A Compressed Sensing Approach To Channel Estimation For Impulse- Radio Ultra-Wideband (IR-UWB) Communication

BY

SYED FARAZ AHMED

A Thesis Presented to the

DEANSHIP OF GRADUATE STUDIES

KING FAHD UNIVERSITY OF PETROLEUM & MINERALS

DHAHRAN, SAUDI ARABIA

In Partial Fulfillment of the
Requirements for the Degree of

MASTER OF SCIENCE

In

ELECTRICAL ENGINEERING

JULY 2011

KING FAHD UNIVERSITY OF PETROLEUM & MINERALS

DHAHRAN- 31261, SAUDI ARABIA

DEANSHIP OF GRADUATE STUDIES

This thesis, written by **Syed Faraz Ahmed** under the direction his thesis advisor and approved by his thesis committee, has been presented and accepted by the Dean of Graduate Studies, in partial fulfillment of the requirements for the degree of **MASTER OF SCIENCE IN ELECTRICAL ENGINEERING.**
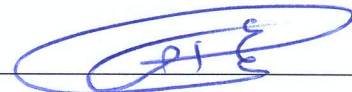
Thesis Committee

_____
Dr. Tareq Y. Al-Naffouri (Advisor)

_____
Dr. Ali H. Muqaibel (Co-Advisor)

_____
Dr. M. Adnan Landolsi (Member)

_____
Dr. Ali A. Al-Shaikhi (Member)

_____
Dr. Wajih A. Abu-Al-Saud (Member)

_____
Dr. Ali Ahmad Al-Shaikhi
Department Chairman

_____
Dr. Salam A. Zummo
Dean of Graduate Studies

1/10/11
_____
Date

Dedicated to my most loving and caring parents

ACKNOWLEDGMENTS

In the name of Allah, the Most Beneficent, the Most Merciful

All praise be to Allah (The One and The Only Creator of everything) for His limitless blessings. May Allah bestow peace and His choicest blessings on His last prophet, Hazrat Muhammad (Peace Be Upon Him), his family (May Allah be pleased with them), his companions (May Allah be pleased with them) and his followers.

I would like to express my profound gratitude and appreciation to my thesis committee chairman and adviser, Dr. Tareq Y. Al-Naffouri, whose expertise, understanding, and patience, added considerably to my graduate experience. I appreciate his vast knowledge and skill in many areas, specifically, stochastic processes, signal processing and above all mathematics. Dr. Tareq is the one professor who truly made a difference in my life. It was under his tutelage that I developed a focus and became interested in research. He provided me with direction, intellectual support and became more of a mentor than a professor. It was through his persistence, understanding and kindness that I completed my degree.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

(B)PPM   (binary) pulse-position modulation

(B)PSK   (binary) phase-shift keying

(MB-)OFDM   (multi band) orthogonal frequency division modulation

(N)DA   (non-)data aided

(N)LOS   (non-)line-of-sight

(O)MP   (orthogonal) matching pursuit

(R)MSE   (root) mean square error

ADC   analog-to-digital converter

AOA   angle-of-arrival

AWGN   additive white Gaussian noise

BER   bit-error rate

c.d.f.   cumulative distribution function

CDMA   code division multiple access

CIR   Channel Impulse Response

AIC	Akaike information criteria
AR	auto regressive
ARMA	auto regressive moving average
BIC	Bayesian information criteria
dB	decibel
EM	expectation maximization
GLRT	generalized likelihood ratio test
IC	information criteria
LTI	linear time invariant
LTV	linear time variant
UE	un-biased estimator
MVUE	minimum variance un-biased estimator
BLUE	best linear un-biased estimator
RBLS	Rao-Blackwell-Lehmann-Scheffe
WSN	wireless sensor network
FCC	Federal Communication Commission
IEEE	Institute of Electrical and Electronics Engineers
SVD	singular value decomposition
EVD	eigenvalue decomposition
PCA	principal component analysis
ICA	independent component analysis
GA	Genetic algorithm
ANN	artificial neural network

CRLB   Cramer-Rao Lower Bound

CS   compressive(compressed) sensing

CSI   channel state information

DARPA   Defense Advanced Research Projects Agency

DCT   discrete cosine transform

DFT   discrete Fourier transform

DS   direct sequence

DSSS   direct sequence spread spectrum

EGC   equal gain combiner

EU   European Union

EVD   eigen value decomposition

FBMP   fast Bayesian matching pursuit

FCC   Federal Communications Committee

FD   frequency domain

FDMA   frequency division multiple access

FFT   fast Fourier transform

FIR   finite impulse response

GA   Genetic Algorithm

GA   singular value decomposition

GPS   global positioning system

i.i.d.	independent and identically distributed
ICI	inter-chip interference
IEEE	Institute of Electrical and Electronics Engineers
IFI	inter-frame interference
ISI	inter-symbol interference
ITU	International Telecommunications Union
LASSO	least absolute shrinkage and selection operator
LMS	least mean-squares
LS	least squares
LTI	linear time invariant
MAI	multiple access interference
MAP	maximum a posteriori probability
MF	matched filter
MIMO	multiple-input and multiple-output
ML(E)	maximum likelihood (estimator)
MMSE	minimum mean square error
MPC('s)	multipath component(s)
MRC	maximal ratio combining
MUD	mulit-user detection
MUSIC	multiple signal classification

NBI	narrow-band interference
NMSE	normalized mean square error
OOK	on-off keying
p.d.f.	probability density function
PDP	power delay profile
PHY	physical
PN	pseudorandom noise
PSD	power spectral density
r.v.	random variable
RF	radio frequency
RIP	restricted isometry property
RLS	recursive least-squares
SNIR	signal-to-noise-and-interference ratio
SNR	signal-to-noise ratio
SV	Saleh-Valenzuela
TD	time domain
TDMA	time division multiple access
TH	time hopping
TOA	Time-of-Arrival
TR	transmitted reference

TX/RX	transmitter / receiver
ULA	uniform linear array
USB	universal serial bus
UWA	under-water acoustic
UWB(-IR)	Ultra-Wideband (Impulse Radio)
WLAN	wireless local area network
WPAN	Wireless Personal Area Network
WPAN	wireless personal area network
WSS	wide sense stationary

THESIS ABSTRACT

NAME: Syed Faraz Ahmed

TITLE OF STUDY: Channel Estimation for Impulse-Radio Ultra-Wideband (IR-UWB) Communications

MAJOR FIELD: Electrical Engineering

DATE OF DEGREE: June 2011

The thesis addresses the problem of channel estimation in Impluse-Radio Ultra-Wideband (IR-UWB) communication system. The IR-UWB communications utilize low duty cycle pulses to transmit data over the wireless channel. The transmitted energy is distributed over a large number of multipath components (MPCs). At the receiver, these MPCs need to be estimated accurately to capture sufficient energy for successful communications. In our work, the IEEE 802.15.4a channel model is used where the channel is assumed to be Linear Time Invariant (LTI) and thus the problem of channel estimation becomes the estimation of the sparse channel taps and their delays. Since, the bandwidth of the signal is very large and the Nyquist rate sampling ($\backsim$ 16 GHz.) is impractical therefore we estimate the channel taps from the subsampled versions of the received signal profile. The

transmitted pulse shape considered is the second derivative of the Gaussian pulse. We decompose the channel estimation problem into two parts: (i) estimation of the channel support, followed by, (ii) estimation of the support co-efficients (channel amplitudes). We exploting the signal sparsity and reduce the search space for the channel support by using three different methods: Genetic Algorithm, Correlation and Compressive Sensing. In the classical estimation approach we develop Low-Complexity Maximum Likelihood (LCML) estiamtor by leveraging the underlying structure of the problem. In the Bayesian framework, first we estimate the decomposed channel by incorporating the a priori multipath arrival time statistics for three different cases of amplitude statistics, namely (i) non-Gaussian, (ii) non-Gaussian with known second order statistics from the IEEE model, and (iii) Gaussian. Second, we jointly estimate the channel support and co-efficients by deveopling an Approximate Minimum Mean Square Error Estimator (AMMSE). We leverage the structure to reduce the computational complexity and propose a Low-Complexity MMSE (LCMMSE) channel estimator. The performance of the various methods in terms of the Normalized Root Mean Square Error (NRMSE) in estimation of MPC arrival times and energy capture were compared in the prescence of AWGN. The novel low-complexity estimators, namely LCML, AMMSE and LCMMSE, presented in the thesis outperform other conventional UWB channel estimators. Furthermore, the computational complexity is much less as compared to that of Compressive Sensing, ML and MMSE estimators.

ملخص الرسالة

الاسم الكامل: سيد فراز احمد

عنوان الرسالة: طريقة استشعار الضغط للحصول على تقدير قناة الاندفاع – الاتصال بالراديو (IR - UWB) فائقة الاتساع
التخصص: هندسة كهربائية.
تاريخ الدرجة العلمية: يونيو 2011

تعالج هذه الرسالة مشكلة تقدير القناة في الاندفاع للإذاعة لنظام الاتصالات فائقة الاتساع (IR - UWB). إن الاتصالات IR - UWB تستفيد من انخفاض نبضات دورة العمل لنقل البيانات عبر قناة لاسلكية. حيث يتم توزيع الطاقة المنقولة عبر عدد كبير من المكونات المتعددة (MPCs). عند الاستقبال نحتاج الى تقدير المكونات المتعددة (MPCs) بدقة لالتقاط ما يكفي من الطاقة للاتصالات ناجحة. تم في هذا العمل استخدام نموذج قناة (IEEE802.15.4a) وعلى افتراض أن القناة خطية ثابتة من الزمن (LTI) تصبح مشكلة تقدير القناة عبارة عن تقدير الصنابير للقناة المتفرقة والتأخير بهم. و بسبب أن عرض النطاق الترددي للإشارة كبيرة جدا وأخذ العينات معدل Nyquist (~ 16 غيغاهيرتز.) غير عملي ولذلك فإننا نقدر الصنابير القناة من الإصدارات subsampled من ملف ردت إشارة. تم اعتبار شكل نبضة المرسلة هو المشتق الثاني للنبض غاوسي (Gaussian). تم تقسيم مشكل تقدير القناة إلى قسمين: (1) التقدير لدعم القناة ويتبع ذلك (2) التقدير لمعاملات الدعم (سعة القناة). تم استغلال تبعثر الإشارة لتقليل مساحة البحث عن دعم القناة باستخدام ثلاث طرق مختلفة: الخوارزمية الوراثية, الارتباط والاستشعار المضغوط. في تقدير النهج الكلاسيكي تم تطوير مقدر قليل التعقيد الأقصى احتمال (ML - LC) من خلال الاستفادة من البنية الأساسية للمشكلة. في الإطار البايزي (Bayesian) أولا نقدر القناة متحللة من خلال دمج وكلما كانت الإحصاءات المتعددة وصول الوقت مسبقا لمدة ثلاث حالات مختلفة من الإحصاءات السعة , وبالتحديد (1) غير الغاوسي (2) غير الغاوسي مع معرفة إحصاءات الدرجة الثانية من نموذج IEEE (3) الغاوسي . ثانيا تم بشكل مشترك تقدير لدعم القناة والمعاملات والاستفادة من هيكل المشكلة للحد من التعقيد الحسابي واقتراح مقدر قناة منخفض التعقيد MMSE (LC-MMSE).

تمت مقارنة الأداء بالأساليب المختلفة من حيث الجذر تطبيع متوسط خطأ سكوير (NRMSE) في تقدير أوقات وصول المكونات المتعددة (MPC) من حيث والتقاط الطاقة في حضور AWGN. وقد تفوق المقدر منخفض التعقيد بالمسمى تحديدا ب (MMSE-LC ML and-LC) والذي تم تطويره في هذه الرسالة على المقدرات التقليدية لقناة UWB. إضافة الى ذلك فان التعقيد الحسابي اقل بكثير مقارنة لاستشعار الضغط (ML و MMSE).

CHAPTER 1

INTRODUCTION AND

MOTIVATION

1.1 Introduction

The Thesis is concerned with the application of state of the art digital signal processing techniques to a problem in wireless communication. Specifically, we develop methods to estimate the channel impulse response for an Impulse-Radio Ultra-Wideband (IR-UWB) communication system.

1.2 Motivation

UWB technology is a promising technology for very high speed short range wireless communication and as well as for precision ranging and positioning applications. UWB systems have attracted renewed attention in recent years and many research work has been directed to solve the issues in UWB communications. Channel

estimation is an important step for successful communication over the wireless channel. It also assists in mitigation of interference from other signals which is an important requirement of UWB systems. Since the wireless channel changes with time, accurate and efficient methods are required to enable the receiver to periodically estimate the channel correctly and quickly to decipher the information from the received signal. UWB channel have some specific characteristics such as large bandwidth, low-power transmission and rich multipath propagation which makes the channel estimation of UWB channels a unique challenge. There are several estimation techniques proposed for estimating UWB channels, but a lot of work is still to be done to arrive at the best estimation tehcnique which provides reliable estimates and is of low complexity. The received UWB signal is sparse and also rich in structure. Our motivation in this thesis is to develop accurate channel estimators for UWB channel which exploits the sparsity of the received UWB signal and the rich underlying structure to reduce the computational complexity of the estimators.

1.3   Thesis Objective

The objective of the thesis is to develop a Low-Complexity Channel estimators for IR-UWB communication systems by taking into consideration the following:

1. Sparsity of the received UWB signal profile

2. Rich structure of the sensing matrix

3. A priori statistical information about the UWB channel

1.4 Thesis Contributions

The main contributions of the thesis are the development of novel channel estimators for IR-UWB communication systems, as follows:

1. Development of channel support estimators by exploiting sparsity where each of the following, (i) Genetic Algorithm, (ii) Correlation and (iii) Compressive Sensing are used to obtain the coarse estimates and hence reduce the search space

2. In the classical estimation framework, development of the Low-Complexity Maximum Likelihood (LCML) channel estimator by leveraging the structure of the model

3. In the Bayesian framework for the decomposed channel, development of Low-Complexity Maximum A Posteriori (LCMAP) estimator for the case of both Gaussian and non-Gaussian channel co-efficients

4. In the Bayesian framework, development of Low-Complexity Minimum Mean Square Error (LCMMSE) estimator for jointly estimating the channel support and co-efficients for the case of both Gaussian and non-Gaussian channel co-efficients

1.5 Thesis Organization

In Chapter 2 background about the UWB channel, IR-UWB communication systems and different UWB receivers is presesnted. Some of the popular UWB channel estimation techniques in the literature are also briefly discussed. A concise review of linear estimation theory is also provided in the chapter. In Chapter 3, the IR-UWB communication model is discussed and the problem of UWB channel estimation is formulated based on our model. We exploit the sparsity of the signal in Chapter 4 where Genetic Algorithm, Correlation, Compressive Sensing and their combinations are employed to estimate the IR-UWB channel. In Chapter 5, the IR-UWB channel estimation is performed in a Classical Estimation framework and a Low-Complexity Maximum Likelihood estiamtor is developed. In Chapter 6, the IR-UWB channel estimaton is performed in a Bayesian framework where both Low-Complexity MAP and MMSE estimators are developed by considering three different statistical priors for the channel fading amplitudes. Chapter 7 provides a discussion and concludes the thesis.

CHAPTER 2

BACKGROUND

2.1 Introduction

In this chapter the background to the work in this thesis is presented. We begin by defining the UWB communication systems in Section 2.2 and the specific characteristics of the UWB channel model in Section 2.3. Thereafter, the different types of receivers for UWB communication are briefly presented in section 2.4 followed by the definition of the UWB channel estimation task in Section 2.5. We also discuss several channel estimation techniques for UWB systems available in the literature. Lastly in Section 2.6 a review of estimation theory is summarized, in particular parametric linear estimation and Bayesian estimation are revisited.

2.2 Ultra-Wideband (UWB) Systems

Ultra-Wideband (UWB) radio is a rapidly evolving technology, which is aimed primarily for indoor wireless communications and precision positioning appli-

cations. UWB technology has been around since 1960s, when it was mainly used for radar and military applications [2]. The United States Federal Communications Commission (FCC) allowed UWB waveforms to overlay over other systems which resulted in an exponential increase in interest towards UWB technology from academia, industry, and global standardization bodies. In 2002, FCC allocated limited use of a huge chunk of spectrum between 3.1 GHz and 10.6 GHz to allow UWB systems to overlay over existing narrowband systems.

The history of UWB can be traced back a century to Guglielmo Marconi's spark gap transmitters which conducted radio communications using an enormous bandwidth. However, modern UWB technology as we know it today has been around since the 1960s and began with the invention of the impulse radars which found strong application in military primarily due to its robusteness to jamming. The academic interest in UWB technology was pioneered by Prof. Scholtz and his group in the 1990's where the focus was on low-rate applications.

With recent developments in high-speed switching and narrowband pulse generation a renewed interst in UWB technology has resulted. These efforts lead to a spread of UWB from military applications to consumer electronics. The principle of UWB is based on transmitting low energy signals over a large bandwidth which results in immunity to frequency flat fading.

UWB technology finds itself applications in wide and diverse areas:

- Wireless networks

- Sensor networks

- Imaging systems

- Vehicular radar systems

UWB transmitter is defined by the FCC as a transmitter that has a fractional bandwidth equal to or greater than 0.2 or has a UWB bandwidth equal to or greater than 500 MHz. The UWB bandwidth is the frequency band bounded by the points that are 10 dB below the highest radiated emission, as based on the complete transmission system including the antenna [3]. To specifically define what is meant by an Ultra-Wideband signal, the following fractional bandwidth definition is often employed:

$$B_f = 2\frac{f_H - f_L}{f_H + f_L} \tag{2.1}$$

where f_L and f_H are the lower and upper end of the signal bandwidth, respectively. The FCC spectral mask for UWB indoor communication is shown in Fig. 2.1. For indoor systems, the average output power spectral density is limited to -41.3 dBm/MHz, which complies with the long standing Part 15 general emission limits to successfully control radio interference. A typical UWB impulse radio employs short pulses with ultra low power for communication and ranging. UWB impulse radio system does have several advantages over other conventional systems:
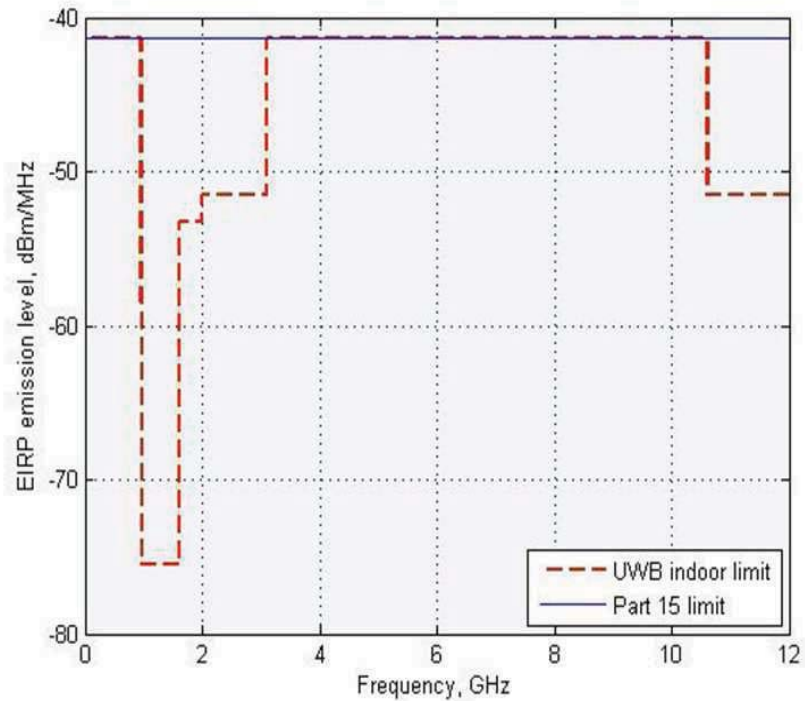
Figure 2.1: FCC Spectral Mask

- High data rate wireless transmission - Due to the ultra-wide bandwidth of several GHz, UWB systems can support more than 500 Mb/s data transmission rate within the range of 10 m, which enables various new services and applications.

