the performance of the annealing algorithm remained

substantially higher than expected. This was attributed to a continued deficiency in the interpolated RP radio maps, where the fingerprint RSS measurements of the 4 RPs used for interpolation did not accurately capture the RSS-location dependency at each ANP location. However, the accuracy of labeling of the most probable sequence of states for each unknown asset trace remained high, therefore the multiple online RSS measurements added into the deficient interpolated radio maps improved their statistical view and RSS-location dependency.

4.3.5   Remarks

The determination of the optimum smoothing factors, for both the interpolated turntable and RP radio maps, were explored in terms of their effects on the MMSE location estimation ADE. Marginal fluctuations in s, with increases in n, indicated that the average of value of s, which resulted from the smallest ADE at each increment in n, could be used as the optimal value of s for each interpolated radio map. These optimal values were fixed for the analysis of the annealing algorithm.

The performance of the annealing algorithm, with the use of both interpolated turntable and PR radio maps, was analyzed for variations in sparseness and in the value of n. The results showed that the annealing algorithm was able to accurately label the most probable sequence of states for each unknown asset traces regardless of the level of deficiency in each of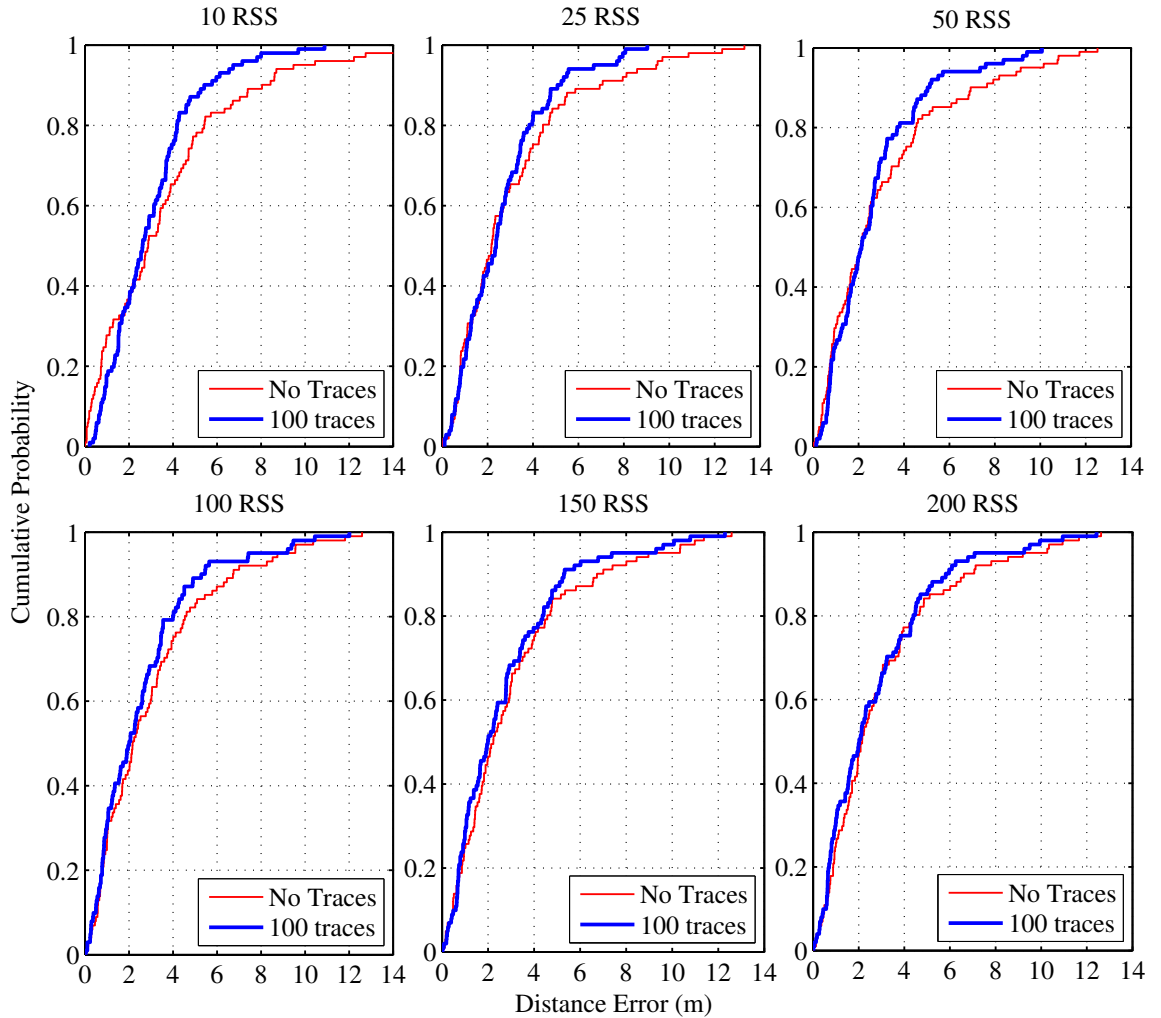 the interpolated radio maps. The annealing of the interpolated RP radio maps provided the largest percentage decreases in the ADE over the largest ranges of n. This was most notable for the 4 meter interpolated RP radio map as shown in Fig. 4.28. The results indicate that the proposed interpolation and annealing algorithms, combined with a sparse RP site survey configuration, can be

used as a viable alternative to traditional fingerprinting techniques while maintaining relatively high location estimation performance.

4.4 Discussion

An analysis of both range-free location estimation approaches was conducted for the determination of optimal values of the parameters K and s utilizing the GT turntable radio map. Two methods for the selection of APs within each location estimation algorithm were compared to determine which resulted in the smallest ADE with variations in n. It was determined that the selection of $d = 5$ APs by the highest frequency of online RSS measurements performed best. This method was used throughout the analysis of radio map annealing via dynamic asset traces.

The interpolation algorithm was utilized on two different schemes: increased sparseness of fingerprint ANPs from the original GT radio map and increased sparseness of RPs affixed to opposing walls of the original ANPs. The interpolation of each radio map, regardless of the scheme, resulted in a complete radio map of fingerprints at each of the 71 original ANP locations. The smoothing factor was analyzed for each of the interpolated radio maps with variation in the initial number of fingerprint RSS measurements. Optimum values of s were then calculated for the use in the annealing algorithm and the MMSE location estimation of all static online locations. It is important to note that the optimum values of K and s, for WNNSS and MMSE, were determined from a single data set specific to the location estimation environment used for this validation. Changes to the locations of APs, building modifications, or the density of personal would alter the influences of shadowing on the fingerprint RSS measurements, thus resulting in inaccurate optimum values.

