Indoor Radio Localization in a DECT Wireless Network using Bayesian Inference

Master's thesis

Kasper Ramsgaard-Jensen

June 6, 2019

Aalborg University
Department of Mathematical Sciences
Skjernvej 4A
DK-9220 Aalborg

AALBORG UNIVERSITY
STUDENT REPORT

Department of Mathematical Sciences
Aalborg University
http://www.aau.dk

Title:
Indoor Radio Localization in a DECT Wireless Network using Bayesian Inference

Theme:
Statistical Signal Processing

Project Period:
September 2018 to June 2019

Project Group:
Group 216e

Participant:
Kasper Ramsgaard-Jensen

Supervisors:
Troels Pedersen
Jakob Gulddahl Rasmussen

Copies: 1

Page Numbers: 113

Date of Completion:
June 6, 2019

Abstract:

This master's thesis is written in collaboration with RTX A/S and addresses indoor radio localization in a DECT network using *received signal strength* (RSS). In order to model the localization problem and the relationship between RSS and distance an extensive system specification has been employed. Based on the findings, an iterative moment matching variational message passing algorithm has been derived. It has proven advantageous to restrict the messages to the exponential family of probability distributions and employing moment matching for deriving simple message approximations. The performance of the algorithm has been tested with simulated and real data through Monte Carlo simulations. If the base stations are assumed to be sector antennas, the algorithm quickly converges to a final position in the vicinity of the true position. Those estimates which do not agree with the true position are caused by inferior RSS measurements and due to the fast convergence, other base stations may not contribute. Sorting the base stations with respect to the RSS does, however, yield mean error distances of approximately 5 meters in simulated environments and less than 5 meters using data from a measurement campaign. As the explored methods are able to infer latent distance information from RSS measurements the results of this project can impact the future direction of indoor localization based on RSS.

Contents

Preface

The present master's thesis is submitted to Aalborg University, Denmark, in fulfilment of the requirements for the Master of Science in Mathematical Engineering. This work presents the results concerning indoor radio localization using message passing algorithms.

The thesis has been made in collaboration with RTX A/S and carried out during the period spanning from September 2018 to June 2019 at the Faculty of Engineering and Science, Aalborg University, Denmark.

I am grateful for the opportunity of experiencing the daily life of an engineering company offered by RTX A/S. A special thanks to Niels Christian Mølgaard Holm and Finn Hebsgaard Andreasen who helped me install the measurement setup and guided me in the limits and resources of the DECT radio technology.

I would like to extend my gratitude for the excellent support and guidance provided by my supervisors: Troels Pedersen and Jakob Gulddahl Rasmussen. A special thanks to Troels Pedersen for guiding me through the theory of message passing algorithms and the long and fruitful discussions.

Aalborg University, June 6, 2019

Kasper Ramsgaard-Jensen
kramsg14@student.aau.dk

Danish abstract

Nutidens teknologiudvikling går i retning af, at antallet af enheder, der kan tilgå internettet eller andre former for trådløs kommunikation, er stigende. Derfor er antallet af områder, hvor der kan være behov for lokaliseringsmetoder, blevet mere udbredt, specielt i områder hvor satellitforbindelse ikke er tilgængelig, hvilket resulterer i, at teknologier såsom GPS forringes betydeligt. På baggrund af dette har RTX A/S efterspurgt en løsning i sådanne miljøer.

Denne specialeafhandling dokumenterer en analyse om, hvorvidt statistiske metoder kan benyttes til at udlede en algoritme til indendørs lokalisering af enheder, der kommunikerer trådløst ved hjælp af DECT radioteknologien. Algoritmen skal på baggrund af modtaget signal styrke fra et antal basestationer, med kendte positioner, estimerer den pågældende enheds position. For at opnå en sådan lokaliseringsmetode vil kendte algoritmer blive studeret og forsøgt omdannet til den pågældende problemstilling.

Rapporten indledes med en gennemgående systemspecifikation af det pågældende DECT netværk for at vurdere, hvilke observerbare størrelser der er til rådighed og kan bruges til positionering. De relevante variable er derefter blevet sammensat til det generelle lokaliseringsproblem i form af et inferensproblem. I de følgende kapitler undersøges det hvorledes sådanne inferensproblemer kan blive løst på en struktureret og effektiv måde ved hjælp af message passing algoritmer, specielt variational message passing.

Den sidste del af rapporten beskæftiger sig med at anvende den gennemgåede teori om inferensmetoder til at skabe en ny algoritme passende til det relevante lokaliseringsproblem. Den endelige algoritme er en kombination af variational message passing og moment matching, der ved at restringere de pågældende beskeder til den eksponentielle familie af fordelingsfunktioner opnår simple beskedapproksimationer. Approksimationerne bliver efterfølgende vurderet og viser stor sammenlignelighed med de sande beskeder.

For at teste algoritmens anvendelighed er denne blevet implementeret i Python. Resultater fra både simuleret og indsamlet data viser, at den udledte algoritme giver gode resultater, hvis de pågældende basestationer antages at være sektorantenner. Algoritmen opnår afstands middelfejl på omkring 5 meter i simulerede

miljøer og under 5 meter i det rigtige miljø. Dog er der åbne aspekter af algoritmen, der kan vise sig fordelagtige at undersøge i videre arbejde med emnet men denne specialeafhandling viser som et proof-of-concept, at statistiske metoder kan benyttes til indendørs lokalisering.

1 Introduction

1.1 The localization problem

In this technological age, the amount of devices that can access the world wide web or otherwise establish wireless communication with surrounding equipment is increasing. Thus, areas in which localization methods can be applied are becoming more widespread, especially in urban environments. A widely used estimation standard for localization of wireless devices is the *Global Positioning System* (GPS). However, the accuracy of GPS in urban environments, such as indoor locations, is limited due to Non-Line-of-Sight (NLOS) propagation paths [1][2]. Meanwhile, both public and private indoor environments have established networks, which can provide wireless connectivity and does not rely on satellite connections. These networks are usually connected in a backbone structure which enables intercommunication between base stations. Thus the network can receive information from a device at different fixed base stations yielding relevant information about the position of the specific device as depicted in Figure 1.1.

Indoor environments, in which a localization method might be needed, include e.g. restaurants or hospitals in which locating tables or equipment may save crucial time. Further examples include estimating the position of employees, especially in

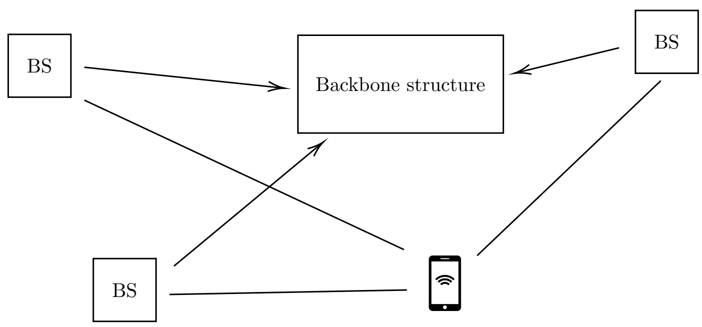


Figure 1.1: A central device or cloud system receives relevant information of a device from fixed base stations (BS) through a backbone structure.

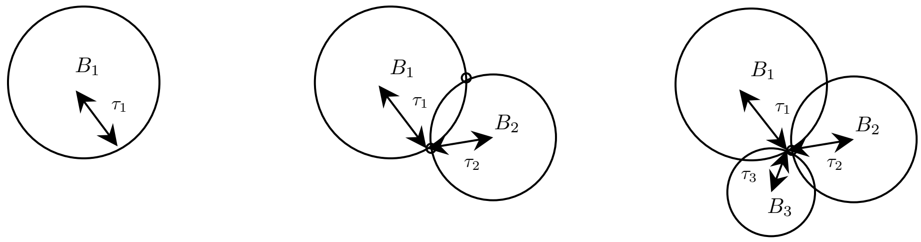


Figure 1.2: Trilateration can be used to find an unambiguous position estimate if precise distance observations from at least three base stations are available.

man-down situations where an employee has fainted in hazardous environments. Common for all such localization problems is the ability to describe the relative position between base stations and non-static devices which is usually done through a measure of relative distance. This can be obtained by using distance dependent link parameters in the radio channel. Some frequently used link parameters for localization are *Received Signal Strength* (RSS) and *Time of Arrival* (TOA) [3][4]. For unambiguous estimation, at least three or four different distance estimates are needed for two- and three-dimensional localization respectively. Two-dimensional localization is employed by intersecting circles with a center in the base station positions and radii given from the distance measurements also known as trilateration. The intersections then correspond to possible locations of the device and given perfect distance measurements only one intersection exists, see Figure 1.2.

Model based distance measurements can be obtained through both RSS and TOA. If the base stations are able to measure the round-trip-time of the signal, a distance estimate can be found by using the relation [5]

$$d_{est} = \frac{\tau_{TOA} \cdot c}{2},$$

where τ_{TOA} is the round-trip-time of the signal and c is the speed of light. Although TOA yields precise distance estimates, the precision deteriorates in environments where *Line of Sight* (LOS) components are obscured, e.g. in an office environment as depicted in Figure 1.3. If the obtained distance estimates are corrupted, the method of trilateration cannot be used to obtain unambiguous position estimates as the drawn circles might not intersect in the true position.

As TOA is affected in such environments, the RSS link parameter is often used instead, since it is less affected by NLOS paths. However, using RSS entails several problems. Radio communication is generally affected by the broadcasting environment in which the effects usually describe the decay in signal strength. The general term for the effects is *pathloss* and is commonly categorized as *free space pathloss*, *shadowing* and *small scale fading* [5]. Free space pathloss represents the decay in signal strength due to LOS propagation. Therefore, the free space pathloss depends

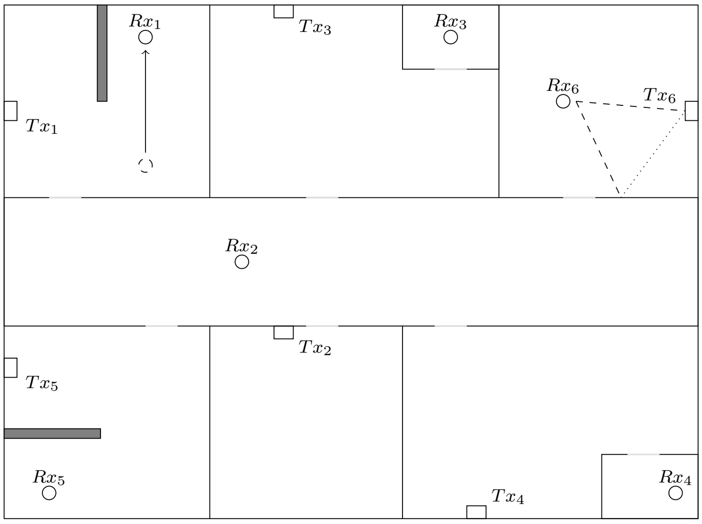


Figure 1.3: Propagation effects are present in the form of free space pathloss and both shadowing and small scale fading due to walls and other scatterers.

on the distance between the transmitter and receiver. The log-distance model [5] is commonly used to describe free space pathloss and is formulated as

$$PL_{\text{dB}}(d) = PL(d_0) + 10 \cdot \eta \cdot \log \frac{d}{d_0}$$

where d is the distance between transmitter and receiver, η is an environment specific pathloss constant and $PL(d_0)$ is the pathloss at a reference distance d_0 from the transmitter. Shadowing occurs when the LOS propagation path between a transmitter and receiver is blocked as depicted by the movement of Rx_1 in Figure 1.3. The blocked LOS path changes the general propagation path and, therefore, also received power. Considering the stationary receiver, R_{x_5}, all propagation paths are blocked and, therefore, shadowing occurs in all communication links. Furthermore, broadcasting into an environment such as the one depicted in Figure 1.3 will introduce *multipath components*. In cluttered environments, a receiver will not only receive the direct signal component of the transmitted signal, but a multitude of multipath components. These components are remnants of the transmitted signal which have traveled along propagation paths different from the direct LOS path, such as the scenario depicted between transmitter Tx_6 and receiver Rx_6. This might entail several problems. If two signal components arrive at the receiver they may be completely out of phase. If their amplitude is the same, then the two impinging signals will cancel each other due to destructive interference and no signal will be received. Similarly, if the signals arrive completely in-phase, the amplitude of the received signal will be double the signal strength of a single component due to constructive interference. Small scale fading will, therefore, alongside shadowing, distort relevant distance information in the signal strength.

1.2 Existing work

In the field of positioning, a vast number of techniques exist and can be divided into three categories [6], *geometric*, *mapping* and *statistical* techniques. Mapping techniques are employed by dividing the propagation environment into a grid in which radio signal characteristics are collected for each grid point. These characteristics are then labeled according to the grid point which creates a *radio map*. Thus, received signal characteristics from a device can be used to estimate the position of the transmitting device according to the acquired radio map [7]. This technique may, however, be problematic due to changing environments and obtaining the radio map is highly time-consuming. The geometric based techniques are employed by solving a number of equations, such as least squares, to estimate the position of a device, e.g. through trilateration as explained above, but may be susceptible to outliers. Lastly, the statistical techniques seek to model errors in the observations and include possible *a priori* knowledge of the devices in order to estimate their positions. The statistical techniques may become rather complex and require a vast amount of computational capabilities but they tend to describe the problem better as more information can be included in the model.

For some wireless networks, it is possible to utilize the network protocol for localization purposes. Different radio communication technologies offer distinct access protocols and, therefore, different communication schemes exist. For access schemes where sensors in a wireless network are able to communicate, cooperative sensor self-localization algorithms have been developed [8][9]. In such scenarios, the sensors share information with neighboring devices in order to estimate their own position while using observations from fixed base stations with known positions. Examples of sensor self-localization algorithms employ approximate inference algorithms on graphical models [10][11] which has proven accurate and has a low communication overhead. However, if the radio communication technology does not allow for inter-communication between non-fixed devices, there may be less and insufficient information, relevant for localization purposes, available.

In order to mitigate problems arising from NLOS propagation paths in indoor environments, and, therefore, inaccurate TOA estimation, it is relevant to utilize the RSS link parameter. However, this parameter does also suffer in precision due to pathloss effects, which have to be accounted for in order to obtain reliable distance estimates. Furthermore, in order to seek a general positioning technique, in which prior knowledge can be inferred, a statistical method is favorable. If the localization problem can be formulated graphically, it might prove advantageous to apply similar methods as employed in [10] and [11] as they have shown to yield trustworthy estimates.

This thesis will explore novel methods for indoor radio localization based on RSS measurements. These are readily available in almost all radio equipment rendering the algorithm applicable in most environments. As inference algorithms on graphical models have proven successful, we will seek to formulate a statistical model that employs this method. This includes deriving a model for RSS while considering the pathloss effects so precise distance estimates can be obtained and used to find the position of a device. When the localization algorithm has been derived, we will discuss the applicability of the model through both simulations and real measurements.

1.3 Problem statement

Utilizing wireless technology for indoor radio localization is inherit challenging, as the propagation environment is rather complex. Therefore, it is relevant to investigate statistical methods in which an inference problem can be derived and solved based on received signal strength which suffers less from NLOS propagation paths. This will be explored by answering the following problem statements

- How can RSS be modeled so that pathloss effects are accounted for?
- What level of precision can be achieved in indoor radio localization through a statistical model based on RSS?
- To what extent can the accuracy of the model be improved?
- To which degree will improvements affect the applicability of the model?

RTX A/S

RTX A/S is one of the leading companies in the field of wireless communication with nearly 300 employees worldwide. The headquarters is located in Nørresundby, Denmark but offices are also located in Hong Kong and USA. The fields in which RTX provides solutions include eSport, professional communication equipment and healthcare.

The role of RTX in this master's thesis is to provide the equipment and environment needed to employ performance testing. Furthermore, RTX has offered a seat at their headquarters for the duration of the project period which has enabled the possibility of experiencing the daily life of an engineering company and fast access to expert knowledge on wireless communication.

Problem definition

Although relevant devices are usually mobile, we will restrict the objects to be stationary while the localization is in progress. Furthermore, we will assume that all calculations needed to obtain position estimates will be done on a central device as we restrict ourselves from considering implementation in embedded devices. Additionally, we will not consider real-time implementation. These restrictions have been made in favor of obtaining a proof-of-concept that a statistical model based on RSS can be used for localization purposes. As the physical topology of realistic propagation environments may have different structures we will assume that our simulation environment is a square room.

We will formulate the model for a wireless network utilizing DECT as this is the type of radio technology available for testing. Thus real-world testing of the derived localization algorithm will be performed using data obtained from base stations and portable devices communicating through the DECT technology. Furthermore, environment specific model parameters will be chosen based on empirical studies, as these are time-consuming to estimate and out of the scope of this work.

1.4 Thesis outline

In Chapter 2 we will specify the used radio technology alongside the information which is available in the network and from these considerations derive the inference problem. Next, we explore methods in Chapter 3 with which the inference problem can be solved such that we in Chapter 4 can model and derive the indoor localization algorithm. The implementation and model choices will be discussed in Chapter 5 such that we in Chapter 6 and Chapter 7 can issue performance tests of the algorithm through simulations and real measurements respectively. Lastly, we will present our conclusions and thoughts on future research.

2 The communication system

In this chapter we will introduce the relevant radio communication technology. This will enable an discussion of the devices in the system and the available information which can be used for positioning.

2.1 The DECT radio technology

In order to introduce the devices in the system, we need to specify the employed radio communication technology. Enabling the possibility of several users sharing the same available resource, often bandwidth, the communication technology must follow a *multiple access* (MA) scheme. A widely used MA scheme is FDMA/TDMA which allocates a set of carrier frequencies to the communication cell. Each user is then assigned one or several time slots at a given carrier for transmission, see Appendix A for more information on FDMA and TDMA. One technology which utilizes the FDMA/TDMA scheme is *Digital Enhanced Cordless Telecommunications* (DECT) which is a high capacity radio access technology [12]. The European DECT standard is allocated the $1880 - 1900$ MHz frequency band with 10 carriers separated by 1728 kHz and a bandwidth of 1 MHz. The TDMA frame in DECT is organized in 24 time slots with an overall duration of 10ms. In the DECT technology the users and base stations are often denoted as portable parts (PP) and fixed parts (FP) respectively.

DECT can be effectively implemented in applications counting both simple cordless phones and larger intercom systems. Furthermore, in addition to FP to PP communication, DECT can also provide direct PP to PP and FP to FP communication. The benefits of DECT become clear when comparing the technology to other mobile radio systems like the *Global System for Mobile communications* (GSM). In GSM the mobile units are only allowed to connect to the unique network which is part of the mobile radio system [12]. DECT, however, provides a substantial set of network protocols which enables the possibility of interworking between different applications and networks. In principle, DECT covers only the air interface between an FP and PP, see Figure 2.1.

This means, that the connection between the local or public network (the Inter-

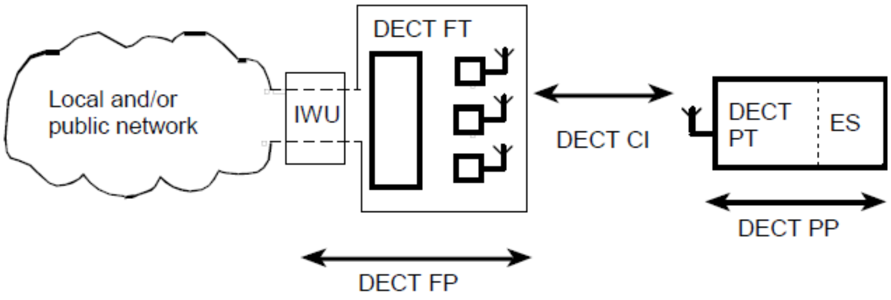


Figure 2.1: DECT common interface [12].

Working Unit) and the DECT system is network specific and, therefore, not part of the DECT Common Interface (CI) specification. The End System (ES) in the DECT PP is similarly excluded. DECT is, therefore, transparent to the services that may be provided by the network [12]. Thus DECT can be viewed as a toolbox with different protocols from which a selection can be made to access a given network.

The DECT TDMA frame structure is depicted in Figure 2.2.

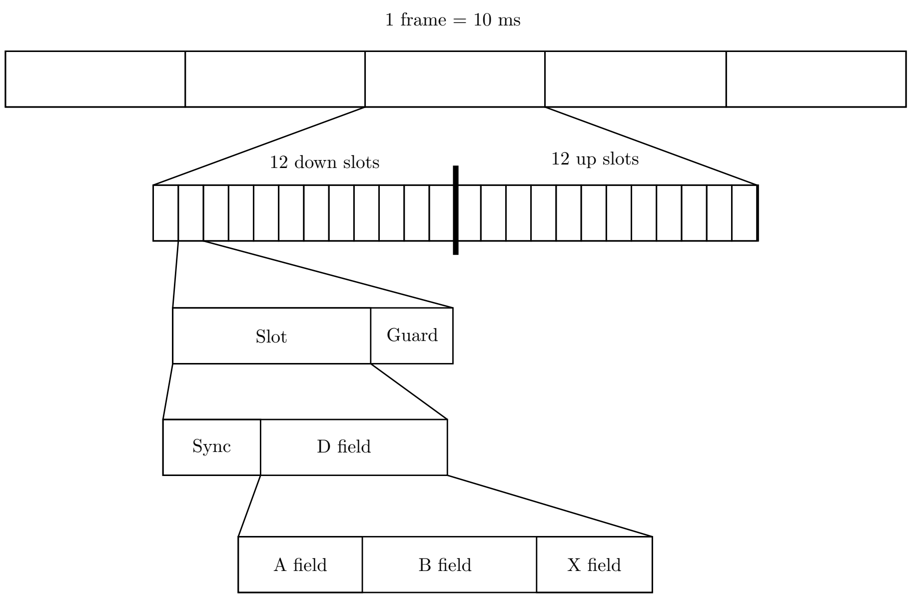


Figure 2.2: DECT TDMA frame structure.

The 24 time slots of a DECT frame is structured such that 12 slots are used for downlink transmission (from FP to PP) and the remaining 12 for uplink (PP to FP). Each slot is equipped with guard bits in order to mitigate phase misalignment problems as discussed in Appendix A. Furthermore, each slot consists of a synchronization field and D-field where the former is used for packet synchronization in the transmission link. The D-field consists of three fields, the A, B and X field. In the A field, an error control code is used to detect possible changes in the data. The B field contains the specific data needed to be transmitted and the X field is used in order to avert collisions with other incoming signals which may lead to destructive interference.

2.2 Information sources and available information

The environment, in which we seek to solve the localization problem, is utilizing the DECT radio technology as described in Section 2.1. We, therefore, know that the relevant devices in the system are the FPs and PPs. The FPs will, at a given time rate, broadcast signal beacons into the environment which the PPs are able to detect. Each PP is always connected to a single FP through a direct link. Through this active link, data is transmitted in an update cycle T_A. However, the PP will also store data from other FPs in the environment, if an FP broadcast is detected by the PP. When the PP is not transmitting through the direct link it will listen for FP broadcasts with an update cycle $T_I > T_A$ of e.g. 10 s (idle state). However, the PP might miss the broadcasts from these due to its update cycle and, therefore, fail to obtain relevant information from them. Therefore, these update cycles can be chosen to be more often or more infrequent if necessary. It should, however, be noted, that the PPs are battery powered and, therefore, altering the update cycle may not be realizable.

As the localization problem is formulated in favor of locating a user (PP) we need to incorporate this device in the inference problem. Similarly, as the FPs have known positions we also desire to include them. We will, therefore, explore the information which can be exchanged between these devices and assess the value of the given information with respect to positioning.

2.2.1 Mails

In the DECT technology variant employed by RTX, the information, which is available in the system, is stored and shared in *mails*. The information received at the PP when it is communicating through its direct link (active state) is listed in a mail named *PP_ACTUAL_RSSI_IND* which is received every T_A as described above. Furthermore, the information received in the PP when it is in idle state is collected in a mail *PP_BEARER_FOUND_IND*. This mail may, however, not be updated every T_I as the PP may not detect any FP besides the one in the direct link. This mail is, therefore, highly asynchronous. The last relevant mail to describe is the *PP_DEBUG_RFPI_LIST_STATUS* mail which is also received in the PP. This mail stores information including RSS of previously detected FPs. This list, therefore, describes possible FPs in the vicinity of the PP. The RSS information stored in this list is downgraded for each update period by a value ϵ_s such that the PP progressively forgets the FP, if it does not detect it again. The information which is available in these mails are summarized in Table 2.1 and are further explained in the following.

Observable	Position	Update period
Idle RSS, R_I	see Section 2.2.2	T_I
Avg. RSS, R_A	see Section 2.2.3	T_A
Background RSS, R_B	see Section 2.2.4	T_A
f_c	No immediate connection	T_A/T_I
# Good syncs	see Section 2.2.6	T_A
A-field check sum	No immediate connection	T_A
X-field check sum	No immediate connection	T_A
Blind slot info	see Section 2.2.8	T_A
Slot	No immediate connection	T_A/T_I
FP ID	see Section 2.2.10	T_I
Phase, ϕ	see Section 2.2.11	T_I

Table 2.1: The first column denotes the specific information contained in the different mails and the second column explains whether or not the information is relevant for position estimation. Lastly, the third column denotes the relevant update cycle.

2.2.2 Idle RSS

When a PP is idle, and, therefore, not transmitting through the direct link, a scan for FP broadcasts is issued every T_I. In this scan the PP might detect one or several FPs different from the direct link FP. If this is the case, the RSS value of the detected signals will be received and stored in the PP.

2.2.3 Average RSS

In the direct link, the PP will receive an average RSS value of 16 DECT frames every T_A. It is possible to get each individual RSS value for each frame if desired but such a data stream should not be stored in the PP due to power and memory constraints.

Both idle and average RSS information are indications of propagated distance as the signal strength decays over distance. We can use the log-distance pathloss model [5] to relate RSS to distance

$$RSS(d) = PL_0 + 10 \cdot \eta \log_{10} \frac{d}{d_0} + X_G \quad [\text{dB}]. \tag{2.1}$$

This model is similar to the one explained above but in (2.1) shadowing effects are modelled through the X_G term. Thus, if we know the signal strength near the transmitter, we can relate the received signal strength to the propagated distance through the pathloss. It will, therefore, be highly relevant to incorporate both active and idle RSS information in the model as these are indicators of the distance between each pair of PP and FP.

2.2.4 Background RSS

There may be non-relevant devices in the environment which also broadcast signals. These signals are not relevant for describing the distance between the FPs and PPs. However, the FP stores information of the "background" RSS levels. This value could potentially be used for quantization or calibration of the system. Thus, this piece of information cannot directly be used to locate a PP, as the distance information in this RSS measurement do not relate the FP to the specific PP.

2.2.5 Carrier frequency

The DECT protocol allows for 10 different carrier frequencies $f_0, \ldots, f_9$ in Europe. The allocated carrier is received in the PP in both active and idle state. The relevance of the carrier frequency for localization purposes emerges when we examine Friis' free-space pathloss equation

$$FSPL = \left(\frac{4\pi \cdot d}{c/f} \right)^2. \tag{2.2}$$

From (2.2) we see that changing the frequency will change the free-space pathloss. However, as DECT occupies the frequency band $1880 - 1900$ MHz with a carrier-carrier spacing of 1728 kHz the difference in pathloss from different carriers is limited. If a signal propagates a distance of 20 m and the carrier is either 1880 MHz or 1900 MHz the difference in FSPL for the two different signals are

$$\frac{||FSPL_{1880} - FSPL_{1900}||}{FSPL_{1900}} \cdot 100 = 2.09\%.$$

Therefore, incorporating the carrier information in the model may not yield significant changes to the result and we have, therefore, chosen not to model it. However, it does contain information which is relevant for the FSPL and, therefore, also RSS.

2.2.6 Number of successful synchronization words

Each DECT frame is composed of 24 time slots as described in Section 2.1. Each of these slots contains a synchronization word which is used for packet synchronization. In the direct link between a PP and FP, this information is received at the

PP thus indicating the quality of the link. Although being an inferior information source of the PP position, this information is dependent on the location of both PP and FP and the distance between them. The number of successful synchronization words should e.g. be zero if the PP and FP are located on different sides of the Earth (as an exaggerated example of course).

2.2.7 A and X field check sum

The A field in a DECT frame time slot handles control and manages signals. It utilizes a *cyclic redundancy check* (CRC) 16 bit error control code which detects changes in the data [13]. The X field occupies 4 bits in a slot and is added in order to ensure, that a signal does not collide with another bearer. The check sum of both A and X field, therefore, describes the quality of the signal. There is no immediate connection between distance and the check sum of the A and X field other than the argument described in the discussion of the number of successful synchronization words.

2.2.8 Blind slot info

In the direct link, the FP registers which slots are in use in each frame. This information is sent to the PP. The blind slot info gives an indication of the number of PPs in some neighbourhood of the FP. There is no immediate connection to position estimation apart from using this information for cooperative localization but as the system do not offer PP to PP communication this information may not be relevant.

2.2.9 DECT frame slot

The PP will in both active and idle state receive information of which time slot it is using (active) or would have used (idle) if it were to connect to the specific FP. This information is updated each T_A and T_I respectively but has no immediate connection to position.

2.2.10 FP ID

Alongside both active and idle RSS information, the PP will also receive an identification of the broadcasting FP. This information is important for localization purposes, as the positions of the FPs are assumed known a priori and thus the latent distance information in the RSS can be inferred with the fixed base station position.

2.2.11 Phase

It is possible to obtain crude TOA information in the system. When an FP emits a signal at time $t = 0$, it will register the delay of the returning signal from the PP. Unfortunately, the delay is registered in bit times. The bit time used in the system is 868.056 ns. Therefore, using the speed of light, the precision in this TOA information is $\frac{c \cdot 868.056 \cdot 10^{-9}\mathrm{s}}{2} \approx 130$ m which is imprecise for practical systems. However, it is possible to record this TOA information in ninths of bit times which renders the distance information to be of approximately 15 m. This is, unfortunately, only possible in the direct link and, therefore, TOA information from other FPs is not relevant. Additionally, the phase information obtained from other FPs is relative to the direct link TOA and they may, therefore, be asynchronous. As the phase information also holds latent distance information it is relevant to incorporate it in the model.

2.3 Probabilistic model for static localization

We seek an expression of the posterior probability of the PP position conditioned on the relevant information described above. For simplicity, let the data available and the position of a PP be denoted by $\mathcal{D}$ and $\boldsymbol{x}_p$ respectively. Let us express the posterior probability

$$\begin{aligned} p(\boldsymbol{x}_p|\mathcal{D}) &= \frac{p(\mathcal{D}|\boldsymbol{x}_p)p(\boldsymbol{x}_p)}{p(\mathcal{D})} \\ &\propto p(\mathcal{D}|\boldsymbol{x}_p)p(\boldsymbol{x}_p). \end{aligned} \tag{2.3}$$

We will shortly identify the components in (2.3).

Posterior PDF, $p(\boldsymbol{x}_p|\mathcal{D})$: The posterior PDF combines the observed data with the prior information of the PP position.

Likelihood function, $p(\mathcal{D}|\boldsymbol{x}_p)$: The likelihood expresses the probability that the data, $\mathcal{D}$, was generated given $\boldsymbol{x}_p$.

Prior PDF, $p(\boldsymbol{x}_p)$: Represents the information about the PP position before data has been observed.

Evidence PDF, $p(\mathcal{D})$: The evidence is independent of the PP position and reflects the probability of observing a particular realization of $\mathcal{D}$. This is often infeasible to compute and its primary objective is to ensure that the posterior probability integrates to one. Therefore it is often neglected as done in (2.3).