- High precision ranging - Due to the sub-nanosecond duration of typical UWB pulses, UWB systems have good time-domain resolution and can provide sub-centimeter accuracy for location and tracking applications.

- Low loss penetration - UWB systems can penetrate obstacles and thus operate under both line-of-sight (LOS) and non-line-of-sight (NLOS) environments.

8

- Fading robustness - UWB systems are immune to multipath fading and capable of resolving multipath components even in dense multipath environments. The resolvable paths can be combined to enhance system performance.

- Security - For UWB signal, the power spectral density is very low. Since UWB systems operate below the noise floor, it is extremely difficult for unintended users to detect UWB signals.

- Coexistence - The unique character of low power spectral density allows UWB system to coexist with other services such as cellular systems, wireless local area networks (WLAN), global positioning systems (GPS), etc.

- Low cost transceiver implementation - Because of low power of UWB signals, the RF and baseband can be integrated into a single chip. The up-converter, down-converter, and power amplifier commonly used in a narrowband system are not necessary for UWB systems.

Industrial standards such as IEEE 802.15.3a Task Group (TG3a) [1] and IEEE 802.15.4a Task Group (TG4a) [4] have been introduced within 802.15 work group to develop standards based on UWB technology. The TG3a group was formed in January 2003 with the objective of providing a higher speed physical layer (PHY) enhancement amendment to IEEE 802.15.3. The group aimed to develop PHY standards to support data rates between $110 - 450$ Mb/s over short ranges (i.e., < 10 m). Among many proposed UWB systems for IEEE 802.15.3a were

two major proposals: the Multi-Band OFDM Alliance (MBOA) proposal and the direct-sequence UWB (DS-UWB) proposal. The MBOA system employs orthogonal frequency-division multiplexing (OFDM) modulation to solve the severe multipath problem. The DS-UWB system uses direct-sequence spread-spectrum technology and relies on the RAKE receiver to capture signal energy dispersed over a large number of paths. After 3 years, TG3a group decided to dissolve the group in 2006 [5]. The TG4a group was formed in March 2004 with the objective of providing an amendment to IEEE 802.15.4 for an alternative PHY. The aim was to provide communications and high precision ranging/location capability, high aggregate throughput, and ultra low power. The baseline consisted of two optional PHYs consisting of a UWB Impulse Radio (operating in unlicensed UWB spectrum) and a Chirp Spread Spectrum (operating in unlicensed 2.4 GHz spectrum). In March 2007, P802.15.4a was approved as a new amendment to IEEE Std 802.15.4-2006 by the IEEE-SA Standards Board [1].

However, there are some technical challenges that remain to be solved in order to develop a UWB system, such as optimum UWB reception, transceiver structure, UWB pulse generation, antenna, low noise amplifiers, ultra-high speed (GHz) analog to digital converter (ADC), coding and modulation, timing acquisition and synchronization, and optimal channel estimation for coherent reception. Generally speaking, the difficulty of UWB system design and development is to handle the ultra-wide bandwidth and to manage the trade-off between low

complexity and high performance.

The initial idea of UWB communication is based on impulse radio communication systems which employ very sharp pulse trains to carry information bits without mixers, oscillators and bandpass filters. There are two main differences between UWB systems and other narrowband or general wideband systems. First, the bandwidth of UWB systems is much greater than the bandwidth used by any current technology for communications. Second, UWB systems are typically implemented as carrierless whereas conventional systems use radio frequency (RF) carriers to move the signal from baseband frequency to the actual carrier frequency region. Conversely, IR-UWB implementations can directly modulate an impulse that has a very sharp rise and fall time, thus resulting in a waveform that occupies a very wide bandwidth.

One of the most attractive property of a UWB system is the ultra high speed communication. From the work of Claude Shannon in the late 1940s, we know that a communication system if subjected only to additive white Gaussian noise (AWGN), then it offers the maximum rate for reliable transfer of information as follows,

$$C = B \log_2 \left(1 + \frac{P}{N_0} \right) \tag{2.2}$$

where, C is the channel capacity (bits/s), B is the transmission bandwidth (Hz), P is the received signal power (W) and N_0 is the single-sided noise power spectral density (W/Hz). Since for UWB systems B is huge therefore, the capacity C of the UWB channel is very high which results potentially in very high data rates. The main limiting factor of UWB wireless systems is power spectral density rather than bandwidth.

UWB signals have certain advantages for communications specifically improved penetration through materials as well as improved performance in dense multipath environments where the UWB signals can be resolved in time making the use of a RAKE receiver possible. Both of these advantages make UWB communication systems well suited for urban and indoor wireless applications where many local objects act as scatterers and absorbers of the transmitted electromagnetic energy. Also, these specific advantages allow for reduced transmitted signal power, which in turn result in low probability of detection or interception.

UWB signals carry data using a low signal level below the thermal noise floor through a dense multipath channel. There has been considerable research on designing suitable (optimal) signal waveforms to satisfy the requirements of the FCC spectral mask. In view of the system design, UWB pulse shape can be chosen for the purpose of simplifying a design. A pulse shape is an important factor affecting the overall system performance. An applicable pulse shape

should be easy to implement and be convenient for theoretical analysis. There are many conceivable signals which will have the required fractional bandwidth to be termed UWB signals. Generally there are three main waveforms used in UWB systems: the Gaussian-like pulse, the monocycle pulse, and the polycycle pulse [6]. Specifically in IR-UWB communication, the pulse shape of choice is a baseband pulse that is shaped as a derivative of the Guassian pulse. Generally the 2nd or the 5th derivative is used due to the radiation properties [7]. The desired order of the pulse comes from the application of a lower order Gaussian derivative pulse to the transmit antenna. The electromagnetic wave radiated by an antenna, for wideband signals, is approximated to be proportional to the time derivative of the antenna's driving current [8] while similarly an additional derivative results from the receive antenna. In narrowband systems employing carriers, this derivative is well approximated as a time-shift [9]. In this thesis we have assumed that the ideal Gaussian pulse is transmitted and therefore, at the receiver we assume the known pulse shape to be given by the second derivative of the Gaussian pulse shape. The second derivative of the Gaussian pulse shape is defined as:

$$p(t) = \sqrt{\frac{4}{3\sigma\sqrt{\pi}}} \left(1 - \left(\frac{t}{\sigma} \right)^2 \right) \exp \left(-\frac{1}{2} \left(\frac{t}{\sigma} \right)^2 \right) \qquad (2.3)$$

The factor $\sqrt{\frac{4}{3\sigma\sqrt{\pi}}}$ ensures that the signal is normalized to unit energy, i.e.,

$$\int_{-\inf}^{+\inf} p^2(t)dt = 1 \qquad (2.4)$$

13

This allows the energy in the received waveform to be stated explicitly, that is, the received energy in $\sqrt{E_p}p(t)$ is simply E_p. The scale factor, σ, determines the effective width of the pulse and is chosen such that it results in the pulse width of approximately 1 nanosecond. The pulse shape in Eq. (2.3) can be considered as the transmitted pulse shape by lumping both derivatives at the transmitter end of the system. This propagation model is very simplistic, but will suffice for the purpose of the work presented here. For detailed propagation studies of UWB signals see [10] or [11] and the references therein.

UWB signals can be modulated in different ways such as pulse position modulation (PPM), phase-shift keying (PSK), pulse amplitude modulation (PAM), and on-off keying (OOK) for binary schemes; M-ary PPM and M-ary PAM for M-ary schemes [12]. Since the UWB transmission is mainly power limited instead of spectrum limited, binary modulation is usually adopted. Initially PPM was exclusively used for UWB communication [7] but as pulse negation became easier to implement, PAM attracted more attention [13].

In a typical UWB system, each information-conveying symbol is represented by a number of (N_f) pulses, each transmitted per frame of duration $T_f \gg T_p$ (here multiple frames comprise a symbol). As the pulse duty cycle is very small, the transmitter is gated off for the bulk of the symbol period. To allow for multi-user access to the UWB channel, mainly two methods have been applied:

- Time-Hopping (TH)

  Time-hopping can be implemented by employing appropriately chosen hop-ping sequences for different users to minimize the probability of collisions due to multiple access. In TH UWB, each frame is subdivided into a number of chips of duration $T_c > T_p$. Each user is assigned a unique pseudo-random time shift pattern called a TH sequence, which provides an additional time shift to each pulse in the pulse train. Therefore each pulse undergoes an additional time shift within the addressable time delay bin.

- Direct-Sequence (DS)

  Direct-sequence codes can also be used with both PAM and PPM modu-lation for multiple access. Since IR-UWB systems are inherently spread spectrum systems, the use of spreading codes in DS-UWB systems is solely for accommodating multiple users where the pseudo-noise (PN) sequence is used to identify the user.

2.3   UWB Channel

This section describes channel models for UWB communications. An accurate model is needed for designing an efficient communication system which includes achieving maximum data rate, adopting suitable modulation scheme, and design-ing efficient algorithm for signal processing. UWB channels are very different from narrowband wireless channels, especially with regard to fading statistics and the prescence of multipath clusters. To accurately appreciate and evaluate UWB sys-

tem designs, it is important to first understand the propagation characteristics of the UWB waveforms and accurately model the channel statistics. Given the wide-band nature of UWB transmissions, the conventional channel models developed for narrowband transmissions are not adequate anymore. In general, the received signal is made up of several components: first, the direct component is commensurate with the portion of the wave travel along a line-of-sight (LOS) between the transmit and receive antennae and; second, the components arrive after having been reflected or diffracted on scattering objects that are part of the propagation environment. The latter is known as multipath propagation. As a result of the rich multipath propagation, the received UWB signal is made up of multiple replicas of the transmitted signal, all of which exhibit different attenuations and delays. Now we examine the channel model recommended by the IEEE 802.15.3a and 4a working groups. In narrowband communication Rayleigh fading channel is widely used, but the UWB channel model is presented by a log-normal fading model in [5] where a modified Saleh-Valenzuela (S-V) model is used for power and delay profile. Four types of UWB channels were defined by the IEEE 802.15.3a group to meet measurement results, namely CM1, CM2, CM3, and CM4, for different channel characteristics.

- CM1: LOS scenario with a separation between transmitter and receiver of less than 4m.

- CM2: the same range as CM1 but non-Line-of-Sight (NLOS).

- CM3: NLOS scenario for distance between $4 - 10$m.

- CM4: a situation with strong delay dispersion, resulting in a delay spread of at least 25ns.

In this thesis we use the more parametrized channel model recommended by the IEEE 802.15.4a WG in [1], which is extracted from a large amount of measurements in different communication environments such as residential, office, industry, and outdoor, covering the frequency range from 2GHz to 10GHz. We will focus on indoor channels since more than 80% of the envisioned commercial UWB applications are for indoor communications. It was noticed that the MPCs arrive in clusters and the mean-square value of the amplitude decays with increasing ray and cluster arrival time. The channel impulse response is modeled as follows

$$h(t) = \sum_{c=0}^{C-1} \sum_{k=0}^{K-1} \alpha_{k,c}\, \delta(t - T_c - \tau_{k,c}) \tag{2.5}$$

where C is the total number of clusters, K is the total number of paths occuring in each cluster, T_c is the arrival time of the cth cluster and $\tau_{k,c}$ is the arrival time (relative to T_c) of the kth path of the cth cluster. The corresponding multipath fading coefficients of the received profile are denoted by $\alpha_{k,c}$. The cluster inter-arrival time, $\Delta T_c = T_c - T_{c-1}$, and the ray inter-arrival time, $\Delta\tau_{k,c} = \tau_{k,c} - \tau_{k,c-1}$, are each exponentially distributed with probability density functions

$$f(\Delta T_c) = \Lambda \exp^{-\Lambda.\Delta T_c} \tag{2.6}$$

$$f(\Delta\tau_{k,c}) = \lambda \exp^{-\lambda.\Delta\tau_{k,c}} \tag{2.7}$$

where Λ is the arrival rate of the Poisson process for the clusters and λ is the arrival rate of the Poisson process for the paths within a cluster and their typical values are given by the IEEE model [1] for various environments. This results in a double Poisson arrival process for the MPCs.

The average power delay profile (APDP) of the received profile from [1] can be expressed as

$$\mathbb{E}(\alpha_{k,c}^2) \propto \exp\left(-\frac{T_c}{\Gamma}\right) \exp\left(-\frac{\tau_k}{\gamma}\right) \quad (2.8)$$

where $\mathbb{E}(.)$ denotes statistical expectation, Γ is the decay factor for the first MPCs of the clusters, and γ is the decay factor for the remaining MPCs of a cluster. Typical values of Γ and γ for various environments such as line-of-sight (LOS), non-line-of-sight (NLOS), indoor, office and resedential environments are given in [1] and shown in Table **??**.

The arrival time of the lth MPC, denoted by τ_l, can be expressed as, $\tau_l = T_c + \tau_{k,c}$, where the lth MPC is the kth path of the cth cluster in the received signal profile . Therefore, the channel impulse response for UWB channel of Eq. (2.5) can now be expressed as

$$h(t) = \sum_{l=0}^{L-1} \alpha_l \delta(t - \tau_l) \quad (2.9)$$

18

where L is the total number of resolvable MPCs, $\delta(t)$ is the Dirac delta function, α_l is the fading gain of the lth MPC and τ_l is its delay relative to the arrival time of the first MPC of the received profile, i.e. $\tau_0 = 0$. As opposed to common baseband models of narrow-band systems α_l is real-valued. Upon synchronization, the receiver adjusts its timing according to the first MPC arrival time τ_0.

It is very difficult to obtain the statistics of α_l from the statistics of $\alpha_{k,c}$ for the double Poisson process. This is because the fading co-efficients α_l's not only depends on the arrival time of the lth MPC but also depends on the arrival of the first MPC of the same cluster. Since, the number of clusters as well as the arival time of clusters is random, it is not possible to ascertain the cluster to which the lth MPC belongs. We overcome this difficulty by approximating the MPCs arrival as a single Poisson process. We assume that $\gamma = \Gamma$ (i.e., the clusters and the paths within each cluster decay at the same rate) in Eq. (2.8) for all the clusters. This implies that the channel impulse response of Eq. (2.9) consists of the arrival of MPCs belonging to a single *large* cluster with a decay rate Γ and MPCs arrival rate λ. Thus the APDP for the channel in Eq. (2.9) is expressed as

$$\mathbb{E}(\alpha_l^2) \propto \exp\left(-\frac{T_c + \tau_{k,c}}{\Gamma}\right) \tag{2.10}$$

$$\propto \exp\left(-\frac{\tau_l}{\Gamma}\right) \tag{2.11}$$

for $l = 0, \dots, (L-1)$. In the matrix form, we can write the APDP as,

$$\mathbf{D} = \mathbb{E}[\mathbf{a}\mathbf{a}^H] \qquad (2.12)$$

where $\mathbf{a}$ is a vector of the fading coefficients of the received signal profile at the sampled instants and $\mathbf{D}$ is a diagonal matrix since the fading of the different MPCs is independent.

Due to the large bandwidth, only a few MPCs fall within a small time bin and therefore, the central limit theorem is not applicable to the UWB channel leading to non-Gaussian fading statistics. Thus, the small-scale fading co-efficients, α_l's, are modelled as independent Nakagami distributed in [1] as,

$$p(\alpha_l) = \frac{2}{\Gamma(m)} \left(\frac{m}{\zeta}\right)^m \alpha_l^{2m-1} \exp\left(-\frac{m}{\zeta}\alpha_l^2\right) \qquad (2.13)$$

where $m \geq 1/2$ is the Nakagami m-factor, $\Gamma(m)$ is the Gamma function and $\zeta = \mathbb{E}(\alpha_l^2)$ corresponds to the mean power and its delay dependence is given by Eq. (2.11).

The model in Eq. (2.9) is known as a specular multipath model where the effect of the channel is assumed to simply sum up many scaled and time-shifted versions of the original transmitted pulse, i.e., there is no pulse waveform distortion. The appropriate multipath model when considering waveform distortion

20

is the diffuse model [14]. The diffuse model can be thought of in a couple different ways. First, the same form as Eq. (2.9) with the summation taken over an uncountable set can be considered. Secondly, the summation can remain countable with the output pulse waveform becoming a function of the index l. Of course, a combination of these two models might also be conceivable. The diffuse model is more accurate but comes with increased complexity. For the purposes of this work, the specular model will suffice as many observed channel response waveforms can be adequately modeled as such.

2.4   UWB Receiver

Although UWB technology can enable many attractive features deploying IR-UWB systems is quite challenging. One of the key challenges for IR-UWB communications system is the construction of low-cost receivers that work well in multipath environments. As a result of high bandwidth of UWB signals, very fine multipath delays are resolvable in such environments. Due to the energy dispersion, a robust receiver that is capable of collecting the rich multipath must be designed to mitigate the performance degradation. The most common UWB receiver designs include Energy Detectors (ED), Transmit-Reference (TR) receivers, and RAKE (correlation) receivers. The ED are simple to implement, with tradeoff on performance, and suitable for UWB radar systems. RAKE correlation receivers coherently detect the received signal and can achieve the optimal performance in theory. Ideally, the RAKE receiver can be used to collect

the multipath components (MPC's). However, a RAKE receiver relies on the maximum ratio combining (MRC) from the accurate channel state information (CSI) to produce reliable decision statistics. Therefore, it is important to devise methods for perfect synchronization and channel estimation. Suboptimal receiver schemes, such as TR and ED which do not need any sophisticated channel estimation and precise synchronization have also been employed [15]. These sub-optimal schemes suffer from performance penalty.

In order to capture a considerable portion of the signal energy scattered in multipath components, a conventional RAKE-based digital receiver not only has to sample and operate at a minimum of hundreds of MHz to even multi-GHz clock rates, but also requires an impractically large number of RAKE fingers. Realizing optimal RAKE reception performance requires accurate channel and timing knowledge, which is quite challenging to obtain as the number of resolvable paths grows. Moreover, the received pulse shapes of resolvable multipath might be distorted differently due to diffraction, which make it suboptimal to use line-of-sight signal waveform as the correlation template in RAKE reception.