Multiple unknown asset traces were utilized within the annealing algorithm to study its effects on the MMSE location estimation distance error for increases in sparseness of both the interpolated turntable and interpolated RP radio maps. It is interesting to note, for both interpolated radio map configurations, the percentage decrease in ADE remained relatively consistent with increases to m, the state-transition distance constraint. Therefore, a system designer could implement different HMMs, for different assets, based on their speculated motion dynamics with pre-selected values of m to capture the assets maximum anticipated speed throughout the environment.

Chapter 5

Conclusion

This thesis presented the formulation of an indoor, range-free, probabilistic location estimation system, with the focus of improving upon the traditional approach to site surveying an environment. This was accomplished through the reduction in the sample time and density of fingerprint locations, which resulted in sparse, inaccurate radio maps. Interpolation was used to construct complete radio maps from the fingerprint RSS measurements collected at the sparse locations. An annealing algorithm was used to dynamically improve these radio maps with multiple unknown asset traces over time. The analysis of the proposed range-free location estimation system provided results that indicate improvements to the sparse, inaccurate radio maps can be achieved over time while also maintaining accurate location estimation. This range-free location estimation system would be ideal for situations, for example a hospital, where traditional site surveying approaches are considered too invasive and unrealistic within the environment.

5.1 Contributions

The objectives for this thesis, as outlined in Section 1.2, were accomplished and a summary of the contributions are as follows:

1. Two range-free location estimation approaches, deterministic WNNSS and probabilistic MMSE, were developed and implemented for a performance comparison, in terms of their ADE, with variations in the initial number of fingerprint RSS measurements, the parameters K and s, and the selection of APs. The results indicated comparable performance in the two approaches, with each relying on optimal values of their selectable parameters for minimum ADE and the optimal selection method of APs. However, due to the useful Bayesian statistics provided by the probabilistic MMSE formulation, this approach was chosen for further development of the location estimation system .

2. A reduction in the density of fingerprint locations, as outlined in Section 3.6.1, was referred to as an increase in *sparseness*. The removal of every 2nd, 3rd, and 4th location resulted in a decrease in the density to 1/2, 1/3, and 1/4 of the original locations, be it ANP locations from the original GT radio map or locations of the RPs.

3. The interpolation was implemented with the IDW algorithm, as outlined in Section 3.6.1.1, which utilized fingerprint RSS measurements from scattered locations. The algorithm weighted the fingerprint RSS measurements from each of the geometrically selected sparse locations solely based on their distance to each ANP of interest within the radio map.

4. The radio map annealing algorithm, as outlined in Section 3.6.2, was derived from a 1st order HMM finite state machine. A GEM algorithm was utilized to train the HMM parameters with multiple unknown asset traces. The trained parameters represented the most probable sequences of states, for each trace, and were used to label the online RSS measurements to their specific ANP lo-

cations. The labeled online RSS measurements were then added to interpolated radio maps at each of their labeled fingerprint ANP locations respectively.

5. An analysis was performed with variations in the initial number of interpolated fingerprint RSS measurements, with and without annealing, to determine if any improvements could be made to the interpolated radio maps, over time. The results indicated, with an initially small number of fingerprint RSS measurements, that the percentage decreases in the ADE can be significant for sparse, interpolated RP radio maps while maintaining relatively accurate location estimation.

Based on the positive outcome of the experimental validation that demonstrated a consistent ADE of approximately 2-3 meters, it is evident that range-free location estimation based on fingerprint RSS measurements are sufficient for enabling practical applications of indoor localization. Furthermore, the caveat of using only a single data set for the experimental validation was addressed and supported by the large, extensive site survey consisting of 6 APs and 71 ANPs adjoined over 3 distinct hallways.

5.2   Future Work

Future work could be explored on many different aspects of the proposed range-free probabilistic location estimation system. Different localization sites should be considered for further validation of the algorithms. Three noteworthy directions will be discussed that could add significant improvements to overall system.

First, the process of constructing initial radio maps, for use with the annealing algorithm, could be based on a blind site survey or static propagation modeling of

the localization environment. In a blind site survey, a user would collect fingerprint RSS measurements while walking pre-defined traces throughout an environment. The RSS measurements would then be inferred to radio map ANP locations along each specific trace. This would eliminate the need for multiple static fingerprint locations and additional hardware, thus reducing invasive site survey procedures throughout an environment. In static propagation modeling, a system could utilize the range equation, outlined in Section 2.4, to calculate RSS measurements using heuristically estimated PL exponents, which would vary depending on the location of each ANP in a radio map relative to each AP in the environment.

Second, the implementation of Bayesian tracking algorithms could be explored to calculate the posterior density location estimate recursively, modeling asset movements with velocity or acceleration parameters. These are typically referred to as dynamic motion models [22]. The formulation of the MMSE location estimate, along with its covariance derived in Section 3.5.2.1, can be used within these posterior density estimates and was successfully demonstrated within an inverse covariance filter in [22].

To try and alleviate the effects of attenuation from a human body, as illustrated in the radiations patterns of Section 3.2, an orientation variable could be incorporated into the location estimate to track an assets trajectory throughout an environment.

Furthermore, the HMM model parameters, which were trained via GEM with multiple unknown asset traces, can be used within a HMM filter to predict a current posterior density estimate based on the previous estimate. Once RSS measurements have been observed, the belief in the current posterior density estimate can be updated. Essentially, a prediction can be computed as to where an asset may be at

time t, based on knowledge of the previous location at $t - 1$, with an update of the prediction computed on new RSS observations.

Third, the integration of RSS measurements with other signals of opportunity from multiple sensors such as radio frequency identification tags or inertial global positioning systems (GPS) to attempt and improve the accuracy of location estimation. Distributed, cooperative sensor networks could also be explored.

Bibliography

[1] Ekahau, "Real-time location systems," 2012. [Online]. Available: http://www.ekahau.com

[2] Cisco, "Wherenet: Real-time location systems," 2012. [Online]. Available: http://www.zebra.com/us/en/solutions/technology-need/wherenet.html

[3] AeroScout, "Real-time location systems," 2012. [Online]. Available: http://www.aeroscout.com

[4] M. Matrix, "Xsight: Real-time location system software," 2010. [Online]. Available: http://www.mobilematrix.com/product-overview/

[5] T. Rappaport, *Wireless Communications: Principles and Practice*, 2nd ed. Prentice Hall, 2002.