In order to compute (2.3), we need an expression of the joint probability, $p(\boldsymbol{x}_p, \mathcal{D})$. However, the joint density might be rather complex and most likely intractable to marginalize for large networks.

From Section 2.2, we have deemed the location of the PPs and FPs, RSS, phase and the ID of the respective FPs important for localization. As information from distinct base stations are independent, we can without loss of generality assume that a given PP only observes a single FP from which it can obtain the aforementioned information. This will relieve the derivation of the inference problem and adding additional observed base stations will scale according to the theory of independent random variables. Furthermore, we assume that the position of the FP is known a priori and thus obtaining the ID information is equivalent to observing the position, $\boldsymbol{x}_a$, of that FP. If the PP only observes a single FP, it will establish a direct link and, therefore, the phase information will be available. We will later allow the PP to obtain information from other FPs from which the phase is not available and make alterations to the inference problem. In order to describe possible fading effects, we will add a latent variable, θ, to the inference problem. We will, however, introduce additional variables to the inference problem. The new variables will represent a radial component and an angle between the PP and a specific base station which we will denote by r and v respectively. The argument for introducing the variables is that in an ideal scenario, if the true distance and angle between a PP and a base station are known, the location of the PP is fully determined. If the distance and angle are known, the PP position can be found through the polar representation

$$\boldsymbol{x}_p = r \cdot \begin{bmatrix} \cos v \\ \sin v \end{bmatrix} + \boldsymbol{x}_a \tag{2.4}$$

where we have added the known position of the FP, $\boldsymbol{x}_a$. If we do not add the FP position, the PP position will be represented in respect to the local coordinate system with origin in the FP position, see Figure 2.3.
Therefore, adding the position will correctly represent $\boldsymbol{x}_p$ in the global system. Although this representation seems attractive, we do, unfortunately, not receive any angle information and, therefore, it might seem insignificant to add the angle variable. However, mapping the one dimensional distance information latent in the RSS or phase into a two dimensional value describing the PP position seems unlikely to produce reliable results. Therefore, adding the angle variable might prove useful if we can update this information conditioned on the other variables.

Thus, the joint probability density of the inference problem for a system consisting of a single PP with observations from a single FP can be expressed as

$$p(\boldsymbol{x}_p, \boldsymbol{x}_a, r, v, \phi, \theta, rss) \tag{2.5}$$

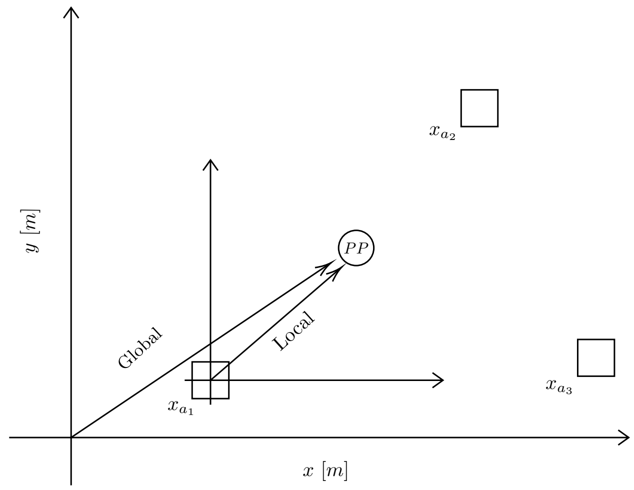


Figure 2.3: Adding the base station position ensures that the PP position is in respect to the global system.

where *rss* and ϕ are the RSS and phase information respectively. However, in order to obtain the marginal posterior PDF of the PP position, $\boldsymbol{x}_p$, we need to perform the integration

$$p(\boldsymbol{x}_p|\boldsymbol{x}_a, r, v, \phi, \theta, rss) \propto \int p(\boldsymbol{x}_a, r, v, \phi, \theta, rss|\boldsymbol{x}_p)p(\boldsymbol{x}_p)d\boldsymbol{x}_p. \tag{2.6}$$

The integral in (2.6) can become rather unwieldy based on the model choices and adding additional observed FPs will only make it more complex.

The considerations and discussions in this chapter have served to specify the working system and present the inference problem needed to be solved for indoor localization. We have seen, that the DECT technology offers relevant information which can be used to obtain a position estimate of a PP. However, in order to compute the posterior probability of the PP position, we need to model the individual variables in (2.6). Additionally, we have made the observation, that this marginalization may become unwieldy, especially if more than one FP is observed. Therefore, we wish to explore inference algorithms which may relieve the marginalization.

3 Probabilistic modeling of indoor localization and inference methods

In this chapter we will explore inference algorithms which can procure marginal posterior probabilities in an efficient and structured manner.

3.1 Graphical models

As already mentioned, certain inference algorithms have shown to produce reliable results while maintaining a low computational overhead. We will, therefore, investigate such algorithms, especially *message passing* algorithms. However, as message passing algorithms are applied to a specific class of inference problems, that can be represented graphically, we need to investigate graphical models.

If a statistical model, in the form of a joint PDF, can be factorized into several factors, e.g. by the use of the product rule

$$p(a, b) = p(b|a)p(a), \tag{3.1}$$

the inference problem can be visualized through a *directed acyclic graph* (DAG), see [14] for more on graphs. Using the graph object, we can represent the inference variables through nodes and the statistical dependencies by edges which creates a structured graphical representation of the problem. Specifically, if a joint PDF can be factorized by the use of the product rule (3.1), the inference problem can be visualized with a *factor graph* [15]. In a factor graph, each component of the factorized joint PDF is represented by square factor nodes and each stochastic variable is represented by a circular node. The nodes are then connected with respect to their statistical dependencies. In Figure 3.1, a factor graph has been visualized for a joint PDF which factorizes as

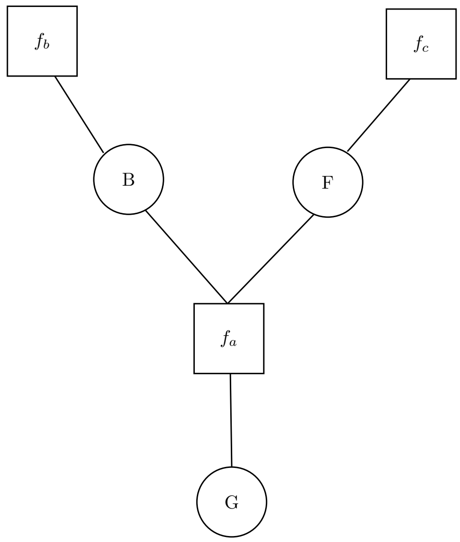


Figure 3.1: A factor graph representing the joint distribution given by $p(B, F, G) = \underbrace{p(G|B, F)}_{f_a}\underbrace{p(B)}_{f_b}\underbrace{p(F)}_{f_b}$.

$$p(B, F, G) = \underbrace{p(G|B, F)}_{f_a}\underbrace{p(B)}_{f_b}\underbrace{p(F)}_{f_b}.$$

For the present inference problem in (2.5) we can derive a similar factor graph by applying the product rule (3.1) to the joint distribution

$$\begin{aligned}
p(\boldsymbol{x}_p, \boldsymbol{x}_a, r, v, \phi, \theta, rss) &= p(\boldsymbol{x}_p, \boldsymbol{x}_a, r, \phi, \theta, rss|v)p(v) \\
&= p(\boldsymbol{x}_p, \boldsymbol{x}_a, r, \theta, rss|v, \phi)p(v) \\
&= p(\boldsymbol{x}_p, \boldsymbol{x}_a, r, rss|v, \phi, \theta)p(v)p(\theta) \\
&= p(\boldsymbol{x}_p, \boldsymbol{x}_a, rss|v, \phi, \theta, r)p(v)p(\theta)p(r) \\
&= p(\boldsymbol{x}_p, \boldsymbol{x}_a|v, r)p(rss|r, \theta)p(r|\phi)p(v)p(\theta)p(r) \\
&= \underbrace{p(\boldsymbol{x}_p|v, r, \boldsymbol{x}_a)}_{a}\underbrace{p(rss|r, \theta)}_{c}\underbrace{(r|\phi)}_{b}p(v)p(\theta)p(r)
\end{aligned} \tag{3.2}$$

where we do not multiply with $p(\boldsymbol{x}_a)$ and $p(\phi)$ as these variables are observed. Thus a factor graph of the obtained factorized joint PDF can be derived, see Figure 3.2.

We will later need a model of the relationship between RSS and distance. Therefore, we propose the following RSS model based on the log-distance pathloss model in (2.1)

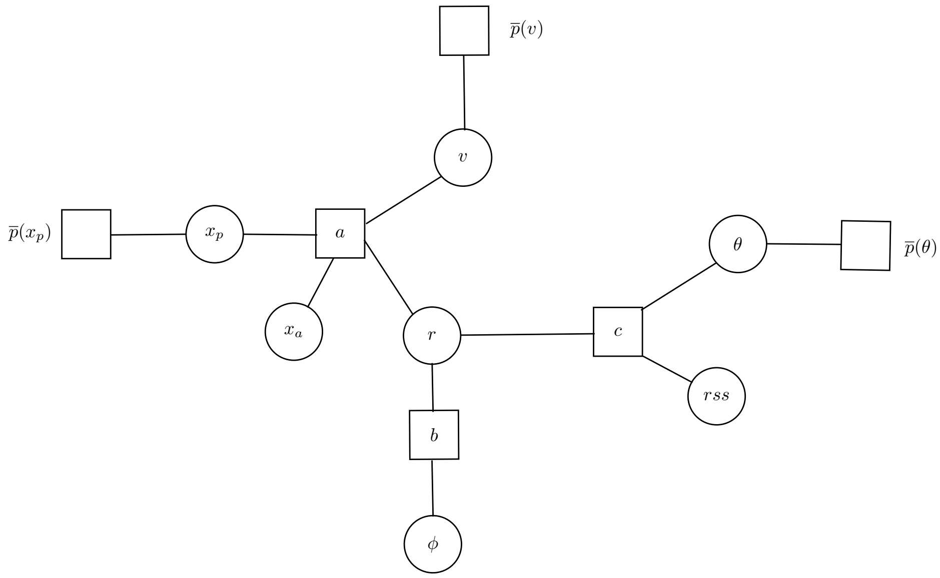


Figure 3.2: The factor graph representing (3.2). Each of the unobserved variables are connected to a factor node containing prior knowledge, $\bar{p}(\cdot)$, of the variable.

$$\begin{aligned} RSS(r) &= RSS_0 - PL(r) \\ &= RSS_0 - PL_0 - 10 \cdot \eta \log_{10} \frac{r}{d_0} - X_G. \end{aligned} \tag{3.3}$$

In (3.3), r is the radial component, RSS_0 is the signal strength at a short range from the transmitter and $X_G \sim \mathcal{N}(0, \theta^{-1})$ is a random variable modeling fading effects where θ is the precision of the normal distribution. Although small scale fading can be well modeled through a Rayleigh distribution [16], the purpose of our model is to represent variations in the RSS observations and not the causality of these variations. Therefore, letting X_G be modeled by a normal distribution might be sufficient.

3.2 Approximate inference

For those models where the joint PDF can be visualized by a factor graph, the inference problem can be solved either exactly or approximately by the use of message passing algorithms. For inference problems which can be represented by a factor graph containing no cycles, i.e. one and only one path exists between each set of nodes, exact inference can be employed through the *sum-product* algorithm if the joint distribution is tractable. This algorithm computes the exact marginal distribution for each unobserved variable. However, as the fading descriptor, θ, in our inference problem will be shared among all communication links, i.e. observed FPs, the factor graph will contain cycles and the sum-product algorithm is, therefore, not guaranteed to produce exact inference. As we have already argued that marginalization may become rather complex in the present inference problem, we will, therefore, seek an alternative inference algorithm but for more information on exact inference and the sum-product algorithm see [15].

If we accept that exact inference may not be required to obtain a sufficient marginal posterior of the PP location, we can explore approximate inference algorithms. Say that we have a probabilistic model which is parametrized by a set of parameters $\mathbf{\Theta}$ and that the model includes both observed and unobserved variables, denoted by $\boldsymbol{x}$ and $\mathbf{z}$ respectively. For simplicity, assuming only one FP is observed, these variables are $\boldsymbol{x} = \{\boldsymbol{x}_a, \phi, rss\}$ and $\mathbf{z} = \{\boldsymbol{x}_p, \theta, r, v\}$ in our model. For such inference problems, we may be interested in finding the maximum a posteriori (MAP) estimate, $\hat{\mathbf{\Theta}}_{MAP}$, given $\boldsymbol{x}$ as we might have prior knowledge of e.g. the PP position. In the given inference problem, $\hat{\mathbf{\Theta}}_{MAP}$ is the estimate of the PP position, fading descriptor θ, radial component r and angle v. Initially, we observe that the log-posterior PDF of $\mathbf{\Theta}$ can be expressed as

$$\ln p(\mathbf{\Theta}|\mathbf{x}) = \ln p(\boldsymbol{x}|\mathbf{\Theta}) + \ln p(\mathbf{\Theta}) - \ln p(\boldsymbol{x}). \tag{3.4}$$

For the first term on the right hand side in (3.4) we can make the following decomposition

$$\ln p(\boldsymbol{x}|\mathbf{\Theta}) = \underbrace{\int q(\mathbf{z}) \ln \left(\frac{p(\boldsymbol{x}, \mathbf{z}|\mathbf{\Theta})}{q(\mathbf{z})} \right) \mathrm{d}\mathbf{z}}_{\mathcal{L}(q,\mathbf{\Theta})} - \underbrace{\int q(\mathbf{z}) \ln \left(\frac{p(\mathbf{z}|\mathbf{x}, \mathbf{\Theta})}{q(\mathbf{z})} \right) \mathrm{d}\mathbf{z}}_{-KL(q||p)} \tag{3.5}$$

where $q(\mathbf{z})$ is an auxiliary density over the unobserved variables $\mathbf{z}$. We identify the last term in (3.5) as the negative Kullback-Leibler divergence between the auxiliary PDF $q(\mathbf{z})$ and $p(\mathbf{z}|\boldsymbol{x}, \mathbf{\Theta})$.

Definition 3.1 (Kullback-Leibler divergence [17])
For distributions p and q of a continuous random variable, x, the negative *Kullback-Leibler divergence* is defined as

$$-KL(p||q) = \int_{-\infty}^{\infty} p(x) \ln \frac{q(x)}{p(x)} \mathrm{d}x. \tag{3.6}$$

The Kullback-Leibler divergence is a measure of how one probability distribution diverges from another. In Bayesian inference the Kullback-Leibler divergence, $KL(p||q)$, is the amount of information which is lost by approximating p by q where p represents the true distribution of the data and q is an auxiliary PDF [18].

Returning to (3.5) we know that $KL(q||p) \geq 0$ which ensures that $\mathcal{L}(q, \boldsymbol{\Theta})$ is a lower bound of the log-likelihood, $\ln p(\boldsymbol{x}|\boldsymbol{\Theta})$, i.e.

$$\mathcal{L}(q, \boldsymbol{\Theta}) \leq \ln p(\boldsymbol{x}|\boldsymbol{\Theta})$$

with equality if and only if $q(\mathbf{z}) = p(\mathbf{z}|\boldsymbol{x}, \boldsymbol{\Theta})$. Thus in order to approach the true MAP estimate, $\hat{\boldsymbol{\Theta}}_{MAP}$, the right hand side of (3.4)

$$\mathcal{L}(q, \boldsymbol{\Theta}) + KL(q||p) + \ln p(\boldsymbol{\Theta}) - \ln p(\boldsymbol{x})$$

can be iteratively maximized separately with respect to q and $\boldsymbol{\Theta}$. This method of finding the MAP estimate is employed by the EM-algorithm which we will not present here, but we direct the interested reader to [15] for more information on the algorithm.

Let us, however, assume that we have a graphical representation in which all unobserved variables are assigned a prior distribution. We once more denote all observed variables by $\boldsymbol{x}$ and all unobserved variables by $\mathbf{z}$. The model specified by this graph can thus be represented by the joint PDF $p(\boldsymbol{x}, \mathbf{z})$. The parameters $\boldsymbol{\Theta}$ no longer appears in the density as the parameters are stochastic variables and, therefore, absorbed into $\mathbf{z}$. As before, we will make a decomposition of the log-evidence (marginal likelihood) function of the model. The decomposition reads

$$\begin{aligned}\ln p(\boldsymbol{x}) &= \int q(\mathbf{z}) \ln \left(\frac{p(\boldsymbol{x}, \mathbf{z})}{q(\mathbf{z})} \right) \mathrm{d}\mathbf{z} - \int q(\mathbf{z}) \ln \left(\frac{p(\mathbf{z}|\boldsymbol{x})}{q(\mathbf{z})} \right) \mathrm{d}\mathbf{z} \\ &= \mathcal{L}(q) + KL(q||p).\end{aligned}$$

It is important to notice that since $\ln p(\boldsymbol{x})$ is constant and $KL(q||p) \geq 0$, maximizing $\mathcal{L}(q)$ with respect to q is equivalent to minimizing $KL(q||p)$. If this is done without any restrictions on q we obtain

$$\arg\max_q \mathcal{L}(q) = \arg\min_q KL(q||p) = p(\mathbf{z}|\boldsymbol{x}).$$

The problem is, however, that working with the joint posterior distribution $p(\mathbf{z}|\boldsymbol{x})$ is impractical for large system with many variables. We are usually interested in computing only some of the marginal posteriors

$$p(\mathbf{z}_i|\boldsymbol{x}) = \int p(\mathbf{z}|\boldsymbol{x}) \mathrm{d}\mathbf{z}_{\setminus i}$$

where $\mathbf{z}_{\setminus i}$ is all the variables in $\mathbf{z}$ except $\mathbf{z}_i$. We have already discussed that these marginals can be derived exactly for tractable factor graphs by applying the sum-product algorithm. However, when exact inference of the marginal posteriors is intractable we can alternatively choose to approximate them by maximizing the lower bound $\mathcal{L}(q)$ (equivalent to minimizing $KL(q||p)$) with respect to a restricted and simpler class of auxiliary PDFs $q(\mathbf{z})$. This can be done by employing *mean field approximation*.

Definition 3.2 (Mean field approximation [15])
Given an auxiliary PDF, $q(\mathbf{z})$, suppose that $\mathbf{z}$ can be partitioned into disjoint groups $\mathbf{z}_i$, $i = 1, \dots, M$. Mean field approximation assumes that $q(\mathbf{z})$ factorizes with respect to these partitions

$$q(\mathbf{z}) = \prod_{i=1}^{M} q_i(\mathbf{z}_i). \tag{3.7}$$

The mean-field approximation does not assume a specific distribution for $q(\mathbf{z})$ nor does it restrict the individual factors $q_i(\mathbf{z}_i)$ to have a specific distribution.

With the factorization defined in Definition 3.2 the lower bound $\mathcal{L}(q)$ can be maximized with respect to each of the factors $q_i(\mathbf{z}_i)$ sequentially by keeping the other factors fixed. This is done by inserting the factorized auxiliary function from (3.7) in the expression of $\mathcal{L}(q)$

$$\begin{aligned}
\mathcal{L}(q) &= \int q(\mathbf{z}) \ln\left(\frac{p(\boldsymbol{x}, \mathbf{z})}{q(\mathbf{z})}\right) \mathrm{d}\mathbf{z} \\
&= \int \left(\prod_{i=1} q_i(\mathbf{z}_i)\right) \ln p(\boldsymbol{x}, \mathbf{z}) \mathrm{d}\mathbf{z} - \sum_{j=1}^{M} \underbrace{\int \left(\prod_{i=1} q_i(\mathbf{z}_i)\right) \ln q_j(\mathbf{z}_j) \mathrm{d}\mathbf{z}}_{\int q_j(\mathbf{z}_j) \ln q_j(\mathbf{z}_j) \mathrm{d}\mathbf{z}_j} \\
&= \int \left(\prod_{i=1} q_i(\mathbf{z}_i)\right) \ln p(\boldsymbol{x}, \mathbf{z}) \mathrm{d}\mathbf{z} - \sum_{i=1}^{M} \int q_i(\mathbf{z}_i) \ln q_i(\mathbf{z}_i) \mathrm{d}\mathbf{z}_i.
\end{aligned}$$

Then, if we want to maximize $\mathcal{L}(q)$ with respect to $q_j(\mathbf{z}_j)$, we can consider terms that do not contain $q_j(\mathbf{z}_j)$ as constants, i.e

$$\mathcal{L}(q) = \int q_j(\mathbf{z}_j) \left\{ \int \left(\prod_{i \neq j} q_i(\mathbf{z}_i)\right) \ln p(\boldsymbol{x}, \mathbf{z}) \mathrm{d}\mathbf{z}_{\setminus j} \right\} \mathrm{d}\mathbf{z}_j - \int q_j(\mathbf{z}_j) \ln q_j(\mathbf{z}_j) \mathrm{d}\mathbf{z}_j + \text{consts.}$$

We can define a distribution $\tilde{p}(\boldsymbol{x}, \mathbf{z}_j)$ such that

$$\ln \tilde{p}(\boldsymbol{x}, \mathbf{z}_j) = \underbrace{\int \ln p(\boldsymbol{x}, \mathbf{z}) \left(\prod_{i \neq j} q_i(\mathbf{z}_i) \mathrm{d}\mathbf{z}_i \right)}_{E_{q_{\setminus j}}[\ln p(\boldsymbol{x}, \mathbf{z})]} + \text{consts.} \tag{3.8}$$

where $E_{q_{\setminus j}}[\ln p(\boldsymbol{x}, \mathbf{z})]$ is the expectation of $\ln p(\boldsymbol{x}, \mathbf{z})$ with respect to $q_{\setminus j}$. With the distribution in (3.8) we can write

$$\mathcal{L}(q) = \underbrace{\int q_j(\mathbf{z}_j) \ln \left(\frac{\tilde{p}(\boldsymbol{x}, \mathbf{z}_j)}{q_j(\mathbf{z}_j)} \right) \mathrm{d}\mathbf{z}_j}_{-KL(q_j(\mathbf{z}_j) || \tilde{p}(\boldsymbol{x}, \mathbf{z}_j))} + \text{consts.}$$

and since $KL(q_j(\mathbf{z}_j) || \tilde{p}(\boldsymbol{x}, \mathbf{z}_j)) \geq 0$, the maximum of $\mathcal{L}(q)$ with respect to q_j can be found as the minimum of $KL(q_j(\mathbf{z}_j) || \tilde{p}(\boldsymbol{x}, \mathbf{z}_j))$. Thus the optimal value for $q_j(\mathbf{z_j})$, for all other factors $q_i(\mathbf{z_i}), \;\; i \neq j$ fixed, is

$$q_j^*(z_j) \propto \exp\{E_{q_{\setminus j}}[\ln p(\boldsymbol{x}, \mathbf{z})]\}. \tag{3.9}$$

Usually, if we have a factorization of $p(\boldsymbol{x}, \mathbf{z})$, many of the factors do not depend on $\mathbf{z}_j$ and will, therefore, not influence $q_j^*(z_j)$. Therefore, the expectation in 3.9 needs only be taken for the log-terms of $\ln p(\boldsymbol{x}, \mathbf{z})$ that contain $\mathbf{z}_j$.

The update equations (3.9) for $q_j^*(\mathbf{z}_j)$ can be iteratively calculated for each of the factors based on the probabilistic model. These updates tend to be simpler if the involved PDFs belong to the *exponential family* of distributions which include e.g. the normal and Gamma distribution. For more information on the exponential family see [15]. As the name suggests, each density in the exponential family is proportional to a exponential term which is why the update equations tend to be more tractable to compute as we apply the logarithm to the density.

To employ the iterative procedure, the parameters of all the factors $q_i(\mathbf{z}_i), \;\; i = 1, \ldots, M$ need to be initialized, e.g. if $q_i(\mathbf{z}_i)$ is chosen to be a normal distribution we initialize the mean and variance of the prior distribution. The update equation is then iteratively applied to each factor which is typically done sequentially but an arbitrary schedule is allowed. Observing that we are maximizing the lower bound, $\mathcal{L}(q_j)$, at each iteration t, we have that

$$\mathcal{L}(q^{t-1}) \leq \mathcal{L}(q^t)$$

which, therefore, guarantees convergence of the procedure. However, it is not guaranteed that the procedure converges to the global maximum $\mathcal{L}(q) = \ln p(\mathbf{z}|\boldsymbol{x})$.

Messages from a variable node v_i to a factor node $g \in \mathcal{N}(v_i)$

$$m_{v_i \to \mathcal{N}(v_i)}(v_i) = \frac{1}{Z} \prod_{h \in \mathcal{N}(v_i)} m_{h \to v_i}(v_i) \tag{3.10}$$

Messages from local factors to variable v_i

$$m_{f \to v_i}(v_i) = p(v_i) \tag{3.11}$$

$$m_{g \to v_i}(v_i) = \exp\left(\int_{v \in \mathcal{N}(g) \neq v_i} \prod_{v \in \mathcal{N}(g) \neq v_i} m_{v \to g}(v) \ln g \, \mathrm{d}v_{\setminus i} \right) \tag{3.12}$$

Marginal update of the PDF estimate of v

$$q(v_i) = m_{v_i \to \mathcal{N}(v_i)}(v_i) \tag{3.13}$$

Figure 3.3: The variational message passing algorithm. In the equations, $\mathcal{N}(v_i)$ denotes the set of factor nodes neighbouring variable node v_i and $\mathcal{N}(g) \neq v_i$ the set of variable nodes neighbouring the factor node g excluding the variable node v_i. The Z constant is a normalization constant as defined in (3.14).

3.3 Approximate inference on factor graphs

Assuming that we have derived a factor graph, we have in the previous sections investigated both exact and approximate inference techniques. However, if the joint PDF is intractable or the graph contains cycles, marginalization may lead to impractical integrals and, therefore, approximate inference is preferred. Therefore, the variational methods explained in the above seem attractive. In order to adopt the factor graph message passing framework, we have to employ the variational methods on the graph structure. From the previous section we know that the variational method uses auxiliary functions to approximate the true probability densities and then updates these with the update equation in (3.9). The *Variational Message Passing* algorithm has been derived such that it employs exactly this update equation. Say that we have a factor graph in which variable nodes are denoted by v and the factor nodes by f if the factor node is the prior density of the variable or g otherwise. Thus the update equations from above can be computed by the VMP algorithm in Figure 3.3 which has been slightly modified according to [10] and [19].

In the VMP algorithm the normalization constant Z is defined as

$$Z = \int_{v_i} \prod_{h \in \mathcal{N}(v_i)} m_{h \to v_i}(v_i) \mathrm{d}v_i. \tag{3.14}$$

The VMP algorithm depicted in Figure 3.3 imposes no restrictions on the messages which are passed between the nodes in the factor graph. For tractability, the messages can be restricted to the exponential family as we have argued that this will produce simpler update equations. If so, the messages in (3.10) and (3.13) must be modified according to the exponential family of distributions, $\mathcal{E}$. This can be done by minimizing the Kullback-Leibler divergence between an auxiliary PDF q and the true message

$$m^{\mathcal{E}}_{v_i \to \mathcal{N}(v_i)}(v_i) = \arg\min_{q \in \mathcal{E}} KL(q(v_i)||\tilde{p}(v_i)) \tag{3.15}$$

where

$$\tilde{p}(v_i) = \frac{1}{Z} \prod_{h \in \mathcal{N}(v_i)} m_{h \to v_i}(v_i)$$

and

$$q^{\mathcal{E}}(v_i) = m^{\mathcal{E}}_{v_i \to \mathcal{N}(v_i)}(v_i).$$

Solving (3.15) requires that we find the parameters of the distribution $q(v_i) \in \mathcal{E}$ which minimize $KL(q(v_i)||\tilde{p}(v_i))$. This will be explore in the next chapter when we derive the model.

A few remarks should be attached to the algorithm and to message passing in general. If a *leaf* node, i.e. a variable node with only one edge attached, is observed, e.g. the RSS measurement, the message from the specific variable node is simply 1. Additionally, in all factors where the variable is present, the variable is fixed to the observed value. Furthermore, if a factor node is a leaf node, thus representing prior knowledge, the message from this factor node to the variable is the prior distribution of the specific variable, e.g. the prior distribution describing the PP location.

In the next chapter we will present and discuss the model choices and derive the needed messages for the above VMP algorithm.

4 Moment matching variational message passing

In this chapter we will apply the system specification from Chapter 2 and the theory of graphical models explored in the previous chapter to derive suitable model choices for the inference problem variables.

4.1 Statistical properties of the inference problem variables

In order to conduct inference on the devices and the available information, we need to assign statistical properties to both the FPs and PPs. Additionally, the relevant information and their statistical relations with the other components in the model have to be modeled. As we seek to employ the VMP algorithm to our inference problem, we will restrict the distributions to the exponential family.

A suitable and simple choice of the PP position prior is a bivariate normal distribution

$$p(\mathbf{x}_p) \sim \mathcal{N}(\boldsymbol{\mu}_{x_p}, \Sigma_{\mathbf{x}_p}).$$

This distribution is a member of the exponential family and is fully specified by its mean and covariance matrix. We will argue, that if the covariance matrix is small, the mean of the PDF is equivalent to the position of the PP. Thus, when we have applied the algorithm and updated the PP location, the mean of the variable will correspond to the position estimate. Consider a Taylor expansion of a scalar function of a vector variable X of which we take the expectation

$$\begin{aligned}\mathbb{E}[g(X)] &\approx \mathbb{E}[g(x_0) + g'(x_0)(x - x_0)] \\ &= g(x_0) + g'(x_0)\mathbb{E}[(X - x_0)].\end{aligned}$$

The function $g(\cdot)$ is continuously differentiable and is used to approximate the moments of the random variable through a linear function. If we choose $x_0 = \mu_X$, then the expectation is $\mathbb{E}[g(X)] \approx g(\mu_X)$. Thus the mean of the distribution seems to be a good approximation of the position if $g(\cdot)$ does not vary rapidly near x_0. In order to argue the precision of this approximation we can make similar observations of the variance

$$\begin{aligned}
\mathrm{Var}(g(X)) &= \mathbb{E}[(g(X) - g(\mu_X))^2] \\
&\approx \mathbb{E}[(g'(\mu_X)(X - \mu_X))^2] \\
&= g'^T(\mu_X)\mathbb{E}\left[(X - \mu_X)^T(X - \mu_X)\right] g'(\mu_X) \\
&= g'^T(\mu_X)\Sigma_X\, g'(\mu_X)
\end{aligned}$$

where Σ_X is the covariance matrix of X. We can thus see, that the exactness of approximating the position of the PP by its mean depends on the covariance matrix. Thus, if $\mathrm{Cov}(X_i, X_i),\ i = 1, 2$ are small, we are fairly certain that the position of the PP is equivalent to the mean of the distribution.