For these reasons, Transmit-Reference, also known as autocorrelation, receivers have drawn significant attention in recent years. TR encodes the data in the phase difference of the two pulses of a pulse pair. The first pulse in that pair does not carry information, but serves as a reference pulse; the second

pulse is modulated by the data and is referred to as the data pulse. The two pulses are separated by a fixed delay. It can be easily shown that the receiver can demodulate this signal by simply multiplying the received signal with a delayed version of itself. In a slow fading environment, TR collects multipath energy efficiently without requiring multipath tracking or channel estimation. Nevertheless, TR autocorrelators entail several drawbacks: the use of reference pulses increases transmission overhead and reduces data rate, which results in reduced transmission power efficiency; the bit-error-rate (BER) performance is limited by the noise term in the reference signal [16]. Finally, the performance of TR receivers relies on the implementation of accurate analog delay lines which can save and delay the reference waveforms for up to tens of nanoseconds. This is still a big challenge to current circuit technology [17].

Optimal energy capture is obtained by a coherent RAKE receiver that has enough fingers to collect all resolvable multipath components (MPCs). A RAKE receiver can be used to exploit the diversity by constructively combining the separable received MPCs. In order to benefit from the optimality of RAKE reception and to make its implementation practical, Selective-RAKE (S-RAKE) scheme has been adopted [18]. In an S-RAKE receiver only the strongest MPC's are estimated and used to detect the data. Such a receiver consists of L correlators/fingers to collect the received signal energy from the L strongest paths. The lth correlator, for $l = 0, 1, 2, ..., L - 1$, is to correlate the received

signal with the receiver locally generated reference signal delayed by τ_l. The output of the correlators can be linearly combined in different ways to form the decision variable. The maximal ratio combining (MRC) approach provides better performance, with the prerequisite of accurate channel information at the receiver [19]. When accurate channel information is not available, equal gain combining (EGC) could be used [20]. Since, the UWB channel is sparse (i.e., several time bins have no MPCs) and randomly distributed, the RAKE receiver searches for the finger locations and positions them at the correlation lags where the impulse response has power. If the receiver fingers are uniformly spaced then some of these would carry only noise and would lead to unnecessary noise enhancement and reduction in detection performance at the receiver. In our work we have incorporated the UWB channel sparsity information.

2.5 UWB Channel Estimation

This multipath diversity of a UWB channel calls for the use of RAKE or S-RAKE receivers for significant energy capture, higher performance and flexibility, despite its complexity over sub-optimal counterparts. The RAKE receiver is a coherent receiver and relies on the accurate channel estimates. Recently, there has been a renewed interest in the use of RAKE in UWB communications, e.g., MIMO systems [14], BAN [21], cooperative BAN [22], and the prerake systems [23]. In essence, the design or enhancement of accurate channel estimator must not be overlooked and is imperative for efficient IR-UWB communications and can

greatly improve the performance of the UWB based positioning and ranging systems [24].

Several estimators have been developed for UWB channel. In [10] Win and Scholtz proposed a maximum likelihood (ML) channel estimator for an isolated UWB pulse, and in [25] Lottici et al. presented both data-aided (DA) and non-data aided (NDA) based ML estimation. Unfortunately, these methods require operating at the formidable Nyquist sampling frequency. Since then, many other variants of the estimator and new receiver types have been proposed to reduce the complexity. Although the ML scheme is shown to be a superior estimator, the tremendous BW of UWB signal renders its implementation difficult because of the Nyquist criterion. Since UWB applications are mostly for high-rate communications, in our work we asssume the channel to be constant only over a single symbol. This implies that fading is assumed to be quasi-static, allowing all channel coefficients α_l's and relative delays τ_l's to be constant over a single symbol period and change independently from one symbol to another.

Sampling rate plays a crucial role in signal processing and communications. With time more and more analog techniques are being replaced by their digital counterparts. It is well known from Nyquist-Shannon sampling theorem that unambiguous reconstruction is possible if the signal is bandlimited and the sampling frequency is greater than twice the signal bandwidth. The error which corre-

sponds to the failure of band limitation is referred to as aliasing. The condition for alias-free sampling at rate F_s called Nyquist sampling frequency is $F_s \gg 2B$, where B is the bandwidth of the signal. UWB signal processing requires much higher sampling rate than general narrowband signal if the Nyquist sampling frequency is observed due to the much wider bandwidth. High Nyquist sampling frequency requires more expensive analog-to-digital converter (ADC) and more power to support high speed signal processing and is thus a design challenge.

Most existing literature on ML complexity reduction tackle the issue by redefining the problem, or eliminating the use of ML altogether. Refs. [26] and [27] approached the complexity issue by way of formulating it as a synchronization or timing recovery problem, respectively. A frame synchronization approach to complexity reduction was addressed in [28], where a search over possible frame delays was performed to maximize the log-likelihood function. In contrast, [10] eliminated the ML formulation and concentrated on timing recovery with LS signal model. In [29], the ML estimator was simplified by recognizing that MPCs arrive in clusters, and executing search only for rays falling into the highest energy clusters. Although they are of low-complexity, their performance implicitly depends on acquiring high-rate samples. Ref. [30] proposed a finite rate of innovation approach which projects a signal into lower dimensional subspace. Unfortunately, due to the closely spaced path arrivals inherent in UWB systems, the solution to rate of innovation is often ill-conditioned. Despite these attempts,

the most critical issue - Nyquist sampling rate reduction - of the ML channel estimator has yet to be addressed.

The limitation due to the high sampling requirements led to leave aside classical conception of sampling (Nyquist) and seek for new techniques that allow more information rate using less sample requirements. The emerging theory of compressed sensing (CS) outlines a novel strategy to jointly compress and detect a sparse signal with fewer sampling resources than the traditional method, opening a new range of possibilities in UWB communication. For a signal $\mathbf{r} \in \mathbb{R}^N$ which is K-sparse, with $K \ll N$ being an integer, compressed sensing shows that with high probability $\mathbf{r}$ can be reconstructed from M compressive measurements when $M \geq CKlog(N/K) \ll N$, where $C > 1$ is the oversampling factor [31]. Compressed sensing for UWB was first proposed in [24] as a generalized likelihood ratio test receiver taking advantage of the signal structure by incorporating pilot assisted modulation. It was later discussed in [32] as an alternative for UWB channel estimation. In both cases the signal was reconstructed using the matching pursuit (MP) algorithm. Unfortunately, how well MP estimates ties directly to the design parameters, such as the number of iterations and residual error for convergence [33], which are subject to change depending on the environment.

In our case the algorithms used for recovering the signal and estimating

27

the channel will address the trade-off between the high sampling rate, the high computational time for reconstruction and the structural and statistical a priori knowledge from the model.

2.6 Linear Estimation

2.6.1 Parametric Models

A parametric model is a mathematical function that depends on the values of some parameters. The aim in parametric modeling is often to adjust the parameters of an appropriate model function such that the model optimizes some criterion, such as fitting the measured signal with a minimum possible error. This task is called *parameter estimation.* The opposite of a parametric model is a non-parametric model and there is no common form for non-parametric models. The performance of the signal processing methods for estimation depends on the chosen model structure and on the quality of the parameter estimates. Let $y(n)$ describe a signal sampled at the time instants $n = 0, 1, ..., N-1$. Then $y(n)$ can be modeled as follows:

$$y(n) = f(\boldsymbol{\alpha}, n) + e(n) \tag{2.14}$$

where $f(.)$ will be termed the parametric model function, $\boldsymbol{\alpha}$ is the parameter vector and $e(n)$ describes the difference between the measured signal and the model function. The term $e(n)$ is often termed as noise, error or residual. In practical applications of signal processing there will always be a non-zero noise

term included in the signal and is commonly treated as an additive noise. It is important to realize that $f(\boldsymbol{\alpha}, n)$ is only a model of reality. There exist many different classes of parametric signal models. Some different types of models are listed and described below.

- Physical and Black-Box Models

  A physical model is derived from knowledge of the physical reality which generate the data. This means that the parameters are interpretable quantities, and their accurate estimation is often the task of interest. A black-box signal model, on the other hand, is not derived from any physical properties of the signal. It is simply a mathematical description that is appropriate for the signal under study, and the model parameters do not necessarily have a physical meaning.

- Deterministic and Stochastic Models

  A deterministic model is a model for which, once the exact parameter values are known, the signal can be reproduced exactly. A stochastic model is inherently random, so exact signal reproduction is not possible.

- Linear and Non-Linear Models

  A signal model is linear if the signal depends linearly on the model parameters and possible inputs and disturbances. Otherwise, the model is non-linear.

In some applications an appropriate model function $f(\boldsymbol{\alpha}, n)$ is known beforehand. However, often only an appropriate function type is known, and not the exact

expression. In those instances, there will be several candidate model functions to choose from. The goal in model selection is to decide which one of the candidates is most appropriate for describing $y(n)$. A very complex model (i.e., a model with many parameters) for describing $y(n)$ is highly flexible and it can likely describe most of the characteristics of the signal. Indeed, models with more parameters will always fit the measured data better than the models with few parameters (due to their increased degrees of freedom). However, there is always a problem with this basic approach: The more complex the model is, the more data we need to estimate its parameters accurately. If we have only, e.g. 5 data samples we might be able to estimate the coefficients of a 0th or 1st degree polynomial with a reasonable accuracy, but hardly the coefficients of a polynomial of degree 4 or 5, unless the noise level is very low. Therefore, for a fixed number of data samples, we must not choose a model that is too complex. Otherwise, the parameter estimates will often be severely affected by the random measurement noise (so called overfitting).

The Occam's razor principle is often discussed in relation to the problem of model selection [34]. This principle states essentially that all things being equal, the simplest solution should be preferred. For model selection purposes, this should be interpreted as follows: When several models are equal in other respects, the model which imposes the least restrictive assumptions and introduces the fewest parameters should be selected. This is a reasoning which is intuitively pleasing. It is also interesting to note that this is essentially how sound model

selection algorithms work: they often include a term which penalizes complex models [35].

A nested set of models is one in which each model in the set can be described as a special case of the models in the set with higher complexity. Finite impulse response (FIR) filters of different lengths are examples of nested models where a single integer valued parameter (i.e., the length of the FIR filter) is sufficient to describe the model complexity. This integer valued parameter is the order of the model, and its estimation is often called *model order selection*. In the general model set, a model can consist of any combination of the considered parameters. This means that if we consider n parameters we get 2^n possible model structures. Prior knowledge about the allowed model structures can, however, reduce this number considerably. Sparse models play an important role in some applications, such as the UWB channel estimation in a multipath propagation environment. The general problem of interest is then detection and estimation of time-delayed reflections of the transmitted signal.

2.6.2 Bayesian Inference

Bayesian inference is a scientific method for finding probabilities (or probability densities) of propositions by combining measured data and information given by the user [36]. Bayesian arguments will be frequently used in parts of this thesis, so it seems worthwhile to provide a short background. The mechanisms

of Bayesian inference operate on distributions and follow a few simple rules from probability theory. The big advantage of the Bayesian framework is that, once the necessary distributions are available, inference is a purely mechanical process which leads to optimal solutions (conditional on the information supplied by the user). The disadvantage is that these solutions are often computationally prohibitive and in this thesis we have specially addressed this issue and reduced the computational complexity of our Bayesian estimators.

In this thesis for Bayesian channel estimation of UWB channels, we make use of the a priori statistical information about the channel from the IEEE 802.15.4a standard [4]. The probability distributions used for the parameters in the standard have been adopted after fitting a large number of physical measurements. To perform Bayesian inference, essentially only two tools are required: namely the product rule and marginalization. The first of these two tools has the following expression:

$$p(A, B|C) = p(A|C)p(B|A, C) \qquad (2.15)$$

The above equation should be read "given the information C, the probability for the propositions A and B to both be true equals the product between the probability that A is true and the probability that B is true given that A is true." Since the propositions A and B are exchangeable, the product rule can also be

written as:

$$p(A, B|C) = p(B|C)p(A|B, C) \tag{2.16}$$

and by combining the above two equations and rearranging the terms, one obtains the famous Bayes' theorem:

$$p(A|B, C) = \frac{p(A|C)p(B|A, C)}{p(B|C)} \tag{2.17}$$

Bayes' theorem is often used to make a hypothesis and data "exchange positions" in $p(.)$ (since it shows how $p(A|B, C)$ relates to $p(B|A, C)$). Say that A stands for a hypothesis and B for some data. Then Eq. (2.17) tells us the relation between the probability of the hypothesis A given the data B and the probability that the data B is observed assuming the hypothesis A is true. The expression $p(A|C)$ is the prior of the hypothesis A and $p(B|C)$ is the prior of the data B. The second tool, i.e., marginalization, describes a way to get rid of nuisance parameters which are of no specific interest for the inference. For a proposition B that takes on discrete values:

$$p(A|C) = \sum_{\mathcal{B}} p(A, B|C) \tag{2.18}$$

In the above expression, $\mathcal{B}$ is the set containing every possible value of B, so the above should be read "the probability for A to be true equals the sum, over all possible values of B, of the probabilities that both A and B are true." If B is

continuous, then the corresponding expression becomes,

$$p(A|C) = \int_{\mathcal{B}} p(A, B|C) dB \qquad (2.19)$$

The Bayesian essentially get its desired information by repeated use of the above mentioned tools.

2.6.3 Linear Regression

Consider the linear regression model,

$$\mathbf{y} = \mathbf{Hr} + \mathbf{w} \qquad (2.20)$$

where $\mathbf{y} \in \mathbb{R}^M$ is a vector of observed data, $\mathbf{H} = [\mathbf{h}_1 \mathbf{h}_2 \dots \mathbf{h}_N] \in \mathbb{R}^{M \times N}$ is a known matrix of N regressors $\{\mathbf{h}_i\}_{i=1}^N$, $\mathbf{r} = [r_1 r_2 \dots r_N]^T \in \mathbb{R}^N$ is the unknown vector of linear regression coefficients ($\mathbf{r}$ is called the parameter vector) and $\mathbf{w} \sim \mathcal{N}(\mathbf{0}, N_0\mathbf{I})$ is a length M vector of zero-mean Gaussian white noise with co-variance matrix $N_0\mathbf{I}$. We call Eq. (2.20) the full model and assume that the data are generated by a model of the form:

$$\mathcal{M}_k : \mathbf{y} = \mathbf{H_k r_k} + \mathbf{w} \qquad (2.21)$$

where the model order is k and that we know k_{min} and k_{max} such that $k_{min} < k < k_{max}$. We consider the following interrelated problems:

1. The model order selection problem: to find the correct order k, given $\mathbf{H}$ and $\mathbf{y}$.

2. The parameter estimation problem: to estimate $\mathbf{r}$ as accurately as possible when the order k is known.

3. The joint model order selection and parameter estimation problem: to estimate $\mathbf{r}$ as accurately as possible when the order k is unknown.

The Least Squares (LS) estimation is commonly used owing to its simplicity and its connection to ML when the noise is Gaussian. However, if something is known about the $\mathbf{r}$ a priori (before the data are collected), then one can do better than the LS estimate. For example, if one knows that $\mathbf{r} \sim \mathcal{N}(\mathbf{0}, \sigma_r^2 \mathbf{I})$ a priori, then the estimate $\hat{\mathbf{r}}$ which has the smallest mean-square error (MSE), i.e., $\mathbb{E}[\|\hat{\mathbf{r}} - \mathbf{r}\|_2^2]$ is given by the conditional mean of $\mathbf{r}$ given that $\mathbf{y}$ was observed. Note that when $\sigma_r^2 \to \infty$, corresponding to the observer having no a priori knowledge of $\mathbf{r}$, then the MMSE and LS estimates coincide. Generally, the minimum MSE (MMSE) estimate is better (in the MSE sense) than the LS estimate, owing to the influence of the a priori knowledge of $\mathbf{r}$.

2.6.4 Sparse Linear Models

Sparse linear models are relevant in a variety of applications. For example, in a statistical data analysis one may know before the measurement that the data are likely to be explained by only a few factors. The estimation of the communication channel impulse responses (CIR) of the UWB channel is also a sparse signal estimation. This is so because the bandwidth of UWB signal is so large that individual multipath components can be resolved, and in general are separated

by more than one sample period. Linear regression for sparse models has been studied both in the statistics community and in the signal processing literature [35]. We next review some of the representative contributions and approaches.

A large class of methods [22] is based on Bayesian maximum a posteriori (MAP) estimation of **r** assuming a prior density for **r** which induces sparsity. The LASSO (Least absolute shrinkage and selection operator) method [37] estimates **r** by minimizing the LS criterion subject to a l_1-norm constraint on the parameter vector. More precisely, LASSO finds **r** as a solution to a linearly constrained quadratic problem, which can be effciently solved. Interestingly, LASSO has a Bayesian interpretation in that the estimate of **r** turns out to be the same as the MAP estimate obtained if **r** has a Laplacian prior density of the form,

$$p(\mathbf{r}|\lambda) = \frac{\lambda}{2} \exp\left(-\frac{\lambda}{2}\|\mathbf{r}\|_1\right) \tag{2.22}$$

Using a small enough value for λ, a user parameter in LASSO typically leads to a parameter vector estimate for which many coeffcients actually are equal to zero (i.e., not only "small"). The Sparse Bayesian Learning (SBL) method of [38] is based on the assumption that **r** is composed of independent zero-mean Gaussian entries with unknown variances. These variances and the noise variance are treated as hyperparameters and can be eliminated from the likelihood by maximizing it using the expectation-maximization (EM) algorithm.

All methods (including the standard MMSE estimate) which use an explicit prior for $\mathbf{r}$ allow the regression problem to be underdetermined, that is $M < N$ and one can think of the use of priors on $\mathbf{r}$ as a way of regularizing the problem.

The methods discussed above are Bayesian (or at least they have a Bayesian interpretation) and as such they are arguably optimal (in the sense of MAP) if the model and the a priori knowledge assumed in the algorithm match perfectly with the process that generates the data. There are many existing methods which primarily concern with determining the structure of $\mathbf{r}$ (i.e., finding out what elements are zero). This problem is the model selection problem as mentioned previously. In the sparse signal reconstruction methods the objective is then to find a (small) set of "basis vectors" (i.e., columns of $\mathbf{H}$) such that the observed vector $\mathbf{y}$ can be expressed as a linear combination of these vectors. Thus the objective is to find the *sparsest* representation of $\mathbf{r}$.

In this thesis our primary objective is to estimate $\mathbf{r}$ (the channel parameter vector) from $\mathbf{y}$ (the received signal) as accurately as possible, although we will obtain a solution to the model selection problem as well. Note that accurate estimation of $\mathbf{r}$ is of interest in the channel estimation problem because the performance of the communication system is directly related to the quality of the channel estimate.

CHAPTER 3

IR-UWB CHANNEL

ESTIMATION

3.1   Introduction

The estimation of the parameters that characterizes the channel is of paramount importance to increase the performance of UWB coherent receivers such as the RAKE receiver. A conceptual UWB communication model is depicted in Fig. **??**. The received UWB waveform $\mathbf{r}$ not only depends on the transmitted symbols $\mathbf{x}$ but also on a set of parameters related to the UWB channel $\boldsymbol{\alpha}$. They are unknown to the receiver and in order to be able to retrieve the symbol sequence $\mathbf{x}$, it must estimate these parameters which can be considered unwanted since they somehow corrupt the signal that transport the sequence $\mathbf{x}$. Once estimated, the parameters are then used as if they were true values. The estimation of the unknown channel parameters, $\boldsymbol{\alpha}$, at the receiver is termed as *channel estimation.*

In Section 3.2 we present the linear model for channel estimation of IR-UWB communication system and highlight its statistical and structural information. In Section 3.3 we discuss the channel estimation problem with respect to the linear model.