[6] B.-C. Liu, K.-H. Lin, and J.-C. Wu, "Analysis of hyperbolic and circular positioning algorithms using stationary signal-strength-difference measurements in wireless communications," *Vehicular Technology, IEEE Transactions on*, vol. 55, no. 2, pp. 499–509, march 2006.

[7] S. Mazuelas, A. Bahillo, R. M. Lorenzo, P. Fernandez, F. A. Lago, E. Garcia, J. Blas, and E. J. Abril, "Robust indoor positioning provided by real-time rssi values in unmodified wlan networks," *IEEE Journal of Selected Topics in Signal*

Processing, vol. 3, no. 5, pp. 821–831, 2009.

[8] G. Delisle, J.-P. Lefevre, M. Lecours, and J.-Y. Chouinard, "Propagation loss prediction: A comparative study with application to the mobile radio channel," *Vehicular Technology, IEEE Transactions on*, vol. 34, no. 2, pp. 86–96, may 1985.

[9] A. Bahillo, S. Mazuelas, J. Prieto, P. Fernandez, R. M. Lorenzo, and E. J. Abril, "Hybrid rss-rtt localization scheme for wireless networks," *International Conference on Indoor Positioning and Indoor Navigation (IPIN)*, pp. 1–7, 2010.

[10] H. Laitinen, S. Juurakko, T. Lahti, R. Korhonen, and J. Lahteenmaki, "Experimental evaluation of location methods based on signal-strength measurements," *Vehicular Technology, IEEE Transactions on*, vol. 56, no. 1, pp. 287–296, jan. 2007.

[11] N. Patwari, A. O. Hero III, M. Perkins, N. S. Correal, and R. J. O'Dea, "Relative location estimation in wireless sensor networks," *IEEE Transactions on Signal Processing*, vol. 51, no. 8, pp. 2137–2148, 2003.

[12] X. Li, "Rss-based location estimation with unknown pathloss model," *IEEE Transactions on Wireless Communications*, vol. 5, no. 12, pp. 3626–3633, 2006.

[13] Y. Qi, "Wireless geolocation in a non-line-of-sight environment," *Ph.D. dissertation, Princeton Univ.*, 2003.

[14] N. Chang, R. Rashidzadeh, and M. Ahmadi, "Robust indoor positioning using differential wi-fi access points," *Consumer Electronics, IEEE Transactions on*, vol. 56, no. 3, pp. 1860–1867, aug. 2010.

[15] H. Hashemi, "The indoor radio propagation channel," *Proceedings of the IEEE*, vol. 81, no. 7, pp. 943–968, jul 1993.

[16] Y. Gwon, R. Jain, and T. Kawahara, "Robust indoor location estimation of stationary and mobile users," in *INFOCOM 2004. Twenty-third AnnualJoint Conference of the IEEE Computer and Communications Societies*, vol. 2, march 2004, pp. 1032–1043.

[17] M. M. Atia, A. Noureldin, and M. J. Korenberg, "Dynamic propagation modeling for mobile users' position and heading estimation in wireless local area networks," *IEEE Wireless Communication Letters*, vol. 1, no. 2, pp. 101–104, 2012.

[18] S. Liu, Y. Chen, W. Trappe, and L. J. Greenstein, "Non-interactive localization of cognative radios based on dynamic signal strength mapping," *International Conference on Wireless On-Demand Network Systems and Services (WONS)*, pp. 85–92, 2009.

[19] A. Kushki, K. N. Plataniotis, and A. N. Venetsanopoulos, "Kernel-based positioning in wireless local area networks," *IEEE Transactions on Mobile Computing*, vol. 6, no. 6, pp. 689–705, 2007.

[20] ——, "Intelligent dynamic radio tracking in indoor wireless local area networks," *IEEE Transactions on Mobile Computing*, vol. 9, no. 3, pp. 405–419, 2010.

[21] A. Kushki, K. Plataniotis, and A. Venetsanopoulos, "Location tracking in wireless local area networks with adaptive radio maps," in *Acoustics, Speech and Signal Processing, 2006. ICASSP 2006 Proceedings. 2006 IEEE International Conference on*, vol. 5, may 2006.

[22] A. Kushki, "A cognitive radio tracking system for indoor environments," *Ph.D. dissertation, University of Toronto*, 2008.

[23] M. McGuire, K. N. Plataniotis, and A. N. Venetsanopoulos, "Location of mobile terminals using time measurements and survey points," *IEEE Transactions on*

Vehicular Technology, vol. 52, no. 4, pp. 999–1011, 2003.

[24] P. Bahl and V. Padmanabhan, "Radar: an in-building rf-based user location and tracking system," in *INFOCOM 2000. Nineteenth Annual Joint Conference of the IEEE Computer and Communications Societies. Proceedings. IEEE*, vol. 2, 2000, pp. 775–784.

[25] K. Connelly, Y. Liu, D. Bulwinkle, A. Miller, and I. Bobbitt, "A toolkit for automatically constructing outdoor radio maps," in *ITCC 2005*, vol. 2, april 2005, pp. 248–253.

[26] X. Chai and Q. Yang, "Reducing the calibration effort for probabilistic indoor location estimation," *IEEE Transactions on Mobile Computing*, vol. 6, no. 6, pp. 649–662, 2007.

[27] J. Bardwell, "You believe you understand what you think i said: The truth about 802.11 signal and noise metrics," Connect802 Corporation, Tech. Rep., 2004.

[28] C. Yeh, H. Lim, and D. Kwon, "Rssi measurements," in *IEEE 802.16 Broadband Wireless Access Working Group*, dec 2003, pp. 1–3.

[29] S. Lee, Y. Song, and S. Nam, "Fast rssi circuit using novel power detector for wireless communication," in *SoC Design Conference, 2008. ISOCC '08. International*, vol. 01, nov. 2008, pp. 9–12.

[30] P. Bahl and V. N. Padmanabhan, "A software system for locating mobile users: Design, evaluation, and lessons," 2000, microsoft Research.

[31] K. Kaemarungsi and P. Krishnamurthy, "Properties of indoor received signal strength for wlan location fingerprinting," in *Mobile and Ubiquitous Systems: Networking and Services, 2004. MOBIQUITOUS 2004. The First Annual International Conference on*, aug 2004, pp. 14–23.