Returning to our model choices, we will use a circular symmetric normal PDF for the FPs. However, as the positions of the FPs are known, the variance is 0 and the mean corresponds to their position

$$p(\mathbf{x}_a) \sim \mathcal{N}(\boldsymbol{\mu}_{\mathbf{x}_a}, \mathbf{0}).$$

Therefore, the PDF of an FP reduces to a Dirac delta function located at $\mu_{\mathbf{x}_a}$ in $\mathbb{R}^2$. Additionally, as we have access to RSS measurements, we are interested in relating these to the distance between the PP and FP, i.e. the radial component r in the inference problem. With the model proposed in (3.3), the PDF of RSS conditioned on r and θ can be described with the following density

$$p(rss|r, \theta) \sim \mathcal{N}(\mu_r(r), \theta^{-1}) \tag{4.1}$$

where

$$\mu_r(r) = RSS_0 - PL_0 - 10 \cdot \eta \log_{10} \frac{r}{d_0}$$

is the mean signal strength which varies with r. We will, furthermore, model the fading effects, θ, such that it can be incorporated in the inference problem. The precision of a normal distribution is commonly modeled as a Gamma distribution. This is due to the fact, that the *conjugate prior* of the variance of a normal distribution is an inverse gamma distribution. If a posterior distribution, $p(\Theta|x)$, and the prior distribution, $p(\Theta)$, are in the same probability distribution family, e.g. the exponential family, the prior and posterior are said to be conjugate distributions and the prior is called the conjugate prior. The reason for using a conjugate prior is that it gives a closed form expression of the posterior [15]. Therefore, it seems appropriate to use a Gamma distribution for θ

$$p(\theta) \sim \Gamma(\alpha_\theta, \beta_\theta)$$

where α_θ and β_θ are the shape and inverse scale parameter of the Gamma distribution respectively.

As we can utilize the phase information, we are also interested in modeling the statistical relationship between the radial component r and the phase. Assuming that the signal may have been reflected or the bit time representing the phase has been rounded wrongly, the phase can be modeled as

$$\phi = \frac{\tau}{T_B} + \mathcal{N}(0, \sigma_\phi^2)$$

where τ is the measured delay of the signal, T_B is the bit time and the normal term is included in order to model possible reflections and rounding errors. The delay, τ, is calculated from the propagated distance

$$\tau = 2 \cdot \frac{r}{c} + T_{pp}$$

where c is the speed of light and T_{pp} is the processing time in the system which arises if the signal has to wait in order to get a slot in the DECT frame. We will, however, for simplicity, assume that T_{pp} is zero such that the PP will always be able to communicate with a specific FP. From this expression we can rearrange the terms in favor of the distance r

$$r = \frac{\phi \cdot T_b \cdot c}{2} - \mathcal{N}(0, \sigma_\phi^2)$$

such that the statistical relationship between the distance and phase can be expressed as

$$p(r|\phi) \sim \mathcal{N}(\mu_\phi(\phi), \sigma_\phi^2)$$

where $\mu_\phi(\phi) = \frac{\phi \cdot T_b \cdot c}{2}$.

Finding a suitable prior model for the radial component r, we need a distribution in the exponential family which is continuous and has support on $\mathbb{R}_+$. The Gamma distribution is once more a candidate and we, therefore, choose a prior distribution on r as

$$p(r) \sim \Gamma(\alpha_r, \beta_r).$$

In order to find a prior distribution of the angle variable v, we seek a distribution which is defined for all inputs between 0 and 2π. The von Mises distribution fulfils this requirement and it also belongs to the exponential family. This distribution is defined by a mean angle, μ, and a concentration parameter κ. We therefore model the prior as

$$p(v) \sim VM(\mu_v, \kappa_v) = \frac{1}{2\pi I_0(\kappa_v)} \exp(\kappa_v \cdot \cos(v - \mu_v)) \tag{4.2}$$

where $I_0(\kappa)$ is the zero order modified Bessel function [20]

$$I_0(x) = \sum_{m=0}^{\infty} \frac{1}{m!\Gamma(m+1)} \left(\frac{x}{2}\right)^{2m}.$$

Finally, from the factorization in (3.2) we observe that the conditional distribution $p(\boldsymbol{x}_p|v, r, \boldsymbol{x}_a)$ is a degenerate distribution, as the position of the PP is fully specified if we know the position of the FP and the distance and angle to the PP. Therefore, this conditional probability is a Dirac delta which has a few complications for the VMP algorithm but will be fixed in the following section.

4.2 Derivation of messages

Until now we have only considered a single FP observation and, therefore, also only a single RSS and phase measurement. However, localization using a single FP will most likely fail to find a good estimate, due to unambiguous estimation, and, therefore, we now assume that more than one FP is observed. The PP will then observe several RSS measurements but still only the phase information from the direct link FP, which we will assume, without loss of generality, is the FP which is observed first. Including more FPs in the inference problem changes the joint distribution which now reads

$$p(\boldsymbol{x}_p, \boldsymbol{x}_{a1}, \ldots \boldsymbol{x}_{aN}, r_1, \ldots r_N, v_1, \ldots v_N, rss_1, \ldots rss_N, \phi, \theta)$$

if we assume that N FPs have been observed. Therefore, the factor graph of the altered inference problem is different from the one depicted in Figure 3.2. The new factor graph can be seen in Figure 4.1 where we observe, that the new factor graph is quite similar to the original as we are just adding new branches (without the phase information) to the $\boldsymbol{x}_p$ node.

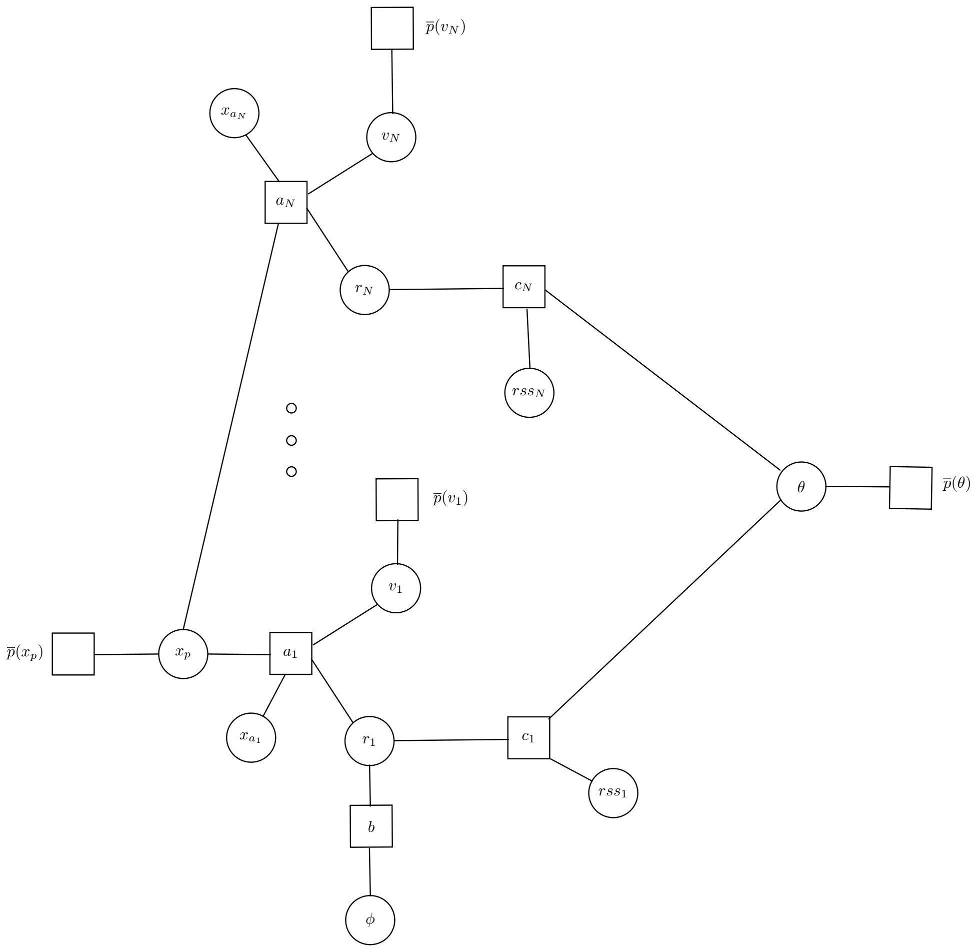


Figure 4.1: The factor graph visualization of the full inference problem. It is seen, that for each observed FP we just add another branch to the PP node.

As the network does not allow for communication between a pair of PPs the marginal posterior of a single PP will not depend on other PPs. Therefore, when we derive the messages, we will do so for a single PP. The messages will have the same form for all PPs in the system. For each observed FP, i.e. each branch in Figure 4.1, we draw a local factor $g_{\boldsymbol{x}_p,r,v,\boldsymbol{x}_a}$ (representing $a_1,\ldots,a_N$) for the node connecting $\boldsymbol{x}_p, r$ and v and a factor $g_{r,\theta,rss}$ (representing $c_1,\ldots,c_N$) for the node connecting r,θ and rss. If the FP is the direct link we also draw a factor $g_{r,\phi}$ (representing b) connecting the r and ϕ variables. These factor nodes are then connected to the variable nodes with which they have statistical relationships. Furthermore, factor nodes representing the prior distribution of the unobserved variables are also drawn, denoted by $\bar{p}(\cdot)$.

For the sake of readability, we will list the messages which we need to derive:

$m_{g_{\boldsymbol{x}_p,r,v,\boldsymbol{x}_a}\to\boldsymbol{x}_p}(\boldsymbol{x}_p)$: Message from r and v to $\boldsymbol{x}_p$

$m_{g_{\boldsymbol{x}_p,r,v,\boldsymbol{x}_a}\to r}(r)$: Message from $\boldsymbol{x}_p$ and v to r

$m_{g_{r,\phi}\to r}(r)$: Message from ϕ to r

$m_{g_{r,\theta,rss}\to r}(r)$: Message from θ and rss to r

$m_{g_{\boldsymbol{x}_p,r,v,\boldsymbol{x}_a}\to v}(v)$: Message from $\boldsymbol{x}_p$ and r to v

$m_{v\to\mathcal{N}(v)}(v)$: Message from v to $\mathcal{N}(v)$ (the neighbourhood of v)

$m_{g_{r,\theta,rss}\to\theta}(\theta)$: Message from r and rss to θ

Updating $\boldsymbol{x}_p$

In this section we derive the message $m_{g_{\boldsymbol{x}_p,r,v,\boldsymbol{x}_a}\to\boldsymbol{x}_p}(\boldsymbol{x}_p)$. We know from Section 3.2, that for approximate inference we update an auxiliary PDF, q, until the Kullback-Leibler divergence between q and the true posterior PDF, $\tilde{p}$, is minimized. For the PP location, we are thus interested in minimizing

$$KL(q(\boldsymbol{x}_p)||\tilde{p}(\boldsymbol{x}_p)) = \int_{\boldsymbol{x}_p} q(\boldsymbol{x}_p)\ln\frac{q(x_p)}{\tilde{p}(\boldsymbol{x}_p)}\mathrm{d}\boldsymbol{x}_p.$$

We do, however, know from the VMP algorithm in Figure 3.3 that $\tilde{p}(\boldsymbol{x}_p)$ can be computed as

$$\tilde{p}(\boldsymbol{x}_p) = m_{x_p\to\mathcal{N}(x_p)}(x_p) \propto \prod_{h\in\mathcal{N}(x_p)} m_{h\to\boldsymbol{x}_p}(x_p). \tag{4.3}$$

where we neglect the normalization, Z. However, the product of these messages will most likely not be a known density and, therefore, intractable for further use in the message updates. Furthermore, each time the VMP algorithm is iterated, these messages will become increasingly complex and, therefore, we seek a simpler method for this update.

The VMP algorithm states that the message from a prior knowledge factor node is simply the prior probability. For the $\boldsymbol{x}_p$ node this message is, therefore, a bivariate normal density. From (4.3) we see that we have to multiply each incoming messages to $\boldsymbol{x}_p$. It will, therefore, be convenient if the messages from each of the local factors $g_{\boldsymbol{x}_p,r,v,\boldsymbol{x}_a}$ also follows a bivariate normal distribution. If so, we can use the fact that the product of two multivariate normal densities, $\mathcal{N}_1(\boldsymbol{\mu}_1, \Sigma_1)$ and $\mathcal{N}_2(\boldsymbol{\mu}_2, \Sigma_2)$, is also a multivariate normal density with parameters

$$\boldsymbol{\mu}_3 = \Sigma_2(\Sigma_1 + \Sigma_2)^{-1}\boldsymbol{\mu}_1 + \Sigma_1(\Sigma_1 + \Sigma_2)^{-1}\boldsymbol{\mu}_2 \tag{4.4}$$

$$\Sigma_3 = \Sigma_2(\Sigma_1 + \Sigma_2)^{-1}\Sigma_2. \tag{4.5}$$

The true messages calculated from the VMP update equations can be projected onto a bivariate normal PDF which is achieved by minimizing the Kullback-Leibler divergence between an auxiliary PDF, $q(\boldsymbol{x}_p) \in \mathcal{G}$, and the true messages from each of the factors where $\mathcal{G}$ denotes the family of bivariate normal distributions. However, a convenient result exists for minimizing Kullback-Leibler divergence between distributions in the exponential family. This result states that the auxiliary PDF which minimizes the Kullback-Leibler divergence is a PDF in the exponential family with the same moments as the true PDF [21], i.e.

$$q = \mathrm{proj}[p] \Leftrightarrow \forall_j \int_x g_j(x)q(x)\mathrm{d}x = \int_x g_j(x)p(x)\mathrm{d}x$$

where $g_j(x) = (1, x, x^2)$ represents the normalization term, first and second moment. For the bivariate normal density we, therefore, need to match the mean and covariance matrix of the incoming messages in order to make the projection. We will denote this method in the rest of the report as *moment matching*. With this in mind, we can proceed by finding the messages needed for the product in (4.3). However, as we have argued that these factors are degenerate, i.e. Dirac deltas, we are not allowed to use the update equation in the VMP algorithm. Fortunately, in the article by Dauwels [19], they present a solution to this problem. If the incoming messages to a degenerate factor node are of the same variable, the message from such a factor node can be computed as

$$m_{g\to x}(x) \propto m_{x_1\to g} \cdot m_{x_2\to g} \cdot \cdots m_{x_N\to g} \tag{4.6}$$

where $m_{x_i\to g}$, $i = 1, \ldots, N$ are the messages entering the factor node g. However, in the present situation the messages entering the factor $g_{\boldsymbol{x}_p,r,v,\boldsymbol{x}_a}$ are not functions of $\boldsymbol{x}_p$ but functions of r and v respectively. Thus we cannot immediately use this result. The argument for adding the r and v variable to the network was that if we know the distance and angle to a base station with known position then we can find the position of the PP through the transformation

$$\boldsymbol{x}_p = r \cdot \begin{bmatrix} \cos v \\ \sin v \end{bmatrix} + \boldsymbol{x}_a.$$

In order to proceed with the message derivations we can make a variable transformation according to the above such that we can compute the message $m_{g_{x_p,r,v,x_a} \to x_p}(\boldsymbol{x}_p)$ using (4.6). By using the Jacobian of the transformation, we can make a variable change while the probability density remains the same if we multiply the variable changed messages by the determinant of the Jacobian. Thus, in this case, the message to $\boldsymbol{x}_p$ can be calculated as

$$\begin{aligned} m_{g_{x_p,r,v,x_a} \to x_p}(\boldsymbol{x}_p) &\propto m_{r \to g_{x_p,r,v,x_a}}(r) \cdot m_{v \to g_{x_p,r,v,x_a}}(v) \cdot \det \mathcal{J} \Big|_{x_p = r \cdot \begin{bmatrix} \cos v \\ \sin v \end{bmatrix} + x_a.} \\ &= m_{r \to g_{x_p,r,v,x_a}}(r) \cdot m_{v \to g_{x_p,r,v,x_a}}(v) \cdot r \\ &= m_{r \to g_{x_p,r,v,x_a}}(||\boldsymbol{x}_p - \boldsymbol{x}_a||) \cdot m_{v \to g_{x_p,r,v,x_a}}\left(\arctan \frac{\boldsymbol{x}_{t_2}}{\boldsymbol{x}_{t_1}}\right) \cdot ||\boldsymbol{x}_p - \boldsymbol{x}_a|| \end{aligned}$$

where $\boldsymbol{x}_t = \boldsymbol{x}_p - \boldsymbol{x}_a = [\boldsymbol{x}_{t_1}, \boldsymbol{x}_{t_2}]^T$ is the vector obtained from subtracting $\boldsymbol{x}_a$ from $\boldsymbol{x}_p$ for which we need the angle, v. Thus the message is simply the product of a Gamma and a von Mises density.

Such a product is not a known density and finding the moments analytically might become rather complex. However, we can make a convenient observation. We know that the von Mises distribution defines a density on the unit circle where the density is concentrated around the mean angle if the concentration parameter κ is large. Additionally, the Gamma distribution describes a radial component with a mean length and a variance. Therefore, the density represented by the von Mises and Gamma product above might resemble the orange cloud drawn in Figure 4.2 with the location and form determined by the mean of the radial and angle component and the variances respectively.

With the observation in Figure 4.2, we can choose to approximate the mean vector and covariance matrix of the von Mises-Gamma product. An approximation of the mean vector of this product is

$$\hat{\boldsymbol{\mu}}_{m_{g_{x_p,r,v,x_a} \to x_p}(x_p)} \approx \mu_{m_{r \to g_{x_p,r,v,x_a}}(r)} \cdot \begin{bmatrix} \cos \mu_{m_{v \to g_{x_p,r,v,x_a}}(v)} \\ \sin \mu_{m_{v \to g_{x_p,r,v,x_a}}(v)} \end{bmatrix} \tag{4.7}$$

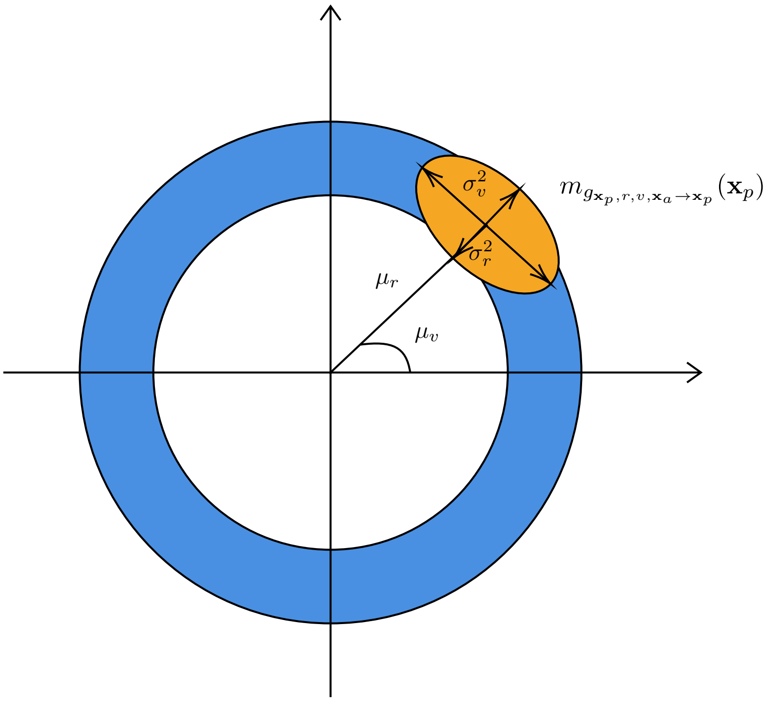


Figure 4.2: The probability density of the product of a Gamma and von Mises distribution is located on a circle with radius defined by the Gamma and an angle defined by the von Mises.

i.e. the mean of the product is located at a length and an angle determined by the Gamma and von Mises term respectively. In order to derive an approximation of the covariance matrix, imagine that the angle is zero, i.e. the probability density shown in Figure 4.2 has been rotated such that it is aligned with the coordinate axes. If this is the case, then an approximation of the covariance matrix is

$$\hat{\Sigma}_{m_{g_{x_p,r,v,x_a}\to x_p}(x_p)} \approx \begin{bmatrix} \sigma_r^2 & 0 \\ 0 & \sigma_v^2 \end{bmatrix}.$$

In order to model those situations where the mean angle is not zero, we can choose to multiply the covariance matrix approximation with a rotation matrix such that the final approximation is

$$\hat{\Sigma}_{m_{g_{x_p,r,v,x_a}\to x_p}(x_p)} \approx \begin{bmatrix} \cos \mu_{m_{v\to g_{x_p,r,v,x_a}}(v)} & -\sin \mu_{m_{v\to g_{x_p,r,v,x_a}}(v)} \\ \sin \mu_{m_{v\to g_{x_p,r,v,x_a}}(v)} & \cos \mu_{m_{v\to g_{x_p,r,v,x_a}}(v)} \end{bmatrix} \cdot \begin{bmatrix} \sigma_r^2 & 0 \\ 0 & \sigma_v^2 \end{bmatrix}. \tag{4.8}$$

Thus we can construct a bivariate normal distribution with these parameters and use it to update the PP location. Using this method of updating the $\boldsymbol{x}_p$ location is quite favorable as we only need to iterate the moments of a bivariate normal density in the network and not compute the product of a number of messages in (4.3) which do not follow the same distribution.

Updating r

In this section we derive the messages $m_{g_{x_p,r,v,x_a}\to r}(r)$, $m_{g_{r,\phi}\to r}(r)$ and $m_{g_{r,\theta,rss}\to r}(r)$. Initially, we will derive $m_{g_{x_p,r,v,x_a}\to r}(r)$.

We will once more seek to apply moment matching in order to mitigate complex message multiplications. Thus, for the r node, we want to project the incoming messages onto Gamma distributions. From the variable transformation we know that $r = ||\boldsymbol{x}_p - \boldsymbol{x}_a||$ and as the angle and radial component are independent, we can simply find the moments of the Euclidean norm $||\boldsymbol{x}_p - \boldsymbol{x}_a||$ and use them in projecting the message. Deriving the first moment, or equivalently the mean, we have to find $\mathbb{E}_{\boldsymbol{x}_p}[||\boldsymbol{x}_p - \boldsymbol{x}_a||]$. For the present properties of $\boldsymbol{x}_p$ and $\boldsymbol{x}_a$, the Euclidean norm does not follow a known density and we, therefore, have to find the first and second moment ourselves. Using the relationship

$$\mathrm{Var}(X) = \mathbb{E}[X^2] - \mathbb{E}[X]^2$$

we can find the mean of the Euclidean norm if we can find the variance and second moment. The variance of the variable can be derived by using the following observation

$$
\begin{aligned}
\mathrm{Var}(||X||) &= \mathbb{E}[||X - \mathbb{E}[X]||^2] \\
&= \mathbb{E}[\sum_{i=1}^{n}(X_i - \mathbb{E}[X_i])^2] \\
&= \sum_{i=1}^{n}\mathbb{E}[(X_i - \mathbb{E}[X_i])^2] \\
&= \sum_{i=1}^{n}\mathrm{Var}(X_i) \\
&= \sum_{i=1}^{n}\Sigma_{ii} \\
&= \mathrm{tr}(\Sigma)
\end{aligned}
$$

where $tr(\cdot)$ is the trace operator. Thus the variance of the Euclidean distance is just the trace of the covariance matrix of $\boldsymbol{x}_p$. We continue by computing the second moment of the variable

$$
\begin{aligned}
\mathbb{E}[||\boldsymbol{x}_p - \boldsymbol{x}_a||^2] &= \mathbb{E}[(\boldsymbol{x}_{p_1} - \boldsymbol{x}_{a_1})^2 + (\boldsymbol{x}_{p_2} - \boldsymbol{x}_{a_2})^2] \\
&= \boldsymbol{x}_{a_1}^2 + \boldsymbol{x}_{a_2}^2 + \mathbb{E}[\boldsymbol{x}_{p_1}^2] + \mathbb{E}[\boldsymbol{x}_{p_2}^2] - 2\cdot\boldsymbol{x}_{a_1}\mathbb{E}[\boldsymbol{x}_{p_1}] - 2\cdot\boldsymbol{x}_{a_2}\mathbb{E}[\boldsymbol{x}_{p_2}] \\
&= \boldsymbol{x}_{a_1}^2 + \boldsymbol{x}_{a_2}^2 + \Sigma_{\boldsymbol{x}_{p_{11}}} + \mu_{\boldsymbol{x}_{p_1}}^2 + \Sigma_{\boldsymbol{x}_{p_{22}}} + \mu_{\boldsymbol{x}_{p_2}}^2 - 2\cdot\boldsymbol{x}_{a_1}\mu_{\boldsymbol{x}_{p_1}} - 2\cdot\boldsymbol{x}_{a_2}\mu_{\boldsymbol{x}_{p_2}}.
\end{aligned}
$$

We can thus find the first moment

$$
\begin{aligned}
\mathbb{E}[||\boldsymbol{x}_p - \boldsymbol{x}_a||] &= \sqrt{\mathbb{E}[||\boldsymbol{x}_p - \boldsymbol{x}_a||^2] - \mathrm{Var}(||\boldsymbol{x}_p - \boldsymbol{x}_a||)} \\
&= \sqrt{\boldsymbol{x}_{a_1}^2 + \boldsymbol{x}_{a_2}^2 + \Sigma_{\boldsymbol{x}_{p_{11}}} + \mu_{\boldsymbol{x}_{p_1}}^2 + \Sigma_{\boldsymbol{x}_{p_{22}}} + \mu_{\boldsymbol{x}_{p_2}}^2 - 2\cdot\boldsymbol{x}_{a_1}\mu_{\boldsymbol{x}_{p_1}} - 2\cdot\boldsymbol{x}_{a_2}\mu_{\boldsymbol{x}_{p_2}} - \mathrm{tr}(\Sigma_{\boldsymbol{x}_p})}.
\end{aligned}
$$

Thus we have found the required moments of the message from g_{x_p,r,v,x_a} which we can convert into Gamma parameters, i.e. α and β, such that we can update the r node. The expected value and variance of a Gamma distribution are given by $\frac{\alpha}{\beta}$ and $\frac{\alpha}{\beta^2}$ respectively. In order to find the parameters we can treat these equations as a system of linear equations with two unknown parameters. We can thus find the parameters through the equations

$$
\begin{aligned}
\beta_{x_p} &= \frac{\mathbb{E}[||\boldsymbol{x}_p - \boldsymbol{x}_a||]}{\mathrm{Var}(||\boldsymbol{x}_p - \boldsymbol{x}_a||)} \\
&= \frac{\sqrt{\boldsymbol{x}_{a_1}^2 + \boldsymbol{x}_{a_2}^2 + \Sigma_{\boldsymbol{x}_{p_{11}}} + \mu_{\boldsymbol{x}_{p_1}}^2 + \Sigma_{\boldsymbol{x}_{p_{22}}} + \mu_{\boldsymbol{x}_{p_2}}^2 - 2\cdot\boldsymbol{x}_{a_1}\mu_{\boldsymbol{x}_{p_1}} - 2\cdot\boldsymbol{x}_{a_2}\mu_{\boldsymbol{x}_{p_2}} - \mathrm{tr}(\Sigma_{\boldsymbol{x}_p})}}{\mathrm{tr}(\Sigma_{\boldsymbol{x}_p})}
\end{aligned}
\tag{4.9}
$$

and

$$\alpha_{\boldsymbol{x}_p} = \mathbb{E}[||\boldsymbol{x}_p - \boldsymbol{x}_a||] \cdot \beta_{\boldsymbol{x}_p}. \tag{4.10}$$

In order to update the r node we make the following observation. Say we have two Gamma distributions over the same variable, $f_1(x) \sim \Gamma(\alpha_1, \beta_1), f_2(x) \sim \Gamma(\alpha_2, \beta_2)$. The product of the densities is

$$\begin{aligned} f_1(x) \cdot f_2(x) &= \frac{\beta_1^{\alpha_1}}{\Gamma(\alpha_1)} x^{\alpha_1 - 1} \cdot e^{-\beta_1 x} \cdot \frac{\beta_2^{\alpha_2}}{\Gamma(\alpha_2)} x^{\alpha_2 - 1} \cdot e^{-\beta_2 x} \\ &= \frac{\beta_1^{\alpha_1} \beta_1^{\alpha_2}}{\Gamma(\alpha_1)\Gamma(\alpha_2)} x^{(\alpha_1 + \alpha_2 - 1) - 1} e^{-(\beta_1 + \beta_2) x} \approx \Gamma(\alpha_1 + \alpha_2 - 1, \beta_1 + \beta_2) \end{aligned} \tag{4.11}$$

where we see that the product is proportional to another Gamma distribution, $\Gamma(\alpha_1 + \alpha_2 - 1, \beta_1 + \beta_2)$. Therefore, updating the r node will not change the original form of the density, only the parameters. In order to fully update the r node, we are also interested in projecting messages from the $g_{\phi,r}$ and $g_{\theta,r,rss}$ factor nodes onto Gamma densities. We can compute the message from the $g_{\phi,r}$ factor by using the VMP message equation

$$m_{g_{\phi,r}}(r) = \exp\left[\int_\phi m_{\phi \to g_{\phi,r}} \cdot \ln \mathcal{N}\left(\frac{\phi \cdot T_b \cdot c}{2}, \sigma_\phi^2\right) \mathrm{d}\phi\right].$$

However, as the phase is observed, this message is simply a normal distribution with known mean

$$m_{g_{\phi,r} \to r}(r) = \mathcal{N}\left(\frac{\phi_{obs} \cdot T_b \cdot c}{2}, \sigma_\phi^2\right).$$

Thus we simply have to convert these parameters to a Gamma density such that we can use the result of products of Gamma distributions to update r. These parameters can be computed by matching the mean of a Gamma distribution to the mean of the normal distribution

$$\beta_\phi = \frac{\frac{\phi \cdot T_b \cdot c}{2}}{\sigma_\phi^2} = \frac{\phi \cdot T_b \cdot c}{2\sigma_\phi^2} \tag{4.12}$$

$$\alpha_\phi = \frac{\phi \cdot T_b \cdot c}{2} \cdot \frac{\phi \cdot T_b \cdot c}{2\sigma_\phi^2} = \frac{(\phi \cdot T_b \cdot c)^2}{4\sigma_\phi^2}. \tag{4.13}$$

The message from the $g_{\theta,r,rss}$ factor node is, however, more complicated. We can derive the message by using the VMP algorithm

$$
\begin{aligned}
m_{g_{\theta,r,rss}\to r}(r) &= \exp\left[\int_{\theta}\int_{rss} m_{\theta\to g_{\theta,r,rss}} \cdot m_{rss\to g_{\theta,r,rss}} \cdot \ln\mathcal{N}\left(\mu_r(r), \theta^{-1}\right) \mathrm{d}rss\,\mathrm{d}\theta\right] \\
&= \exp\left[\mathbb{E}_{\theta}\left[\ln\mathcal{N}\left(\mu_r(r), \theta^{-1}\right)\right]\right] \\
&= \exp\left[\mathbb{E}_{\theta}\left[\ln\frac{1}{\sqrt{2\pi}} + \ln\theta - \frac{\theta^2}{2}(rss_{obs} - \mu_r(r))^2\right]\right] \\
&= \exp\left[\ln\frac{1}{\sqrt{2\pi}} + \psi(\alpha_\theta) - \ln\beta_\theta - \frac{1}{2}(rss_{obs} - \mu_r(r))^2\left(\frac{\alpha_\theta}{\beta_\theta^2} + \frac{\alpha_\theta^2}{\beta_\theta^2}\right)\right]
\end{aligned}
$$

where we have used that the mean of a log-Gamma variable, X, is given by [22]

$$\mathbb{E}[X] = \psi(\alpha) - \ln\beta.$$

in which $\psi(\cdot)$ is the digamma function. The derived message is clearly not a known density and we, therefore, need to find the first and second moments analytically which requires solving the integrals

$$\mu^{(1)} = \int_0^\infty r \cdot m_{g_{\theta,r,rss}\to r}(r)\mathrm{d}r \tag{4.14}$$

$$\mu^{(2)} = \int_0^\infty r^2 \cdot m_{g_{\theta,r,rss}\to r}(r)\mathrm{d}r. \tag{4.15}$$

However, as $m_{g_{\theta,r,rss}\to r}(r)$ is rather complex, solving the integrals may show cumbersome and we may need to employ numerical methods. In order to employ numerical methods we need to truncate the integration interval as we cannot integrate numerically from 0 to ∞. If the true density can be shown to be concentrated in a finite interval, we can employ numerical integration to find the moments. Using a numerical method for finding the parameters will be explored in the next chapter. Thus we have derived all the messages arriving at the r node and we can then update the Gamma parameters of the node through the observation in (4.11).