3.2 IR-UWB Communication Model

The received UWB signal profile is given by the following,

$$r(t) = p(t) * h(t) + \omega(t) \tag{3.1}$$

$$= \sum_{l=0}^{L-1} \alpha_l p(t - \tau_l) + \omega(t) \tag{3.2}$$

where L is the total number of MPCs, $p(t)$ is the transmitted pulse, $h(t)$ is the channel impulse response and $*$ denotes linear convolution while $\omega(t)$ is the additive noise at the receiver which is assumed to be white Gaussian (AWGN).

In this thesis, we assume that the pulse shape $p(t)$ is known at the receiver and so the task of the UWB channel estimator of Fig. ?? is to estimate a set of unknown parameters α_l's and τ_l's in a known signal corrupted by noise.

Consider the signal profile in Eq. (3.2), which we would like to express in matrix form. We can represent $r(t)$ using its Nyquist rate (F_N). Thus, the

samples are taken every $\delta_t = \frac{1}{F_N}$ seconds which is much less than the pulse duration T_p and we can write,

$$r(n\delta_t) = \sum_{l=0}^{L-1} \alpha_l p(n\delta_t - l\Delta\delta_t) + \omega(n\delta_t) \tag{3.3}$$

$$r(n) = \sum_{l=0}^{L-1} \alpha_l p(n - l\Delta) + \omega(n) \tag{3.4}$$

where we assume that the delays τ_l's can be represented as integral multiples of δ_t, i.e. $\tau_l = l\Delta\delta t$ (Δ is the number of samples of the basic shift of the pulse as shown in Eq. (3.12)) and where we drop δ_t from the argument in Eq. (3.4) for notational convenience.

Since the MPCs arrival is a Poisson process and its rate is given by λ in Eq. (2.7) the expected number of paths occuring in a time bin of duration δ_t seconds is given by $\lambda\delta_t$. It thus follows that the probability of having k paths in a duration of δ_t seconds is given by the Poisson distribution as follows:

$$P(k) = \frac{(\delta_t\lambda)^k \exp^{-(\delta_t\lambda)}}{k!} \tag{3.5}$$

Therefore, the probability of having no path (i.e., $k = 0$) during δ_t is $\exp^{-(\delta_t\lambda)}$ and of having a single path (i.e., $k = 1$) is $(\delta_t\lambda) \exp^{-(\delta_t\lambda)}$. If the duration δ_t is small enough then the probability of having multiple paths during δ_t is much smaller

than having a single path or no path, that is,

$$P(1), P(2) \gg P(k) \quad for \quad k = 3, 4, \ldots, L-1 \tag{3.6}$$

When δ_t is very small Eq. (3.6) is satisfied and $\exp^{-(\delta_t \lambda)} \approx 1$, therefore,

$$P(1) = (\delta_t \lambda) \exp^{-(\delta_t \lambda)} \tag{3.7}$$

$$\approx \delta_t \lambda \tag{3.8}$$

Now the occurance of an MPC in a bin of duration δ_t can be approximated by a Bernoulli trial [39], where the probability of success (i.e., a single path occurs during δ_t) is $\mathrm{p}_b = \lambda \delta_t$ and probability of failure (i.e., no path occurs during δ_t) is $1 - \mathrm{p}_b$. This approximation results from the well-known approximation of a Poisson process by a Binomial distribution. This implies that if we divide a certain time span of T seconds into N small durations of δ_t seconds each, then the probability of having MPCs in k of these N bins is given by,

$$P_b(k) = \mathrm{p}_b^k (1 - \mathrm{p}_b)^{N-k} \tag{3.9}$$

$$= (\lambda \delta_t)^k (1 - \lambda \delta_t)^{N-k} \tag{3.10}$$

In practice, the number of multipath components L is generally large but only the L_{max} strongest MPCs capture the significant portion of the transmitted signal energy [40]. This leads to a practical RAKE (Selective-RAKE) Receiver

implementation where estimates of only L_{max} τ_l's and the corresponding α_l's are required.

Now, while we can represent $r(t)$ using its Nyquist rate samples, we sub-sample it at a lower rate $F_S = \mu F_N$ where $\mu = \frac{M}{N}$ and $M < N$. We represent this in the matrix form as,

$$\mathbf{y} = \mathbf{\Psi}\boldsymbol{\alpha} + \boldsymbol{\omega} \tag{3.11}$$

where

$$\mathbf{\Psi} = \begin{bmatrix} p(n-\Delta) & \dots & p(n-N\Delta) \\ p(n+1-\Delta) & \dots & p(n+1-N\Delta) \\ . & \dots & . \\ . & \dots & . \\ p\,(n+(M-1)-\Delta) & \dots & p\,(n+(M-1)-N\Delta) \end{bmatrix} \tag{3.12}$$

The matrix $\mathbf{\Psi}$ consists of, as its columns, the discretized shifted versions of the pulse $p(t)$ of Eq. (2.3). Note that in Eq. (3.11) $\mathbf{y}$ is the $M \times 1$ received vector, and $\boldsymbol{\omega}$ is the $M \times 1$ AWGN vector with zero-mean and $M \times M$ covariance matrix $\mathbf{C}_\omega = N_0 \mathbf{I}$. The vector $\boldsymbol{\alpha}$ is the $N \times 1$ channel parameter vector. Moreover, $\boldsymbol{\alpha}$ is sparse and its active elements correspond to the channel taps and so we decompose $\boldsymbol{\alpha}$ as

$$\boldsymbol{\alpha} = \mathbf{a} \odot \underline{\boldsymbol{\alpha}} \tag{3.13}$$

where $\odot$ denotes elelment by element multiplication. In this equation, $\underline{\boldsymbol{\alpha}}$ is an $N \times 1$ binary vector that represents the support of $\boldsymbol{\alpha}$ (i.e., $\underline{\boldsymbol{\alpha}} = supp(\boldsymbol{\alpha})$), and $\mathbf{a}$

42

is an $N \times 1$ vector of the amplitudes of $\boldsymbol{\alpha}$. Now, if we set $\mathbf{A} = diag(\boldsymbol{\alpha})$, then we can rewrite eq. (3.11) as,

$$\mathbf{y} = \boldsymbol{\Psi} \mathbf{A} \underline{\boldsymbol{\alpha}} + \boldsymbol{\omega} \tag{3.14}$$

Suppose the Hamming weight of a certain $\underline{\boldsymbol{\alpha}}$ is l (i.e., $l = |\underline{\boldsymbol{\alpha}}|_0$) then $\mathbf{a}_{\underline{\boldsymbol{\alpha}}}$ is the $l \times 1$ vector which contains the non-zero amplitudes of $\boldsymbol{\alpha}$, and $\boldsymbol{\Psi}_{\underline{\boldsymbol{\alpha}}}$ represents the $N \times l$ sub-matrix of $\boldsymbol{\Psi}$ formed by collecting those columns of $\boldsymbol{\Psi}$ which are indicated by $\underline{\boldsymbol{\alpha}}$. Therefore, when $\underline{\boldsymbol{\alpha}}$ is known Eq. (3.14) can be written as,

$$\mathbf{y} = \boldsymbol{\Psi}_{\underline{\boldsymbol{\alpha}}} \mathbf{a}_{\underline{\boldsymbol{\alpha}}} + \boldsymbol{\omega} \tag{3.15}$$

First we assume the unknown channel parameters $\underline{\boldsymbol{\alpha}}$ and $\mathbf{a}_{\underline{\boldsymbol{\alpha}}}$ to be deterministic and develop estimators using classical estimation techniques. Secondly, we assume these channel parameters to be random and apply Bayesian estimation. We also exploit the sparsity of the received UWB signal profile and the structure of the matrix $\boldsymbol{\Psi}$ of our model Eq. (3.11) to develop low-complexity estimators. In the following sub-sections we highlight the statistical information about the channel and the rich structure of $\boldsymbol{\Psi}$.

3.2.1 Statistical Information

The statistical a priori information about the UWB channel used in this thesis is basically form the IEEE 802.15.4a model where the MPCs arrival is modeled as the double Poisson process. We approximate this as a single Poisson process

with the MPC arrival rate λ as discussed previously. for the Indoor Resedential Line-of-Sight (LOS) channel model. The MPCs amplitudes are modeled to be Nakagami-m distributed as discussed previously. with an exponentially decaying APDP. Under this approximation the received signal profile is a single cluster as descussed with the decay rate Γ. The Table 3.1 shows the typical values of λ and Γ for the Indoor Resedential LOS model. The approximation is made in order to incorporate the statistical information from the model [14] into the design of our Bayesian estimators. The double Poisson process is very complex to handle and to the best of our knowledge no reserch work has been able to do so in its entirety. The difficulty arises from the fact that in the double Poisson process, the probability of the arrival of MPCs in a given time bin depends on which cluster that MPC belongs to. The start of each cluster itself belongs to another Poisson process and hence to determine the probability of MPC arrival in a certain time bin cannnot be expressed in a closed form. The MPCs time of arrival statistics for the double Poisson process is intractable and cannot be used and so we assume it as a single large cluster which leads us to the Bernoulli assumption on the the MPCs arrivals as discussed. Moreover, the fading statistics are also a function of the MPCs' arrival time and so the assumption of a single Poisson process makes this statistical information to be useful. Since it is difficult to obtain closed form

Parameter	Value
Λ	$0.047ns^{-1}$
Γ	$22.61ns$

Table 3.1: Parameters for Resedential LOS Channel Model from [1]

44

expressions for the density functions of the Nakagami distributed channel coeffi-
cients, we consider the following three cases for the channel amplitudes statistics
in order to develop estimators in the Bayesian framework:

1. Non-Gaussian Amplitudes

2. Non-Gaussian Amplitudes with Known 2nd Order Statistics from the APDP

3. Gaussian Amplitudes with zero mean and known 2nd Order Statistics from
   the APDP

3.2.2 Structural Information

The sensing matrix $\boldsymbol{\Psi}$ in Eqs. (3.11) - (3.14) is rich in structure. It is not only
Toeplitz but is a banded diagonal matrix. If the length of the support of the basic
pulse $p(n)$ at Nyquist rate is denoted by $|p|$, then for a given sub-sampling ratio,
μ, the bandwidth of the matrix $\boldsymbol{\Psi}$ is given by $\beta = |p|\mu$. This implies that

$$\boldsymbol{\psi}_i^H \boldsymbol{\psi}_j = \begin{cases} f(|i - j|), & |i - j| \leq \beta \\ 0, & |i - j| > \beta \end{cases}$$

where $f(.)$ is a function denoting the correlation between the columns and $\boldsymbol{\psi}_i$ is
the ith column of $\boldsymbol{\Psi}$ which is correlated only to few of its neighbouring columns,
as shown in Fig. 3.1. This is true for all the columns of $\boldsymbol{\Psi}$ because of its Toeplitz
structure. Therefore we can group consecutive columns of $\boldsymbol{\Psi}$ into a number of
custers of width $\nu = s\beta$, where s is an integer. Since these clusters are mutually

nearly orthogonal, this technique is refered to as *orhtogonal clustering* (OC). We remark here that for a received vector **y** only a few of these clusters will be active due to the multipath sparsity and those active clusters are more likley to be mutually orthogonal. This is illustrated in Fig. 4.1. We note here that clusters thus formed are refered to as 'othogonal clusters' in the thesis to differentiate them from the clusters in the received profile due to the UWB channel.
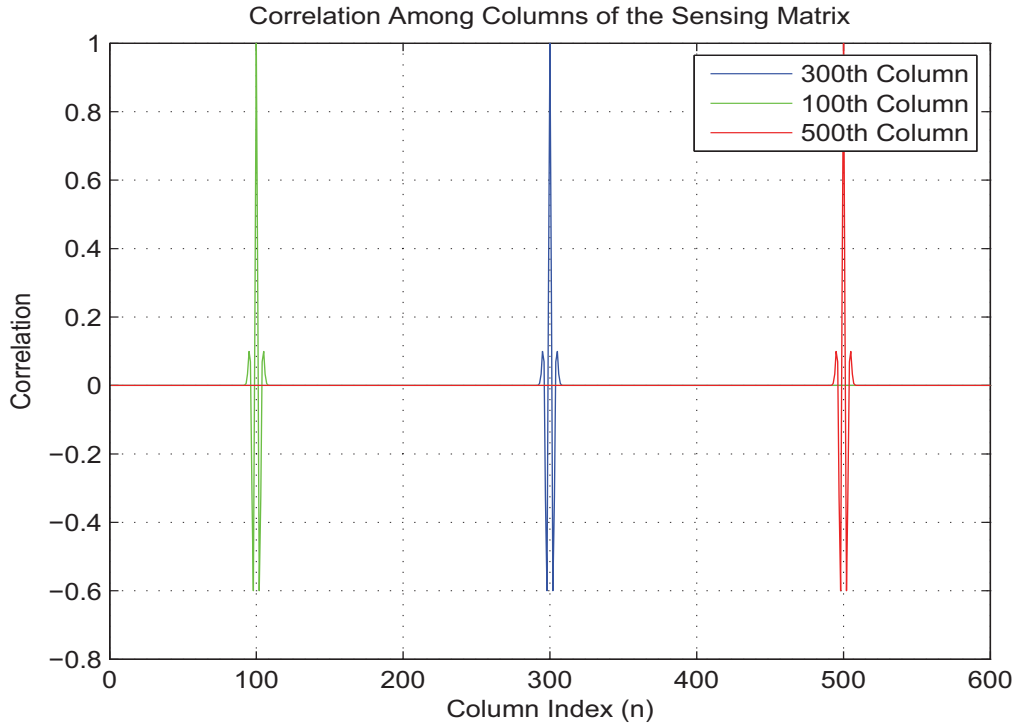
Figure 3.1: Correlations Among the Columns of $\boldsymbol{\Psi}$

3.3   Channel Estimation Problem

In order to estimate the UWB channel we need to estimate the channel parameter vector $\boldsymbol{\alpha}$ of Eq. (3.11). As described in the previous section, $\underline{\boldsymbol{\alpha}}$ is the support of the channel and $\mathbf{a}_{\underline{\boldsymbol{\alpha}}}$ is the vector of channel fading coefficients. Therefore, in

46

order to estimate the channel we need to jointly estimate $\underline{\boldsymbol{\alpha}}$ and $\mathbf{a}_{\underline{\boldsymbol{\alpha}}}$. Thus we develop the joint MMSE estimate of the channel. We also estimate the channel by decomposing the channel estimation problem into two parts: (i) estimation of the support vector $\underline{\boldsymbol{\alpha}}$, followed by (ii) estimation of the corresponding amplitude vector $\mathbf{a}_{\underline{\boldsymbol{\alpha}}}$. The channel decomposition is explained in the next sub-section.

3.3.1 Channel Decomposition

From the formulation in Eq. (3.14) and Eq. (3.15) the estimation of the MPCs delays, τ_l's, and their amplitudes, α_l's, of Eq. (3.2) translates into the estimation of the vectors $\underline{\boldsymbol{\alpha}}$ and $\mathbf{a}_{\underline{\boldsymbol{\alpha}}}$ respectively. Furthermore, based on the model in Eq. (3.14) we can first estimate the channel support and then for the estimated support estimate the corresponding channel amplitudes from Eq. (3.15). In this way we have decomposed the channel estimation problem.

The estimation of the channel support vector $\underline{\boldsymbol{\alpha}}$ is a model selection problem and where we need to find the best model from the 2^N possible ones. We denote by $\hat{\underline{\boldsymbol{\alpha}}}$ the estimate of the channel support, and develop the estimate in the classical estimation framework (i.e., assuming deterministic unknown parameters) which reduces the square error is the solution to the following search problem

$$\hat{\underline{\boldsymbol{\alpha}}} = \arg\min_{\underline{\boldsymbol{\alpha}} \in \aleph} \|\mathbf{y} - \boldsymbol{\Psi}\mathbf{A}\underline{\boldsymbol{\alpha}}\|^2 \tag{3.16}$$

where $\aleph$ is the set consisting all the 2^N possible support vectors $\underline{\boldsymbol{\alpha}}$. The fading co-efficients are then estimated using the Least Squares (LS).

In the case of Bayesian framework, channel estimators are developed for the different fading statistics of the channel. As such, we develop channel estimators for the following three fading statistics:

- Non-Gaussian

- Non-Gaussian Amplitudes with Known Second Order Statistics

- Gaussian

3.3.2 LS Amplitudes Estimation

When the amplitudes are assumed to be unknown deterministic (classical estimation) or random but non-Gaussian (Bayesian estimation) then the Least-Squares (LS) is the best solution. For a certain support vector $\underline{\boldsymbol{\alpha}}$ we have,

$$\mathbf{y} = \boldsymbol{\Psi}_{\underline{\boldsymbol{\alpha}}} \mathbf{a}_{\underline{\boldsymbol{\alpha}}} + \boldsymbol{\omega} \tag{3.17}$$

where $\boldsymbol{\Psi}_{\underline{\boldsymbol{\alpha}}}$ is the matrix formed by those columns of $\boldsymbol{\Psi}$ which are indicated by $\underline{\boldsymbol{\alpha}}$. The amplitudes vector corresponding to the support $\mathbf{a}_{\underline{\boldsymbol{\alpha}}}$ can now be estimated using as

$$\hat{\mathbf{a}}_{\underline{\boldsymbol{\alpha}}_{LS}} = \left(\boldsymbol{\Psi}_{\underline{\boldsymbol{\alpha}}}{}^H \boldsymbol{\Psi}_{\underline{\boldsymbol{\alpha}}} \right)^{-1} \boldsymbol{\Psi}_{\underline{\boldsymbol{\alpha}}}{}^H \mathbf{y} \tag{3.18}$$

3.3.3 MMSE Amplitudes Estimation

When the amplitudes are assumed to be Gaussian then the a posteriori probability density is also Gaussian in the prescence of AWGN. Therefore, we can estimate the amplitudes using the Minimum Mean Square Error (MMSE) estimator which is optimal in the sense of minimizing the mean-square estimation error. The MMSE estimate of the amplitudes corresponding to a known support vector is given by,

$$\hat{\mathbf{a}}_{\underline{\boldsymbol{\alpha}}_{MMSE}} = \mathbb{E}(\mathbf{a}_{\underline{\boldsymbol{\alpha}}}|\mathbf{y}, \underline{\boldsymbol{\alpha}}) \tag{3.19}$$

3.3.4 Estimation Performance Metrics

The performance metrics used to evaluate the performance of the various channel estimation methods developed in the thesis are:

- Normalized Root Mean Square Error (NRMSE) in the estiamtion of the MPC's arrival times

- Energy Capture (EC)

First, we discuss the estimation of the arrival times of the MPCs. The absolute arrival time of the lth MPC is denoted by τ_l as in Eq. (3.2). The estimate of the channel support vector $\hat{\underline{\boldsymbol{\alpha}}}$ corresponds to the estimate of the τ_l's. Without loss of generality we assume that the arrival time of the first MPC is $\tau_0 = 0$. Therefore, the estimate of the absolute value of the lth MPC's arrival time, for