[32] K. El-Kafrawy, M. Youssef, A. El-Keyi, and A. Naguib, "Propagation modeling for accurate indoor wlan rss-based localization," in *Vehicular Technology Conference Fall (VTC 2010-Fall), 2010 IEEE 72nd*, sept. 2010, pp. 1–5.

[33] P. Krishnan, A. Krishnakumar, W.-H. Ju, C. Mallows, and S. Gamt, "A system for lease: location estimation assisted by stationary emitters for indoor rf wireless networks," in *INFOCOM 2004. Twenty-third AnnualJoint Conference of the IEEE Computer and Communications Societies*, vol. 2, march 2004, pp. 1001–1011.

[34] L. Peng, J. Han, W. Meng, and J. Liu, "Research on radio-map construction in indoor wlan positioning system," in *Pervasive Computing Signal Processing and Applications (PCSPA), 2010 First International Conference on*, sept. 2010, pp. 1073–1077.

[35] T. Roos, P. Myllymaki, H. Tirri, P. Misikangas, and J. Sievanen, "A probabilistic approach to wlan user location estimation," *International Journal of Wireless Information Networks*, vol. 9, no. 3, pp. 155–164, 2002.

[36] M. Youssef, A. Agrawala, and A. Udaya Shankar, "Wlan location determination via clustering and probability distributions," in *Pervasive Computing and Communications, 2003. (PerCom 2003). Proceedings of the First IEEE International Conference on*, march 2003, pp. 143–150.

[37] M. Youssef and A. Agrawala, "The horus wlan location determination system," in *Proceedings of the 3rd international conference on Mobile systems, applications, and services*, ser. MobiSys '05, 2005, pp. 205–218.

[38] D. W. Scott, *Multivariate Density Estimation*. Wiley-Interscience, 1992.

[39] B. W. Silverman, *Density Estimation for Statistics and Data Analysis*. Chapman

& Hall/CRC, 1998.

[40] H. Wang, "Bayesian radio map learning for robust indoor positioning," in *Indoor Positioning and Indoor Navigation (IPIN), 2011 International Conference on*, sept. 2011, pp. 1–6.

[41] L. Peng, J. Han, W. Meng, and J. Liu, "Research on radio-map construction in indoor wlan positioning system," in *Pervasive Computing Signal Processing and Applications (PCSPA), 2010 First International Conference on*, sept. 2010, pp. 1073–1077.

[42] J. Krumm and J. Platt, "Minimizing calibration effort for an indoor 802.11 device location measurement system," Tech. Rep.

[43] Data Sheets, *Wifly GSX Super Module: RN-134*, Roving Networks Inc., Dec. 2009.

[44] C. Furse, *Basic Introduction to Bioelectromagnetics*, 2nd ed. Chapman & Hall/CRC, 2009.

[45] P. L. Ryan, "Radio frequency propagation differences through various transmissive materials," *M.Sc. dissertation, University of North Texas*, 2002.

[46] Data Sheets, *Rufa 2.4 GHz SMD Antenna*, Antenova Ltd., 2008.

[47] M. Brunato and R. Battiti, "Statistical learning theory for location fingerprinting in wireless lans," *Computer Networks*, vol. 47, no. 6, pp. 825–845, Apr. 2005.

[48] Y. Bar-Shalom, X. R. Li, and T. Kirubarajan, *Estimation with Applications to Tracking and Navigation: Theory, Alogrithms, and Software.* Wiley-Interscience, 2001.

[49] D. Shepard, "A two-dimensional interpolation function for irregularly-spaced data," *Proceedings of the 1968 23rd ACM national conference*, pp. 517–524, 1968.

[50] C. M. Bishop, *Pattern Recognition and Machine Learning.* Springer Science, 2006.

[51] L. R. Rabiner, "A tutorial on hidden markov models and selected application in speech recognition," *Proceedings of the IEEE*, vol. 77, no. 2, pp. 257–286, 1989.

[52] J. Bilmes, "A gentle tutorial of the em algorithm and its application to parameter estimation for gaussian mixture and hidden markov models," International Computer Science Institute, U.C. Berkeley, Tech. Rep., 1998.

[53] Y. Chen and M. R. Gupta, "Em demystified: An expectation-maximization tutorial," *University of Washington Electrical Engineering Technical Report*, vol. UWEETR-2010, no. 2, pp. 1–17, 2010.

[54] A. P. Dempster, N. M. Laird, and D. B. Rubin, "Maximum likelihood from incomplete data via the em algorithm," *Journal of the Royal Statistical Society*, vol. 39, no. 1, pp. 1–38, 1977.

[55] R. O. Duda, P. E. Hart, and D. G. Stork, *Pattern Classification*, 2nd ed. Wiley-Interscience, 2001.

[56] G. J. McLachlan and T. Krishnan, *The EM Algorithm and Extensions*, 2nd ed. Wiley-Interscience, 2008.

[57] J. Bilmes, "EE515 - computer speech processing: Lecture notes 11," 2005. [Online]. Available: http://ssli.ee.washington.edu/~bilmes/ee516/

[58] G. D. Forney, "The viterbi algorithm," *Proceedings of the IEEE*, vol. 61, no. 3, pp. 268–278, 1973.

Appendix A

Accerometer Daughter Board

Standard 802.11 modules can easily be configured to periodically scan pre-selected channels and collect the available RSS measurements from neighboring APs and transmit these readings back to the location system. In our proposed indoor localization system, it is important that the system be able to discern between RSS measurements taken at static locations or during dynamic motion. The stock WiFly® RN-134 802.11b modules did not support this feature, therefore it was necessary to develop additional circuitry to allow for the determination between static and dynamic RSS measurements.

A.1 Specifications and Components

During the initial circuit development stage, the preferred specifications and design requirements were outlined.

- planer, bi-directional motion detection
- no additional battery requirement
- minimal current drain during operation
- minimal number of SMD components
- small Printed Circuit Board (PCB) form factor

A three axis, Low-g micro machined accelerometer, manufactured by Freescale Semi-conductor, was chosen for motion detection. Notable specifications include selectable sensitivity of $\pm 3g$ or $\pm 9g$, low operating voltage of $2.2V - 3.6V$, low power consumption at $400\mu A$, sleep mode at $3\mu A$, built in signal conditioning via low-pass filter, LGA-14 packaging ($3.0mm \times 5.0mm \times 1.0mm$), and low cost. A low voltage quad comparator, manufactured by Texus Instruments, was chosen based on its low power consumption at $140\mu A$, required supply voltage of $2.7V - 5.0V$, TSSOP-14 packaging ($5.0mm \times 6.4mm \times 1.1mm$), open collector output, and low cost. The stock WiFly® RN-134 802.11b modules supply on-board power regulation to $3.3V$ which was utilized for the daughter board via additional pin headers. The list of required components based on Fig. A.3 are shown in Table A.1.