Updating v

In this section we derive the message $m_{g_{x_p,r,v,x_a}\to v}(v)$ and $m_{v\to\mathcal{N}(v)}(v)$. As we have assumed a von Mises prior on the angle, we wish to project the message from g_{x_p,r,v,x_a} to v onto a von Mises distribution which requires finding the mean angle and concentration parameter. An approximation of the mean angle can be found by assuming that we stand in the base station position and wish to find the angle between this position and the PP. If so, an approximation of the mean is simply

$$\mathbb{E}_{x_p}[v] \approx \arctan\left(\frac{\mu_{x_{p_2}} - \mu_{x_{a2}}}{\mu_{x_{p_1}} - \mu_{x_{a1}}}\right). \tag{4.16}$$

In order to find the concentration parameter, κ, we will explore estimation techniques. A simple approximation of this parameter is [23]

$$\hat{\kappa} = \frac{R(p - R^2)}{1 - R^2}, \tag{4.17}$$

where p is the dimension (here $p = 2$) and $R = \frac{\sum_i^N x_i}{N}$ with $x_i,\ i = 1, \ldots, N$ being samples drawn from the von Mises distribution. As we do not have the distribution, we can approximate angle samples by sampling from $\boldsymbol{x}_p$, i.e. draw samples from the bivariate normal distribution, and then convert these into angles by using the FP position. Thus the required parameters have been found in order to project the message $m_{g_{x_p,r,v,x_a} \to v}(v)$ onto a von Mises distribution. For the $\boldsymbol{x}_p$ and r nodes, we have utilized that products of Gaussian and Gamma distributions are proportional to a Gaussian and Gamma distribution respectively. However, such results do not exist for the product of two von Mises distributions. We, therefore, need to rely on different methods for updating the v node

$$m_{v \to \mathcal{N}(v)}(v) \propto \bar{p}(v) \cdot m_{g_{x_p,r,v,x_a} \to v}(v).$$

The derivations needed in order to update the $\boldsymbol{x}_p$ node were initiated by employing minimization of the Kullback-Leibler divergence. We will do the same here. Thus we want to solve

$$\underset{q \in VM}{\arg\min}\ KL(q(v)||\tilde{p}(v))$$

where q is an auxiliary density in the family of von Mises distributions and $\tilde{p}(v) = \bar{p}(v) \cdot m_{g_{x_p,r,v,x_a}}(v)$. In order to find the auxiliary density which minimizes the divergence we will use a result from [24]. The result is stated here.

Theorem 4.1 (von Mises Kullback-Leibler Divergence [24])
Consider a von Mises distribution, $p(x; \mu, \kappa)$, with parameters μ and κ and an arbitrary density, $q(x)$, on the unit circle which is nowhere zero. Then

$$[\mu, \kappa] = \arg\min_{[\mu,\kappa]} KL(q(x)||p(x))$$

yields the same result as matching the first trigonometric moment m_1 of $p(x)$

$$m_1 = \int_0^{2\pi} p(x) \cdot \exp(ix)\mathrm{d}x.$$

The authors of [24] also provide a result of how to map the trigonometric moment to von Mises parameters through

$$\mu = \arctan(\mathrm{Im}(m_1), \mathrm{Re}(m_1)), \tag{4.18}$$

$$\kappa = A^{-1}(|m_1|) \tag{4.19}$$

where $A(\kappa) = \frac{I_1(\kappa)}{I_0(\kappa)}$ with I_i the modified Bessel function of order i. Thus to update the v node, we can simply find the first trigonometric moment of $\tilde{p}(v)$ and express a new von Mises density with these parameters. We might encounter issues with this approximation. The Theorem states that a von Mises distribution can be approximated by an arbitrary density on the unit circle. However, we use the theorem somewhat backward, i.e. we want to match a single von Mises distribution to a product of two von Mises densities. We will still apply the result and investigate the applicability in the next chapter. As $\tilde{p}(v)$ is a product of two von Mises densities, the trigonometric moment is intractable to integrate analytically. Therefore, we once more resort to numerical methods which in this case is straightforward as the integration interval is finite. When we have solved the integral, we can use (4.18) and (4.19) to find the updated von Mises distribution of the v node.

Updating θ

We now only need to derive the message from the $g_{\theta,r,rss}$ factor to the θ node. Using the VMP algorithm, we can derive the message

$$m_{g_{\theta,r,rss}\to\theta}(\theta) = \exp\Big[\ln\frac{1}{\sqrt{2\pi}} + \ln\theta - \frac{\theta^2}{2}\Big\{rss_{obs}^2 + B^2 + \left(\frac{10\eta}{\ln 10}\right)^2\left(\psi^{(1)}(\alpha_r) + (\psi(\alpha_r) - \ln\beta_r)^2\right) - 2B\frac{10\eta}{\ln 10}\left(\psi(\alpha_r) - \ln\beta_r\right) - 2rss_{obs}B + 2rss_{obs}\frac{10\eta}{\ln 10}\left(\psi(\alpha_r) - \ln\beta_r\right)\Big\}\Big]$$

where $B = RSS_0 - PL_0 + 10 \cdot \eta \cdot \log_{10}(d_0)$. The full derivation can be found in Appendix B. As we have assumed that the prior of the θ node is a Gamma distribution it is convenient to project $m_{g_{\theta,r,rss} \to \theta}(\theta)$ onto a Gamma density and then once more use the result in (4.11). However, the message above is clearly not a Gamma density. Once again, the integrals needed in order to find the moments

$$\mu^{(1)} = \int_0^\infty \theta \cdot m_{g_{\theta,r,rss} \to \theta}(\theta) \mathrm{d}\theta \tag{4.20}$$

$$\mu^{(2)} = \int_0^\infty \theta^2 \cdot m_{g_{\theta,r,rss} \to \theta}(\theta) \mathrm{d}\theta. \tag{4.21}$$

seem intractable to solve analytically and we will, therefore, once more utilize numerical methods.

4.3 The MMVMP algorithm

With the derived messages above we have found all the needed components to apply the VMP algorithm. As we have modified the original VMP algorithm in favor of moment matching, we will in Algorithm 1 present our version which we will name the *moment matching* VMP (MMVMP) algorithm.

In the initialization loop in Line 3 we assign prior knowledge to all unobserved variables, i.e. $\boldsymbol{x}_p, r, v$ and θ. This includes choosing a prior mean and covariance matrix for $\boldsymbol{x}_p$, shape and inverse scale parameter of r and θ and lastly a mean angle and concentration parameter of v. When the prior knowledge has been initialized, the update messages derived above can be subsequently applied to each variable node in the estimation loop in Line 11.

We will shortly discuss how the algorithm is scaled with the number of PPs, available base stations and message passing iterations. As each loop in Line 12,13 and 14 are all linear in time complexity, the total time complexity of the MMVMP algorithm is $\mathcal{O}(I \cdot P \cdot F)$ where I, P, F are the total number of message passing iterations, PPs and FPs respectively. Therefore, doubling each of these parameters will double the runtime of the algorithm.

In order to test the performance of Algorithm 1, we require a simulation environment in which the relevant devices and information can be generated. Furthermore, we observed that resorting to numerical methods, in order to compute some of the moments, may be needed. We will, therefore, in the next chapter present our implementation choices and discuss the validity of the message approximations.

Algorithm 1 MMVMP

Input: Factor graph with $\boldsymbol{x}_p, \boldsymbol{x}_a, r, v, \phi, rss$ and θ nodes
Output: Updated posterior probabilities of $\boldsymbol{x}_p, r, v$ and θ
Initialization
for each $\boldsymbol{x}_p$ node **do**:
  Initialize $\bar{p}(\boldsymbol{x}_p),\ \bar{p}(\theta)$
  for each observed $\boldsymbol{x}_{ai}$ **do**:
    Initialize $\bar{p}(r_i),\ \bar{p}(v_i)$
  end for
end for

Estimation
for number of message passings **do**
  for each $\boldsymbol{x}_p$ node **do**:
    for each observed $\boldsymbol{x}_{ai}$ **do**:
      Compute moment matched $m_{g_{r_i,v_i,x_p} \to v_i}(v_i)$ using (4.16) and (4.17)
      Compute moment matched $m_{v_i \to \mathcal{N}(v_i)}(v_i)$ using (4.18) and (4.19)
      Compute moment matched $m_{g_{r_i,v_i,x_p} \to r_i}(r_i)$ using (4.9) and (4.10)
      Compute moment matched $m_{g_{r_i,\theta,rss_i} \to r_i}(r_i)$ using (4.14) and (4.15)
      if $\boldsymbol{x}_{ai}$ is first observation **then**:
        Compute moment matched $m_{g_{r_1,\phi} \to r_1}(r_1)$ using (4.12) and (4.13)
      end if
      Compute $m_{r_i \to \mathcal{N}(r_i)}(r_i)$ using (4.6) and (4.11)
      Compute $m_{g_{r_i,v_i,x_p} \to \boldsymbol{x}_p}(\boldsymbol{x}_p)$ using (4.7) and (4.8)
      Compute $m_{\boldsymbol{x}_p \to \mathcal{N}(\boldsymbol{x}_p)}(\boldsymbol{x}_p)$ using (4.4), (4.5) and (4.6),
    end for
    Compute moment matched $m_{g_{r_i,\theta,rss_i} \to \theta}(\theta)$ using (4.20) and (4.21)
    Compute $m_{\theta \to \mathcal{N}(\theta)}(\theta)$ using (4.6) and (4.11)
  end for
end for

5 Implementation of the MMVMP algorithm

In this chapter we will present and discuss the implementation choices which are relevant for the MMVMP algorithm. This includes the assumptions of the devices and the information which is available in the network. Furthermore, we will also describe how we instantiate the simulation environment and how we establish communication between the devices. We will additionally discuss how to initialize the algorithm and the message approximations.

5.1 Simulation platform

In order to employ numerical and non-real-time scientific computing, different simulation environments exist. The choice usually depends on previous experience and possible licensing. Among popular choices are Python and Matlab. The implementation of the MMVMP algorithm and the simulations have been performed in Python. This programming language is an open source, high-level language which offers a broad range of optimized and pre-compiled libraries. Python also offers object oriented and functional programming which have been utilized in the implementation. Although Python is considered to be less efficient for simulations, access to free libraries and a large community combined with previous experience have been the motives for choosing this platform as this reduces the development time.

5.2 Practical considerations

In this thesis, we will conduct both simulations and real-world testing of the localization algorithm and we will briefly mention our assumptions here. As we have assumed that all calculations are done on a centralized unit we will assume that the mails described in Section 2.2.1 can be collected in this unit. Furthermore, we consider the PPs in the environment to be stationary during the measurements in order to mimic man down scenarios or equipment localization. Additionally, we will assume that the connectivity of a PP is given by a fixed radius which is a crude assumption on radio communication but induces great reduction in implementation complexity.

5.2.1 Communication and observations

We consider a communication system utilizing the DECT radio technology as described in Section 2.1. Although communication is necessary for the localization algorithm, it is not relevant how the units communicate and, therefore, other radio technology standards might be used, as long as the relevant information discussed in Section 2.2 can be obtained. In the inference problem, we have included a variable, θ, which seeks to describe fading effects but we have not done so for the phase information. The phase information, however, will also be distorted e.g. by multipath propagation. The phase observation noise is difficult to estimate and will most certainly change rapidly for different environments. We will, therefore, assume a fixed variance. Lastly, different radio frequency equipment may be used for the units in the environment which may change how communication is performed. This is, however, not important for the localization algorithm.

5.2.2 Message scheduling

The localization algorithm relies on being able to receive information from more than one FP. Therefore, there might be complications, if the PP misses the broadcast from other FPs than the direct link. Furthermore, if a PP has detected an FP at $t = t_1$ but does not hear it again, it will downgrade the RSS information. However, it should be considered to exclude the specific FP from the localization algorithm e.g. after some time t_{gone}. Considering the message update scheme in the localization algorithm there may exist message update schedules that are optimal in some scenarios and unsuitable in others. If some of the equipment has more frequent update cycles, the messages containing information from these should be prioritized. As the VMP algorithm allows for an arbitrary update schedule, bottle neck problems might be mitigated, if a certain update scheme is chosen. We will, however, refrain from exploring which update scheme is optimal and simply update the PP position for each updated FP information, i.e. each time we have updated a pair of r and v nodes.

5.3 Overview of the implementation

The implementation can be considered as three independent, but consecutive, blocks which are shown in Figure 5.1.

We will in the following sections discuss the blocks in Figure 5.1. The *Generator* and *Estimator* blocks have several aspects and have some differences when used for

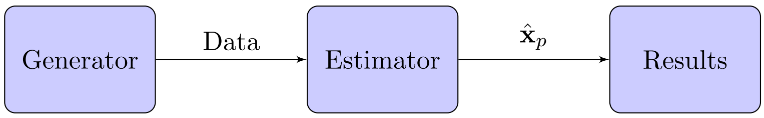


Figure 5.1: Overview of the implementation illustrated as a block diagram. The data includes the observations in the system and $\hat{x}_p$ are the estimates of the PP positions.

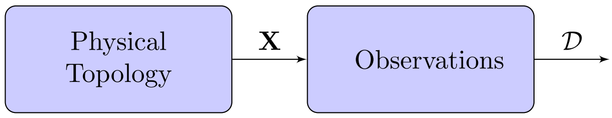


Figure 5.2: The generator module consists of two blocks. The first block initializes the network and the second block generates (or relate in real world testing) RSS and phase observations if a PP observes a base station.

simulations in contrast to real world testing. However, the *Results* block is identical for both scenarios, as this module will present the relevant results.

5.3.1 Generator

The *Generator* block is different when using it for simulations and real world testing respectively. For both scenarios, we can, however, expand the generator block, as seen in Figure 5.2.

Physical Topology

In order to model both PPs and FPs, we have employed the graphical representation induced by a graph. In effect, the physical topology is instantiated as a number of nodes representing either PPs, FPs, angle, distance or a fading descriptor. The position of the FPs can either be chosen to be systematic or random. We will, however, argue, that the results obtained for systematic FPs are a relevant scenario, as installation of base stations is assumed to be done intelligently. Choosing the positions in the pure simulations can be done without any constraints on the environment. However, for real-world testing, instantiating the nodes will be constrained by the possible placements of base stations and PPs in the environment but the physical topology is generally initialized similarly for the two scenarios.

Observations

For simulation purposes we need to be able to simulate the observations of RSS and phase which are used in the localization algorithm. In order to complete the graphical representation of the environment, we need to represent the edges in the graph. If an $\boldsymbol{x}_p$ node has incoming edges, FPs are present within the communication radius. We have already mentioned, that the connectivity of a node is given by a fixed radius. Therefore, when we have instantiated the PPs and FPs in the network, we only generate observations between those pairs of PPs and FPs which fulfil $||PP_i - FP_j|| \leq r_c$ where r_c is a chosen communication radius. As the PPs in the real world network do not have the possibility of communicating with each other, we only generate observations between PPs and FPs. There might be more elaborate ways of limiting the connectivity but the scheme chosen in this work simplifies how and when to establish communication links. If a connection is established between a PP and FP, we need to generate observations of both RSS and phase. Generating an observation of RSS is done by using the RSS model in (3.3). Thus in the simulation environment, the true distance between the PP and each of the FPs within the communication radius is calculated and used in the model to obtain an RSS value. The RSS_0 and PL_0 value have been measured by using a real base station and mobile device. For the pathloss exponent, η, we will not estimate it but use empirical data which show that for an office environment this exponent can be chosen to be 3 [25] for a frequency of 1.9 GHz which is close to the frequency used in DECT. As we are aiming to simulate an environment which is probably cluttered like an office, we will assume that the pathloss exponent is 3. Similarly, we will not estimate the variance of the X_G term. We will again rely on empirical data from [25] where the standard deviation arising from shadowing can be assumed to be $\sigma = 10$ dB. Therefore, we will generate RSS values with $X_G \sim \mathcal{N}(0, 10)$.

It is relevant to discuss to which degree the fading term affects the relation between signal strength and distance in the proposed model. If the fading term does not introduce a significant change to this relationship, the message passing framework (and statistical modeling in general) seems somewhat excessive in order to solve the localization algorithm. If we can achieve relatively precise distance estimates through the RSS model (3.3), we might as well employ trilateration to find the position. To see the fading effects, we have plotted four different realizations of the RSS model for a range of distances and with a standard deviation of $\sigma = 10$ dB in Figure 5.3.

From Figure 5.3 we can see that if we were to neglect the influence of the fading term and use the RSS model to relate observed RSS values to the propagated distance, we will get different results. We can e.g. observe that realization 1 relates an RSS value of approximately -50 dB to a distance of 10 m where realization 3 and 4 relate an RSS value of approximately -30 dB and -25 dB respectively. Finally, realization 2 relates a value of approximately 15 dB.

In order to simulate the phase information we use the model described in Section 4.1

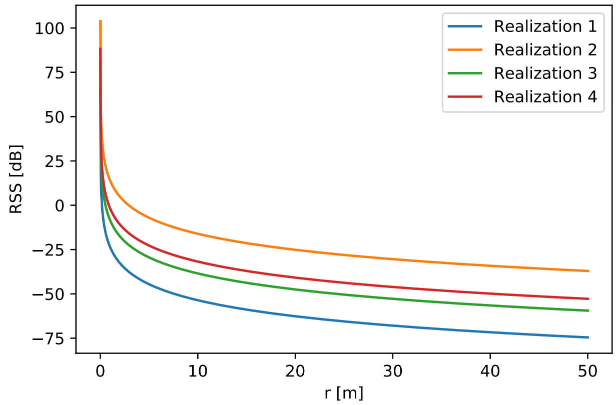


Figure 5.3: Several realizations of the relationship between distance and RSS with $\sigma_{X_G} = 10$ dB.

$$\phi = \frac{\tau}{T_B} + \mathcal{N}(0, \sigma_\phi^2).$$

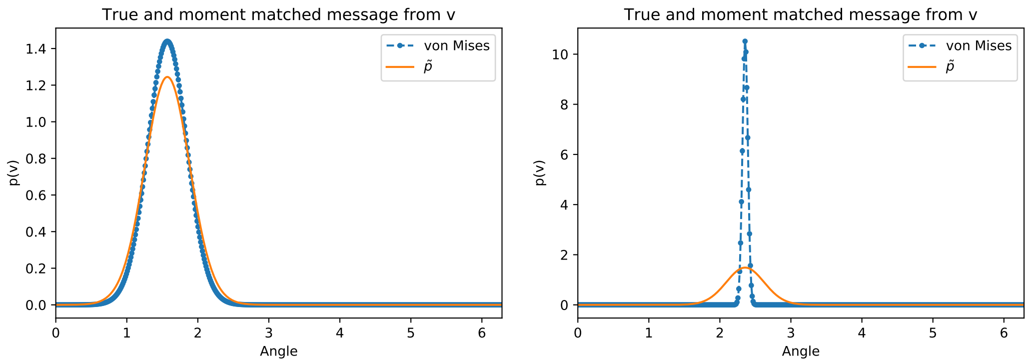
The phase is only generated for the first observed FP, which we assume to be the direct link, in order to simulate real conditions. We use the true distance between the PP and FP and the bit time available in the system to convert the distance into ninths of bit time. As the phase information is observed in discrete bit times where $1,\ 2,\ 3, \ldots$ bits represent approximately $15,\ 30,\ 45, \ldots$ meters we add noise to the true distance through a normal distribution in order to model the rounding errors in the bit times. As the bits represent distances with a resolution of ≈ 15 m we choose a standard deviation of $\sigma = 15$ m of the noise.

5.3.2 Estimator

With an instantiated network, we can use the MMVMP algorithm to estimate the position of a PP given the generated information. In Chapter 4, we have derived the needed messages for the localization algorithm. Some of the messages can be directly computed, e.g. converting the phase information into a Gamma density. However, for some of the messages, computing the needed moments analytically is intractable. Instead, we integrate numerically using the trapezoidal rule, see Appendix C, which is both simple and already implemented in Python. We will later discuss the validity of the message approximations using this rule.

In order to infer information from the network, we instantiate a measurement list for each PP. For each observed FP, we add a list of information to the PP measurement list. The first index of the information list is the FP position and the second index is the generated RSS value. For each observed FP we also generate each of the variables, r and v. These are similarly represented by nodes with α_r and β_r parameters for the r node and μ_v and κ_v for the v node. We add these nodes when an FP is observed in order to associate the right distance and angle to the right pair of PP and FP. As the θ node is general for all established communication links, we will only add a single node with the statistical properties of the fading descriptor to the network. Finally, the generated phase information is also added to the measurement list of the specific PP.

As each of the unobserved variable nodes in the factor graph are connected to a node describing prior knowledge, we need to discuss prior initialization of these. For the prior knowledge of the PP position, $\boldsymbol{x}_p$, we need to initialize a mean and covariance matrix of the bivariate normal distribution. A simple choice of the mean position is a random position in the simulation environment, i.e. the prior mean can be drawn uniformly on the simulation environment or simply chosen to be the center of the room. For the prior covariance matrix, we need the standard deviations in both planar directions to be large enough to cover the entire environment. This assures that we do not infer significant knowledge of the PP position through its prior. Therefore, choosing the variance of each direction to be e.g. $\sigma_x^2 = \sigma_y^2 = 10^6$ m^2, the bivariate density should be wide enough to cover the simulation environment and, therefore, all coordinates are equally likely to be the true PP position. As for the fading descriptor, θ, we need to initialize the α_θ and β_θ parameter. We can choose to set the parameters such that they match the variance of X_G in (3.3), i.e. choosing $\alpha_\theta, \beta_\theta$ such that $\mathbb{E}_{\tilde{p}(\theta)}[\theta]^{-1} = \sigma_{X_G}$. In order to initialize the prior distribution of an r node, we can specify the Gamma parameters of the node such that the mean distance is e.g. 100 m and with a high variance. Lastly, for the v node, we have to initialize a mean angle and concentration parameter κ. Assuming that knowing the positions of the base stations also include knowing the environment around them, we can use the environment constraints to set prior mean angles for the v nodes. Thus if a base station is mounted on an outer wall, and we are only interested in locating employees or equipment inside the building, the prior mean angles can be initialized with this constraint in mind. Say that an FP is mounted in a corner in a square room and relevant signals may only arrive from inside the room. Then the mean angle can be set to 45° if the base station is located in the south west corner. Similar considerations can be made for different placements of the base stations. As for the concentration parameters these have to be chosen such that all angles in line of sight are equally likely, i.e. the prior distribution of the angle node v for a corner base station should have its density concentrated around 45° but with a concentration parameter incorporating angles from 0° to 90°. However, if the information of the environment is not available, the prior mean angles can be drawn uniformly from 0 to 2π and the concentration parameters can be set to 0. Setting the concentration parameter to 0 reduces the von Mises distribution to a uniform distribution [26].



(a) $\mu_1 = \pi$, $\kappa_1 = 0$ and $\mu_2 = \frac{\pi}{2}$, $\kappa_2 = 10$. **(b)** $\mu_1 = \pi$, $\kappa_1 = 10$ and $\mu_2 = \frac{\pi}{2}$, $\kappa_2 = 10$.

Figure 5.4: True (yellow) and approximated (blue) von Mises messages from a v node to its neighbours. The true message is normalized such that the area is one.

5.4 Discussion of implementation choices

As already mentioned, the practical considerations have been made in order to simplify the implementation in favor of proof-of-concept. There might exist ways to represent the environment and the factor graph e.g. through matrices which yields a lower memory demand of the algorithm. Furthermore, we have already discussed, that establishing a fixed communication radius is not realistic, especially not in the form of a circle. We will, however, focus on the validity of the message approximations, as these impact the inference algorithm.

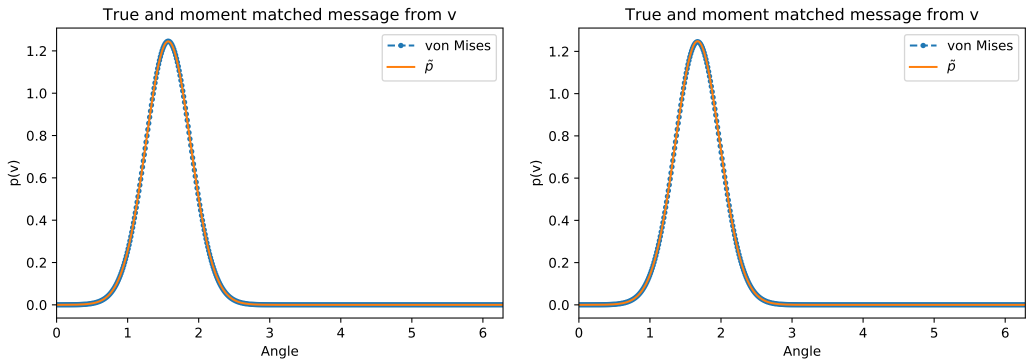
5.4.1 Message approximations

We will in the following explore the precision of approximating the true messages $\tilde{p}$ by the moment matched auxiliary densities q.

Angle approximation

Initially, we will investigate the approximation of the message $\tilde{p}(v)$, which we saw from Section 4.2 is a product of two von Mises distributions. In Figure 5.4a we have plotted the true and moment matched density for a product of two von Mises distributions, $p_1(v) \sim VM(\pi, 0)$ and $p_2(v) \sim VM(\frac{\pi}{2}, 10)$. The approximation has been made according to Theorem 4.1 where we have used the trapezoidal rule to approximate the first trigonometric moment. We choose to normalize the true message such that the area under the graph is one.

In Figure 5.4a we observe that the approximation and true density are agreeing. However, as the concentration parameter of $p_1(v)$ is 0 we only get a contribution from the second von Mises distribution. This is expected, due to the formulation



(a) $\mu_1 = \pi,\ \kappa_1 = 0$ and $\mu_2 = \frac{\pi}{2},\ \kappa_2 = 10$. **(b)** $\mu_1 = \pi,\ \kappa_1 = 1$ and $\mu_2 = \frac{\pi}{2},\ \kappa_2 = 10$.

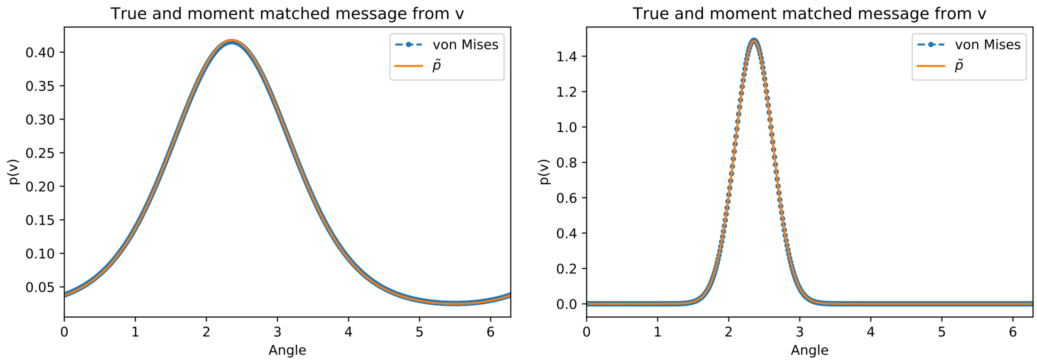
Figure 5.5: True (yellow) and approximated (blue) von Mises messages from a v node to its neighbours. The true message is normalized such that the area is one.

of the von Mises distribution in (4.2). Thus, for $\kappa = 0$, the distribution is not defined and reduces to the uniform distribution. Investigating the behaviour in Figure 5.4b we validate the suspicion from the last chapter where we argued that using Theorem 4.1 backwards might yield insufficient approximations. We do, however, see that the mean of the true density is represented in the approximation but we have to estimate the concentration parameter through alternative methods. As we seek an auxiliary density in the family of von Mises distributions which is closest to the product of two von Mises densities we wish to compute

$$\arg\min_{q \in VM} KL(q||\tilde{p}) = \arg\min \int_v q \cdot \ln \frac{q}{\tilde{p}} \mathrm{d}v. \tag{5.1}$$

As we saw from Figure 5.4 that the mean is approximated well with the first trigonometric moment, (5.1) reduces to a finite integral of one unknown variable, κ. We can, therefore, numerically integrate the KL-divergence for a range of κ values and choose the concentration parameter which yields the lowest integral. This has been done for the remaining figures of this subsection.

We observe in Figure 5.5a, that if the concentration parameter of one of the distributions is zero, we obtain the same result as before. In Figure 5.5b, we have changed the concentration parameters of each of the distributions. We observe, that the mean of both the true and approximated message is now shifted from the mean of p_2 towards the mean of p_1, however, only slightly, as the concentration parameter of p_1 is still low compared to p_2. This is a nice feature, as we are interested in estimating the mean of the approximation closest to the distribution with the highest concentration parameter. As the true and approximated density are perfectly agreeing, it seems appropriate to use moment matching for the mean



(a) $\mu_1 = \pi,\ \kappa_1 = 1$ and $\mu_2 = \frac{\pi}{2},\ \kappa_2 = 1$. **(b)** $\mu_1 = \pi,\ \kappa_1 = 10$ and $\mu_2 = \frac{\pi}{2},\ \kappa_2 = 10$.

Figure 5.6: True (yellow) and approximated (blue) von Mises messages from a v node to its neighbours. The true message is normalized such that the area is one.

angle and minimizing the KL-divergence for the concentration parameter. Finally, we will investigate the density of the von Mises product, if the concentration parameters are equal, see Figure 5.6.

In Figure 5.6a, we observe that the densities once more agree and that the peak is located at the average of the means, i.e. $\frac{\pi + \pi/2}{2} \approx 2.36$. For even higher concentration parameters we see the same behavior in Figure 5.6b where the densities are more concentrated as expected. Therefore, on the basis of the above, we argue that using the first trigonometric moment to approximate the mean of the true angle message and minimizing the KL-divergence to find the concentration parameter are valid. However, there may arise complications with this method. The modified Bessel function mentioned above only exist as a numerical implementation in Python. This implementation will return *nan* values if the input is larger than 709. We will, therefore, investigate the difference in the true densities if we manually set the value of the concentration parameter if it becomes larger than 709. In Figure 5.7 we have plotted the true densities, $p_1(v) \sim VM(\pi, \kappa_1)$ and $p_2(v) \sim VM(\frac{\pi}{2}, \kappa_2)$ where κ_1, κ_2 are chosen to be either 709 and 200.

We can see from the figure, that the densities almost agree if we neglect the magnitude of both densities. Thus we will argue, that manually changing the concentration parameter, if it exceeds 709, will not affect the statistical properties of the true density apart from a smaller magnitude.