$l = 1, 2, \ldots, L - 1$ is given as,

$$\hat{\tau}_l = \hat{\underline{\alpha}}^{(l)}(\Delta\delta_t) \tag{3.20}$$

where $\hat{\underline{\alpha}}^{(l)}$ is the index of the lth non-zero element of the estimated support vector $\hat{\underline{\alpha}}$. The $L \times 1$ vector $\boldsymbol{\tau}$ is composed of τ_l's as its elements. Thus the Normalized Root Mean Square Error (NRMSE) in the estiamtion of the MPC's arrival times which is expressed in number of samples, is calculated from Z runs of Monte Carlo simulations as,

$$\widetilde{\boldsymbol{\tau}}_{NRMSE} = \frac{1}{\delta_t}\sqrt{\frac{1}{Z}\sum_{z=1}^{Z} \|\boldsymbol{\tau} - \hat{\boldsymbol{\tau}}(z)\|^2} \tag{3.21}$$

The amplitude of the lth MPC is denoted by α_l as in Eq. (3.2). The estimate of the $L \times 1$ channel amplitude vector $\hat{\mathbf{a}}_{\underline{\alpha}}$ corresponds to the estimate of the α_l's. Therefore, the estimate of the lth MPC's amplitude, for $l = 1, 2, \ldots, L - 1$ is given as,

$$\hat{\alpha}_l = \hat{\mathbf{a}}_{\underline{\alpha}}^{(l)} \tag{3.22}$$

where $\hat{\mathbf{a}}_{\underline{\alpha}}^{(l)}$ is the lth element of the estimated support vector $\hat{\mathbf{a}}_{\underline{\alpha}}$. Thus the Energy Capture (EC) at the receiver based on the estimates is given as,

$$EC = \left(1 - \frac{\|x(t) - \hat{x}(t)\|^2}{\|x(t)\|^2}\right) \times 100\% \tag{3.23}$$

where,

$$x(t) = \sum_{l=0}^{L-1} \alpha_l p(t - \tau_l) \tag{3.24}$$

$$\hat{x}(t) = \sum_{l=0}^{L-1} \hat{\alpha}_l p(t - \hat{\tau}_l) \tag{3.25}$$

CHAPTER 4

SPARSITY BASED

ESTIMATION

4.1 Introduction

In this chapter we present the UWB channel estimation by exploiting the sparsity of the received UWB signal profile. As we have seen in that due to the large bandwidth, the MPCs can be finely resolved in UWB systems and the received UWB signal profile consists of MPCs arrival in only a few of the time bins. This is termed as *mulitpath sparsity* and depicted in Figure 4.1. We present Genetic Aglorithm based search method for channel estimation in Section 4.2 and Correlation based support estimation in Section 4.3. We also employ Compressive Sensing in Section 4.4 to estimate the UWB channel. In Section 4.5, we present the two-step estimation approaches by exploiting the multipath sparsity of the channel.
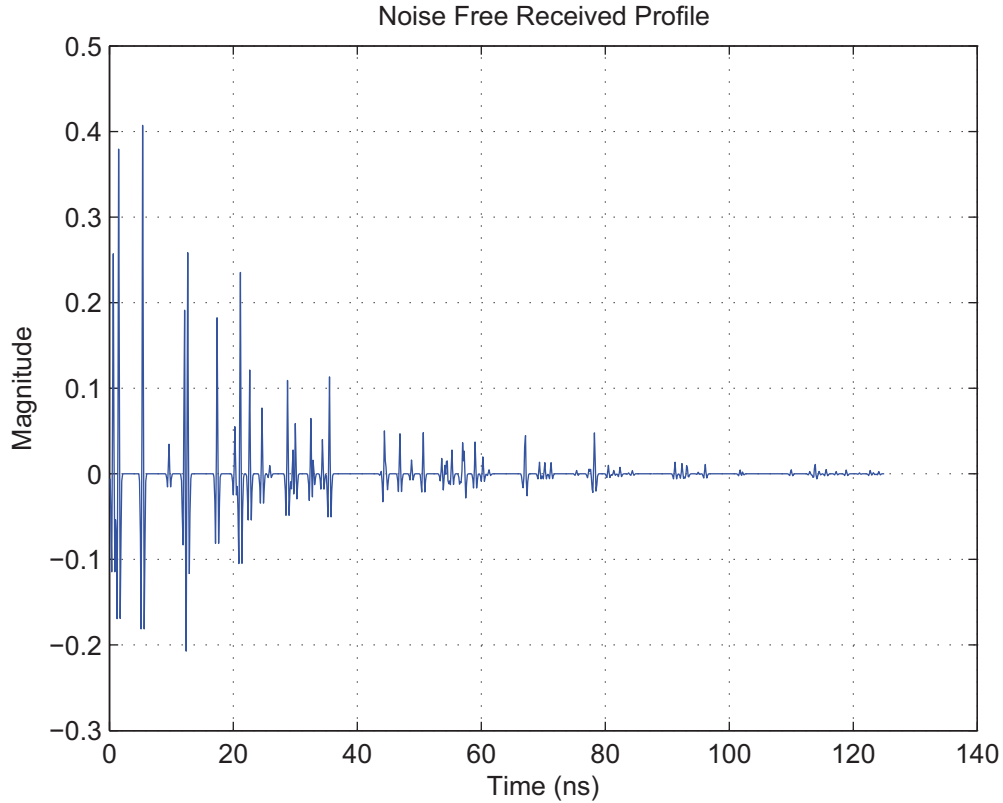
Figure 4.1: Noise Free Received Typical UWB Received Signal

4.2 Genetic Algorithm (GA) Based Search

Genetic Algorithms (GAs) belong to the class of guided random search techniques. These are evolutionary computing techniques in which a randomly chosen initial population of potential solutions (chromosomes) is evolved using evolutionary operations and the next generation is selected based on the principles of the surrvival of the fittest. A fitness function is used to assign the fitness values to the individuals of the popoulation at every generation. Genetic algorithms are specifically used to find the global optima of a given objective function that may or may not be subject to constraints [41]. GAs take several evolution cycles to reach to the solution and are attractive as they do not get stuck in or around a

local optima and are thus more likely to reach the global optima. GAs should be useful if they have a higher rate of convergence, good quality of solution and reasonable computaional requirements.

In our channel estimation problem we employ GA to search for the $L_{max} = 20$ MPCs that minimize the square error. The fitness function $g(.)$ assigns the normalized squared error as the fintness value. If $\hat{\boldsymbol{\alpha}}$ denotes the estimate of $\boldsymbol{\alpha}$ the fitness function is

$$g\left(\hat{\boldsymbol{\alpha}}\right) = 1 - \left(\frac{\|\hat{\boldsymbol{\alpha}} - \boldsymbol{\alpha}\|^2}{\|\boldsymbol{\alpha}\|^2}\right) \tag{4.1}$$

and the GA estimate of $\boldsymbol{\alpha}$ is

$$\hat{\boldsymbol{\alpha}}_{GA} = \arg\max_{\boldsymbol{\alpha}} g\left(\boldsymbol{\alpha}\right) \tag{4.2}$$

Now the estimate of the channel support using the Genetic algorithm is given by

$$\underline{\hat{\boldsymbol{\alpha}}}_{GA} = supp(\hat{\boldsymbol{\alpha}}_{GA}) \tag{4.3}$$

We note here that the fitness function of Eq. (4.1) is also the fraction of $\boldsymbol{\alpha}$'s energy present in $\hat{\boldsymbol{\alpha}}_{GA}$ and as such is a good choice for a fitness function. We select the population size to be more than the L_{max} and use binary encoding for the chromosomes of the population. The new population is generated by applying the evolutionary operations such as selection, crossover and mutation.

The simulations show that with a larger population size the estimates are better but the computaional time is increased manifolds. The rate of convergence improves slightly with the increase in the population size. We observe that the high computational complexity and slow convergence is due to the large search space and undirected random initial population. In the follwoing we propose two-step estimation methods where the initial population for the GA based search is first found by applying Compressive Sensing and Correlation, respectively.

4.3 Correlation Based Support Estimation

Now we present the estimation of the channel support by correlating the receivied vector $\mathbf{y}$ with all the columns of $\mathbf{\Psi}$. Based on the sparsity of the received UWB signal profile we expect only a few of these correlations to be significant. Therefore, we use thresholding to find the significant of these correlations to estimate of the channel support. The true value of the threshold η is a function of the receiver SNR.

$$\hat{\underline{\boldsymbol{\alpha}}}_{CR}(i) = \begin{cases} 0 & \mathbf{y}^H \boldsymbol{\psi}_i < \eta \\ 1 & \mathbf{y}^H \boldsymbol{\psi}_i > \eta \end{cases}$$

for $i = 1, 2, \ldots, N$ and $\hat{\underline{\boldsymbol{\alpha}}}_{CR}(i)$ is the ith element of the vector $\hat{\underline{\boldsymbol{\alpha}}}_{CR}$.

Once we obtain the estimate of the channel support, $\hat{\underline{\boldsymbol{\alpha}}}_{CR}$, we can use LS to estimate the channel fading co-efficients. The Figure 4.2 shows the normalized RMSE (NRMSE) in estimation of the channel support using Correlation and

comparing it with CS based estimation. It is clear that the Correlation method for estimating the channel support is sensitive to the choice of η but it is very simple to implement. Therefore, we describe how we use Correlation based estimates as the initial coarse estimates in a two-step estimation approach.

4.4 Compressive Sensing (CS) Based Estimation

Compressive sensing (CS) is a revolutionary and rapidly growing field in signal processing. The core idea in compressive sensing, or also known as compressed sensing, is to exploit the sparsity of the unknown data to reconstruct it from fewer observations. In a few years CS has found applications in all the major areas of signal processing where it is used for many tasks such as signal reconstruction, signal estimation, signal de-noising etc. In the following a brief background is presented on CS as an estimation technique since we employ CS to the UWB channel estimation problem.

Suppose we have a vector $\mathbf{r}$ of size $N \times 1$ that we need to estimate but are able to sense only $M < N$ observations of $\mathbf{r}$. We denote these observations by the $M \times 1$ vector $\mathbf{y}$. Thus mathematically

$$\mathbf{y} = \mathbf{\Phi r} \tag{4.4}$$

where $\mathbf{\Phi}$ is the $M \times N$ sensing matrix. Now, we wish to obtain $\mathbf{r}$ from $\mathbf{y}$ that is to solve the under-determined system of Eq. (4.4). Suppose we also know that $\mathbf{r}$ can be expressed as

$$\mathbf{r} = \mathbf{D}\boldsymbol{\alpha} \qquad (4.5)$$

where $\mathbf{D}$ is a known $N \times N$ matrix, also commonly known as the *dictionary* for $\mathbf{r}$ in the CS literature and $\boldsymbol{\alpha}$ is the $N \times 1$ vector. The theory of compressive sesning reveals that if the matrices $\mathbf{D}$ and $\mathbf{\Phi}$ are mutually incoherent (i.e., a vector cannot be simultaneously sparse in both $\mathbf{D}$ and $\mathbf{\Psi}$, for more details see [42]) and known, and $\boldsymbol{\alpha}$ is sparse (i.e., only few of its entries are non-zero), then we can reconstruct $\mathbf{r}$ from $\mathbf{y}$ successfully with a very high probability [43]. Combining Eqs. (4.4) and (4.5), we can write

$$\mathbf{y} = \mathbf{\Phi}\mathbf{D}\boldsymbol{\alpha} \qquad (4.6)$$

$$= \mathbf{\Psi}\boldsymbol{\alpha} \qquad (4.7)$$

Therefore, in order to apply CS for estimating $\mathbf{r}$ all we need to know is the dictionary $\mathbf{D}$, in which $\mathbf{r}$ can be represented by a sparse vector $\boldsymbol{\alpha}$ and use a suitable sensing martix $\mathbf{\Phi}$ to obtain the observation vector $\mathbf{y}$. Then CS reconstruction methods find the *sparsest* $\boldsymbol{\alpha}$ that satisfies Eq. (4.6).

It has been shown [44] that for any fixed deterministic matrix $\mathbf{D}$ a random sensing matrix $\mathbf{\Phi}$ with *i.i.d.* entries can be used for CS. This has lead to

many applications of CS in signal reconstruction from randomly sub-sampled observations. Suitable sensing matrices such as random matrices have been shown to satisfy the Restricted Isometry Property (RIP) [45]. RIP is widely used to determine whether a certain matrix can be used as a sensing matrix in CS. Recently, certain class of deterministic matrices such as Toeplitz matrices, have also been shown to be suitable choices for sensing matrices in CS [46].

The optimal estimate of $\boldsymbol{\alpha}$ from the CS perspective (i.e., the sparsest solution) is given as

$$\hat{\boldsymbol{\alpha}}_{CS} = \arg\min_{\boldsymbol{\alpha}} \|\boldsymbol{\alpha}\|_0 \quad s.t. \quad \mathbf{y} = \boldsymbol{\Psi}\boldsymbol{\alpha} \tag{4.8}$$

where $\|.\|_0$ represents the l_0 norm.

In our UWB channel estimation problem, the under-determined system is described by Eq. (3.11), which we reproduce here

$$\mathbf{y} = \boldsymbol{\Psi}\boldsymbol{\alpha} + \boldsymbol{\omega} \tag{4.9}$$

We need to estimate the chanel parameter vector $\boldsymbol{\alpha}$ from the observation $\mathbf{y}$. In Eq. (4.9) the matrix $\boldsymbol{\Psi}$ is a uniformly sub-sampled Toeplitz banded matrix. We apply CS to recover $\boldsymbol{\alpha}$ from the uniformly sub-sampled received vector $\mathbf{y}$. We note that Eq. (4.9) is different from the model in Eq. (4.4) as it also contains a noise term.

There are several methods for reconstructing $\boldsymbol{\alpha}$ from $\mathbf{y}$ used in the CS literature. These reconstruction methods can be broadly classified as:

1. Convex Relaxation Based Reconstruction

2. Greedy Algorithm Based Reconstruction

4.4.1   CS Based on Convex Relaxation

The optimal estimate of Eq. (4.8) is a solution to a combinatorial problem and is NP-hard. Therefore, convex relaxation of the problem using the l_1-norm is widely used in the sparse signal recovery literature. Thus the CS estimate of $\boldsymbol{\alpha}$ based on the l_1 relaxation in the prescence of noise Eq. (4.9) is given as

$$\hat{\boldsymbol{\alpha}}_{CS} = \arg\min_{\boldsymbol{\alpha}} \|\boldsymbol{\alpha}\|_1 \quad s.t. \quad \|\mathbf{y} - \boldsymbol{\Psi}\boldsymbol{\alpha}\|_2^2 < \epsilon \tag{4.10}$$

where ϵ is the amount of tolerable residual error in the estimate and depends on the received SNR. This is known as the Basis Pursuit De-Noising (BPDN) algorithm in the CS literature. The optimization problem of Eq. (4.10) is equivalent to the famous Least Absolute Shrinkage and Selection Operator (LASSO) in the statistical community. Interestingly, an equivalent form of Eq. (4.10) also has a Bayesian interpretation which we will discuss. In general, the l_1 relaxation based reconstruction methods are computationally extensive and therefore alternate reconstruction methods such as Greedy algorithms have also been adopted

[47].

4.4.2 CS Based on Greedy Reconstruction

Several Greedy algorithms have been developed and applied to solve convex optimization poblems [43]. These algorithms have also been used for sparse signal approximation and therefore are used for signal reconstruction in CS problems. Greedy methods are a fast alternative to l_1 based convex optimization problems. The greedy algoritms follow a problem solving heuristic of choosing the local optima at every stage with the hope of finding the global optima. In certain circumstances greedy strategies have been shown to perform better or nearly as good as the convex methods. The popular greedy methods are:

1. Matching Pursuit (MP)

   Matching Pursuit finds the best matching projections of $\mathbf{y}$ onto the dictionary $\mathbf{D}$. MP iteratively generates a sorted list of indices of those atoms of the dictionary (i.e., columns of $\mathbf{D}$) onto which the projection of the residuals in each iteration is the largest. After every iteration, the contribution of the selected column is subtracted from the previous residual to obtain the next residual. In every subsequent iteration the best projections of the residual is sought. The iterations stop when the projections fall below a pre-defined threshold value [48].

2. Orthogonal Matching Pursuit (OMP)

   Orthogonal Matching Pursuit was developed as an improvement to the MP

[47]. It shares some properties of MP and inherits the selection procedure. The difference is that, in OMP the residual at each iteration is calculated such that it is always orthogonal to the subspace formed by the previously selected columns of **D** [49].

3. Gradient Pursuit (GP)

   Gradient Pursuit is an approximation to the OMP developed with reduced computational requirements. It uses a gradient term to update the direction for the pursuit of the GP in the next iteration [50].

We denote the estimate of $\boldsymbol{\alpha}$ obtained by applying the greedy algorithms on the received vector of Eq. (4.9) as $\hat{\boldsymbol{\alpha}}_G$. The corresponding estimate of the support of $\boldsymbol{\alpha}$ is given by

$$\underline{\hat{\boldsymbol{\alpha}}}_G = supp(\hat{\boldsymbol{\alpha}}_G) \tag{4.11}$$

and corresponds to the estimate of the MPCs arrival times. Figure 4.3 shows the NRMSE in the estimation of the MPCs arrival times using CS with various greedy algorithms for reconstruction. Figure 4.4 shows the comparison in the performance of the greedy algorithms in terms of the resulting Energy Capture where the amplitudes are estimated using LS.

4.5   Two-Step Estimation

We observed that Genetic algorithm performs better in estimating the channel when the number of population generations is increased but requires a large

amount of time to converge to the estimates. On the other hand, CS and Correlation based estimation methods provide the most likely locations of the channel support with a far fewer number of computations and has less computational complexity. Therefore, in the following we combine the CS and Correlation based estimation techniques in a two-step estimation approach.

4.5.1   CS followed by GA

In the first step, we apply CS to obtain the likely support of the channel. From the CS estimate of the channel we retain the $L_{cs} > L_{max}$ largest estimates and the corresponding support is used as the initial population for the GA in the second step. Figure 4.5 shows the NRMSE performance of this two-step channel estimation approach.

4.5.2   Correlation followed by GA

In the first step, we apply Correlation technique to obtain the likely support of the channel. From the Correlation based estimate of the channel we retain the $L_{cr} > L_{max}$ largest estimates and the corresponding support is used as the initial population for the GA in th second step. Figure 4.5 shows the NRMSE performance of this two-step channel estimation approach.