Daughter Board Component List		
Part	Value	Description
R1, R4	100 $k\Omega$	resistor
R2	2.9 $k\Omega$	resistor
R8, R9	10 $k\Omega$	resistor
R5, R6, R7, R10, R11, R12, R13	0 Ω	short resistor
P1	500 $k\Omega$	potentiometer
P2	100 $k\Omega$	potentiometer
C1, C2	33 nF	capacitor
C3	1 μF	capacitor
Accel1	MMA7341	XYZ accelerometer
4COMP	LMV339	4× comparator

Table A.1: Daughter Board Component List

A.2 Overview of Operation

The quad comparator is used to check the analog accelerometer output voltage of each axis against a reference threshold voltage, V_{thr}, which translates to either positive or negative motion on each axis. The accelerometer output voltage will contain numerous spikes and considerable noise, as shown in Fig. A.1, yet the comparator will trigger a module wake-up every time the accelerometer output voltage reaches or exceeds the threshold voltage, with RSS measurements collected every 1 second. This is ideal for user motion where sporadic movements will never result in a constant acceleration, allowing the module and the localization system to be sequentially alerted to continued motion through the multiple wake-up triggers. The threshold voltage is configurable through two potentiometers and allows the designer to set the amount of positive or negative motion required on any axis to trigger a wakeup of the WiFly® module. To account for user actions that do not result in a continued motion, such as bumping the module, the localization software tracks the number of individual wake-up triggers over a given time period and registers a unique motion event with 3 or more sequential triggers.

The circuit schematic in Fig. A.2 illustrates the interconnections between the WiFly® module and the accelerometer daughter board. The PCB gerber layers of the accelerometer daughter board can be seen in Fig. A.3 and were sent to a local manufacturing company for fabrication. The SMD components were reflow soldered in house with the overall fabricated design shown in Fig. A.4. It can be seen that the accelerometer daughter board connects directly to the stock WiFly® RN-134 module via existing pin headers. If the potentiometers were replaced with fixed SMD resistor values, the overall increase in thickness to the stock module would be 1 cm with no
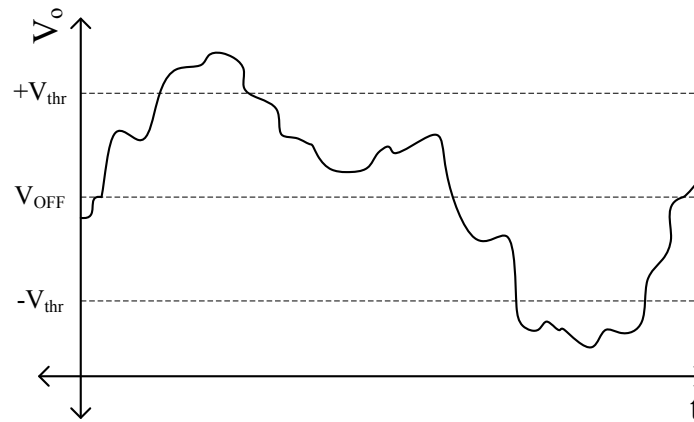
increases to its height or width.

Figure A.1: Pictorial Accelerometer Output Voltage for 1 Axis
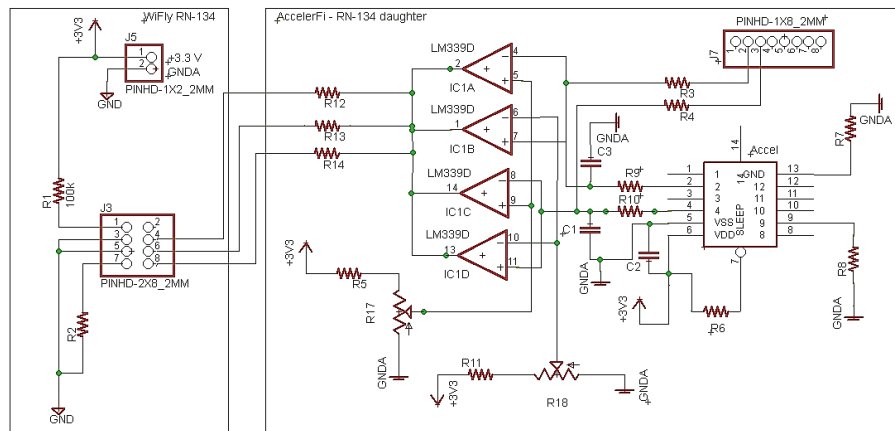

Figure A.2: Accelerometer Daughter Board Schematic
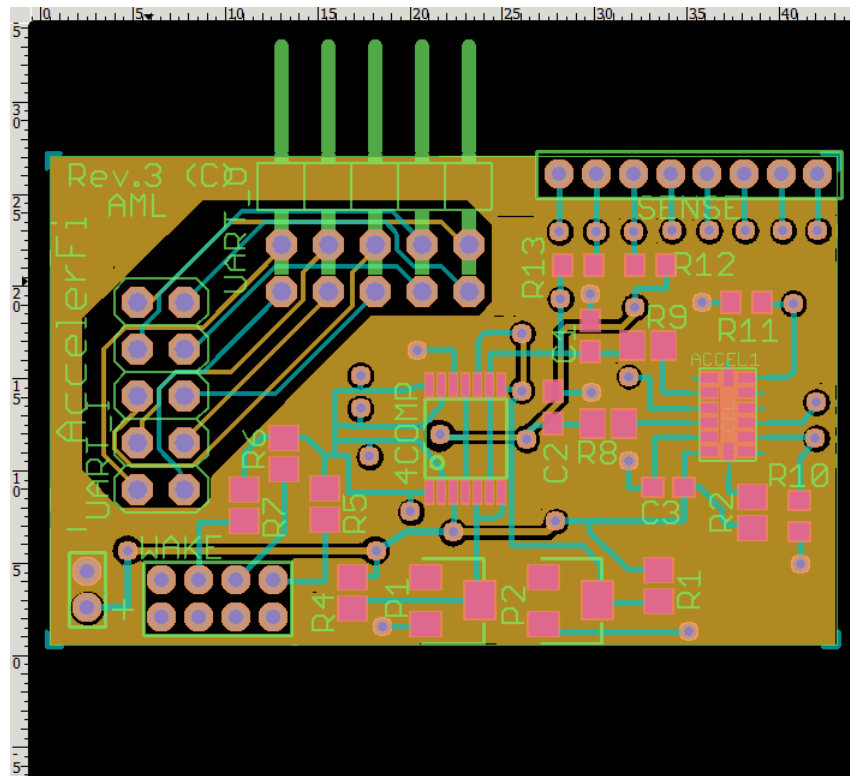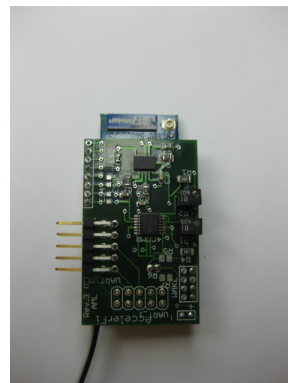
Figure A.3: Accelerometer Daughter Board PCB Gerber Layers [mm]