Phase approximation

We can conduct similar validations of the phase message approximation. However, we expect that the approximation is relatively valid as we are simply converting a normal density into a Gamma distribution. Furthermore, as the phase information is positive, there should be no problem in using the Gamma distribution. In Fig-

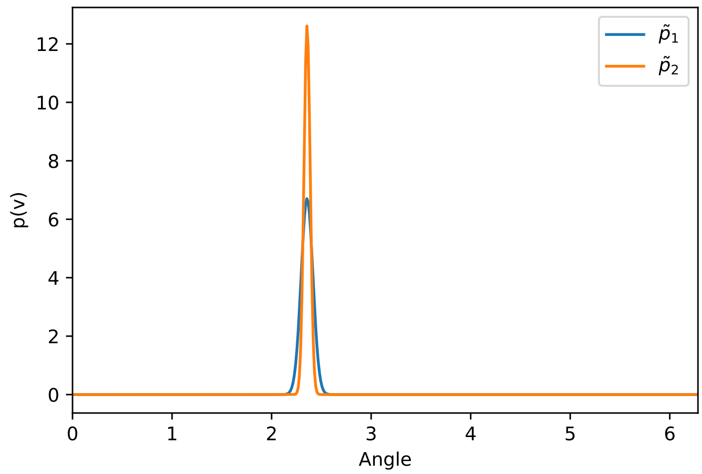


Figure 5.7: The difference in the true densities for $\kappa_1 = \kappa_2 = 709$ (yellow) and for $\kappa_1 = \kappa_2 = 200$ (blue).

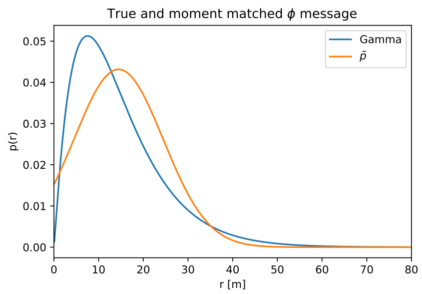
ure 5.8, we have plotted the true and approximated message for several observed phases.

It is relevant to observe, that the approximation in Figure 5.8a is not quite consistent with the normal density. The reason for this is that we are trying to represent a density with support on $\mathbb{R}$ with a density supported on $\mathbb{R}_+$. Therefore, the Gamma density is more concentrated and its mean is shifted. However, for the approximations in Figures 5.8b, 5.8c and 5.8d, we see that the approximations follow the true densities. Furthermore, as the standard deviation has been set to a large value in order to model true distances, the messages are also able to represent phase information ± 1 bit time away from the true value. Thus, from the above observations, we will argue that a moment matched Gamma approximation of the phase message will carry the same information about the distance as the normal distribution.

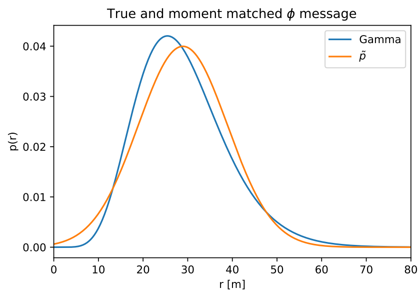
RSS to distance approximation

We also need to verify the validity of approximating the message from the *rss* and θ node to the distance variable, r. We choose to set $\alpha_\theta = 2$ and $\beta_\theta = 20$. Doing so, we assume that the fading descriptor have a mean of $\frac{\alpha_\theta}{\beta_\theta} = 0.1$. As the θ node is representing the precision in (4.1) these model parameters represent a standard deviation of $\sigma = 10$ dB which is the value we use to generate RSS measurements. Initially, we will investigate the moment matching approximation for an RSS value of -40 dBm, see Figure 5.9.

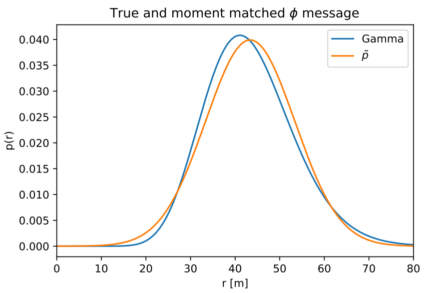
We can see in Figure 5.9 that matching the mean of the true message with a



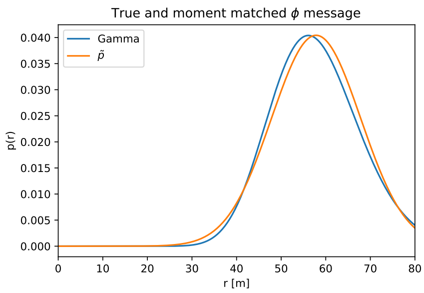
(a) With $\phi_{obs} = 1$.



(b) With $\phi_{obs} = 2$.



(c) With $\phi_{obs} = 3$.



(d) With $\phi_{obs} = 4$.

Figure 5.8: Normal (yellow) and Gamma (blue) messages for several phase observations with $\sigma_\phi = 10$.

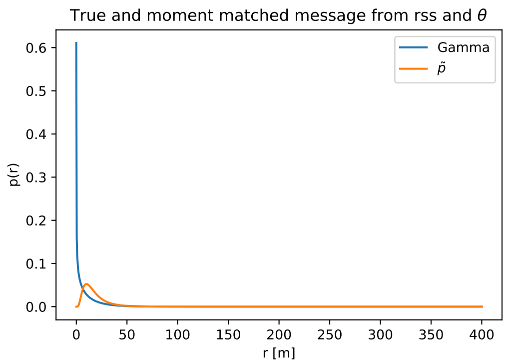
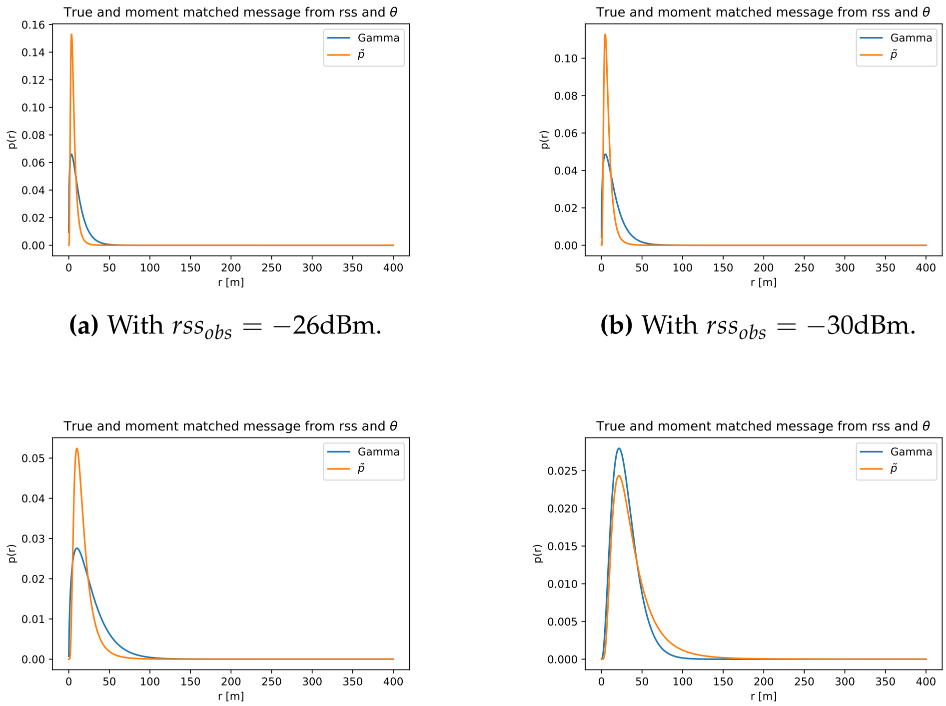


Figure 5.9: True (yellow) and moment matched Gamma (blue) messages for an RSS value of -40 dBm and $\alpha_\theta = 2, \beta_\theta = 20$.



(a) With $rss_{obs} = -26$dBm.

(b) With $rss_{obs} = -30$dBm.

(c) With $rss_{obs} = -40$dBm.

(d) With $rss_{obs} = -50$dBm.

Figure 5.10: True (yellow) and moment matched Gamma (blue) messages for several RSS observations and $\alpha_\theta = 2, \beta_\theta = 20$.

Gamma mean is not a valid approximation. We will instead make sure that the mode of the true message and Gamma distribution is matched. We will assume that the β parameter is well approximated and then find the α parameter through the relationship [22]

$$\text{mode} = \frac{\alpha - 1}{\beta} \Leftrightarrow \alpha = \beta \cdot \text{mode} + 1. \tag{5.2}$$

Thus we can find the r value which yields the mode of the true message and use it in (5.2). The newly found parameters will be used to argue the validity of the message approximations for the rest of this subsection.

We choose to show the true and approximated densities for several RSS values in Figure 5.10 where both densities have been normalized as before.

We see that the approximated messages all capture the same mean of the true density. They do, however, not fully capture the variance. Like the Bessel function mentioned above, the digamma function is also only available as a numerical implementation in Python and we might, therefore, expect some error from this. We will, however, argue, that using this approximation is not unreasonable as it does

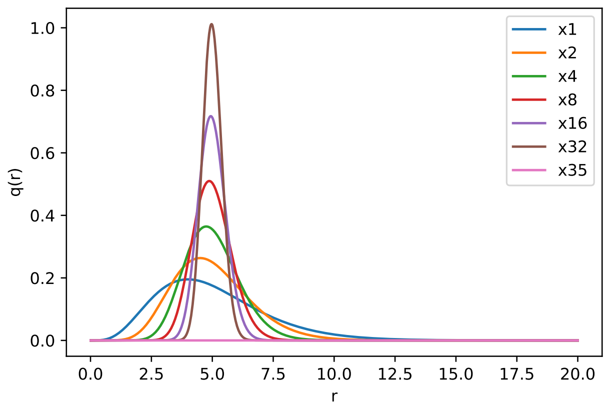


Figure 5.11: The Gamma distribution for $\alpha = 5 \cdot t,\ \beta = 1 \cdot t$ for $t = 1, 2, 4, 8, 16, 32, 35$. For $\alpha = 175,\ \beta = 35$ the distribution degenerates.

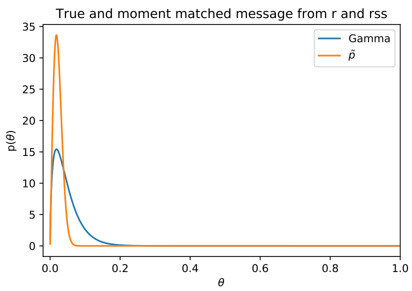
capture the mean of the true density. Additionally, using this message alongside the phase information might provide the needed variability in the distance information and thus these messages will provide sufficient information of the true distance between a PP and FP. As before, using numerical integration is once more allowed, as the true density is concentrated in a finite interval. A final remark shall be given to the Gamma parameters. In the algorithm, we seek to pass information from an r node to the $\boldsymbol{x}_p$ node which hopefully describes the true distance. This information is contained in the update r node parameters. However, as we do not constrain the α and β parameters, the Gamma distribution representing r might degenerate, as seen in Figure 5.11.

From the observation in Figure 5.11 we observe that we might need to manually correct the Gamma parameters if they increase a certain threshold. This can be done by scaling the parameters with the same factor, e.g. with a factor 10. Doing so, the mean will remain the same but the variance of the Gamma distribution will increase.

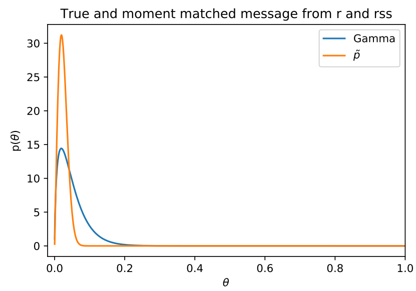
Distance to θ approximation

For the message approximation from an r and *rss* node to the fading descriptor, θ, we will also show the validity of the approximation. We will assume that the r node represents a distance of 10 m, which can be achieved if the Gamma parameters of the node are $\alpha_r = 10$ and $\beta_r = 1$. In Figure 5.12, we have plotted the densities for the same RSS observations as in the previous subsection. Argued by the observation in Figure 5.9 we will also find the Gamma parameters for this approximation through the relationship in (5.2).

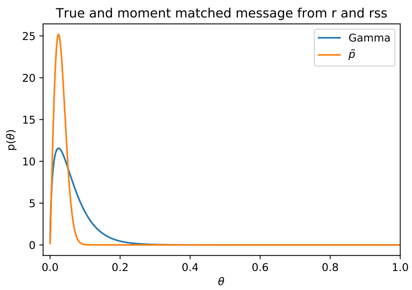
In Figure 5.12 we once more see that the approximation captures the mean



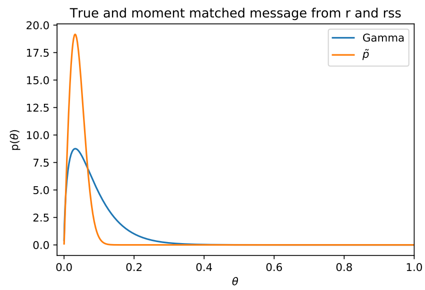
(a) With $rss_{obs} = -26$dBm.



(b) With $rss_{obs} = -30$dBm.

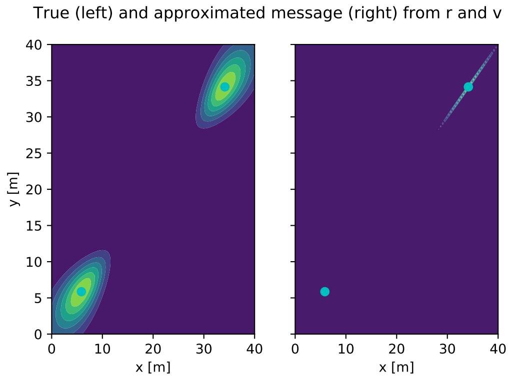


(c) With $rss_{obs} = -40$dBm.

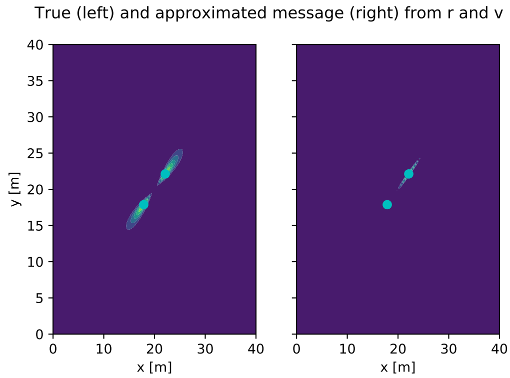


(d) With $rss_{obs} = -50$dBm.

Figure 5.12: True (yellow) and moment matched Gamma (blue) messages for several RSS observations and $\alpha_r = 10, \beta_r = 1$.



(a) With $\alpha_r = 20,\ \beta_r = 1$.



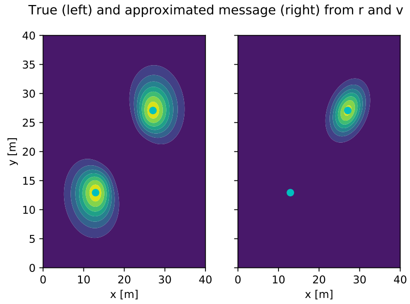
(b) With $\alpha_r = 3,\ \beta_r = 1$.

Figure 5.13: True (left) and approximated (right) message from an r and v node to the $\boldsymbol{x}_p$ node. The message we try to approximate is bimodal and varies more than the approximation which is unimodal.

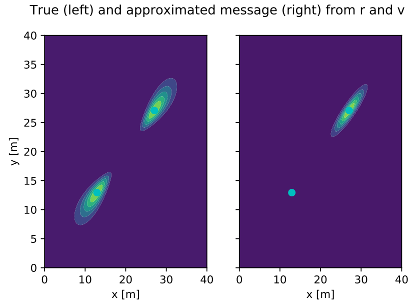
but not the variance. We will, however, assume that the approximation will capture enough of the information in order to update the fading descriptor. The true density is concentrated in a finite interval so numerical integration is still allowed.

Angle and distance to x_p

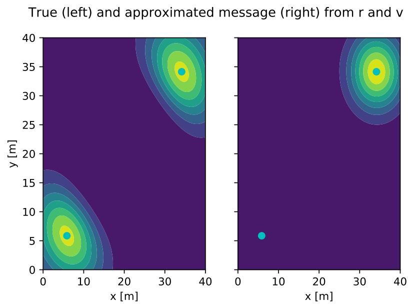
In Chapter 4 we made approximations of the mean and covariance matrix from an r and v node to the PP position, $\boldsymbol{x}_p$. In Figure 5.13 we have plotted the true and approximated density for different Gamma parameters of an r node and a v node parametrized with $\mu_v = \frac{\pi}{4}$, $\kappa = 100$. The base station is positioned in the coordinate $[20, 20]^T$ and we, therefore, suspect the true density to be located in a cloud at an angle of 45° and distance given by the mean of the Gamma distribution. In Figure 5.13 we observe two problems of our approximation. We see that the true density is bimodal which is not represented in the approximation. In [27] the author argues, that faulty distance measurements from a base station may result in bimodal densities and the mode representing the true position may be discarded. Therefore, the author models the device position with a mixture of Gaussian densities and employs the message passing scheme. A variant of this procedure might have corrected the issue in the approximation but the bimodal behaviour has been observed too late in the project period and has, therefore, not been implemented. However, as the two modes in Figure 5.13 seem to be reflections of each other and mirrored in the base station position the problem may be mitigated by setting the prior of the PP to be zero outside the simulation environment. In the next chapter we will test the performance of the algorithm with this observation in mind. How-



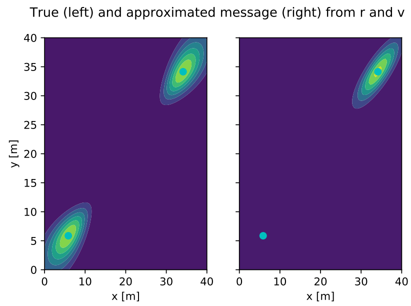
(a) With $\alpha_r = 10, \beta_r = 1$ and $\kappa = 10$.



(b) With $\alpha_r = 10, \beta_r = 1$ and $\kappa = 100$.



(c) With $\alpha_r = 20, \beta_r = 1$ and $\kappa = 10$.



(d) With $\alpha_r = 20, \beta_r = 1$ and $\kappa = 100$.

Figure 5.14: True and approximated message from r and v to x_p for several distances and concentration parameters.

ever, we also observe that the approximation in Figure 5.13 does not capture the angular variance. We observe, that the angular variance of the true message varies with distance which is not present in the approximation. Therefore, we need to make alterations to the covariance matrix from the previous chapter. A choice of modification is to scale σ_v^2 with the mean distance, $\left(\frac{\alpha_r}{\beta_r}\right)^2$, such that the variance becomes more pronounced for longer distances. Therefore, the new covariance matrix approximation reads

$$\hat{\Sigma}_{m_{g_{x_p,r,v,x_a} \to x_p}(x_p)} \approx \begin{bmatrix} \cos \mu_{m_{v \to g_{x_p,r,v,x_a}}(v)} & -\sin \mu_{m_{v \to g_{x_p,r,v,x_a}}(v)} \\ \sin \mu_{m_{v \to g_{x_p,r,v,x_a}}(v)} & \cos \mu_{m_{v \to g_{x_p,r,v,x_a}}(v)} \end{bmatrix} \cdot \begin{bmatrix} \sigma_r^2 & 0 \\ 0 & \sigma_v^2 \cdot \left(\frac{\alpha_r}{\beta_r}\right)^2 \end{bmatrix}.$$

For the remainder of this section, we will use the new covariance matrix to validate the message approximation.

In Figure 5.14 we have plotted the true and approximated message for different settings of distances and an angle of 45° where the observed FP is located in the coordinate $[20, 20]^T$.

From Figure 5.14, we see that with the new covariance matrix the angular variance of the approximation seems more similar to that of the true density. It does, however, still not model the variance completely. We also observe that the approximation correctly uses the mean angle and distance to represent a single mode of a bivariate normal distribution which agrees with the corresponding mode in the true message. Furthermore, we see that increasing the concentration parameter will concentrate the density around the chosen angle. This is a nice feature, as we want to project the density along the correct angle. On the basis of these observations, we will argue that the approximations mentioned in Chapter 4 with the new covariance matrix are valid for this message.

In this Chapter, we have presented and discussed our implementation choices, with focus on the message approximations. For most of the moment matched approximations we see that they have the desired properties and resemble the true densities. We will, therefore, argue that using the approximations for performance testing is a valid choice instead of using the true densities which are more complex to use in the VMP algorithm. In the next chapter, we will investigate several simulation scenarios in order to validate the applicability and performance of the derived algorithm.

6 | Performance assessment

We will in the following chapter test the performance of the MMVMP algorithm with the implementation choices described in the previous chapter.

6.1 Simulation scenarios

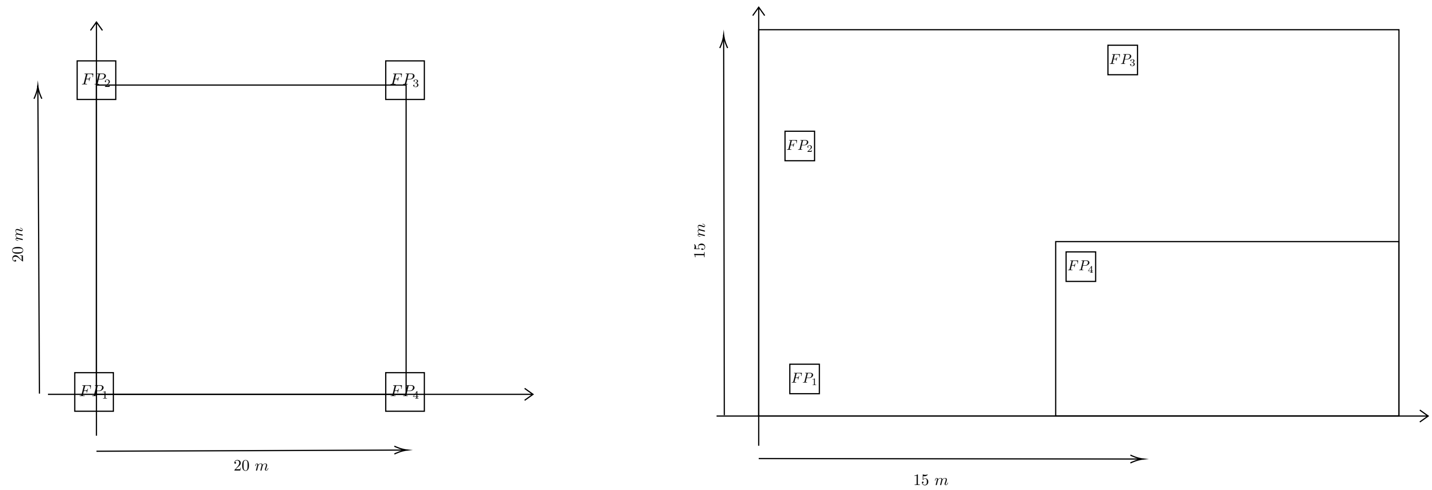
The simulations will aim at representing realistic environments such that we can discuss the applicability of the algorithm. We consider two scenarios:

Scenario 1 **(*S1*)** In this scenario we will represent a square room with four base stations mounted in each of the corners, see Figure 6.1a. The true position of the PP is drawn uniformly in the environment. In order to simulate RSS and phase information we choose a standard deviation in the RSS model and in the phase model to be $\sigma_{X_G} = 10$ dBm and $\sigma_\phi = 15$ m respectively. In Table 6.1 we have listed the specifications of scenario 1 which for the remainder of this chapter will be denoted *S1*.

Scenario 2 **(*S2*)** This scenario will be formulated such that it resembles the environment in which we will conduct a measurement campaign. The base stations are installed as illustrated in Figure 6.1b. We will more thoroughly present the measurement campaign in Chapter 7 but for simulation purposes we only need the positions. We use the same standard deviation of the fading term and phase information as in *S1*. The simulation environment specifications are listed in Table 6.1. Due to the positions of the base station, the dimension of the simulation environment is difficult to specify as a square. We will, however, let the environment be the square with a dimension of 15×15 m^2. This scenario will for the rest of the chapter be denoted *S2*. For both *S1* and *S2* we will randomize which base station is observed first in order to simulate that the direct FP may not be the closest base station.

	Dim [m^2]	FP_1	FP_2	FP_3	FP_4
Scenario 1	20×20	$[0, 0]^T$	$[0, 20]^T$	$[20, 20]^T$	$[20, 0]^T$
Scenario 2	15×15	$[0.8, 1.20]^T$	$[0.8, 10.52]^T$	$[15.31, 13.31]^T$	$[14.28, 6.57]^T$

Table 6.1: Specifications of *S1* and *S2* including environment dimensions and FP positions. The standard deviation of the fading and phase noise is $\sigma_{X_G} = 10$ dBm and $\sigma_\phi = 15$ m respectively.



(a) *S1* environment.

(b) *S2* environment.

Figure 6.1: Illustrations of the simulation environments, *S1* and *S2*.

MC runs	# FPs.	Mess. Itt.	Runtime [s]
1	4	1	0.46
1	4	5	2.25
2	4	1	0.85
2	4	5	4.43
1	8	1	0.88
1	8	5	4.36
2	8	1	1.81
2	8	5	8.82

Table 6.2: Runtimes for several setups. The simulations have been run on a laptop with the following specs: Intel Core i7-7820HQ CPu @ 2.90 GHz with 16 GB RAM 64-bit running Windows 10 Enterprise.

6.1.1 Run time

As we have derived the message approximations using moment matching, we do not need to rely on optimizing the Kullback-Leibler divergence through an optimization algorithm. Therefore, we suspect that the runtime of a single Monte Carlo simulation with one message iteration will not change significantly for each run. We will compute the runtime of a few setups in order to verify the time complexity in Chapter 4. Furthermore, computing the runtime in a few scenarios enables the possibility of predicting runtimes of more complex simulations. The runtimes are presented in Table 6.2.

From Table 6.2 we see that the time complexity indeed behaves linearly, as the runtime doubles when we e.g. double the amount of PPs (running 2 Monte Carlo simulations is equivalent to locating two PPs).

6.2 Simulations, *S1*

As the MMVMP algorithm is iterative, we rely on messages being passed between the nodes in the factor graph such that the marginal posterior probabilities of the unobserved variables can be updated. We will, therefore, investigate how many iterations are needed in order to obtain convergence in the error distance. Thus we seek to estimate the sample mean of the distance error

$$\mathbb{E}\left[||\hat{\boldsymbol{x}}_p - \boldsymbol{x}_p||\right].$$

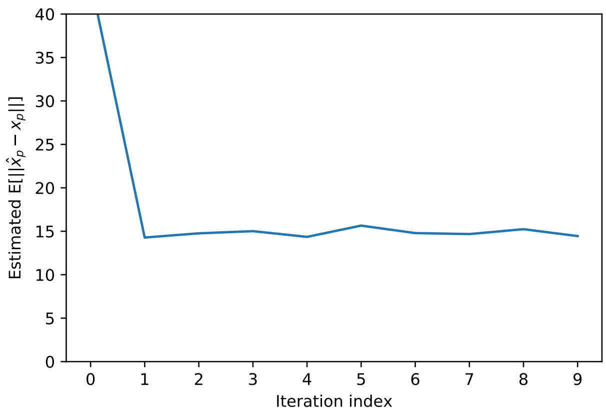
We will do so on the basis of 100 Monte Carlo simulations and use the average distance error as the empirical mean error distance. Initially, we consider a maximum of 9 message iterations as we from Table 6.2 see that the predicted runtime of 100 Monte Carlo simulations for just 5 iterations will be 225 s. We, therefore, hope that the mean error converges before reaching 9 message iterations. In this case, we chose a prior of the PP position as a bivariate normal distribution with a mean drawn uniformly on a scaled version of the simulation environment and with a large covariance matrix, i.e.

$$\bar{\mu}_{x_p} \sim \mathcal{U}\left([-5 \cdot \text{dim},\ 5 \cdot \text{dim}] \times [-5 \cdot \text{dim},\ 5 \cdot \text{dim}]\right)$$
$$\bar{\Sigma}_{x_p} = \begin{bmatrix} 10^6 & 0 \\ 0 & 10^6 \end{bmatrix}.$$

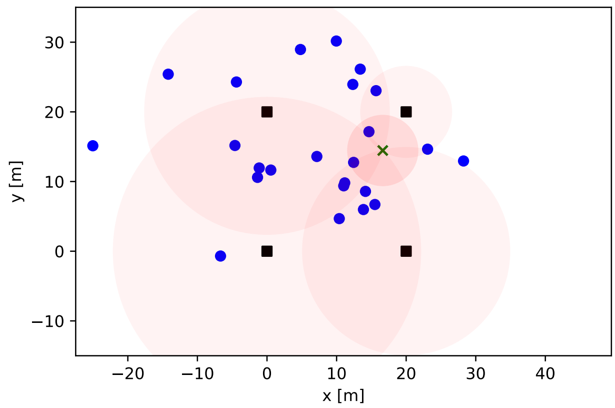
Thus we do not provide the algorithm with any significant prior knowledge of the PP position. The angles between the PP and observed base stations are drawn uniformly between 0 and 2π. The mean error distance results for *S1* can be seen in Figure 6.2.

In Figure 6.2a we see that the mean error distance converges after only a few iterations. We also see that the mean distance error is highly imprecise with a mean of $\hat{\mathbb{E}}[||\hat{\boldsymbol{x}}_p - \boldsymbol{x}_p||] \approx 15$ m which is insufficient for localization purposes. This behaviour becomes clear in Figure 6.2b where we have used four message iterations to obtain the position estimates. Each blue dot is a new estimate from a different PP prior and the black squares represent observed base stations. The red circles with center in the base stations represent the true distance between each FP and the PP where the circle around the PP is a distance of 5 m. As the algorithm does not receive any significant prior knowledge of the angle between the base stations and the PP, the angle is solely updated on the basis of the prior position of the PP. The prior position may be drawn in a completely different direction than the true position which is why we see the behaviour in Figure 6.2b. In order to better visualize the probability of observing errors smaller than some value ξ, we also present the Empirical Cumulative Distribution Function (ECDF) of the estimates. In Figure 6.2c we have visualized the ECDF of *S1* where we see that it is improbable to observe small error distances. We will, therefore, investigate to which degree the precision of the algorithm is improved if we apply some knowledge of the angles.

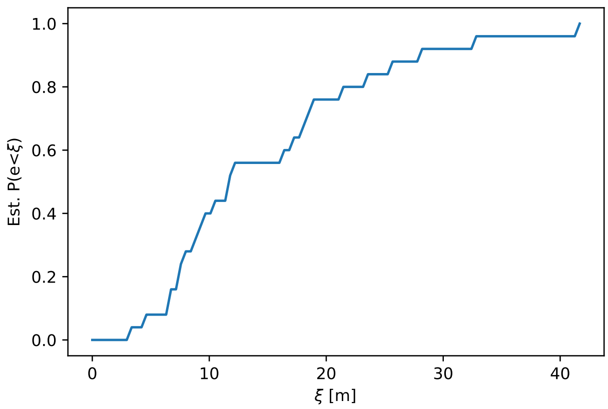
If we imagine that the room represents e.g. a restaurant and that we are only interested in locating objects within the same room, we can provide the algorithm with prior knowledge of the angles by simulating the base stations as sector antennas. This can be done by assuming that the prior knowledge of a v node in the factor graph can be initialized based on the position of the respective base station as discussed in Section 5.3.2. For the positions of the base stations in *S1*, the mean angles are chosen to be $45°, -45°, 225°$ and $135°$ for base station FP_1, FP_2, FP_3 and FP_4 respectively. However, we wish to represent all possible angles from within



(a) Iterationplot.



(b) 25 estimates (blue) of the same true position (green cross).



(c) ECDF of the estimates.

Figure 6.2: Performance of the algorithm for *S1* without any prior knowledge. Due to no angle information the algorithm provides imprecise estimates.