4.6   Results
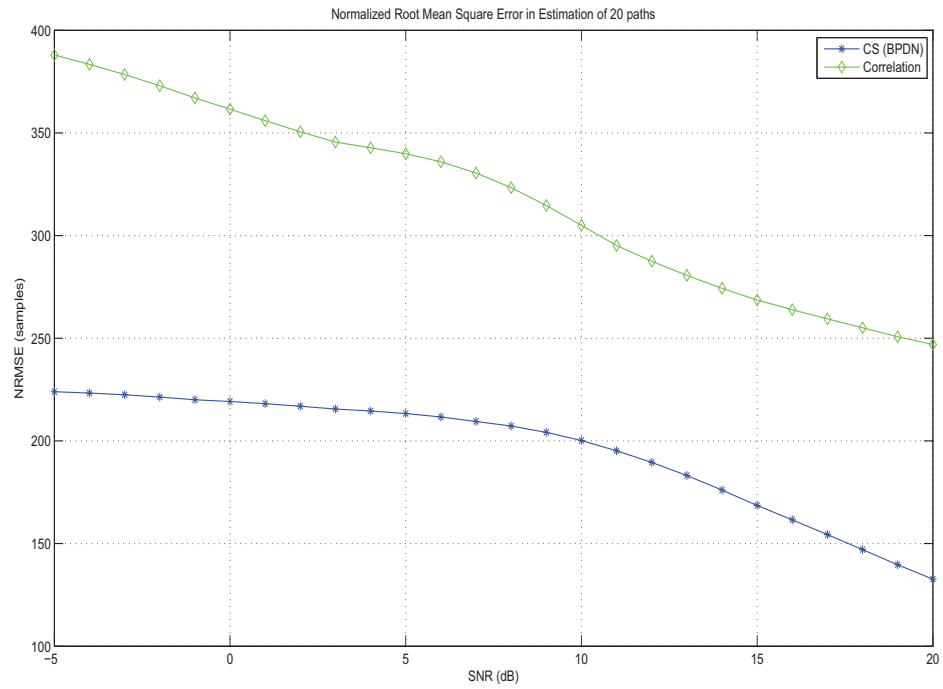
Figure 4.2: Performance of Correlation Based Estimation and CS
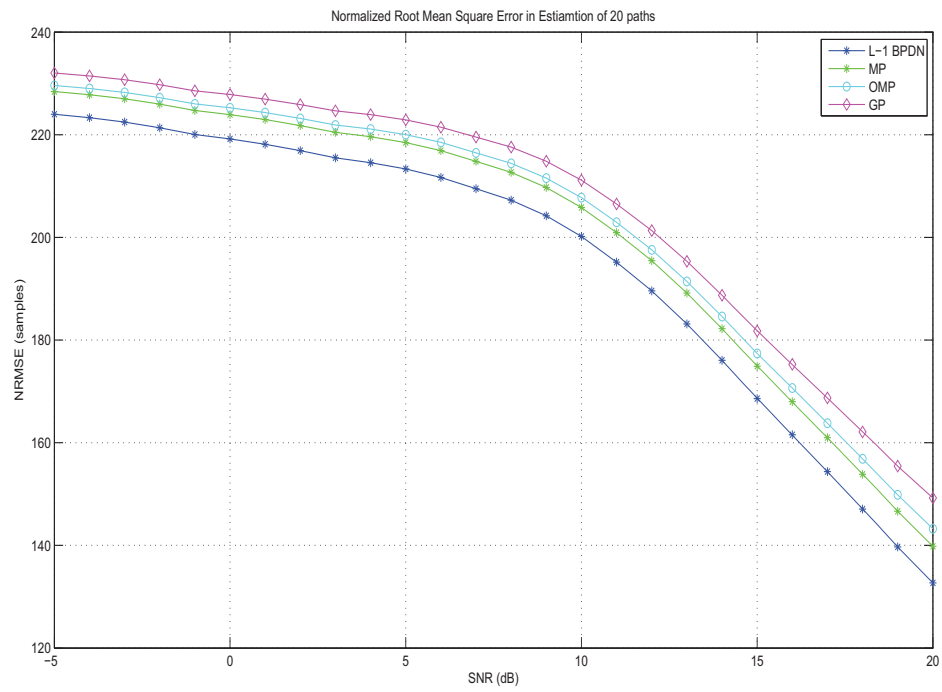
Figure 4.3: Performance Comparison in MPCs Arrival Time Estimation Error of Different CS Methods
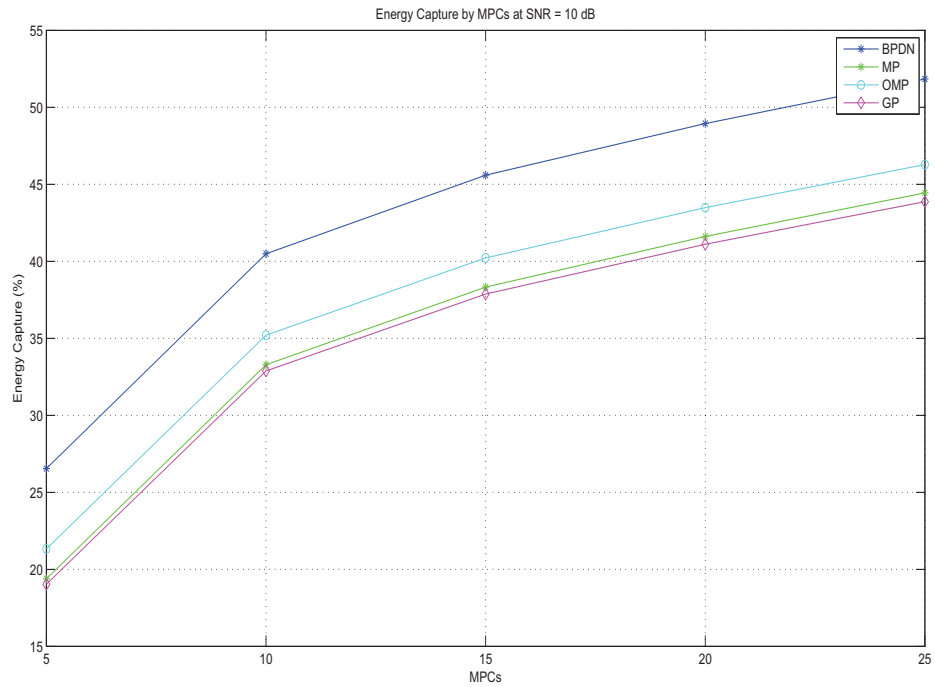
Figure 4.4: Performance Comparison in Energy Capture using Different CS Methods + LS for Amplitudes Estimation
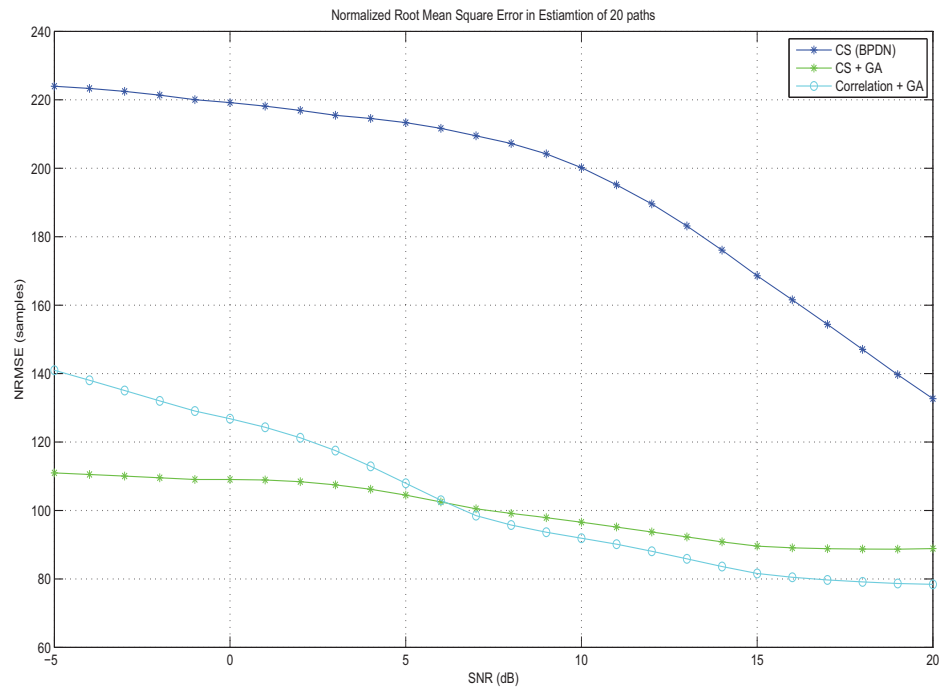
Figure 4.5: Comparison of CS+GA and Correlation+GA Estimation Techniques

CHAPTER 5

CLASSICAL ESTIMATION

5.1 Introduction

In this chapter we present the channel estimation of the UWB channel in a classical estimtion framwork where the unknown quantities to be estimated are assumed to be deterministic. There are several estimators in the classical estimation theory among whcih the Minimum Variance Un-biased Estimator (MVUE) is known to be the optimal estimator in the sense that it has the minimum variance among all the un-biased estimators. MVUE achieves the Cramer-Rao Lower Bound (CRLB) and are thus efficient estimators. But we are not always able to determine the MVUE and even if we can, in some cases, it could not be implemented [36]. Therefore, if an efficient estimator does not exist then the approach of the CRLB to find an MVUE fails. In such cases, Rao-Blackwell-Lehmann-Scheffe (RBLS) theorem is sometimes used to determine the MVUE, but this requires to first find the sufficient statistic which is not

always possible. All these methods of estimation require the data model and are thus parametric estimation methods. Moreover, these classical estimation techniques require that the knowledge of the data is summarized in the parametric probability density function (pdf) $p(y; \boldsymbol{\alpha})$ where $\mathbf{y}$ is the observation vector and $\boldsymbol{\alpha}$ the parameter vector to be estimated and pdf $p(y; \boldsymbol{\alpha})$ is functionally dependent on $\boldsymbol{\alpha}$.

When the data model is a linear model, then the MVUE is given by the Best Linear Un-biased Estimator (BLUE). BLUE has the minimum varianec among all the un-biased estimators that are linear in data. The Least-Squares Estimator (LSE) is widely used in various estimation problems. LSE is very straight forward to implement and it minimizes the LS error criterion. In general, the LSE does not minimize the estimation error. It is particularly useful when the signal explicitly depends on the unknown parameter. LSE satisfies no optimality criteria in general but is the Maximum Likelihood Estimator (MLE) in the case when observations are made in the prescence of additive Gaussian noise.

In Section 5.2 the Maximum-Likelihood based channel estimation is described and in Section 5.3 the two-step estimation approaches are described where the search space for MLE is reduced in the first step. Lastly, Section 5.4 describes the novel Low-Complexity MLE technique.

5.2 Maximum-Likelihood (ML) Estimation

Maximum Likelihood Estimator (MLE) is perhaps the most widely used estimator. In order to find the MLE, we have a well-defined method. The MLE of the parameter $\boldsymbol{\alpha}$ is the one that will maximize the likelihood $p(y; \boldsymbol{\alpha})$. If MVUE exists then the maximum likelihood procedure gaurantees to produce it. For the case of additive Gaussian noise the MLE is asymptotically (i.e., for large data records) the MVUE.

The linear model in our UWB channel estimation problem is

$$\mathbf{y} = \boldsymbol{\Psi}\boldsymbol{\alpha} + \boldsymbol{\omega} \qquad (5.1)$$

where $\boldsymbol{\omega} \sim \mathcal{N}(0, N_0\mathbf{I})$. Therefore, we can write the likelihood of $\boldsymbol{\alpha}$ as

$$p(\mathbf{y}|\boldsymbol{\alpha}) = \frac{1}{(2\pi)^{M/2}} \frac{\exp\left(-\frac{1}{2N_0}(\mathbf{y} - \boldsymbol{\Psi}\boldsymbol{\alpha})^H (\mathbf{y} - \boldsymbol{\Psi}\boldsymbol{\alpha})\right)}{det(N_0\mathbf{I})^{1/2}} \qquad (5.2)$$

$$= \frac{1}{(2\pi N_0^2)^{M/2}} \exp\left(-\frac{1}{2N_0^2}\|\mathbf{y} - \boldsymbol{\Psi}\boldsymbol{\alpha}\|^2\right) \qquad (5.3)$$

Equivalently the log-likelihood upto an irrelevant proportionality factor is given by

$$\ln p(\mathbf{y}|\boldsymbol{\alpha}) = \|\mathbf{y} - \boldsymbol{\Psi}\boldsymbol{\alpha}\|^2 \qquad (5.4)$$

Therefore, the MLE estiamte of $\boldsymbol{\alpha}$ denoted by $\hat{\boldsymbol{\alpha}}_{ML}$ is found by maximizing Eq. (5.2) or equivalently minimizing Eq. (5.4).

$$\hat{\boldsymbol{\alpha}}_{ML} = \arg\min_{\boldsymbol{\alpha}} \|\mathbf{y} - \boldsymbol{\Psi}\boldsymbol{\alpha}\|^2 \qquad (5.5)$$

Based on the channel decomposition, the estimation of the channel support and the amplitudes are de-coupled and we can re-write the model as

$$\mathbf{y} = \boldsymbol{\Psi}_{\underline{\alpha}}\mathbf{a}_{\underline{\alpha}} + \boldsymbol{\omega} \qquad (5.6)$$

Since $\boldsymbol{\omega}$ is Gaussian the MLE of $\mathbf{a}_{\underline{\alpha}}$ is given by as a LSE

$$\hat{\mathbf{a}}_{\underline{\alpha}_{ML}} = \left(\boldsymbol{\Psi}_{\underline{\alpha}}^H\boldsymbol{\Psi}_{\underline{\alpha}}\right)^{-1}\boldsymbol{\Psi}_{\underline{\alpha}}^H\mathbf{y} \qquad (5.7)$$

Since $\mathbf{a}_{\underline{\alpha}}$ depends on the estimate of the support $\underline{\alpha}$, therefore, we first need to estimate the support and then use Eq. (5.7) to estimate the amplitudes. The estimation of the support is a non-linear optimization problem as

$$\hat{\underline{\boldsymbol{\alpha}}}_{ML} = \arg\min_{\underline{\boldsymbol{\alpha}}\in\aleph} \quad \mathbf{y}^H\Pi^{\perp}_{\boldsymbol{\Psi}_{\underline{\alpha}}}\mathbf{y} \qquad (5.8)$$

where $\aleph$ is the set over which the search is performed and $\Pi^{\perp}_{\boldsymbol{\Psi}_{\underline{\alpha}}}$ is the orthogonal projector onto the column space of $\boldsymbol{\Psi}_{\underline{\alpha}}$

$$\Pi^{\perp}_{\boldsymbol{\Psi}_{\underline{\alpha}}} = \mathbf{I} - \boldsymbol{\Psi}_{\underline{\alpha}}\left[\boldsymbol{\Psi}_{\underline{\alpha}}^H\boldsymbol{\Psi}_{\underline{\alpha}}\right]^{-1}\boldsymbol{\Psi}_{\underline{\alpha}}^H \qquad (5.9)$$

69

We note here that the solution to Eq. (5.8) is computationally very complex as it requires a search over the entire set $\aleph$ which contains all the 2^N possible support vectors. Therefore, in the next section we attempt at overcoming this problem by adopting a two-step estimation approach. The first step is to reduce the search space and in the second step the MLE is found within the reduced space.

5.3   Two Step Estimation

5.3.1   Correlation followed by ML Estimation

In the first step, we apply Correlation technique to obtain the likely support of the channel. From the Correlation based estimate of the channel we retain the $L_{cr} > L_{max}$ largest estimates and the corresponding support is used to form the reduced search space $\aleph^{cr}$. The MLE of the support is thus found as

$$\hat{\underline{\boldsymbol{\alpha}}}_{ML} = \arg \min_{\underline{\boldsymbol{\alpha}} \in \aleph^{cr}} \quad \mathbf{y}^H \Pi^{\perp}_{\boldsymbol{\Psi}_{\underline{\boldsymbol{\alpha}}}} \mathbf{y} \tag{5.10}$$

We note that now the search is over a reduced space which contains $2^{L_{cr}}$ possible support vectors. Figure 5.1 and Figure 5.2 show the performance of this two-step channel estimation approach.

5.3.2   CS followed by ML Estimation

In the first step, we apply CS to obtain the likely support of the channel. From the CS estimate of the channel we retain the $L_{cs} > L_{max}$ largest estimates and the

corresponding support is used to form the reduced search space $\aleph^{cs}$. The MLE of the support is thus found as

$$\hat{\underline{\alpha}}_{ML} = \arg \min_{\underline{\alpha} \in \aleph^{cs}} \quad \mathbf{y}^H \Pi_{\underline{\Psi}_{\underline{\alpha}}}^{\perp} \mathbf{y} \tag{5.11}$$

We note that now the search is over a reduced space which contains $2^{L_{cs}}$ possible support vectors. Figure 5.1 and Figure 5.2 show the performance of this two-step channel estimation approach.

5.3.3 GA followed by ML Estimation

In the first step, we apply GA to obtain the likely support of the channel. From the GA estimate of the channel we retain the $L_{ga} > L_{max}$ largest estimates and the corresponding support is used to form the reduced search space $\aleph^{ga}$. The MLE of the support is thus found as

$$\hat{\underline{\alpha}}_{ML} = \arg \min_{\underline{\alpha} \in \aleph^{ga}} \quad \mathbf{y}^H \Pi_{\underline{\Psi}_{\underline{\alpha}}}^{\perp} \mathbf{y} \tag{5.12}$$

We note that now the search is over a reduced space which contains $2^{L_{ga}}$ possible support vectors. Figure 5.1 and Figure 5.2 show the performance of this two-step channel estimation approach.

71

5.4 Low-Complexity (LC) ML Estimation

The MLE of the support involves the minimization of the metric $\mathbf{y}^H \Pi_{\mathbf{\Psi}_{\underline{\alpha}}}^\perp \mathbf{y}$ in Eq. (5.8) where,

$$\mathbf{y}^H \Pi_{\mathbf{\Psi}_{\underline{\alpha}}}^\perp \mathbf{y} = \|\mathbf{y}\|^2 - \mathbf{y}^H \mathbf{\Psi}_{\underline{\alpha}} \left[\mathbf{\Psi}_{\underline{\alpha}}^H \mathbf{\Psi}_{\underline{\alpha}} \right]^{-1} \mathbf{\Psi}_{\underline{\alpha}}^H \mathbf{y} \tag{5.13}$$

Therefore, the log-likelihood of $\underline{\alpha}$ is to be maximized to obtain the MLE of $\underline{\alpha}$, given by

$$\hat{\underline{\alpha}}_{ML} = \arg \max_{\underline{\alpha}} \mathcal{L}(\underline{\alpha}) \tag{5.14}$$

where,

$$\mathcal{L}(\underline{\alpha}) = \mathbf{y}^H \mathbf{\Psi}_{\underline{\alpha}} \left[\mathbf{\Psi}_{\underline{\alpha}}^H \mathbf{\Psi}_{\underline{\alpha}} \right]^{-1} \mathbf{\Psi}_{\underline{\alpha}}^H \mathbf{y} \tag{5.15}$$

In the following we leverage the structure of $\mathbf{\Psi}$ to reduce the computational complexity in evaluating Eq. (5.15). It was shown that the matrix $\mathbf{\Psi}$ can be divided into a number of nearly orthogonal clusters. Let there be C such orthogonal clusters denoted by $\mathbf{\Psi}_{\Theta_r}$ for $r = 1, 2, \ldots, C$. The matrix $\mathbf{\Psi}_{\Theta_r}$ is composed of the columns belonging to the rth orthogonal cluster and is of size $N \times \nu$ since the length of each cluster is ν. Now we can express $\mathbf{\Psi}_{\underline{\alpha}}$ in a block matrix form as,

$$\mathbf{\Psi}_{\underline{\alpha}} = [\mathbf{\Psi}_{\Theta_1} \quad \mathbf{\Psi}_{\Theta_2} \ldots \mathbf{\Psi}_{\Theta_C}] \tag{5.16}$$

where each sub-matrix $\boldsymbol{\Psi}_{\Theta_r}$ consists of those columns of $\boldsymbol{\Psi}$ at the corresponding locations which are indicated by $\underline{\boldsymbol{\alpha}}$. The locations in $\boldsymbol{\Psi}_{\Theta_r}$ which are not indicated by $\underline{\boldsymbol{\alpha}}$ contain zero cloumns.