(a)Size Comparison   (b)Attached - Front View   (c)Attached - Side View
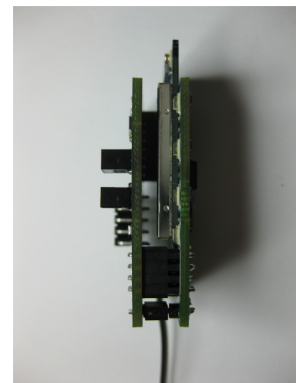
Figure A.4: Fabricated Accelerometer Daughter Board

Appendix B

Derivations for Hidden Markov Model Training

B.1 Maximization of Model Parameters using Lagrange Multipliers

As outlined in Sec. 3.6.2.2, the $\mathcal{Q}$ function can be rewritten as the sum of each parameter allowing for the maximization to be performed individually. With knowledge of the joint distribution for the current iteration

$$P(R, Q|\theta) = P(q_1|\boldsymbol{\pi}) \left[\prod_{t=2}^{T} P(q_t|q_{t-1}, \mathbf{A}) \right] \prod_{m=1}^{T} P(\mathbf{r}_m|q_m, \boldsymbol{\lambda}) \tag{B.1}$$

the $\mathcal{Q}$ function expands to

$$
\begin{aligned}
\mathcal{Q}\left(\theta, \theta^{(k)}\right) = {} & \sum_{Q \in S} \log P(q_1|\boldsymbol{\pi}) \frac{P\left(R, Q|\theta^{(k)}\right)}{P\left(R|\theta^{(k)}\right)} \\
& + \sum_{Q \in S} \left(\sum_{t=2}^{T} \log P(q_t|q_{t-1}, \mathbf{A}) \right) \frac{P\left(R, Q|\theta^{(k)}\right)}{P\left(R|\theta^{(k)}\right)} \\
& + \sum_{Q \in S} \left(\sum_{t=1}^{T} \log P(\mathbf{r}_t|q_t, \boldsymbol{\lambda}) \right) \frac{P\left(R, Q|\theta^{(k)}\right)}{P\left(R|\theta^{(k)}\right)} \tag{B.2}
\end{aligned}
$$

where $\theta^{(k)}$ is the previous iteration of the HMM parameters. The first term in Eq. B.2 can be evaluated further since by selecting all $Q \in S$, the values of q_0 are just repeatedly selected, or the marginal probabilities for $t = 1$.

$$\sum_{Q \in S} \left(\sum_{t=2}^{T} \log P\left(q_t | q_{t-1}, \mathbf{A}\right) \right) \frac{P\left(R, Q | \theta^{(k)}\right)}{P\left(R | \theta^{(k)}\right)} = \sum_{i=1}^{N} \log \pi_i^{q_{1,i}} \frac{P\left(R, q_1 = s_i | \theta^{(k)}\right)}{P\left(R | \theta^{(k)}\right)} \quad \text{(B.3)}$$

Therefore, to maximize the current iteration π_i given the previous iteration $\theta^{(k)}$ under the constraint $\sum_{i=1}^{K} \pi_i = 1$ we make use of Lagrange multipliers γ. Constructing the Lagrangian and taking the partial derivative with respect to π_i and setting the result equal to zero

$$\frac{\partial}{\partial \pi_i} \left[\sum_{i=1}^{N} \log \pi_i \frac{P\left(R, q_1 = s_i | \theta^{(k)}\right)}{P\left(R | \theta^{(k)}\right)} + \gamma \left(1 - \sum_{i=1}^{N} \pi_i \right) \right] = 0 \quad \text{(B.4)}$$

$$\frac{1}{\pi_i} \frac{P\left(R, q_1 = s_i | \theta^{(k)}\right)}{P\left(R | \theta^{(k)}\right)} - \gamma = 0 \quad \text{(B.5)}$$

$$\pi_i = \frac{1}{\gamma} \frac{P\left(R, q_1 = s_i | \theta^{(k)}\right)}{P\left(R | \theta^{(k)}\right)} \quad \text{(B.6)}$$

Now the partial derivative is taken with respect to the Lagrange multiplier γ and the result set to zero

$$\frac{\partial}{\partial \gamma} \left[\sum_{i=1}^{N} \log \pi_i \frac{P\left(R, q_1 = s_i | \theta^{(k)}\right)}{P\left(R | \theta^{(k)}\right)} + \gamma \left(1 - \sum_{i=1}^{N} \pi_i \right) \right] = 0 \quad \text{(B.7)}$$

$$1 - \sum_{i=1}^{N} \pi_i = 0 \quad \text{(B.8)}$$

which is our original constraint. The result from Eq. B.6 is substituted into Eq. B.8

$$1 - \sum_{i=1}^{N} \frac{P\left(R, q_1 = s_i | \theta^{(k)}\right)}{\gamma P\left(R | \theta^{(k)}\right)} = 0 \tag{B.9}$$

$$\gamma = \sum_{i=1}^{N} \frac{P\left(R, q_1 = s_i | \theta^{(k)}\right)}{P\left(R | \theta^{(k)}\right)} = 1 \tag{B.10}$$