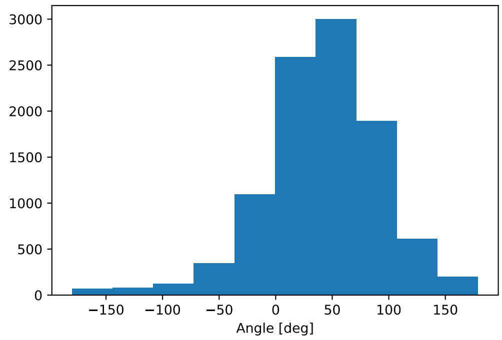


Figure 6.3: Probability distribution of the von Mises density with $\mu_v = \frac{\pi}{4}$ and $\kappa = 2$.

	FP_1	FP_2	FP_3	FP_4
μ_v	$\frac{\pi}{4}$	$-\frac{\pi}{4}$	$-\frac{3\pi}{4}$	$\frac{3\pi}{4}$
κ	2	2	2	2

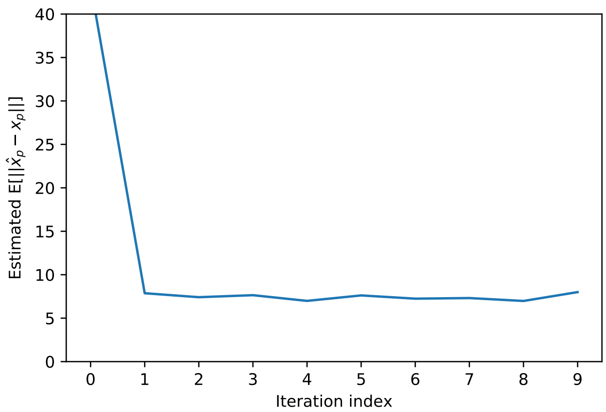
Table 6.3: Prior angle knowledge of the base stations in *S1*.

the room which can be achieved by choosing the concentration parameter of the von Mises prior densities accordingly. In Figure 6.3 we have visualized a histogram showing a von Mises distribution, $p(v) \sim VM(\frac{\pi}{4}, 2)$.

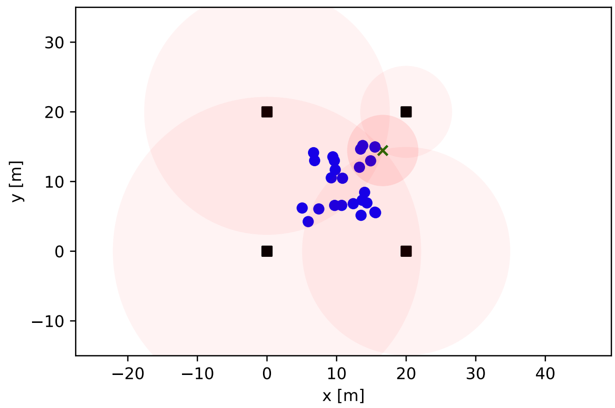
Choosing the concentration parameter $\kappa = 2$ will incorporate the desired variance in the distribution, i.e. include all angles from within the respective quadrant. The prior angle knowledge has been summarized in Table 6.3.

In Figure 6.4a we have provided the FPs with their respective prior knowledge of the possible impinging signals and computed the mean error distance.

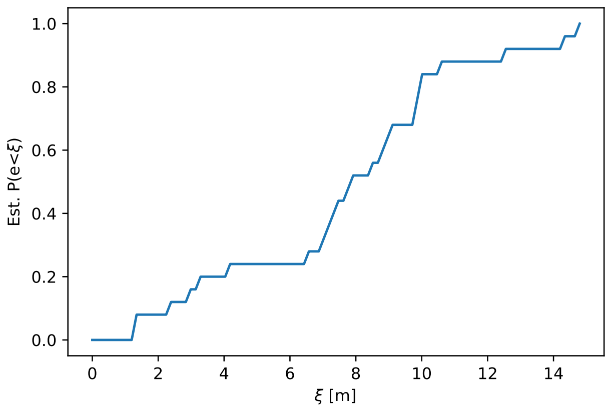
We observe in Figure 6.4a that the mean error distance still converges after a few iterations but the mean value has been almost halved. In Figure 6.4b we have once more simulated 25 estimates for the same true position as in Figure 6.2b. We use the same position in order to see the improvement in the precision. The amount of message passings has once more been set to four. We see that providing the algorithm with sector antennas enables the algorithm to estimate inside the simulation environment. We do, however, see, that some of the positions are estimated towards other base stations than the closest FP which we will investigate later. The ECDF in Figure 6.4c also shows improvement as we are now more likely to observe errors below 10 m. There is, however, still room for improvement, if we seek to estimate the true position of e.g. a table in a restaurant. In Figure 6.5 we show the approximations for each iteration in the algorithm for a single Monte Carlo simulation and a single branch in the factor graph.



(a) Iterationplot.



(b) 25 estimates (blue) of the same true position (green cross).



(c) ECDF.

Figure 6.4: Performance of the algorithm for *S1* assuming base stations to be sector antennas. Using prior knowledge of the angles has a high impact on the precision.

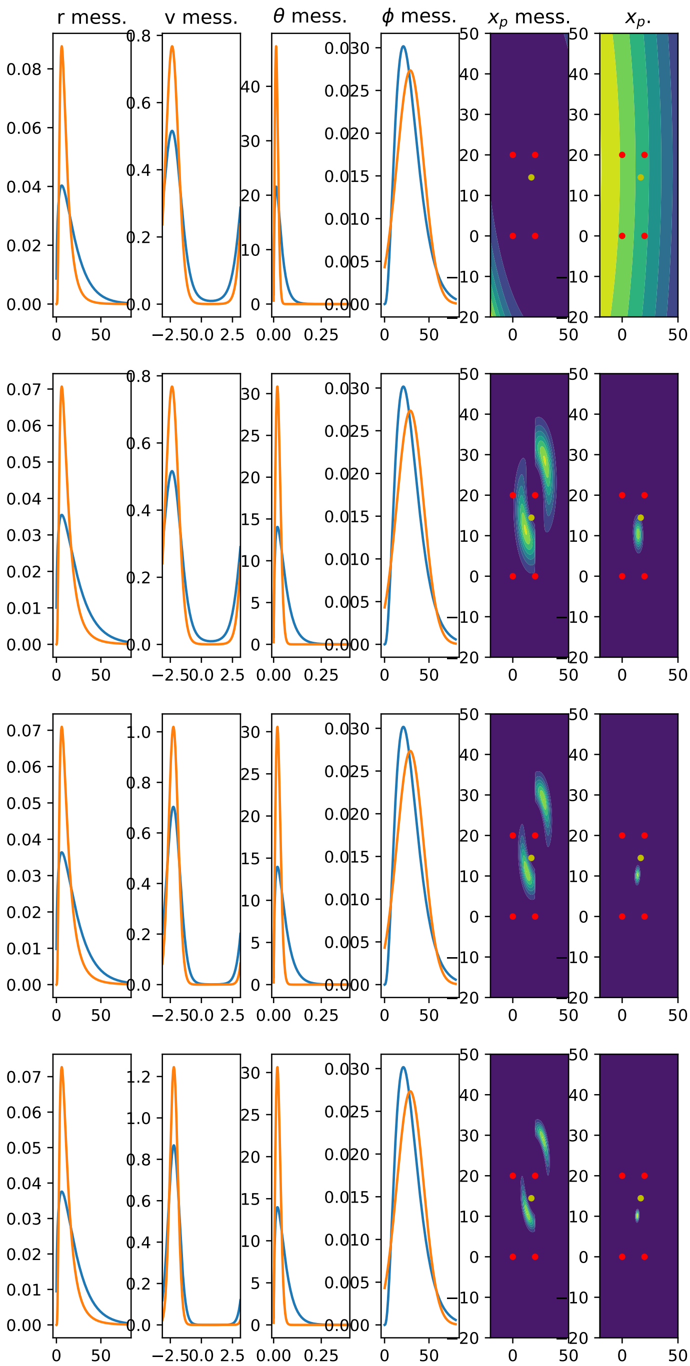


Figure 6.5: Message approximations in *S1* for each iteration when prior knowledge of the angle is included. In the right most plots, the red and yellow dots represent base stations and the true PP position respectively.

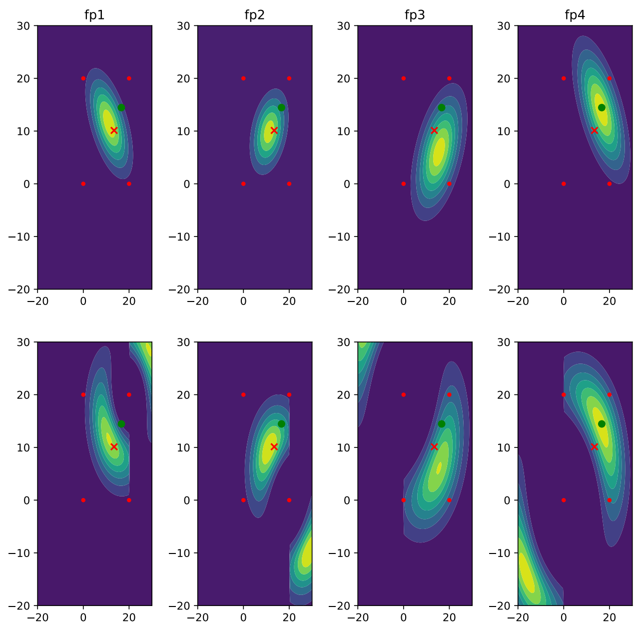


Figure 6.6: Approximated (top) and true (bottom) messages from each branch in the factor graph to the $\boldsymbol{x}_p$ node after a single message iteration. We see that the true mode of the bimodal message is chosen when applying sector antennas. The green dot and red cross represent the true and estimated position respectively while the red dots represent the base stations.

We see that the approximations agree with the true messages. In the fifth and sixth column we represent the true message from an r and v node to the $\boldsymbol{x}_p$ node and the PP position update respectively. Even though the true message is bimodal we see that the mean of the PP position is shifted towards the true position while the covariance matrix is reduced. In Figure 6.6 we have visualized the true and approximated message to the $\boldsymbol{x}_p$ node from all observed base stations after a single message iteration.

From Figure 6.6 we see that the approximation chooses the right mode when the base stations are assumed sector antennas. However, from the above we have seen that the covariance of the PP position is reduced quickly and, therefore, also converges fast to the final estimate which may be wrong. In Figure 6.6 we see that a wrong estimate may be computed if inferior RSS and angle information is used. In the present case, in Figure 6.6, it might have proven advantageous to update the PP position with the information from FP_4 first, as the mode of this message is located at the right position.

	FP_1	FP_2	FP_3	FP_4
μ_v	$\frac{\pi}{4}$	0	$-\frac{3\pi}{4}$	π
κ	2	1	2	1

Table 6.4: Prior angle knowledge of the base stations in *S2*.

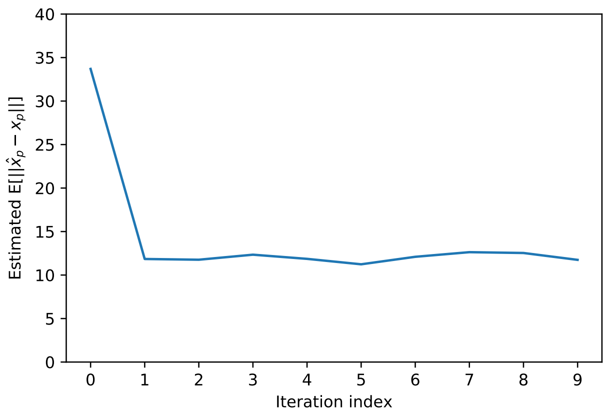
6.3 Simulations, *S2*

As already mentioned, the simulation environment in *S2* has a disadvantage compared to *S1*. In *S1* we considered a square room in which the base stations are mounted in the corners such that any true position drawn in the environment will be located inside the square created by the base stations. This is, however, not the case for *S2* and we might, therefore, expect true positions to be drawn outside the square created by the FPs which in turn might yield erroneous estimates. For the simulations, we will draw true positions uniformly on the dimension specified in Table 6.1. Initially, we will investigate if the absence of angle information also provide a large mean error distance in this scenario, see Figure 6.7.

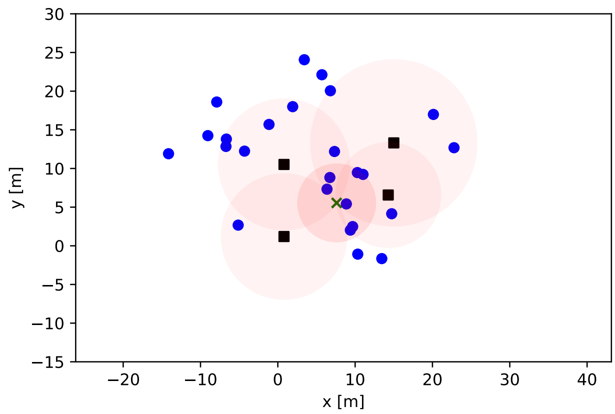
For this scenario we observe a similar behaviour as in *S1*. The error distance quickly converges but the mean is still high as seen in Figure 6.7a. This is once more caused by the angle being updated based on the PP prior which may be chosen completely wrong. Visualizing the estimates returned by the algorithm in Figure 6.7b, we once more see that providing no angle information to the algorithm fails to produce reliable estimates. This is also apparent from the ECDF in Figure 6.7c where any useful errors are highly improbable. Therefore, we will once more assume prior knowledge of the angle information by treating the base stations as sector antennas. Similar considerations about the mean angles and concentration parameters from *S1* have been done and are listed in Table 6.4. The mean angles are, however, somewhat difficult to initialize in this scenario as the base stations are not installed in the corners.

The results where the angle prior knowledge has been incorporated are shown in Figure 6.8.

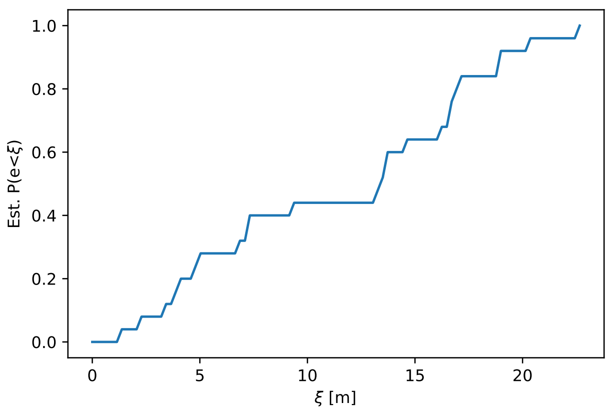
We once more observe a fast convergence in the mean error distance. Using sector antennas has a much higher impact on the mean error distance for this scenario. Providing the algorithm with prior knowledge of the angles more than halves the mean error. The environment investigated here is of course relatively smaller than *S1* and, therefore, we should expect smaller errors. From the ECDF in Figure 6.8c we can see that for this specific position, observing errors less than 5 m is highly probable compared to *S1* which might be induced by the fact that the base stations in *S2* are located closer to each other. We will later explore the effects of adding additional base stations to the simulation environment which



(a) Iterationplot.

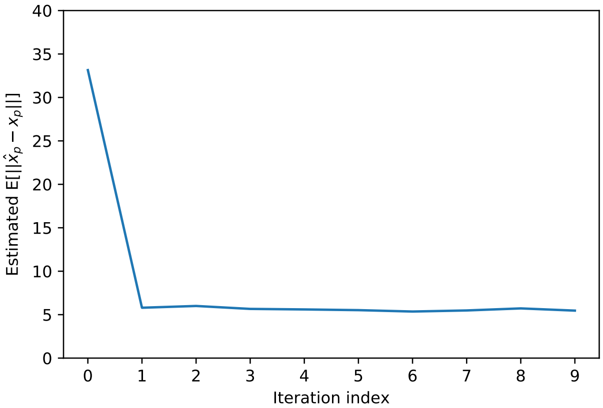


(b) 25 estimates (blue) of the same true position (green cross).

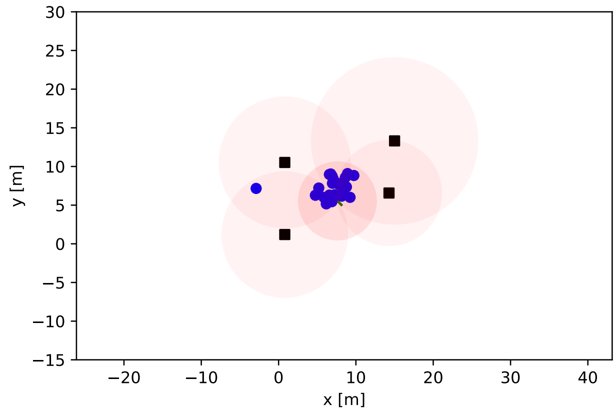


(c) ECDF.

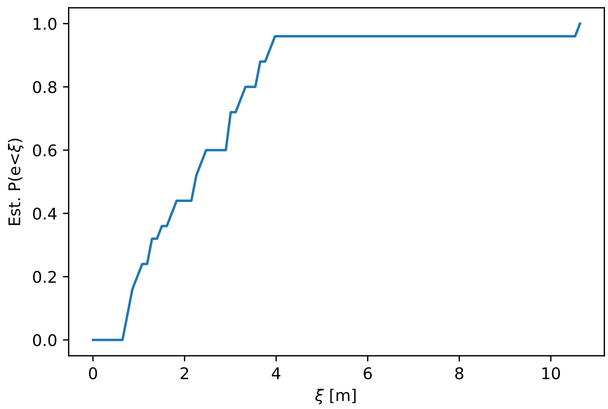
Figure 6.7: Performance of the algorithm for *S2* without any prior knowledge. Due to no angle information the algorithm provides imprecise estimates.



(a) Iterationplot.



(b) 25 estimates (blue) of the same true position (green cross).



(c) ECDF.

Figure 6.8: Performance of the algorithm for *S2* assuming base stations to be sector antennas. Using prior knowledge of the base stations has a high impact on the precision.

may improve the precision. In Figure 6.9 we show the message approximations for the number of message passing iterations in this scenario. We still observe that the message approximations fit the true densities. The position of the PP is also updated towards the true position and the covariance is reduced. Similarly to *S1* we will investigate the initial updates of the messages from the factor graph branches to the $\boldsymbol{x}_p$ node, see Figure 6.10.

We once more see that the bimodal problem is mitigated by assuming prior knowledge of the angles. However, for this scenario, it had seemed advantageous to update the PP position with respect to FP_3 first as the mode of the approximation is located in the true position.

We have now made initial performance tests of the algorithm and from the above it clearly shows that providing the algorithm with some directional information improves the precision. Furthermore, the bimodal problem observed in Chapter 5 is also mitigated by applying angle prior knowledge. However, it seems that the algorithm favors the information from the first observed base station and as the algorithm quickly converges other base stations providing more reliable information may be disregarded. The improvement observed in *S2* might also be an indication that receiving information from base stations closer to the PP will provide an increased precision. It might, therefore, increase the performance if more FPs are added to the environment. Both of these suspicions will be explored and tested in the following.

6.4 Modifications

Although using sector antennas increases the algorithm's performance, there is still room for improvement. Here we investigate how sorting the observations and adding base stations can improve the algorithm.

6.4.1 Sorting the observations

Although we have restricted ourselves from deriving an ideal update scheme for the MMVMP algorithm, we will make a small exception. As we are updating the PP position for each updated FP branch, i.e. for each time we have updated a v and r node in the factor graph, we might suspect that using RSS information from an inferior base station may yield imperfect estimates. In this context, we classify a base station as inferior if it provides a relatively attenuated RSS measurement. Highly attenuated RSS measurements might be an indication that the base station is either relatively far away from the PP or behind an object creating shadowing effects. Thus, due to the update scheme of the PP, we may be interested in updating the position based on the FPs which provide the least attenuated measurements. If we do so, the PP position is updated sequentially according to the base station

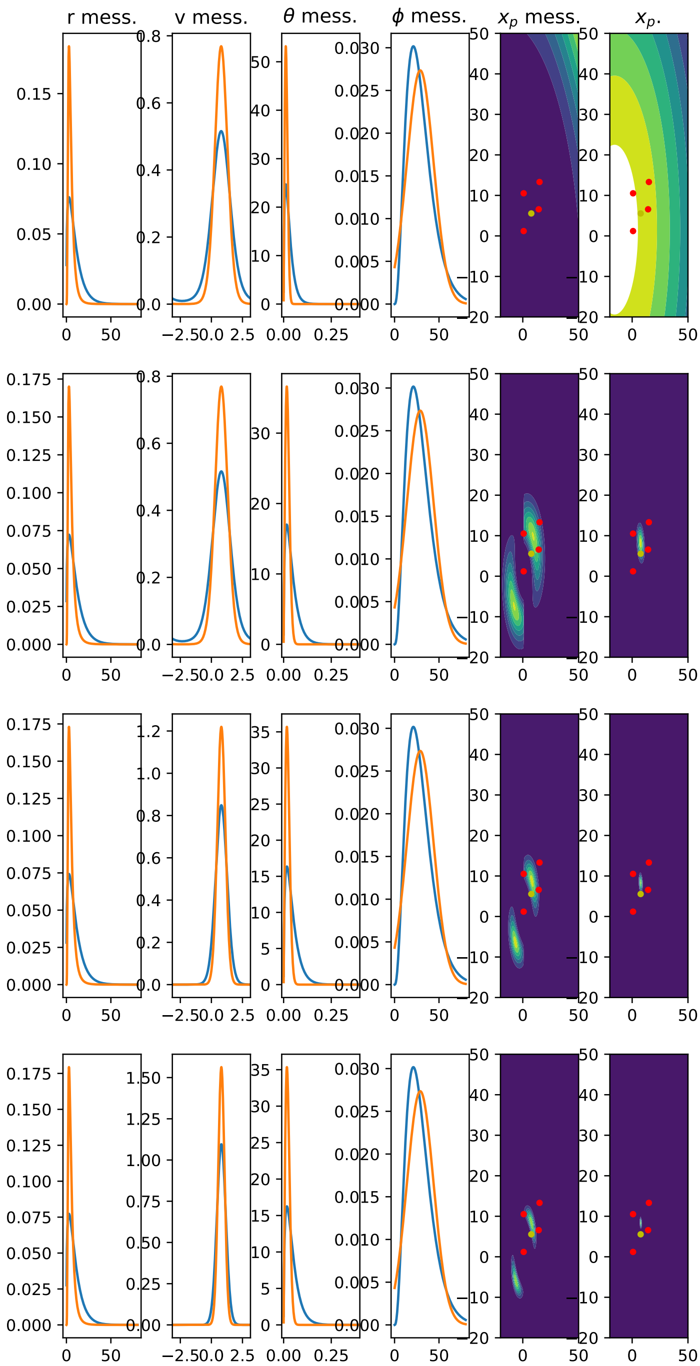


Figure 6.9: Message approximations for each iteration when prior knowledge of the angle is included. We see that the approximations agree with the true densities.

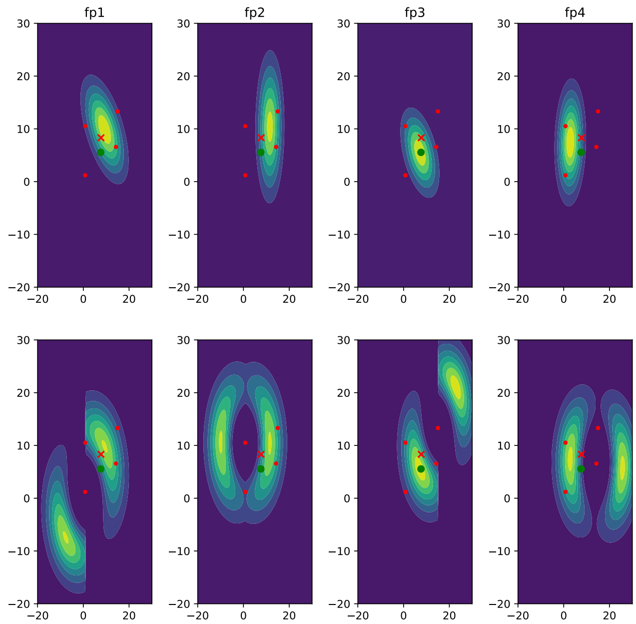
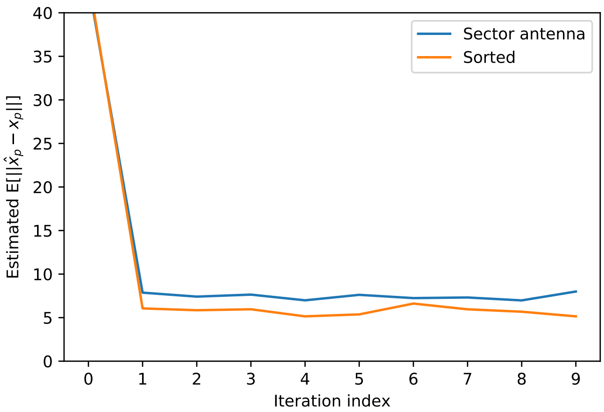
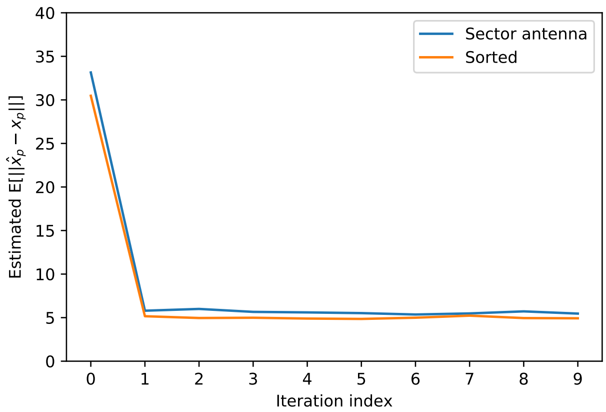


Figure 6.10: Approximated (top) and true (bottom) messages from each branch in the factor graph to the x_p node after a single message iteration. We see that the true mode of the bimodal message is chosen when applying sector antennas. The green dot and red cross represent the true and estimated position respectively while the red dots represent the base stations.



(a) Iterationplot for *S1*



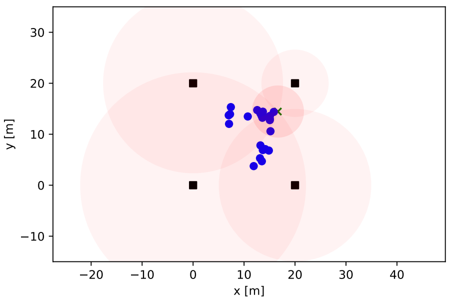
(b) Iterationplot for *S2*

Figure 6.11: Iterationplots for both investigated scenarios where the base stations have been sorted according to their RSS measurement. We see that sorting the RSS information yields an improvement.

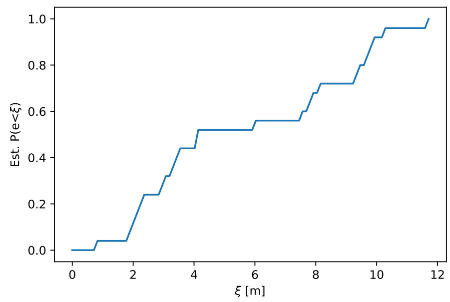
which is, ideally, closest. In Figure 6.11 we have employed this consideration and made similar investigations of the needed iterations for error distance convergence as in the previous sections.

We see in Figure 6.11 that sorting the RSS information has different impact on the precision in the two scenarios. In Figure 6.11a we see a clear improvement in the precision. However, for *S2* the precision is only improved slightly. As the base stations are all relatively close in this scenario, sorting the RSS information may not provide significant changes in the PP update. Although the improvement for *S2* may not be significant, the improvement for *S1* suggests that sorting the RSS is relevant for the algorithm. In order to visualize the improvement we once more simulate 25 Monte Carlo simulations of the same true positions investigated in the above. The results can be seen in Figure 6.12 along with their respective ECDF.

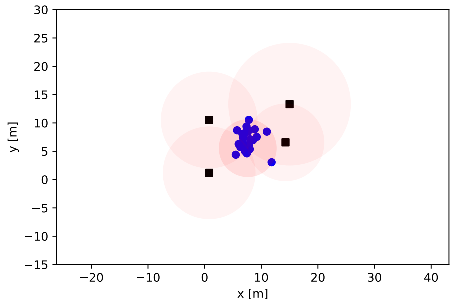
We see in Figure 6.12a that, compared to the estimation in Figure 6.4b, the algorithm now provides estimates which are much closer to the true position. However, we still obtain estimates which are not consistent with the PP. Those estimates, which do not fall into the cluster around the true position are in fact those where the first RSS measurement used in the algorithm do not belong to the closest base station, i.e. the one in the north east corner. Therefore, the algorithm updates the mean of the bivariate normal distribution towards false positions which, in the eyes of the algorithm, fits better with the RSS measurements. Meanwhile, we still see a clear improvement which is validated by the ECDF in Figure 6.12b. Turning our attention to *S2* we see that the estimates are now more clustered around the true position. The ECDF also shows promising results when sorting the RSS measurements, as observing errors larger than 5 m is now highly unlikely where



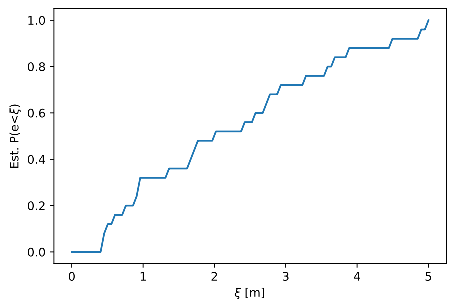
(a) 25 estimates (blue) in *S1* of the same true position (green cross).



(b) ECDF of the estimates in *S1*

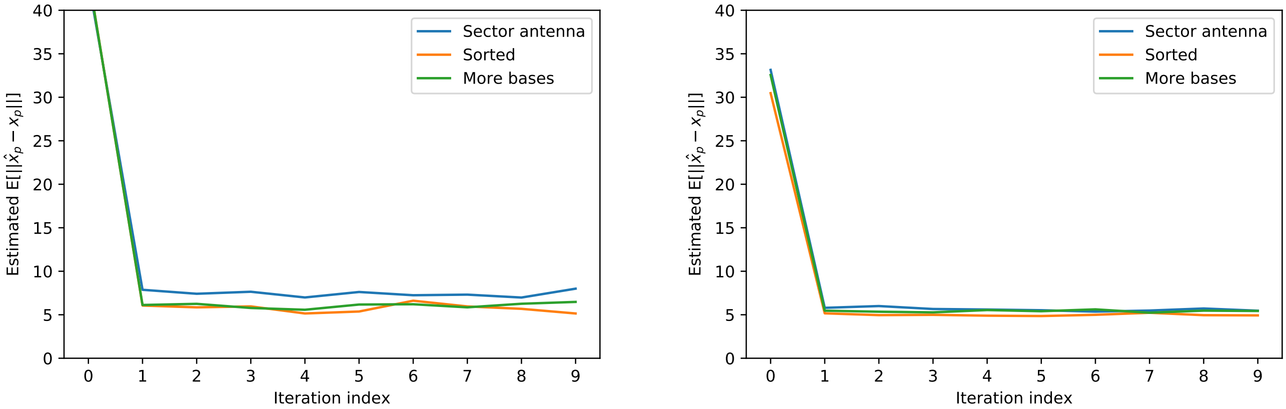


(c) 25 estimates (blue) in *S2* of the same true position (green cross).



(d) ECDF of the estimates in *S2*

Figure 6.12: Estimates and their respective ECDF for the same true positions investigated above where the RSS measurements have been sorted.



(a) Iterationplot for *S1*

(b) Iterationplot for *S2*

Figure 6.13: Iterationplots for both investigated scenarios where we add additional base stations. We see that the precision seems almost unaffected.

there were some probability of observing larger errors if we do not sort the RSS as seen in Figure 6.8c. Therefore, we will argue, that sorting the measurements with respect to attenuation will improve the precision of the algorithm.