Due to the orhtogonality of the clusters, the inverse term that appears in Eq. (5.15) becomes the inverse of the block diagonal matrix,

$$
\left[\boldsymbol{\Psi}_{\underline{\boldsymbol{\alpha}}}^{H}\boldsymbol{\Psi}_{\underline{\boldsymbol{\alpha}}}\right]^{-1} = \begin{bmatrix} (\boldsymbol{\Psi}_{\Theta_1}^{H}\boldsymbol{\Psi}_{\Theta_1})^{-1} & 0 & \cdots & 0 \\ 0 & (\boldsymbol{\Psi}_{\Theta_2}^{H}\boldsymbol{\Psi}_{\Theta_2})^{-1} & \cdots & \vdots \\ \vdots & 0 & \ddots & 0 \\ 0 & \vdots & \cdots & (\boldsymbol{\Psi}_{\Theta_C}^{H}\boldsymbol{\Psi}_{\Theta_C})^{-1} \end{bmatrix} \tag{5.17}
$$

Furthermore, using the Toeplitz structure of $\boldsymbol{\Psi}$ it is also easy to show that, for $r = 1, 2, \ldots, C$;

$$
(\boldsymbol{\Psi}_{\Theta_r}^{H}\boldsymbol{\Psi}_{\Theta_r})^{-1} = (\boldsymbol{\Psi}_{\Theta_1}^{H}\boldsymbol{\Psi}_{\Theta_1})^{-1} \tag{5.18}
$$

Therefore, we can write the log-likelihood as,

$$
\mathcal{L}(\underline{\boldsymbol{\alpha}}) = \mathcal{L}^{\boldsymbol{\Psi}_{\Theta_1}}(\underline{\boldsymbol{\alpha}}) + \mathcal{L}^{\boldsymbol{\Psi}_{\Theta_2}}(\underline{\boldsymbol{\alpha}}) + \cdots + \mathcal{L}^{\boldsymbol{\Psi}_{\Theta_C}}(\underline{\boldsymbol{\alpha}}) \tag{5.19}
$$

where $\mathcal{L}^{\boldsymbol{\Psi}_{\Theta_r}}(\underline{\boldsymbol{\alpha}})$ is the likelihood over the rth cluster and given by,

$$
\mathcal{L}^{\boldsymbol{\Psi}_{\Theta_r}}(\underline{\boldsymbol{\alpha}}) = \mathbf{y}^{H}\boldsymbol{\Psi}_{\Theta_r}(\boldsymbol{\Psi}_{\Theta_r}^{H}\boldsymbol{\Psi}_{\Theta_r})^{-1}\boldsymbol{\Psi}_{\Theta_r}^{H}\mathbf{y} \tag{5.20}
$$

$$
= \mathbf{y}_{\Theta_r}^{H}\overline{\boldsymbol{\Psi}}_{\Theta_r}(\boldsymbol{\Psi}_{\Theta_r}^{H}\boldsymbol{\Psi}_{\Theta_r})^{-1}\overline{\boldsymbol{\Psi}}_{\Theta_r}^{H}\mathbf{y}_{\Theta_r} \tag{5.21}
$$

where $\mathbf{y}_{\Theta_r}$ is a $\nu \times 1$ vector and is the masked version of $\mathbf{y}$ which includes only those elements of $\mathbf{y}$ which correspond to the rth cluster, $\overline{\mathbf{\Psi}}_{\Theta_r}$ is the $\nu \times \nu$ sub-matrix of $\mathbf{\Psi}_{\Theta_r}$ composed of its non-zero portion for the rth cluster. We note here that due to the Toeplitz structure, we have

$$\overline{\mathbf{\Psi}}_{\Theta_r} = \overline{\mathbf{\Psi}}_{\Theta_1} \tag{5.22}$$

for $r = 1, 2, \ldots, C$. Therefore, we need to perform the matrix inversion only for the first cluster to find $(\mathbf{\Psi}_{\Theta_1}^H \mathbf{\Psi}_{\Theta_1})^{-1}$ and can use it to calculate the log-likelihoods for all the clusters in Eq. (5.19). Thus the Low-Complexity MLE of the support is,

$$\hat{\underline{\boldsymbol{\alpha}}}_{LCML} = \arg \max_{\underline{\boldsymbol{\alpha}} \in \aleph} \quad \mathcal{L}(\underline{\boldsymbol{\alpha}}) \tag{5.23}$$

where,

$$\mathcal{L}(\underline{\boldsymbol{\alpha}}) = \mathbf{y}_{\Theta_1}^H \mathbf{\Sigma} \mathbf{y}_{\Theta_1} + \mathbf{y}_{\Theta_2}^H \mathbf{\Sigma} \mathbf{y}_{\Theta_2} + \cdots + \mathbf{y}_{\Theta_r}^H \mathbf{\Sigma} \mathbf{y}_{\Theta_r} \tag{5.24}$$

$$= \|\mathbf{y}_{\Theta_1}\|_{\mathbf{\Sigma}}^2 + \|\mathbf{y}_{\Theta_2}\|_{\mathbf{\Sigma}}^2 + \cdots + \|\mathbf{y}_{\Theta_r}\|_{\mathbf{\Sigma}}^2 \tag{5.25}$$

and

$$\mathbf{\Sigma} = \overline{\mathbf{\Psi}}_{\Theta_1} (\mathbf{\Psi}_{\Theta_1}^H \mathbf{\Psi}_{\Theta_1})^{-1} \overline{\mathbf{\Psi}}_{\Theta_1}^H \tag{5.26}$$

We note that in computing the LCML estimate of the support, the term Σ needs to be computed just once and that for the first cluster. After which we can compute the log-likelihoods for all the remaining orthogonal clusters by simply masking $\mathbf{y}$ and calculating its weighted norm, as shown in Eq. (5.25). Figure 5.1 shows the estimation performance of LCML and Figure 5.2 shows the energy capture.

5.5 Results
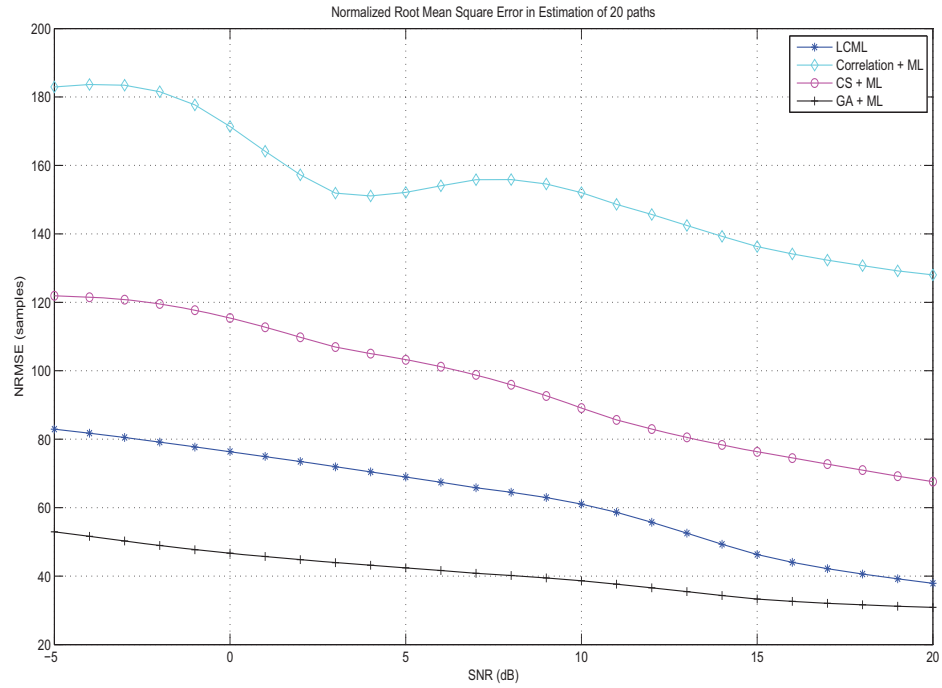
Figure 5.1: NRMSE Performance in Estimation of MPCs Arrival Time
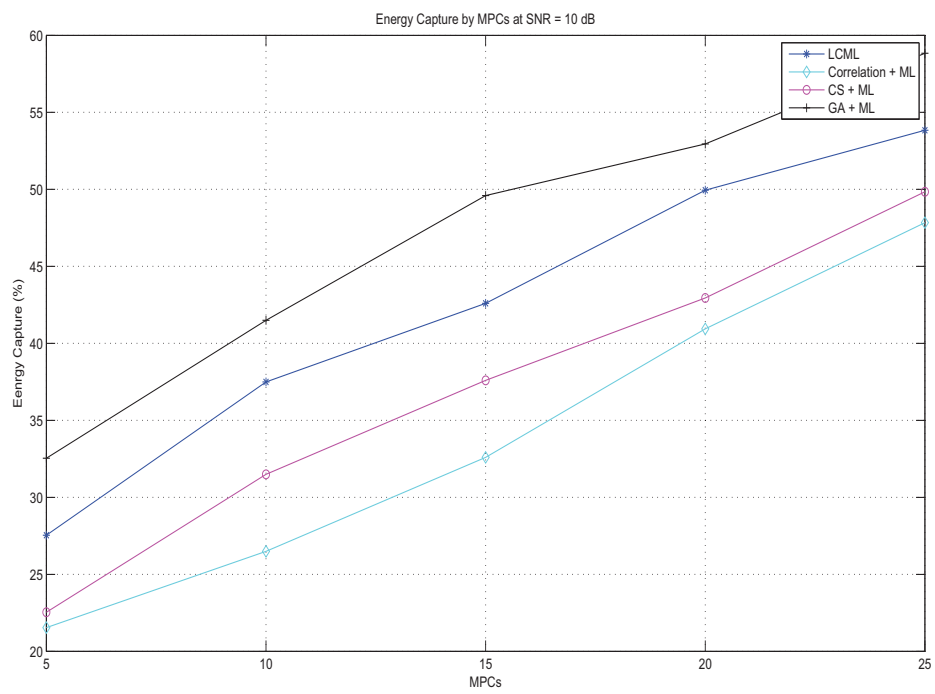
Figure 5.2: Performance in terms of Energy Capture

CHAPTER 6

BAYESIAN ESTIMATION

6.1 Introduction

In this chapter we present the estimation of the UWB channel in a Bayesian Framework. In the Bayesian framework the unknown channel parameters, the channel support vector $\underline{\boldsymbol{\alpha}}$ and the fading co-efficient vector $\mathbf{a}_{\underline{\alpha}}$ are assumed to be random quantities. We remark here that for Bayesian estimation of the UWB channel we need to know the start of the received signal profile in order to incorporate the amplitudes' statistical information from the Average Power Delay Profile (APDP) of the channel given in the IEEE 802.15.4a Standard [1]. Therefore, for all the estimation methods presented in this Chapter we assume that the arrival of the first MPC is estimated (known) prior to performing the channel estimation. The estimation of the arrival time of the first MPC is known as Time-of-Arrival (TOA) estimation. There are several methods reported in literature that estimate the TOA of a received UWB signal profile in various

channel environments such as [51], [52] and [53].

We adopt two approaches to estimate the UWB channel in the Bayesian framework:

1. Decomposed Channel Estimation

   In this approach the channel is estimated for the decomposed channel , where first the channel support is estimated and then the corresponding channel fading co-efficients are estimated.

2. Joint Channel Estimation

   In this approach the channel support and the channel fading co-efficients are jointly estimated.

6.2   A Priori Information

In Bayesian estimation of the UWB channel, a priori statistical information of the channel is incorporated into the estimation process. In the following we look into the statistical prior knowledge about the UWB channel.

6.2.1   Channel Support

We first consider the channel support, denoted by $\underline{\boldsymbol{\alpha}}$ which is the unknown random vector with a prior probability density function $p(\underline{\boldsymbol{\alpha}})$. We expect to improve the estimate by incorporating this *a priori* information. From Eq. (3.9), it follows

that the elements of $\underline{\boldsymbol{\alpha}}$ are independent Bernoulli trials, with

$$P_b\left(\underline{\alpha}_i = 1\right) = \mathsf{p}_b \tag{6.1}$$

$$P_b\left(\underline{\alpha}_i = 0\right) = \left(1 - \mathsf{p}_b\right) \tag{6.2}$$

for $i = 1, 2, \ldots, (N - 1)$ and $\underline{\alpha}_i$ is the ith element of $\underline{\boldsymbol{\alpha}}$. Therefore, for a certain support with $\|\underline{\boldsymbol{\alpha}}\|_0$ number of non-zero elements, $p(\underline{\boldsymbol{\alpha}})$ can be calculated as,

$$p(\underline{\boldsymbol{\alpha}}) = \mathsf{p}_b^{\|\underline{\boldsymbol{\alpha}}\|_0}\left(1 - \mathsf{p}_b\right)^{N - \|\underline{\boldsymbol{\alpha}}\|_0} \tag{6.3}$$

where $\|.\|_0$ deontes the l_0-norm (i.e., the number of non-zero entries) of the vector. We note that $p(\underline{\boldsymbol{\alpha}})$ depends on the l_0-norm of $\underline{\boldsymbol{\alpha}}$ which corresponds to the number of MPCs present in the received profile.

We need to determine the likelihood function of $\underline{\boldsymbol{\alpha}}$ from the model given in Eq. (3.14) which we reproduce here,

$$\mathbf{y} = \boldsymbol{\Psi}\mathbf{A}\underline{\boldsymbol{\alpha}} + \boldsymbol{\omega} \tag{6.4}$$

It is clear that the likelihood function $p(\mathbf{y}|\underline{\boldsymbol{\alpha}})$ depends on the statistics of $\mathbf{A}$ or equivalently $\mathbf{a}_{\underline{\alpha}}$. Therefore, we consider the different statistics of the amplitude vector $\mathbf{a}_{\underline{\alpha}}$ to determine $p(\mathbf{y}|\underline{\boldsymbol{\alpha}})$.

6.2.2 Channel Amplitudes

The vector of unknown channel amplitudes or fading co-efficients denoted by $\mathbf{a}_{\underline{\alpha}}$ is assumed to be random with a prior probability density function $p(\mathbf{a}_{\underline{\alpha}})$. Conditioned on the support the channel amplitudes are estimated from the model given in Eq. (3.15), which we reproduce here,

$$\mathbf{y} = \mathbf{\Psi}_{\underline{\alpha}}\mathbf{a}_{\underline{\alpha}} + \boldsymbol{\omega} \tag{6.5}$$

The likelihood function $p(\mathbf{y}|\underline{\alpha})$ is needed to perform the channel support estimation, the correpsonding channel amplitudes estimation and also the joint channel estimation, as will be discussed in the subsequent Sections of this Chapter. We consider three cases for the statistics of the channel fading amplitudes and determine the likelihood function $p(\mathbf{y}|\underline{\alpha})$ for each of three cases.

1. Non-Gaussian Amplitudes

   We first consider the case when the fading amplitudes do not follow a Gaussian distribution. This means that the elements of $\mathbf{a}_{\underline{\alpha}}$ are non-Gausian and corresponds to the channel fading statistics of the IEEE UWB channel model [5], where the small-scale fading co-efficients are modeled as Nakagami distributed, as given in Eq. (2.13). Therefore, the vector $\mathbf{a}_{\underline{\alpha}}$ in Eq. (6.5) is now non-Gaussian and so it is difficult to obtain the expression for the joint probability distribution of $\mathbf{a}_{\underline{\alpha}}$ and hence the likelihood $p(\mathbf{y}|\underline{\alpha})$ in a closed form. For a certain support vector $\underline{\boldsymbol{\alpha}}$, the best we can say is that the re-

ceived vector $\mathbf{y}$ lies in the subspace spanned by the columns of $\boldsymbol{\Psi}_{\underline{\boldsymbol{\alpha}}}$, plus an AWGN vector $\boldsymbol{\omega}$. Thus, the orthogonal projection of $\mathbf{y}$ onto the orthogonal complement of $\boldsymbol{\Psi}_{\underline{\boldsymbol{\alpha}}}$ is Gaussian. Specifically, the vector $\Pi^{\perp}_{\boldsymbol{\Psi}_{\underline{\boldsymbol{\alpha}}}}\mathbf{y}$ is Gaussian. Therefore, the likelihood $p(\mathbf{y}|\underline{\boldsymbol{\alpha}})$ can be approximated by,

$$p(\mathbf{y}|\underline{\boldsymbol{\alpha}}) \propto \exp\left(-\frac{1}{2N_0}\|\Pi^{\perp}_{\boldsymbol{\Psi}_{\underline{\boldsymbol{\alpha}}}}\mathbf{y}\|_2^2\right) \tag{6.6}$$

where,

$$\Pi^{\perp}_{\boldsymbol{\Psi}_{\underline{\boldsymbol{\alpha}}}} = \mathbf{I} - \boldsymbol{\Psi}_{\underline{\boldsymbol{\alpha}}}\left[\boldsymbol{\Psi}_{\underline{\boldsymbol{\alpha}}}^H\boldsymbol{\Psi}_{\underline{\boldsymbol{\alpha}}}\right]^{-1}\boldsymbol{\Psi}_{\underline{\boldsymbol{\alpha}}}^H \tag{6.7}$$

2. Non-Gaussian Amplitudes with Known 2nd Order Statistics

   Now, we consider the case where the amplitude vector $\mathbf{a}_{\underline{\boldsymbol{\alpha}}}$ is non-Gaussian but its second order statistics are known.