The solution to γ is finally substituted back into Eq. B.6 to reveal

$$\pi_i^{(k+1)} \leftarrow \pi_i = \frac{P\left(R, q_1 = s_i | \theta^{(k)}\right)}{P\left(R | \theta^{(k)}\right)} = P\left(q_1 = s_i | R, \theta^{(k)}\right) \tag{B.11}$$

which is maximized based on the previous iteration $\theta^{(k)}$ and is evaluated for each S. The second term in Eq. B.2 will become the sum of the joint-marginal probabilities for $t-1$ and t since for each t we are progressing over all state transitions from i to j and weighting by the corresponding probability.

$$\sum_{Q \in S} \left(\sum_{t=1}^{T} \log P\left(\mathbf{r}_t | q_t, \boldsymbol{\lambda}\right) \right) \frac{P\left(R, Q | \theta^{(k)}\right)}{P\left(R | \theta^{(k)}\right)} =$$
$$\sum_{i=1}^{N} \sum_{j=1}^{N} \sum_{t=2}^{T} \log a_{i,j}^{q_{t,j} q_{t-1,i}} \frac{P\left(R, q_{t-1} = s_i, q_t = s_j | \theta^{(k)}\right)}{P\left(R | \theta^{(k)}\right)} \tag{B.12}$$

Constructing the Lagrangian with constraint $\sum_{j=1}^{N} a_{ij} = 1$ and taking the partial derivative with respect to a_{ij} and setting the result equal to zero

$$\frac{\partial}{\partial a_{i,j}} \left[\sum_{i=1}^{N} \sum_{j=1}^{N} \sum_{t=2}^{T} \log a_{i,j} \frac{P\left(R, q_{t-1} = s_i, q_t = s_j | \theta^{(k)}\right)}{P\left(R | \theta^{(k)}\right)} + \gamma \left(1 - \sum_{j=1}^{N} a_{i,j} \right) \right] = 0 \tag{B.13}$$

$$\sum_{t=2}^{T} \frac{1}{a_{i,j}} \frac{P\left(R, q_{t-1} = s_i, q_t = s_j | \theta^{(k)}\right)}{P\left(R | \theta^{(k)}\right)} - \gamma = 0 \tag{B.14}$$

$$a_{i,j} = \sum_{t=2}^{T} \frac{1}{\gamma} \frac{P\left(R, q_{t-1} = s_i, q_t = s_j | \theta^{(k)}\right)}{P\left(R | \theta^{(k)}\right)} \tag{B.15}$$

Now the partial derivative is taken with respect to the Lagrange multiplier γ and the result set to zero

$$\frac{\partial}{\partial \gamma} \left[\sum_{i=1}^{N} \sum_{j=1}^{N} \sum_{t=2}^{T} \log a_{i,j} \frac{P\left(R, q_{t-1} = s_i, q_t = s_j | \theta^{(k)}\right)}{P\left(R | \theta^{(k)}\right)} + \gamma \left(1 - \sum_{j=1}^{N} a_{i,j} \right) \right] = 0 \tag{B.16}$$

$$1 - \sum_{j=1}^{N} a_{i,j} = 0 \tag{B.17}$$

which is our original constraint. The result from Eq. B.15 is substituted into Eq. B.17

$$1 - \sum_{t=2}^{T} \frac{1}{\gamma} \frac{P\left(R, q_{t-1} = s_i, q_t = s_j | \theta^{(k)}\right)}{P\left(R | \theta^{(k)}\right)} = 0 \tag{B.18}$$

$$\gamma = \sum_{t=2}^{T} \frac{P\left(R, q_{t-1} = s_i, q_t = s_j | \theta^{(k)}\right)}{P\left(R | \theta^{(k)}\right)} = \sum_{t=2}^{T} \frac{P\left(R, q_t = s_i | \theta^{(k)}\right)}{P\left(R | \theta^{(k)}\right)} \tag{B.19}$$

The solution to γ is finally substituted back into Eq. B.15 to reveal

$$a_{i,j}^{(k+1)} \leftarrow a_{i,j} = \frac{\sum_{t=2}^{T} P\left(R, q_{t-1} = s_i, q_t = s_j | \theta^{(k)}\right)}{\sum_{t=2}^{T} P\left(R, q_t = s_i | \theta^{(k)}\right)} \tag{B.20}$$

$$= \frac{\sum_{t=2}^{T} P\left(q_{t-1} = s_i, q_t = s_j | R, \theta^{(k)}\right)}{\sum_{t=2}^{T} P\left(q_t = s_i | R, \theta^{(k)}\right)} \tag{B.21}$$

The third term in Eq. B.2 will become the sum of the marginal probabilities for t since for each t we are progressing over the emission probabilities for all states and

weighting each possible emission by the corresponding probability.

$$\sum_{Q\in S} \log P\left(q_1|\boldsymbol{\pi}\right) \frac{P\left(R,Q|\theta^{(k)}\right)}{P\left(R|\theta^{(k)}\right)} = \sum_{i=1}^{N}\sum_{t=1}^{T}\sum_{h=1}^{H} \log \lambda_{i,h}^{q_{t,i}\mathbf{r}_{t,h}} \frac{P\left(R,q_t=s_i|\theta^{(k)}\right)}{P\left(R|\theta^{(k)}\right)} \quad \text{(B.22)}$$

Constructing the Lagrangian with constraint $\sum_{h=1}^{H}\lambda_{i,h}=1$ and taking the partial derivative with respect to $\lambda_{i,h}$ and setting the result equal to zero

$$\frac{\partial}{\partial \lambda_i}\left[\sum_{i=1}^{N}\sum_{t=1}^{T}\sum_{h=1}^{H}\log \lambda_{i,h}\frac{P\left(R,q_t=s_i|\theta^{(k)}\right)}{P\left(R|\theta^{(k)}\right)} + \gamma\left(1-\sum_{h=1}^{H}\lambda_{i,h}\right)\right]=0 \quad \text{(B.23)}$$

$$\sum_{t=1}^{T}\frac{1}{\lambda_{i,h}}\frac{P\left(R,q_t=s_i|\theta^{(k)}\right)\delta_{\mathbf{r}_t,v_h}}{P\left(R|\theta^{(k)}\right)} - \gamma = 0 \quad \text{(B.24)}$$

$$\lambda_{i,h}=\sum_{t=1}^{T}\frac{1}{\gamma}\frac{P\left(R,q_t=s_i|\theta^{(k)}\right)\delta_{\mathbf{r}_t,v_h}}{P\left(R|\theta^{(k)}\right)} \quad \text{(B.25)}$$

where only the observations $\mathbf{r}_t$ that are equal to v_h contribute to the hth probability value. Now the partial derivative is taken with respect to the Lagrange multiplier γ and the result set to zero