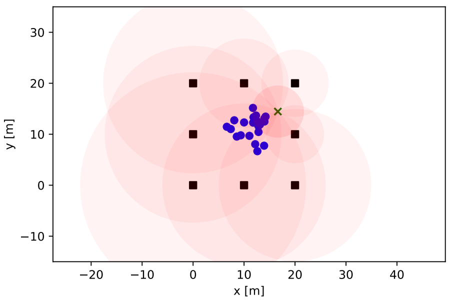
Based on the above it seems that the algorithm favors estimates closer to the base station which yields the least attenuated RSS measurement and as the covariance matrix of the x_p node is quickly reduced, information from other base stations do not contribute.

6.4.2 Additional base stations

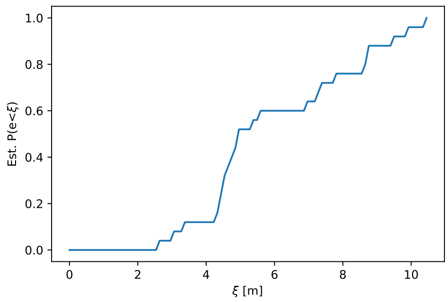
We will in this section explore the effects of adding additional base stations to the environment while still sorting the RSS measurements as this has proven effective for the algorithm. Once more, we have simulated the mean error convergence, see Figure 6.13.

In both scenarios we see that adding additional base stations to the environment does not improve the precision. This is, however, not unexpected as the estimate of the PP position still converges fast and, therefore, the extra information provided by the additional FPs do not contribute. We once more show the estimates and ECDF of each scenario when we add more base stations, see Figure 6.14.

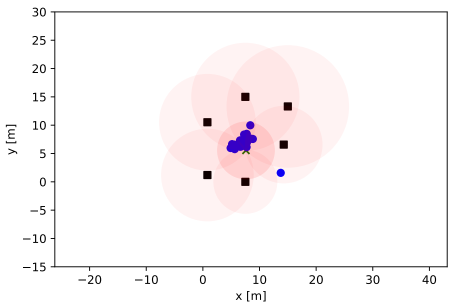
In Figure 6.14a we see that the estimates are not agreeing with the true position of the PP. The reason for this might once more be caused by the algorithm favouring updates from base stations with less attenuated RSS. Therefore, in this environment, we have increased this source of error by adding more stations which might provide better measurements than the close FP in the north eastern corner.



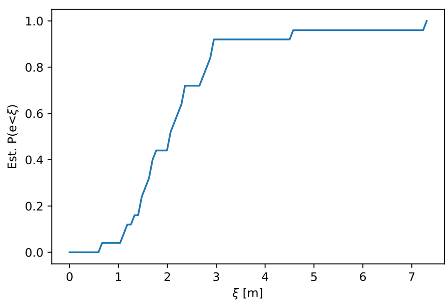
(a) 25 estimates (blue) in *S1* of the same true position (green cross).



(b) ECDF of the estimates in *S1*



(c) 25 estimates (blue) in *S2* of the same true position (green cross).



(d) ECDF of the estimates in *S2*

Figure 6.14: Estimates and their respective ECDF for the same true positions investigated above when adding additional base stations.

This is also evident in the ECDF. However, in *S2*, the estimates and ECDF show a small improvement. The difference between *S1* and *S2* is that in the latter environment the base stations are closer and more encircling of the true position.

With the above performance tests we have explored the precision of the algorithm. We have seen that providing information about the angles has a high impact on the results and the bimodal problem from Chapter 5 is mitigated. Furthermore, as the algorithm quickly converges it may be needed to sort the base stations according to their respective signal strength such that the PP position is updated first according to the least attenuated base station. It may, however, also be the fact that the fading effects are not correctly incorporated and described in the inference problem. Furthermore, it might have proven advantageous to use more observations from each base station, i.e. use RSS information at different time steps. Thus, fluctuations in the signal strength due to fading may have been diminished. This has, however, not been investigated in the present work.

In the next chapter we will present and discuss the measurement campaign which has been conducted in order to test the performance of the algorithm on real data.

7 Measurement campaign and testing

In order to test the performance of the algorithm when applied to real data, we have conducted a measurement campaign. This campaign sought to acquire RSS and phase measurements between a real PP and FPs which can be used in the MMVMP algorithm. In this chapter we will discuss the measurement setup and data acquisition and use the data for testing. We will discuss the performance through the same tests as in the previous chapter such that we can better relate the contrast between pure simulations and field testing.

7.1 Data acquisition

In this section we will describe the setup and how the data was acquired. The environment in which the measurement campaign was done is a warehouse/production facility at the RTX A/S headquarters. A picture of the room can be seen in Figure 7.1.

This environment offers a smaller storage room, marked with yellow in Figure 7.1, which is partitioned from the larger room by a thin plaster wall. Mounting a base station in this room will enable the possibility of receiving NLOS components if the PP is positioned outside the storage room.

For the measurement campaign we have four base stations at our disposal which we are to position in the environment. With the FPs installed, we will then position the PP at different locations and collect RSS and phase information according to the mails discussed in Section 2.2.1. In Figure 7.2 we have depicted the measurement environment with the positions of the FPs and the grid lines on which we position the PP. As we will view the environment as a local coordinate system, we have added coordinate axes which align with the walls.

Base stations FP_1, FP_2 and FP_3 are all installed as depicted in Figure 7.3a just above the ceiling panels. The ceiling panels are made of plaster and, therefore, some attenuation might be expected. The last base station, FP_4, has been mounted inside the small storage room. It has, however, been mounted on a storage shelve

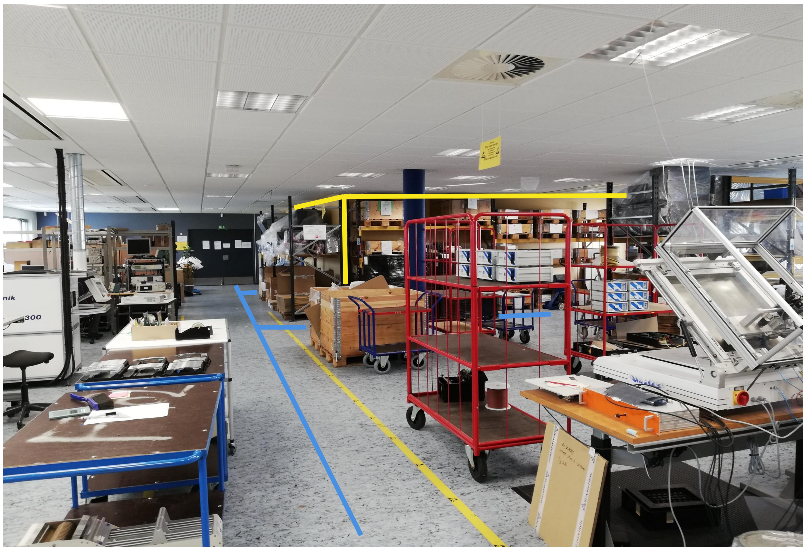


Figure 7.1: The measurement campaign environment. The smaller storage room is marked with yellow and the grid lines used for the PP positions have been depicted by blue lines.

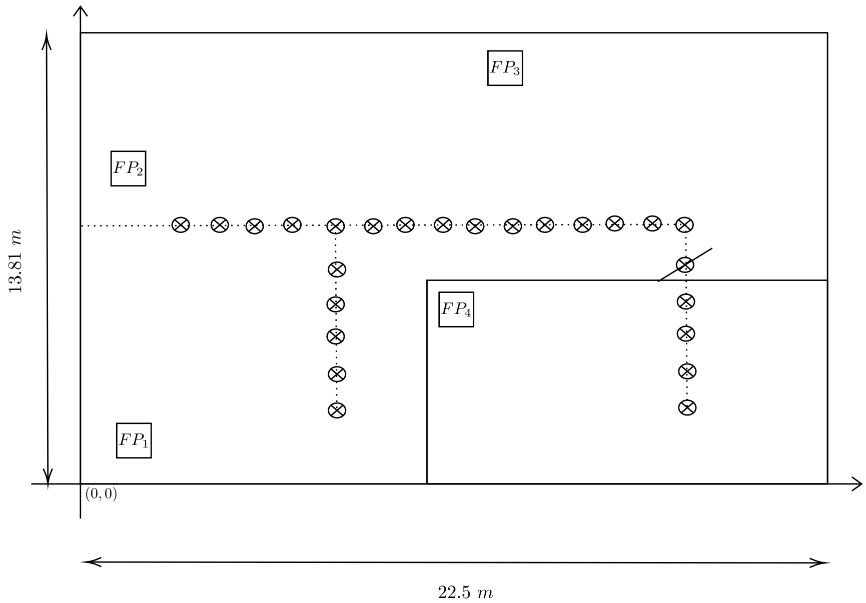
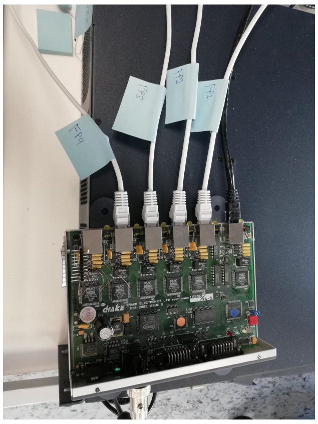


Figure 7.2: Illustration of the measurement setup. FP_1, FP_2 and FP_3 are all located 2.74 m above the ground. The last base station, FP_4, is located at a height of 2.32 m.

(a) FP located in the ceiling.

(b) FP mounted in the storage room.



(c) Splitter.

(d) Central unit.

Figure 7.3: Different positions of the FPs and the central unit allowing communication between the units.

as seen in Figure 7.3b. All base stations are connected with Ethernet cables through a splitter to a central device which allows communication between the FPs and the PP, see Figure 7.3c and 7.3d.

In Table 7.1 we have listed the positions of the base stations. Although the localization algorithm has been developed for planar positioning, we have also listed the height of each base stations in order to include it as a source of error.

In total, we have acquired measurements from 30 different PP positions in the environment with 18 positions on the horizontal grid line in Figure 7.2 and 6 positions on each of the vertical lines thus representing a broad range of LOS and NLOS components to each of the base stations. All positions are separated by one meter. For each position, the PP was mounted on a tripod, see Figure 7.4, with a height of 1.05 m in order to represent a portable device worn by an employee. This also reduces the height difference between the PP and FPs.

FP	**Position** [m]	**Height** [m]
FP_1	$[0.8, 1.20]^T$	2.74
FP_2	$[0.8, 10.52]^T$	2.74
FP_3	$[15, 13.31]^T$	2.74
FP_4	$[14.28, 6.57]^T$	2.32

Table 7.1: Positions and heights of the four base stations.

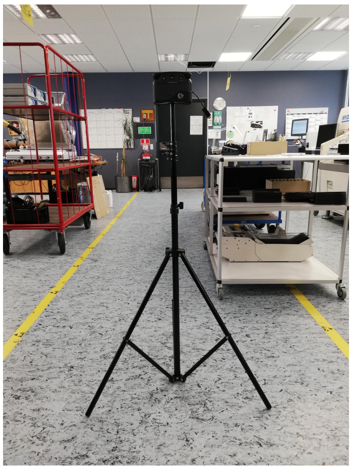
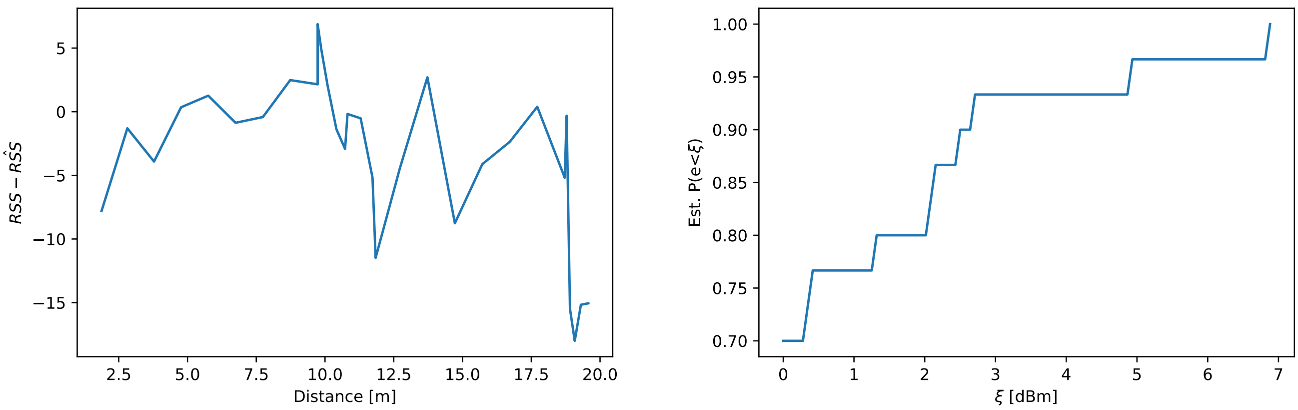


Figure 7.4: The PP is mounted on a tripod which is then positioned at each of the grid line positions. The height of the PP is 1.05 m.



(a) Residuals of the derived model.

(b) ECDF of the residuals.

Figure 7.5: Residuals of the derived model alongside the ECDF.

7.2 Comparing the RSS model to real data

In this section we will compare the relationship between RSS and distance obtained from the measurements to the derived RSS model in (3.3). The constants in the derived model have been chosen based on empirical studies. The RSS_0 and PL_0 term have been estimated in another scenario by measuring the RSS close to the FP and 5 meters away (d_0). Doing so, we obtain the empirical values, $RSS_0 = -26$ dBm and $PL_0 = -5$ dBm. The pathloss constant, η, and the standard deviation of the fading term, X_G, have been discussed in Chapter 5 and chosen to be 3 and 10 dBm respectively. We choose to focus on FP_2 and, therefore, use the RSS and distances between this base station and each of the 30 PP positions. In Table 7.2 we have listed the information between the specific PP position and FP_2.

The RSS values in Table 7.2 are averages of several mails. In order to compare our model, we have generated mean RSS values for the same distances. Thus, to see if we capture the mean signal decay, we have plotted the residuals of the model and true data alongside the ECDF in Figure 7.5.

We can see that the overall tendency in the residuals seems fine. However, some of the residuals differ from 0 and for some of the last measurements, the RSS is highly attenuated which is not captured by the derived model. These fluctuations are also apparent in the ECDF of the residuals. The standard deviation of the residuals is $\sigma = 6.23$ dBm. It should, however, be stated that only 30 measurements have been used to acquire Figure 7.5. The conclusions based on these figures will be more statistical sound if more data points had been acquired and included.

Assuming that the RSS model captures the RSS and distance relationship, we will in the following test the performance of the algorithm for a small range of the

PP$_i$	**RSS** [dBm]	**Distance** [m]
PP_1	-26.02	1.87
PP_2	-24.81	2.81
PP_3	-31.28	3.78
PP_4	-30.03	4.77
PP_5	-31.57	5.75
PP_6	-35.77	6.75
PP_7	-37.11	7.74
PP_8	-35.79	8.74
PP_9	-37.53	9.73
PP_{10}	-43.87	10.73
PP_{11}	-47.28	11.73
PP_{12}	-47.59	12.72
PP_{13}	-41.45	13.72
PP_{14}	-53.83	14.72
PP_{15}	-50.04	15.72
PP_{16}	-49.09	16.72
PP_{17}	-47.10	17.72
PP_{18}	-53.38	18.72
PP_{19}	-32.80	9.73
PP_{20}	-34.94	9.86
PP_{21}	-38.09	10.09
PP_{22}	-41.93	10.41
PP_{23}	-41.23	10.82
PP_{24}	-42.14	11.30
PP_{25}	-53.71	11.84
PP_{26}	-48.56	18.79
PP_{27}	-63.81	18.91
PP_{28}	-66.45	19.08
PP_{29}	-63.77	19.30
PP_{30}	-63.83	19.58

Table 7.2: RSS and distance from the 30 different PP positions to the FP_2 base station.

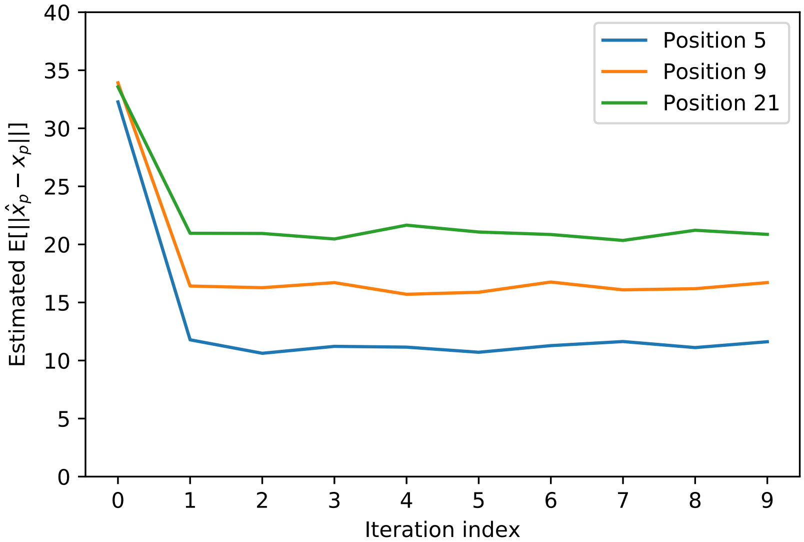


Figure 7.6: Mean error distance for position 5 (blue), position 9 (yellow) and position 21 (green). As no angle information is provide the precision of the algorithm is insufficient.

measured positions.

7.3 Performance with real data

We will issue similar tests as done in Chapter 6. Initially, we will test the algorithm while providing no prior knowledge. In Figure 7.6 we derive the mean error distance. This has been done for the true positions PP_5, PP_9 and PP_{21}. For these positions, the PP is located "inside" the square suspended by the base stations.

The results in Figure 7.6 show the same behaviour as in the simulations. We see that the mean error distance quickly converges but to a high mean for the considered positions. This is, however, expected, as the angle between the PP and each FP is updated based on the prior position of the PP. The prior knowledge of the PP has once more been initialized as a bivariate normal distribution with parameters

$$\bar{\mu}_{x_p} \sim \mathcal{U}([-5 \cdot \mathrm{dim},\ 5 \cdot \mathrm{dim}] \times [-5 \cdot \mathrm{dim},\ 5 \cdot \mathrm{dim}])$$
$$\bar{\Sigma}_{x_p} = \begin{bmatrix} 10^6 & 0 \\ 0 & 10^6 \end{bmatrix}$$

and, therefore, the angle variables may be updated in a completely wrong direction. We have visualized the precision by providing 25 estimates, each obtained

	FP_1	FP_2	FP_3	FP_4
μ_v	$\frac{\pi}{4}$	0	$-\frac{3\pi}{4}$	π
κ	2	1	2	1

Table 7.3: Prior angle knowledge of the base stations in the measurement setup.

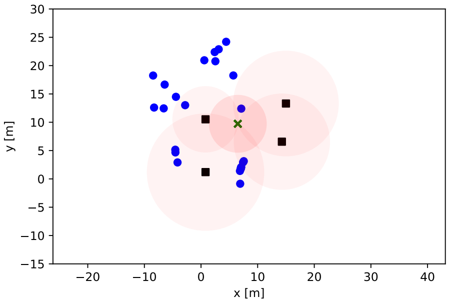
on the basis of a different prior of the PP position. The ECDF of the respective estimates for the positions has also been computed and presented. The results can be seen in Figure 7.7.

Not surprisingly, we receive estimates at random angles from 0 to 2π due to the prior of the PP. From the ECDFs we see that observing a somewhat useful error distance is highly improbable. However, if we imagine that we rotate the estimates towards the true angle, the distance estimates seem to have been inferred correctly from the RSS and phase information just at a wrong angle. Therefore, providing some angle information may provide an improved precision. We will, therefore, use the same alterations from Chapter 6 by applying prior knowledge of the angles. The prior information of the angles has been summarized in Table 7.3 which is identical to the information provided in scenario *S2* in Chapter 6.

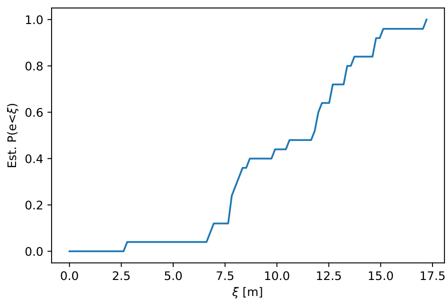
If the environment had allowed for a different base station installation, FP_3 and FP_4 would have been mounted in their respective corner, i.e. the north and south eastern corners to provide a symmetric square room. Doing so, the prior angle knowledge could have been chosen similarly to scenario *S1* in Chapter 6. To determine the mean error distance when providing angle information, we have once more computed the iteration plots, see Figure 7.8.

In Figure 7.8 we see a great improvement in the mean error distance for position 5 and 9, both of which have more than halved due to the prior angle knowledge. The algorithm also quickly converges. However, for position 21, the precision worsens after a single iteration but then stabilizes at a high mean error. A reason for this may be that both position 5 and 9 are located such that the base stations encircle the true positions which is less true for position 21 which is somewhat closer to the boundary created by the FPs. In order to visualize the effects of the prior angle knowledge we once more show 25 estimates of the positions with their respective ECDFs in Figure 7.9.

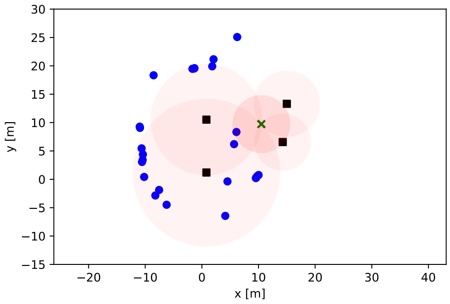
We see that for position 5 and 9 the estimates provided by the algorithm group together in close vicinity to the true position. For position 5 the estimates are especially good and from the ECDF we can see that observing an error less than 2.5 m is highly probable. This is, however, not the case for position 9 but the errors are all within 5 m. The precision of the estimates of position 21 is validated here. We see that the estimates fall into a cluster which is close to the intersection point



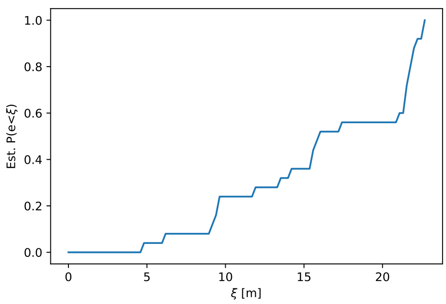
(a) Estimates of position 5.



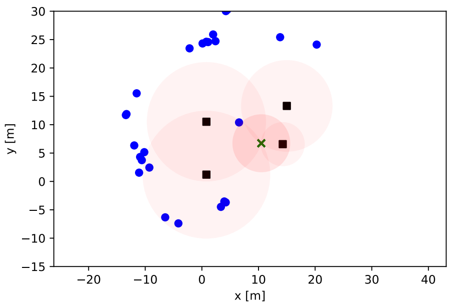
(b) ECDF of position 5.



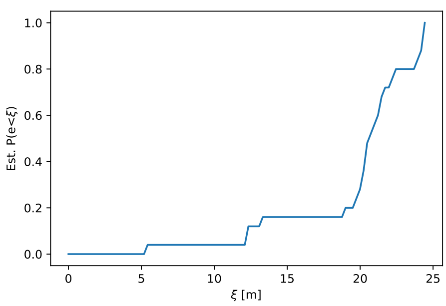
(c) Estimates of position 9.



(d) ECDF of position 9.



(e) Estimates of position 21.



(f) ECDF of position 21.

Figure 7.7: Estimates and ECDFs of the chosen positions in the real setup. We observe the same behaviour in the estimates as in the simulations.

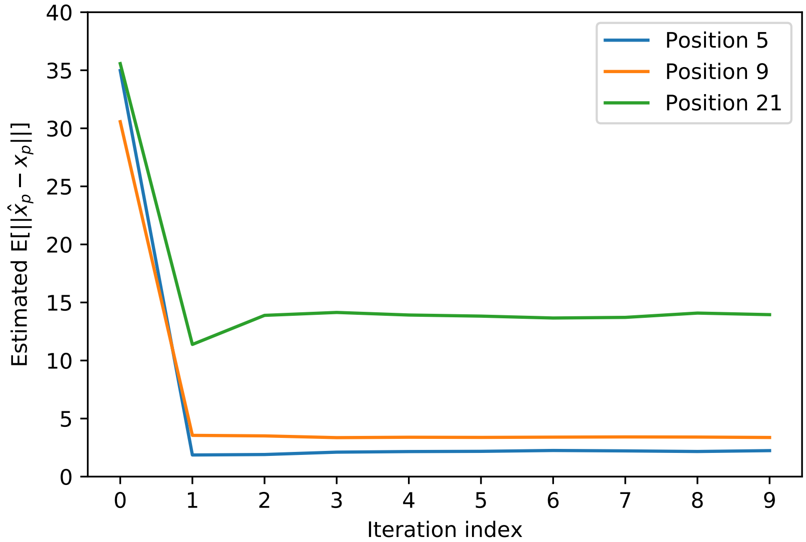


Figure 7.8: Mean error distance when we apply prior knowledge of the base stations.

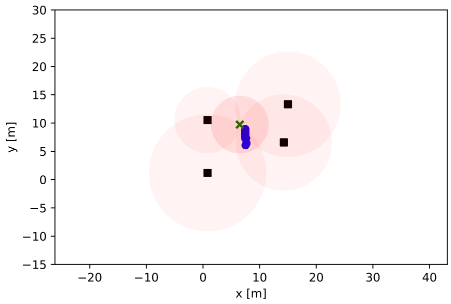
of the circles originating in FP_2 and FP_3. This might be caused by inferior RSS measurements which favors a different intersection point than the true position as discussed in [27]. Before we issue modifications to the algorithm, as done in Chapter 6, we will show the message approximations when using real data. However, we only show it for position 21, see Figure 7.10.

We see that the message approximations do follow the true densities, even for real data. However, we see that the PP position is updated towards a position outside the environment which may be caused by inferior base station measurements. In Figure 7.11 we have presented the approximated and true message from the r and v node to the $\boldsymbol{x}_p$ node after a single message iteration.

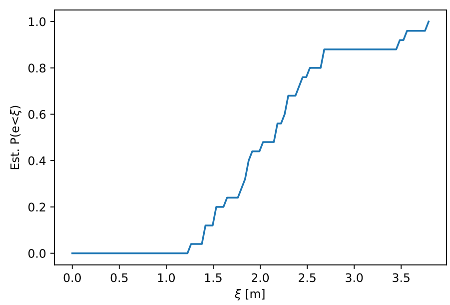
We see that even though the measurement from FP_2 has a mode right next to the true position, the measurement from FP_4 pushes the estimate outside the room. Therefore, using the measurement from FP_3 as the second update instead of FP_4 might prove advantageous.

For the above performance tests we have updated the PP position with the direct link FP first and then the other bases subsequently. As the base station providing the least attenuated RSS measurement may not be used first we will issue the same test from Chapter 6, where we sort the observations with respect to their signal strength. Doing so, we wish to explore the impact on the mean error distance which has been presented in Figure 7.12 alongside the results obtained with sector antennas.

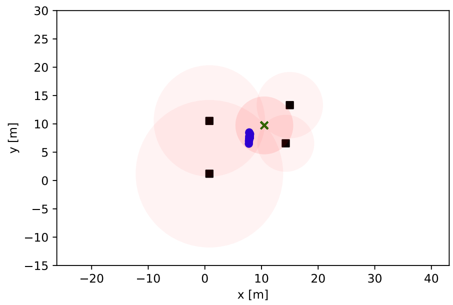
From Figure 7.12 we do not see any improvement in the precision of position



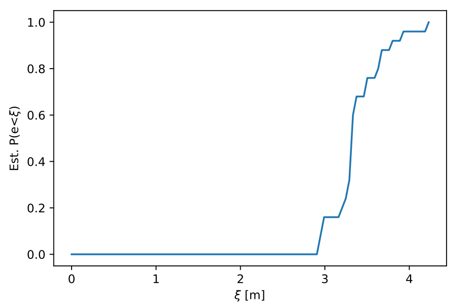
(a) Estimates of position 5.



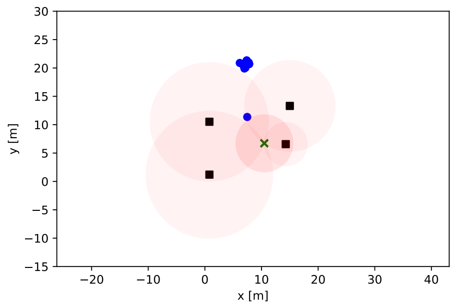
(b) ECDF of position 5.



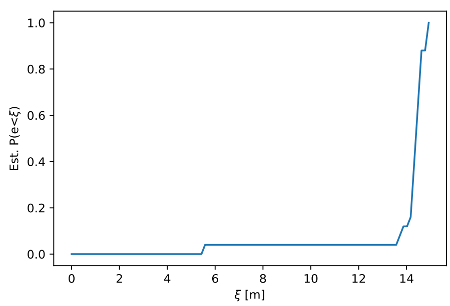
(c) Estimates of position 9.



(d) ECDF of position 9.



(e) Estimates of position 21.



(f) ECDF of position 21.

Figure 7.9: Estimates and ECDFs of the positions in the real setup. We see that using sector antennas has a high impact on the precision.

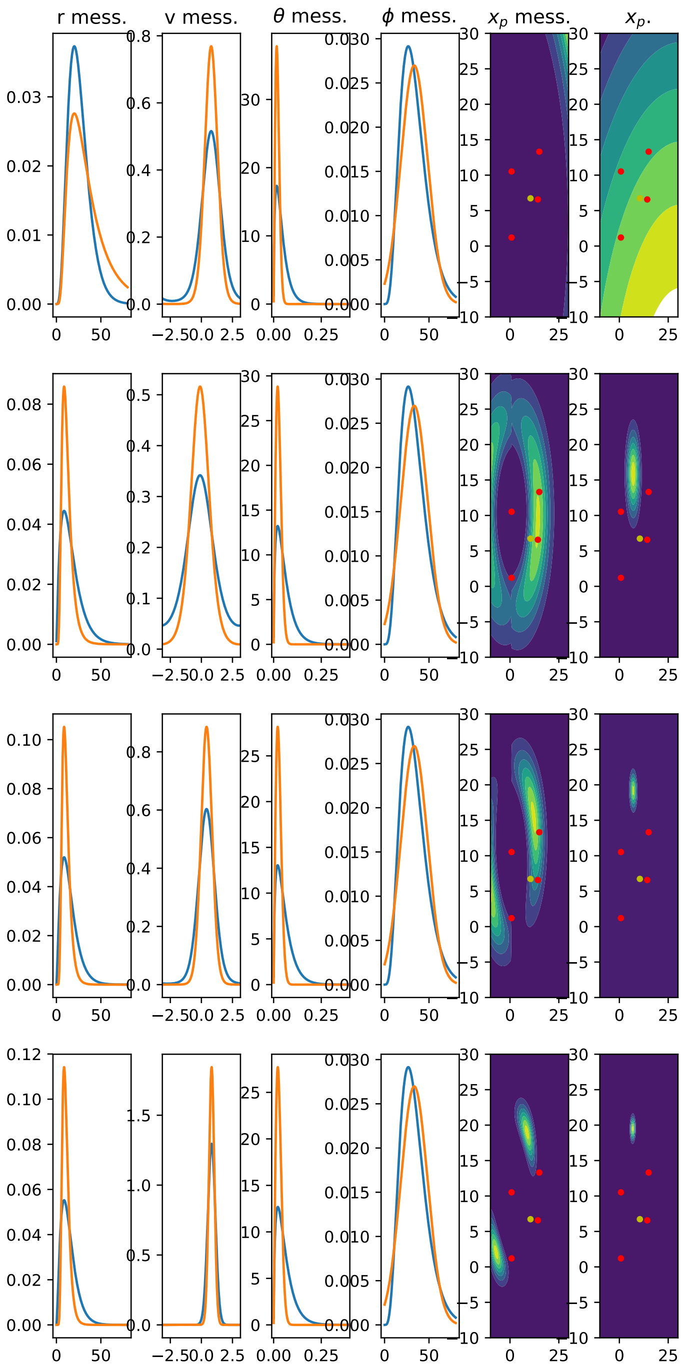


Figure 7.10: Message approximations for position 21 for each iteration when prior knowledge of the angles is included. We see that the approximations agree with the true densities.