$$\frac{\partial}{\partial \gamma}\left[\sum_{i=1}^{N}\sum_{t=1}^{T}\sum_{h=1}^{H}\log \lambda_{i,h}\frac{P\left(R,q_t=s_i|\theta^{(k)}\right)}{P\left(R|\theta^{(k)}\right)} + \gamma\left(1-\sum_{h=1}^{H}\lambda_{i,h}\right)\right]=0 \quad \text{(B.26)}$$

$$1-\sum_{j=1}^{N}\lambda_{i,h}=0 \quad \text{(B.27)}$$

which is our original constraint. The result from Eq. B.25 is substituted into Eq. B.27

$$1-\sum_{t=1}^{T}\frac{1}{\gamma}\frac{P\left(R,q_t=s_i|\theta^{(k)}\right)}{P\left(R|\theta^{(k)}\right)}=0 \quad \text{(B.28)}$$

$$\gamma = \sum_{t=1}^{T} \frac{P\left(R, q_t = s_i | \theta^{(k)}\right)}{P\left(R | \theta^{(k)}\right)} \tag{B.29}$$

The solution to γ is finally substituted back into Eq. B.25 to reveal

$$\lambda_{i,h}^{(k+1)} \leftarrow \lambda_{i,h} = \frac{\sum_{t=1}^{T} P\left(R, q_t = s_i | \theta^{(k)}\right) \delta_{\mathbf{r}_t, v_h}}{\sum_{t=1}^{T} P\left(R, q_t = s_i | \theta^{(k)}\right)} \tag{B.30}$$

$$= \frac{\sum_{t=1}^{T} P\left(q_t = s_i | R, \theta^{(k)}\right) \delta_{\mathbf{r}_t, v_h}}{\sum_{t=1}^{T} P\left(q_t = s_i | R, \theta^{(k)}\right)} \tag{B.31}$$

B.2   Derivation of Partial Probabilities for Viterbi Algorithm

B.2.1   Forward Partial Probability

Utilizing the Markov assumptions of Eq. 3.35 - 3.36, along with the probability product rule, the forward and backward partial probabilities become

$$\alpha_i\left(t\right) = P\left(\mathbf{r}_1, \mathbf{r}_2, \ldots, \mathbf{r}_t, q_t = s_i | \theta^{(k)}\right) \tag{B.32}$$

$$= \sum_{j=1}^{N} P\left(\mathbf{r}_1, \mathbf{r}_2, \ldots, \mathbf{r}_t, q_t = s_i, q_{t-1} = s_j | \theta^{(k)}\right) \tag{B.33}$$

$$= \sum_{j=1}^{N} P\left(\mathbf{r}_1, \mathbf{r}_2, \ldots, \mathbf{r}_{t-1}, q_{t-1} = s_j | \theta^{(k)}\right) P\left(\mathbf{r}_t, q_t = s_i | q_{t-1} = s_j, \mathbf{r}_1, \mathbf{r}_2, \ldots, \mathbf{r}_{t-1} \theta^{(k)}\right)$$
$$\tag{B.34}$$

$$= \sum_{j=1}^{N} \alpha_j\left(t-1\right) P\left(\mathbf{r}_t, q_t = s_i | q_{t-1} = s_j, \theta^{(k)}\right) \tag{B.35}$$

$$= \sum_{j=1}^{N} \alpha_j\left(t-1\right) P\left(\mathbf{r}_t | q_t = s_i, q_{t-1} = s_j, \theta^{(k)}\right) P\left(q_t = s_i | q_{t-1} = s_j, \theta^{(k)}\right) \tag{B.36}$$

$$= \sum_{j=1}^{N} \alpha_j \left(t-1\right) P\left(\mathbf{r}_t | q_t = s_i, \theta^{(k)}\right) P\left(q_t = s_i | q_{t-1} = s_j, \theta^{(k)}\right) \tag{B.37}$$

$$= \lambda_i^{(k)} \left(\mathbf{r}_t\right) \sum_{j=1}^{N} \alpha_j \left(t-1\right) a_{j,i}^{(k)} \qquad 1 \leq i \leq N \qquad 1 \leq t \leq T \tag{B.38}$$

B.2.2   Backward Partial Probability

$$\beta_i \left(t\right) = P\left(\mathbf{r}_{t+1}, \mathbf{r}_{t+2}, \ldots, \mathbf{r}_T | q_t = s_i, \theta^{(k)}\right) \tag{B.39}$$

$$= \sum_{j=1}^{N} P\left(\mathbf{r}_{t+1}, \mathbf{r}_{t+2}, \ldots, \mathbf{r}_T, q_{t+1} = s_j | q_t = s_i | \theta^{(k)}\right) \tag{B.40}$$

$$= \sum_{j=1}^{N} P\left(q_{t+1} = s_j | q_t = s_i | \theta^{(k)}\right) P\left(\mathbf{r}_{t+1}, \mathbf{r}_{t+2}, \ldots, \mathbf{r}_T | q_{t+1} = s_j, q_t = s_i, \theta^{(k)}\right) \tag{B.41}$$

$$= \sum_{j=1}^{N} a_{i,j}^{(k)} P\left(\mathbf{r}_{t+1}, \mathbf{r}_{t+2}, \ldots, \mathbf{r}_T | q_{t+1} = s_j, \theta^{(k)}\right) \tag{B.42}$$

$$= \sum_{j=1}^{N} a_{i,j}^{(k)} P\left(\mathbf{r}_{t+1} | q_{t+1} = s_j, \theta^{(k)}\right) P\left(\mathbf{r}_{t+2}, \ldots, \mathbf{r}_T | q_{t+1} = s_j, \mathbf{r}_{t+1}, \theta^{(k)}\right) \tag{B.43}$$

$$= \sum_{j=1}^{N} a_{i,j}^{(k)} P\left(\mathbf{r}_{t+1} | q_{t+1} = s_j, \theta^{(k)}\right) P\left(\mathbf{r}_{t+2}, \ldots, \mathbf{r}_T | q_{t+1} = s_j, \theta^{(k)}\right) \tag{B.44}$$

$$= \beta_i \left(t\right) = \sum_{j=1}^{N} a_{i,j}^{(k)} \lambda_i^{(k)} \left(\mathbf{r}_{t+1}\right) \beta_j \left(t+1\right) \qquad 1 \leq t \leq T \tag{B.45}$$