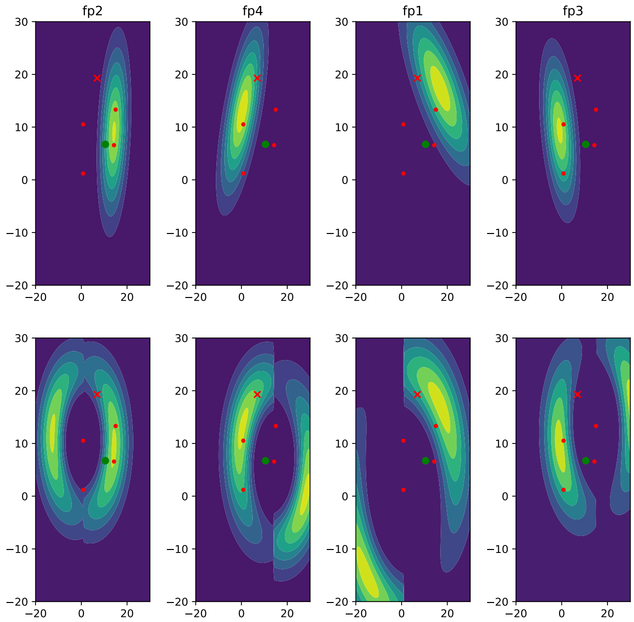


Figure 7.11: Approximated (top) and true (bottom) messages from each branch in the factor graph to the x_p node after a single message iteration. The green dot and red cross represent the true and estimated position respectively while the red dots represent the base stations.

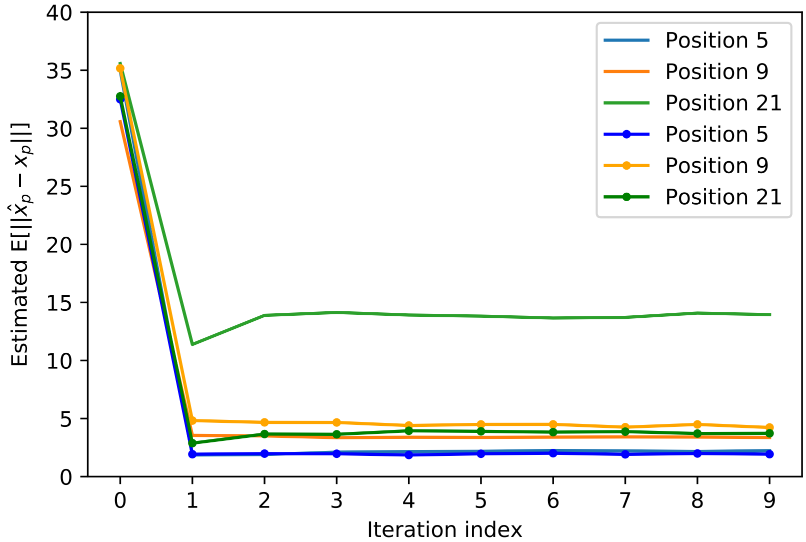
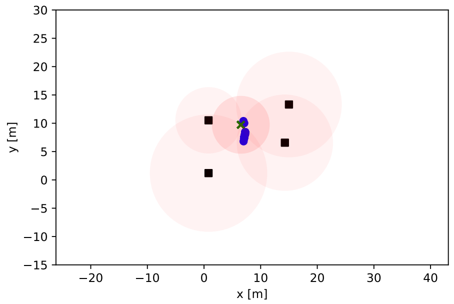


Figure 7.12: Mean error distance when we apply prior knowledge of the base stations and sort the RSS information. The dotted lines represent the mean errors when the FPs are sorted.

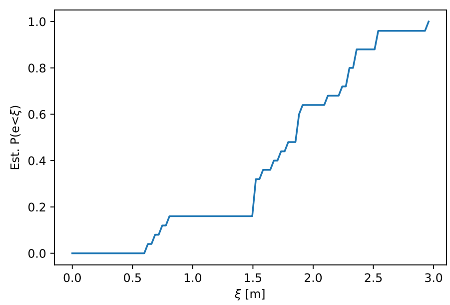
5 and for position 9 we see that the mean error distance has increased. This may be caused by the PP position being updated with a base station providing wrong RSS measurements and that the fading descriptor in the algorithm does not infer fading correctly to the distance. This may also be combined by the fact that, due to the base stations positions, the angle information is somewhat difficult to initialize and this problem may have been reduced if the base stations were mounted in the true corners. Meanwhile, the mean error distance for position 21 has improved by almost a factor three. This might imply that one of the base stations provided inferior RSS measurements combined with a poor angle prior to the PP at this position. In order to visualize the effects of sorting the FPs we show 25 estimates of the positions with their respective ECDFs in Figure 7.13.

We see that sorting the RSS measurements has a positive effect on the estimates of position 21. These are now closer to the true position but some of the estimates do seem to cluster around another mode. This behaviour is also apparent in the estimate of position 9 which seem to be drawn towards the intersection of the circles originating in FP_2 and FP_3. For position 5 the estimates seem somewhat unaffected but the respective ECDF shows some degradation in the precision. However, due to Figure 7.13, we might suspect that the difference in the ECDF may be based on the used priors.

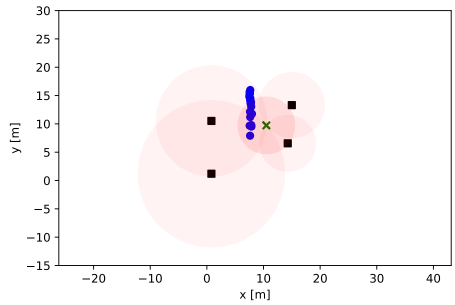
As we only had four FPs available we are not able to add additional base station to this scenario. However, we saw in Chapter 6 that receiving information from



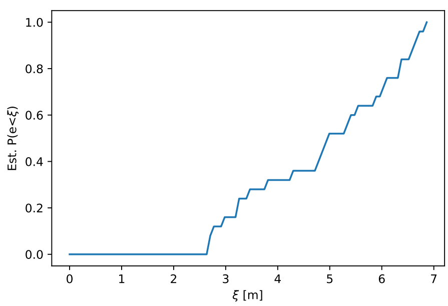
(a) Estimates of position 5.



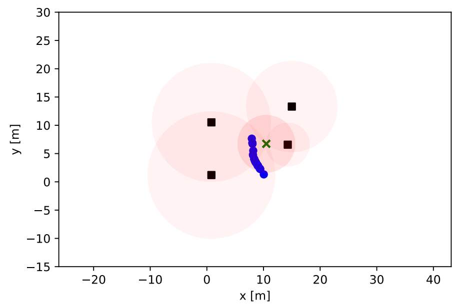
(b) ECDF of position 5.



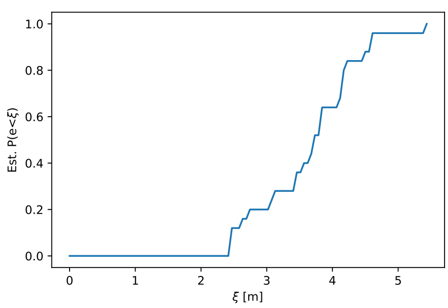
(c) Estimates of position 9.



(d) ECDF of position 9.



(e) Estimates of position 21.



(f) ECDF of position 21.

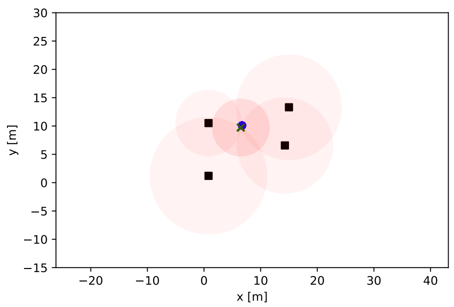
Figure 7.13: Estimates and ECDFs of the positions in the real setup. We see that using sorted measurements has a positive effect on position 21 but not on position 9.

more bases in this setup did not provide a significant increase in precision.

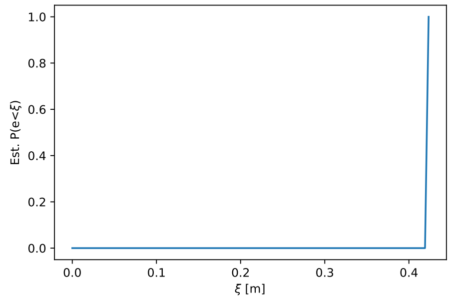
For the sake of completeness, we will for a last performance test, show the results of the algorithm if the true angles are provided and the prior mean of the PP is chosen uniformly on the room with a low variance, i.e. we provide almost all available information except distance, see Figure 7.14.

We now see that the algorithm provides almost perfect estimates of position 5 where the ECDF depicts that observing an error less than 1 m is now highly probable. However, for positions 9 and 21 the algorithm does not capture the true distance between the PP and FPs through the RSS. The estimates are, however, still within a radius of 5 m. This might, therefore, be an indication that the fading effects have not been modelled correctly and the latent distance information in the RSS is, therefore, distorted.

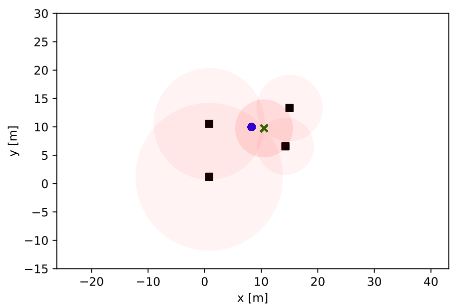
Based on the above performance tests we argue that the MMVMP algorithm returns similar results to those obtained in Chapter 6. However, we do also experience some problems with the algorithm. We have noted that due to the base station placements, the angle priors are somewhat difficult to initialize and, therefore, a more symmetric installation might be favorable. Furthermore, we do also observe erroneous estimates if inferior base stations are used first in the update scheme. We did, however, see that sorting the RSS measurements improved the precision.



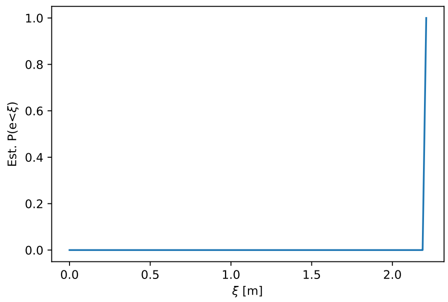
(a) Estimates of position 5.



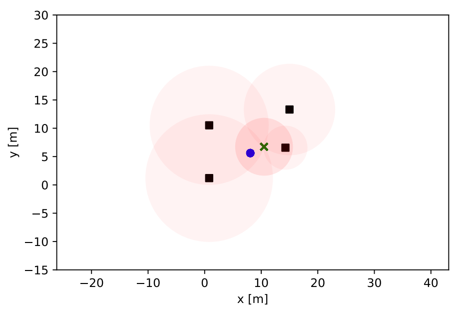
(b) ECDF of position 5.



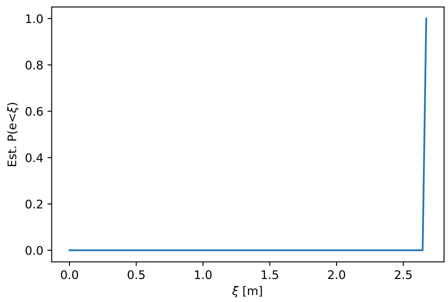
(c) Estimates of position 9.



(d) ECDF of position 9.



(e) Estimates of position 21.



(f) ECDF of position 21.

Figure 7.14: Estimates and ECDFs of the positions in the real setup. We see that providing the true angle and initializing the PP position within the room returns almost precise estimates.

Conclusion and Outlook

This thesis presents the derivation of a Received Signal Strength (RSS) based localization algorithm for indoor positioning. Statistical methods have been explored and an extensive system specification was employed. Based on the findings of this initial exploration, a message passing algorithm has been derived. The algorithm combines variational message passing and moment matching in which the messages are restricted to the exponential family of probability distributions. Adopting moment matching proved to be advantageous, as numerical optimization of the Kullback-Leibler divergence can be avoided. Thereby it is also possible to avoid time consuming derivations of the optimization problem. Some of the approximated messages do, however, express deficient variability compared to the true messages and, therefore, further work should be invested in this area.

Furthermore, an RSS model has been derived based on the log-distance pathloss model in which fading effects are addressed through a Gaussian random variable. The model has shown promising results for describing mean signal strength compared to true data. However, only a small amount of data points were available and, therefore, future research should investigate the model through a larger data set.

To facilitate performance assessment of the algorithm, a simulation environment has been implemented for both simulated and real data. If the algorithm is not provided any angle information, non-reliable estimates are obtained but assuming base stations to be sector antennas highly increases the performance. Therefore, the algorithm can be applied in scenarios where prior knowledge assumes the object to be located inside the environment. However, performance tests depict that the algorithm converges quickly based on the first observed base stations and may, therefore, favor false position estimates. Sorting the RSS information with respect to the least attenuated signal reduces this problem.

Future research should include investigations in the convergence of the PP position and the statistical relationship between RSS and distance. Thus more reliable distance estimates may be obtained. However, when prior angle knowledge is applied and the measurements are sorted the algorithm shows reliable results. In the investigated scenarios mean error distances of approximately 5 meters in simulated environments and below 5 meters for the inspected true positions in the measurement campaign are obtained. Therefore, we argue that the moment matching variational message passing algorithm is applicable in scenarios including restaurants or large warehouses where prior knowledge of the angles can be assumed.

The results of this project can impact the future direction of indoor localization methods based on RSS. Due to the obtained mean error distances the methods explored here are able to infer latent distance information in the signal strength from local base stations and based on these yield relatively reliable position estimates.

Bibliography

[1] G. Dedes and A. G. Dempster, "Indoor gps positioning", 2005.

[2] Y. Gao, S. Liu, M. M. Atia, and A. Noureldin, "Ins/gps/lidar integrated navigation system for urban and indoor environments using hybrid scan matching algorithm", *Sensors*, 2015.

[3] H. Liu, H. Darabi, P. Banerjee, and J. Liu, "Survey of wireless indoor positioning techniques and systems", *IEEE TRANSACTIONS ON SYSTEMS, MAN, AND CYBERNETICS*, 2007.

[4] N. A. Alsindi, B. Alavi, and K. Pahlavan, "Measurement and modeling of ultrawideband toa-based ranging in indoor multipath environments", *IEEE TRANSACTIONS ON VEHICULAR TECHNOLOGY*, 2009.

[5] S. A. R. Zekavat and R. M. Buehrer, *Handbook of Position Location: Theory, Practice, and Advances*. John Wiley & Sons, 2012.

[6] J. Bosse, A. Ferréol, and P. Larzabal, "A glance on geographical positioning", *Computational Advances in Multi-Sensor Adaptive Processing (CAMSAP)*, vol. 4th, 2011.

[7] S. Yiu, M. Dashti, H. Claussen, and F. Perez-Cruz, "Wireless rssi fingerprinting localization", *Signal Processing*, vol. 131, pp. 235–244, 2017.

[8] H. Wymeersch, J. Lien, and M. Z. Win, "Cooperative localization in wireless networks", *Proceedings of the IEEE*, 2009.

[9] A. Ihler, J. Fisher, R. Moses, and A. Willsky, "Nonparametric belief propagation for self-localization of sensor networks", *IEEE Selected Areas in Communications*, 2005.

[10] C. Pedersen, T. Pedersen, and B. H. Fleury, "A variational message passing algorithm for sensor self-localization in wireless networks", *IEEE International Symposium on Information Theory Proceedings (ISIT)*, pp. 2158–2162, 2011.

[11] D. N. Urup, "Distributed localization in dynamic cooperative wireless sensor networks using the mean field approximation", Master's thesis, Aalborg University, 2014.

[12] *Digital Enhanced Cordless Telecommunications (DECT); A High Level Guide to the DECT Standardization*, 2005. [Online]. Available: https://www.etsi.org/deliver/etsi_tr/101100_101199/101178/01.05.01_60/tr_101178v010501p.pdf.

[13] Wikipedia, *Cyclic redundancy check*. [Online]. Available: https://en.wikipedia.org/wiki/Cyclic_redundancy_check.

[14] G. Chartrand, L. Lesniak, and P. Zhang, *Graphs & Digraphs*. CRC Press, 2016.

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

[16] B. Sklar, "Rayleigh fading channels in mobile digital communication systems part i: Characterization", *IEEE Communications Magazine*, 1997.

[17] T. M. Cover and J. A. Thomas, *Elements of Information Theory*, Second. John Wiley & Sons, 2006.

[18] J. C. Baez and T. Fritz, "A bayesian characterization of relative entropy", *Theory and Applications of Categories*, vol. 29, no. 16, pp. 422–456, 2014. [Online]. Available: http://www.tac.mta.ca/tac/volumes/29/16/29-16.pdf.

[19] J. Dauwels, "On variational message passing on factor graphs", *IEEE International Symposium on Information Theory*, 2007.

[20] Wikipedia, *Bessel function*. [Online]. Available: https://en.wikipedia.org/wiki/Bessel_function#Modified_Bessel_functions:_I%CE%B1,_K%CE%B1.

[21] T. Minka, "Divergence measures and message passing", 2005.

[22] Wikipedia, *Gamma distribution*. [Online]. Available: https://en.wikipedia.org/wiki/Gamma_distribution.

[23] ——, *Von mises fisher distribution*. [Online]. Available: https://en.wikipedia.org/wiki/Von_Mises%E2%80%93Fisher_distribution.

[24] G. Kurz and U. D. Hanebeck, "Trigonometric moment matching and minimization of the kullback–leibler divergence", *IEEE Transactions on Aerospace and Electronic Systems*, Oct. 2015.

[25] ITU-R, "Propagation data and prediction methods for the planning of indoor radiocommunication systems and radio local area networks in the frequency range 300 mhz to 100 ghz", 2017.

[26] Wikipedia, *Von mises distribution*. [Online]. Available: https://en.wikipedia.org/wiki/Von_Mises_distribution.

[27] M. B. Koldsgaard, "Implementation, performance assessment, and refinement of selected distributed localization algorithms for cooperative wireless networks", Master's thesis, Aalborg University, 2013.

[28] L. D. Nardis, *TDMA, FDMA and CDMA*, 2007. [Online]. Available: http://acts.ing.uniroma1.it/courses/comelet/Slides/20071217_TEL_lecture_2.pdf.

[29] P. R. Turner, *Guide to Scientific Computing*. MacMillan Press LTD, 2000.

A | FDMA/TDMA Multiple Access

In order to efficiently utilize the communication resource, bandwidth, a *multiple access* (MA) scheme needs to be used. This will ensure the possibility of allowing a number of users in the environment to share the bandwidth and, therefore, access the same base station. One of these MA schemes is the *frequency division multiple access* or FDMA where each transmitter is allocated a different carrier frequency. Therefore, each user can transmit with no limitations in time but only using a portion of the bandwidth, see Figure A.1.

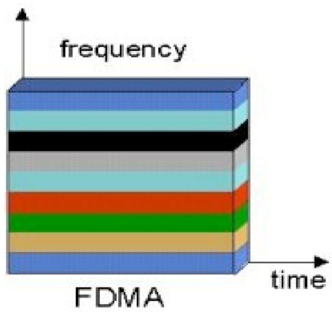


Figure A.1: Illustration of FDMA transmission planning [28]. The different colors represent a different carrier frequency which is allocated to a user.

If time is shared instead of frequency, *time division multiple access* or TDMA is used. In TDMA the users share a common carrier frequency in order to communicate with the base station. In order to distinguish the different users, each user is allocated a single or multiple time slots in which they are allowed to transmit information. Thus, when a user transmits data, they occupy the entire frequency bandwidth and separation among users is determined in the time domain, see Figure A.2.
The mentioned time slots are organized in a frame, see Figure A.3.

Before a frame is transmitted, a reference burst is emitted, which forces a measure of synchronization of all the users. They are then allowed to fill their respec-

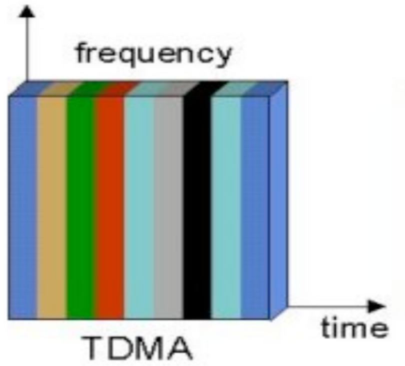


Figure A.2: Illustration of TDMA transmission planning [28]. The different colors identifies a single user occupying a time slot. In each time slot the user is allowed to transmit data in the entire frequency band. Each time slot is seperated in time.

tive time slot(s). However, as there may be significant delays between users, due to propagation effects, each of the users will receive the reference burst with a different phase, see Figure A.4. In order to compensate for the misalignment, each time slot can be equipped with a guard time. In practice, this means that each time slot is longer than the time needed for each user to transmit data. Thus, overlap in data transmission can be avoided [28]. Furthermore, to overcome uncertainty in phase relative to the reference burst, the user will send a preamble prior to the relevant information. This preamble acts as an identification between base station and user.

The advantages of the TDMA method is that common radio modem equipment can be used at a given frequency while being shared among N users. Furthermore, the bit rates to and from the user terminals are readily changed by allocating more or fewer time slots to the specific user and thus altering the time in which the user can transmit data.

TDMA does, however, have some disadvantages. If the system consists of N user terminals with equal bit rates the receiver of each individual terminal operates with a cycle of $\frac{1}{N}$. This entails that the receiver terminals have a periodically pulsating power envelope. Also, as high bit rates can be obtained with TDMA it may be required to perform equalization against multipath propagation.

Usually, a communication channel can be allocated by combining both FDMA and TDMA. In the FDMA/TDMA scheme a channel is a combination of a carrier frequency and a TDMA frame, see Figure A.5

The channels derived through this model are called *orthogonal channels* as each user have a unique combination of frequency and time slot.

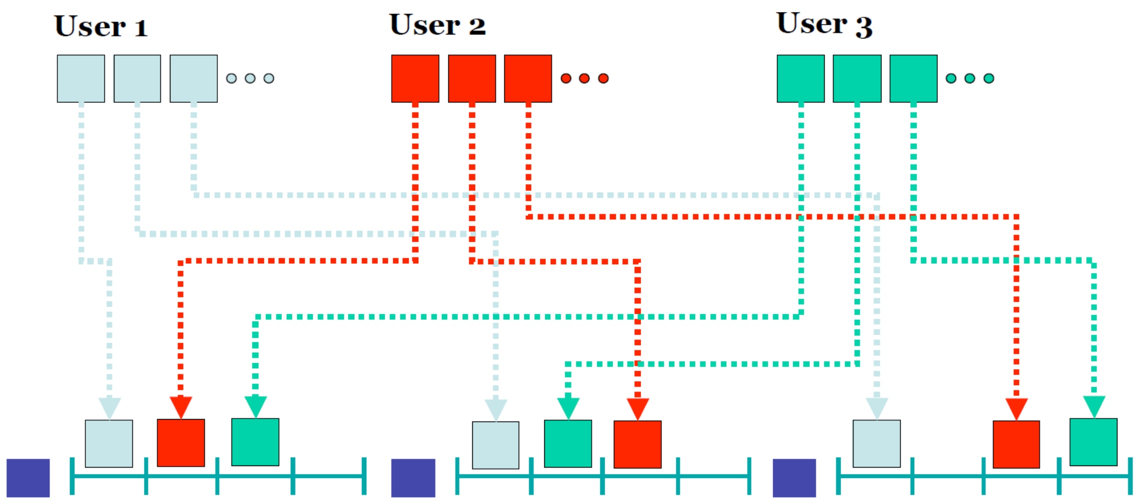


Figure A.3: Illustration of TDMA frames [28]. Prior to each frame a reference burst (blue box) is transmitted. Next, the users occupy their designated time slot(s).

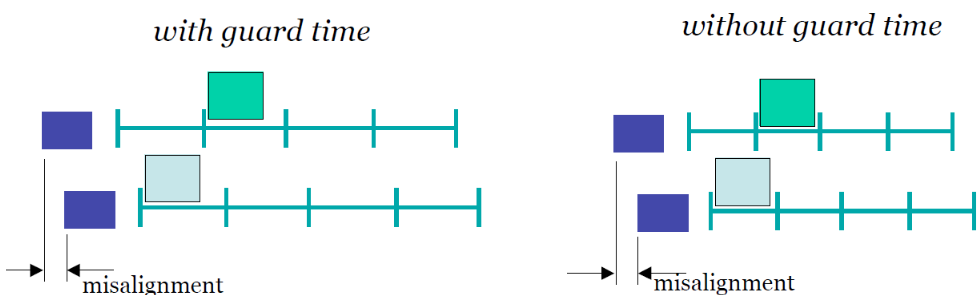


Figure A.4: Illustration of phase misalignment due to delays [28]. Delays among the users introduce phase misalignment in the reference bursts. Transmission overlap is circumvented by adding guard times in each time slot thus compensating for phase differences.

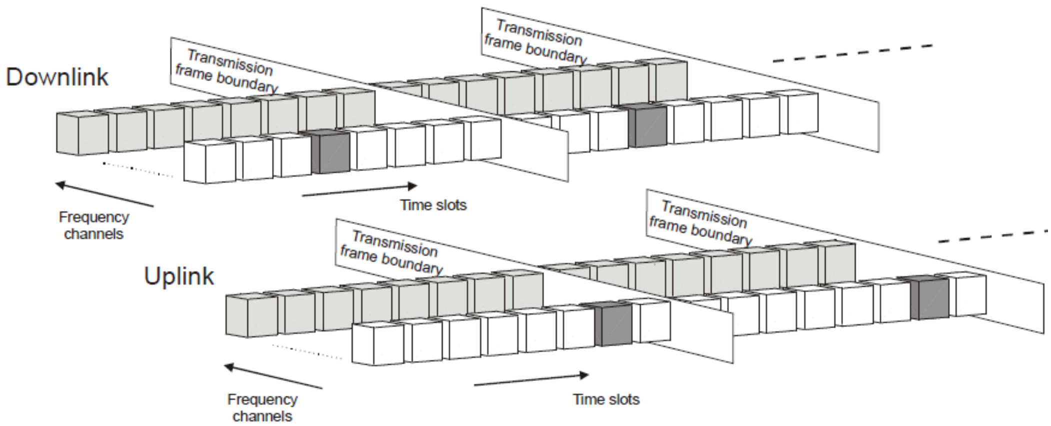


Figure A.5: Illustration of the FDMA/TDMA scheme [28]. A channel is a combination of a carrier frequency and associated TDMA frame.

B Message derivation

We derive the message $m_{g_{\theta,r,rss}\to\theta}(\theta)$.

$$
\begin{aligned}
m_{g_{\theta,r,rss}\to\theta}(\theta) &= \exp\left[\int_r\int_{rss} m_{r\to g_{\theta,r,rss}}(r)\cdot m_{rss\to g_{\theta,r,rss}}(rss)\cdot \ln\mathcal{N}(\mu_r(r),\theta^{-1})\mathrm{d}rss\ \mathrm{d}r\right]\\
&= \exp\left[\int_r q(r)\ln\mathcal{N}(\mu_r(r),\theta^{-1})\mathrm{d}r\right]\\
&= \exp\left[\mathbb{E}_{q(r)}\left[\ln\frac{1}{\sqrt{2\pi}}+\ln\theta-\frac{\theta^2}{2}\cdot(rss_{obs}-\mu_r(r))^2\right]\right]\\
&= \exp\left[\ln\frac{1}{\sqrt{2\pi}}+\ln\theta-\frac{\theta^2}{2}\cdot\mathbb{E}_{q(r)}\left[(rss_{obs}-\mu_r(r))^2\right]\right]\\
&= \exp\left[\ln\frac{1}{\sqrt{2\pi}}+\ln\theta-\frac{\theta^2}{2}\cdot\mathbb{E}_{q(r)}\left[rss_{obs}^2+B^2+\left(\frac{10\eta}{\ln 10}\right)^2\cdot(\ln r)^2-2B\frac{10\eta}{\ln 10}\cdot\ln r\right.\right.\\
&\left.\left.-2rss_{obs}\cdot\left(B-\frac{10\eta}{\ln 10}\ln r\right)\right]\right]\\
&= \exp\left[\ln\frac{1}{\sqrt{2\pi}}+\ln\theta-\frac{\theta^2}{2}\cdot\left(rss_{obs}^2+B^2+\left(\frac{10\eta}{\ln 10}\right)^2\cdot\mathbb{E}_{q(r)}\left[(\ln r)^2\right]-2B\frac{10\eta}{\ln 10}\cdot\mathbb{E}_{q(r)}\left[\ln r\right]\right.\right.\\
&\left.\left.-2rss_{obs}B+2rss_{obs}\frac{10\eta}{\ln 10}\mathbb{E}_{q(r)}[\ln r]\right)\right]\\
&= \exp\left[\ln\frac{1}{\sqrt{2\pi}}+\ln\theta-\frac{\theta^2}{2}\left(rss_{obs}^2+B^2+\left(\frac{10\eta}{\ln 10}\right)^2\left(\psi^{(1)}(\alpha_r)+(\psi(\alpha_r)-\ln\beta_r)^2\right)\right.\right.\\
&\left.\left.-2B\frac{10\eta}{\ln 10}\left(\psi(\alpha_r)-\ln\beta_r\right)-2rss_{obs}B+2rss_{obs}\frac{10\eta}{\ln 10}\left(\psi(\alpha_r)-\ln\beta_r\right)\right)\right].
\end{aligned}
$$

In the deriviation we have used that the mean and variance of a log-Gamma random variable [22], X, are given by

$$
\begin{aligned}
\mathbb{E}[X] &= \psi(\alpha)-\ln\beta\\
\mathrm{Var}(X) &= \psi^{(1)}(\alpha)
\end{aligned}
$$

where ψ is the digamma function and $\psi^{(1)}$ is the polygamma function.

C | Trapezoidal rule

Computing the moments of some of the derived messages in Chapter 4 requires the need of numerical integration. In this appendix we present the trapezoidal rule [29]. This rule seeks to approximate a finite integral as follows

$$\int_a^b f(x)\mathrm{d}x \approx \frac{b-a}{2} \cdot [f(a) + f(b)].$$

Several numerical integration methods exist such as Monte Carlo integration. Monte Carlo integration samples a large amounts of points uniformly on the interval and approximates the specific integral as

$$\int_a^b f(x)\mathrm{d}x \approx V \cdot \frac{1}{N} \sum_{i=1}^{N} f(\bar{x}_i)$$

where $\bar{x}_i, i = 1, \dots, N$ are the samples on the interval and V is the volume of the interval. From the law of large numbers this approximation returns the true integral for $N \to \infty$. We are, however, not interested in drawing a large number of samples in order to compute the integral and, therefore, use the trapezoidal rule.