In the following we motivate one way of incorporating the second order statistics by modifying the sensing matrix. Since the vector $\boldsymbol{\alpha}$ is sparse, therefore we can assume a sparsity inducing prior on $\boldsymbol{\alpha}$, as is widely used in the sparse signal recovery literature. One such popular prior is the Laplacian prior which is used in the famous Basis Pursuit De-noising (BPDN) and the Least Absolutle Shrinkage and Selection Operator (LASSO) algorithms for sparse signal recovery. The LASSO or

the equivalent BPDN for the model in Eq. (3.11) is given as,

$$\hat{\boldsymbol{\alpha}} = \arg\min_{\boldsymbol{\alpha}} \quad \frac{1}{N_0}\|\mathbf{y} - \boldsymbol{\Psi}\boldsymbol{\alpha}\|_2^2 + \rho\|\boldsymbol{\alpha}\|_1 \qquad (6.8)$$

From the Bayesian perspective $\hat{\boldsymbol{\alpha}}$ in Eq. (6.8) is the MAP estimate of $\boldsymbol{\alpha}$ in the prescence of AWGN and with a Laplacian prior on $\boldsymbol{\alpha}$ having $i.i.d$ entries distributed with zero-mean and variance $\frac{2N_o}{\rho}$. In the same spirit if we assume a general sparsity inducing prior on $\boldsymbol{\alpha}$ as,

$$p(\boldsymbol{\alpha}) \backsim \exp\left(-\frac{1}{2}\sum_{j=0}^{L-1} \frac{|\alpha_j|}{b_j}\right) \qquad (6.9)$$

where the b_j's are positive user parameters. The MAP estimate of $\boldsymbol{\alpha}$ in AWGN becomes,

$$\hat{\boldsymbol{\alpha}} = \arg\min_{\boldsymbol{\alpha}} \quad \frac{1}{N_0}\|\mathbf{y} - \boldsymbol{\Psi}\boldsymbol{\alpha}\|_2^2 + \mathbf{B}\|\boldsymbol{\alpha}\|_1 \qquad (6.10)$$

where $\mathbf{B}$ is a diagonal matrix $\mathbf{B} = diag[\frac{1}{b_0} \frac{1}{b_1} \cdots \frac{1}{b_{L-1}}]$. By applying a change of variable $\boldsymbol{\beta} = \mathbf{D}\boldsymbol{\alpha}$, we obtain,

$$\hat{\boldsymbol{\beta}} = \arg\min_{\boldsymbol{\beta}} \quad \frac{1}{N_0}\|\mathbf{y} - \boldsymbol{\Psi}\boldsymbol{\beta}\|_2^2 + \|\boldsymbol{\beta}\|_1 \qquad (6.11)$$

$$= \arg\min_{\boldsymbol{\beta}} \quad \frac{1}{N_0}\|\mathbf{y} - \boldsymbol{\Psi}\mathbf{B}\boldsymbol{\alpha}\|_2^2 + \|\boldsymbol{\beta}\|_1 \qquad (6.12)$$

If we set $\mathbf{B} = \mathbf{D}$, where $\mathbf{D}$ is given by Eq. (2.12), then the MAP estimate

of $\boldsymbol{\beta}$ from Eq. (6.11) becomes

$$\hat{\boldsymbol{\beta}} = \arg\min_{\boldsymbol{\beta}} \quad \frac{1}{N_0}\|\mathbf{y} - \boldsymbol{\Psi}\mathbf{D}\boldsymbol{\alpha}\|_2^2 + \|\boldsymbol{\beta}\|_1 \tag{6.13}$$

where $\boldsymbol{\beta}$ is now given by,

$$\boldsymbol{\beta} = \mathbf{D}\boldsymbol{\alpha} \tag{6.14}$$

The above suggests that one way to incorporate the second order statistics is to modify the sensing matrix $\boldsymbol{\Psi}$ by absorbing the effect of the variance into the sensing matrix, that is,

$$\mathbf{H} = \boldsymbol{\Psi}\mathbf{D} \tag{6.15}$$

Thus the matrix $\mathbf{H}$ absorbs the effect of the variance of the MPC's amplitudes into the respective columns.

From the IEEE 802.15.4a model, we have the a priori information about the average power delay profile of the UWB channel and the second order statistics of the amplitudes are given as shown in Eq. (2.11). Therefore, conditioned on the support $\underline{\boldsymbol{\alpha}}$ we can write,

$$\mathbf{D}_{\mathbf{a}_{\underline{\alpha}}} = \mathbb{E}\left[\mathbf{a}_{\underline{\alpha}}\mathbf{a}_{\underline{\alpha}}^H\right] \tag{6.16}$$

where $\mathbf{D}_{\mathbf{a}_{\underline{\alpha}}}$ is a diagonal matrix. The elements of $\mathbf{a}_{\underline{\alpha}}$ are independent but

non-identically distributed (i.e., their variance are different). Thus the model conditioned on the support vector $\underline{\boldsymbol{\alpha}}$ is,

$$\mathbf{y} = \mathbf{H}_{\underline{\boldsymbol{\alpha}}} \mathbf{a}_{\underline{\boldsymbol{\alpha}}} + \boldsymbol{\omega} \tag{6.17}$$

Now the best we can say is that the received vector $\mathbf{y}$ lies in the subspace spanned by the columns of $\mathbf{H}_{\underline{\boldsymbol{\alpha}}}$, plus an AWGN vector $\boldsymbol{\omega}$. Thus, the orthogonal projection of $\mathbf{y}$ onto the orthogonal complement of $\mathbf{H}_{\underline{\boldsymbol{\alpha}}}$ is Gaussian. Specifically, the vector $\Pi^{\perp}_{\mathbf{H}_{\underline{\boldsymbol{\alpha}}}} \mathbf{y}$ is Gaussian. Therefore, the likelihood $p(\mathbf{y}|\underline{\boldsymbol{\alpha}})$ can be approximated by,

$$p(\mathbf{y}|\underline{\boldsymbol{\alpha}}) \propto \exp\left(-\frac{1}{2N_0} \|\Pi^{\perp}_{\mathbf{H}_{\underline{\boldsymbol{\alpha}}}} \mathbf{y}\|_2^2 \right) \tag{6.18}$$

where,

$$\Pi^{\perp}_{\mathbf{H}_{\underline{\boldsymbol{\alpha}}}} = \mathbf{I} - \mathbf{H}_{\underline{\boldsymbol{\alpha}}} \left[\mathbf{H}_{\underline{\boldsymbol{\alpha}}}^H \mathbf{H}_{\underline{\boldsymbol{\alpha}}} \right]^{-1} \mathbf{H}_{\underline{\boldsymbol{\alpha}}}^H \tag{6.19}$$

3. Gaussian Amplitudes

Now, we consider the case where the channel amplitudes are independent and Gaussian. For UWB communication in a dense multipath environment, such as the industrial environment where there are a large number of scatters, the assumption that there are many MPCs falling in the same bin is fairly valid [39]. Therefore, we assume the vector $\mathbf{a}_{\underline{\boldsymbol{\alpha}}}$ to be Gaussian with

independent but non-identically distributed entries. Therefore, conditioned on the support the amplitudes are Gaussian. This results from Eq. (6.5) and the fact that linear combination of independent Gaussian vectors is also a Gaussian. Thus, $\mathbf{a}_{\underline{\alpha}}$ is a Gaussian vector and the likelihood is given by,

$$p(\mathbf{y}|\underline{\alpha}) = \frac{\exp\left(-\frac{1}{2}\mathbf{y}^H \mathbf{\Sigma}_{\underline{\alpha}}^{-1}\mathbf{y}\right)}{\sqrt{\det(\mathbf{\Sigma}_{\underline{\alpha}})}} \tag{6.20}$$

where $\mathbf{\Sigma}_{\underline{\alpha}}$ is the co-variance matrix conditioned on the support vector $\underline{\alpha}$ and given by,

$$\mathbf{\Sigma}_{\underline{\alpha}} = \mathbb{E}[\mathbf{y}\mathbf{y}^H|\underline{\alpha}] \tag{6.21}$$

$$= N_0\mathbf{I} + \mathbf{\Psi}_{\underline{\alpha}}\mathbb{E}[\mathbf{a}_{\underline{\alpha}}\mathbf{a}_{\underline{\alpha}}^H]\mathbf{\Psi}_{\underline{\alpha}}^H \tag{6.22}$$

$$= N_0\mathbf{I} + \mathbf{\Psi}_{\underline{\alpha}}\mathbf{D}_{\underline{\alpha}}\mathbf{\Psi}_{\underline{\alpha}}^H \tag{6.23}$$

where $\mathbf{D}_{\underline{\alpha}}$ is as given in Eq. (6.16). $\mathbf{D}_{\underline{\alpha}}$ is a diagonal matrix of the variances of the amplitudes corresponding to the support $\underline{\alpha}$.

6.3 Decomposed Channel Estimation

In this section we present the channel estimation for the decomposed channel in a Bayesian framework where we first present the estimation of the channel support followed by the estimation of the channel amplitudes.

6.3.1 Support Estimation

In order to find the most probable support based on the observation $\mathbf{y}$, the maximum of the a posteriori probability (MAP) of $\underline{\boldsymbol{\alpha}}$ is sought. For the MAP estimate, we need the posteriori pdf $p(\underline{\boldsymbol{\alpha}}|\mathbf{y})$ which is given using the Baye's Rule as,

$$p(\underline{\boldsymbol{\alpha}}|\mathbf{y}) = \frac{p(\mathbf{y}, \underline{\boldsymbol{\alpha}})}{p(\mathbf{y})} \tag{6.24}$$

$$= \frac{p(\mathbf{y}|\underline{\boldsymbol{\alpha}})p(\underline{\boldsymbol{\alpha}})}{p(\mathbf{y})} \tag{6.25}$$

The MAP estimate of $\underline{\boldsymbol{\alpha}}$ maximizes $p(\underline{\boldsymbol{\alpha}}|\mathbf{y})$ and is equivalent to the maximization of the numerator in Eq. (6.25), therefore, we need to search for the best estimate of $\underline{\boldsymbol{\alpha}}$ over the entire set $\aleph$. We use the a priori probability $p(\underline{\boldsymbol{\alpha}})$ to obtain the MAP estimate of $\underline{\boldsymbol{\alpha}}$ as,

$$\underline{\hat{\boldsymbol{\alpha}}}_{MAP} = \arg \max_{\underline{\boldsymbol{\alpha}} \in \aleph} p(\mathbf{y}|\underline{\boldsymbol{\alpha}})p(\underline{\boldsymbol{\alpha}}) \tag{6.26}$$

$$= \arg \max_{\underline{\boldsymbol{\alpha}} \in \aleph} p(\mathbf{y}|\underline{\boldsymbol{\alpha}})\, \mathrm{p}_b^{\|\underline{\boldsymbol{\alpha}}\|_0}\, (1 - \mathrm{p}_b)^{N - \|\underline{\boldsymbol{\alpha}}\|_0} \tag{6.27}$$

Equivalently the MAP estimate can be obtained by maximizing the logarithm,

$$\underline{\hat{\boldsymbol{\alpha}}}_{MAP} = \arg \max_{\underline{\boldsymbol{\alpha}} \in \aleph}\ \ln\left[p(\mathbf{y}|\underline{\boldsymbol{\alpha}})\right] + \ln\left(\frac{\mathrm{p}_b}{1 - \mathrm{p}_b}\right)^{\|\underline{\boldsymbol{\alpha}}\|_0} \tag{6.28}$$

where $\ln\left[p(\mathbf{y}|\underline{\boldsymbol{\alpha}})\right]$ is the log-likelihood function of $\underline{\boldsymbol{\alpha}}$. Since the search space $\aleph$ is large we perform a two-step estimation where we first reduce the serch space by applying Compressive Sesnsing and Correlation to obtain the reduced search

86

spaces $\aleph_{cs}$ and $\aleph_{cr}$, respectively. Thus the MAP estimate of the channel support for the three cases of channel amplitudes is found by using the respective likelihood function $p(\mathbf{y}|\underline{\boldsymbol{\alpha}})$ as follows:

1. Non-Gaussian Amplitudes:

$$\hat{\underline{\boldsymbol{\alpha}}}_{MAP} = \arg\max_{\underline{\boldsymbol{\alpha}} \in \aleph^\star} -\frac{1}{2N_0} \|\Pi^{\perp}_{\boldsymbol{\Psi}_{\underline{\boldsymbol{\alpha}}}} \mathbf{y}\|_2^2 + \ln\left(\frac{p_b}{1 - p_b}\right)^{\|\underline{\boldsymbol{\alpha}}\|_0} \tag{6.29}$$

2. Non-Gaussian Amplitudes with Known 2nd Order Statistics:

$$\hat{\underline{\boldsymbol{\alpha}}}_{MAP} = \arg\max_{\underline{\boldsymbol{\alpha}} \in \aleph^\star} -\frac{1}{2N_0} \|\Pi^{\perp}_{\mathbf{H}_{\underline{\boldsymbol{\alpha}}}} \mathbf{y}\|_2^2 + \ln\left(\frac{p_b}{1 - p_b}\right)^{\|\underline{\boldsymbol{\alpha}}\|_0} \tag{6.30}$$

3. Gaussian Amplitudes:

$$\hat{\underline{\boldsymbol{\alpha}}}_{MAP} = \arg\max_{\underline{\boldsymbol{\alpha}} \in \aleph_\star} -\mathbf{y}^H \boldsymbol{\Sigma}_{\underline{\boldsymbol{\alpha}}}^{-1} \mathbf{y} - \ln\left[\det(\boldsymbol{\Sigma}_{\underline{\boldsymbol{\alpha}}})\right] + \ln\left(\frac{p_b}{1 - p_b}\right)^{\|\underline{\boldsymbol{\alpha}}\|_0} \tag{6.31}$$

where in Eq. (6.29) - Eq. (6.31) the search space $\aleph_\star$ is $\aleph_{cs}$ when the reduction in the search space in the first step is done using Compressive Sensing and the support estimate is called as CSMAP (Compressive Sensing based MAP) estimate. Similarly, $\aleph_\star$ is $\aleph_{cr}$ when the reduction in the search space in the first step is done using Correlation and the support estimate is called as CRMAP (Correlation based MAP) estimate. Figure 6.1 shows the performance of CRMAP and CSMAP in estimating the channel support for the three amplitudes cases.

In general the number of MPCs, i.e., $\|\underline{\boldsymbol{\alpha}}\|_0$, is not known a priori at the receiver and so Eq. (6.28) is maximized over $\|\underline{\boldsymbol{\alpha}}\|_0 = 0, 1, \ldots, L_{max}$. When the number of MPCs is known, or, we want to estimate a fixed number of MPCs, then $p(\underline{\boldsymbol{\alpha}})$ is uniform and the MAP estimate of the support reduces to the Maximum Likelihood (ML) estimate of Eq. (5.8).

6.3.2 Amplitudes Estimation

Once the channel support has been estimated the corresponding amplitudes of the channel are estimated for each of the three cases of the amplitudes.

1. Non-Gaussian Amplitudes:

   When the amplitudes are non-Gaussian and we have the MAP estimate of the channel support $\hat{\underline{\boldsymbol{\alpha}}}_{MAP}$, we have an over-determined system. Since the amplitudes are non-Gaussian the best estimate of the amplitudes is the Least-Squares Estimation (LSE).

   $$\hat{\mathbf{a}}_{\underline{\boldsymbol{\alpha}}_{LS}} = (\boldsymbol{\Psi}_{\underline{\boldsymbol{\alpha}}}^H \boldsymbol{\Psi}_{\underline{\boldsymbol{\alpha}}})^{-1} \boldsymbol{\Psi}_{\underline{\boldsymbol{\alpha}}}^H \mathbf{y} \tag{6.32}$$

   where $\boldsymbol{\Psi}_{\underline{\boldsymbol{\alpha}}}$ represents the matrix comprised of the columns of $\boldsymbol{\Psi}$ which are indicated by $\underline{\boldsymbol{\alpha}} = \hat{\underline{\boldsymbol{\alpha}}}_{MAP}$. Figure 6.5 shows the performance in terms of energy capture.

2. Non-Gaussian Amplitudes with Known 2nd Order Statistics

   When the amplitudes are non-Gaussian with second order statistics, and we

have the MAP estimate of the channel support $\hat{\underline{\alpha}}_{MAP}$, we have the over-determined system where the second order statistics are effectively absorbed in the model as described previously. Since the amplitudes are non-Gaussian the best estimate of the amplitudes is the Least-Squares Estimation (LSE)

$$\hat{\mathbf{a}}_{\underline{\alpha}_{LS}} = (\mathbf{H}_{\underline{\alpha}}^H \mathbf{H}_{\underline{\alpha}})^{-1} \mathbf{H}_{\underline{\alpha}}^H \mathbf{y} \tag{6.33}$$

where $\mathbf{H}_{\underline{\alpha}}$ represents the matrix comprised of the columns of $\mathbf{H}$ which are indicated by $\underline{\alpha} = \hat{\underline{\alpha}}_{MAP}$. Figure 6.5 shows the performance in terms of energy capture.

3. Gaussian Amplitudes

When the amplitudes are Gaussian and we have the MAP estimate of the channel support $\hat{\underline{\alpha}}_{MAP}$, we have the over-determined system. Since the amplitudes are Gaussian the best estimate of the amplitudes is the Minimum Mean Square Error (MMSE) estimate.

$$\hat{\mathbf{a}}_{\underline{\alpha}_{MMSE}} = \mathbf{D}_{\underline{\alpha}} \mathbf{\Psi}_{\underline{\alpha}}^H \, \mathbf{\Sigma}_{\underline{\alpha}}^{-1} \, \mathbf{a}_{\underline{\alpha}} \tag{6.34}$$

where $\mathbf{\Psi}_{\underline{\alpha}}$ represents the matrix comprised of the columns of $\mathbf{\Psi}$ which are indicated by $\underline{\alpha}$ and $\mathbf{\Sigma}_{\underline{\alpha}}$ is the co-variance matrix given in Eq. (??) for $\underline{\alpha} = \hat{\underline{\alpha}}_{MAP}$. Figure 6.5 shows the performance in terms of energy capture.

6.4   Joint Channel Estimation

Now we discuss, in a Bayesian framework, the joint estimation of the support and the corresponding amplitudes of the channel which from now onwards we call simply channel estimation. The channel estimation refers to the estimation of the vector $\boldsymbol{\alpha}$. We determine the estimate of the channel for the model given in Eq. (3.11) which we reproduce here,

$$\mathbf{y} = \boldsymbol{\Psi}\boldsymbol{\alpha} + \boldsymbol{\omega} \tag{6.35}$$

The Minimum Mean Square Error (MMSE) estimate of the channel is given as follows,

$$\hat{\boldsymbol{\alpha}}_{MMSE} = \sum_{\underline{\boldsymbol{\alpha}} \in \aleph} p(\underline{\boldsymbol{\alpha}}|\mathbf{y})\mathbb{E}(\mathbf{a}_{\underline{\boldsymbol{\alpha}}}|\,\mathbf{y}) \tag{6.36}$$

In calculating the above MMSE estimate of the channel we need to evaluate $\mathbb{E}(\mathbf{a}_{\underline{\boldsymbol{\alpha}}}|\mathbf{y})$ corresponding to every support vector $\underline{\boldsymbol{\alpha}}$ and weight it by that support's posterior probability $p(\underline{\boldsymbol{\alpha}}|\mathbf{y})$ and eventually sum these to obtain the MMSE estimate of the channel. Since, the space $\aleph$ is large and contains 2^N vectors we proceed to find the approximate MMSE estimate of the channel by reducing the space to the most likely candidates obtained from Compressive Sensing and Correlation based support estimation. Thus the approximate MMSE of the channel

is given by,

$$\hat{\boldsymbol{\alpha}}_{AMMSE} = \sum_{\underline{\boldsymbol{\alpha}} \in \aleph_\star} p(\underline{\boldsymbol{\alpha}}|\mathbf{y}) \mathbb{E}(\mathbf{a}_{\underline{\boldsymbol{\alpha}}}|\mathbf{y}) \qquad (6.37)$$

where the search space $\aleph_\star$ is $\aleph_{cs}$ when the reduction in the search space in the first step is done using Compressive Sensing and the channel estimate is called as CSAMMSE (Compressive Sensing based Approximate MMSE) estimate. Similarly, $\aleph_\star$ is $\aleph_{cr}$ when the reduction in the search space in the first step is done using Correlation and the channel estimate is called as CRAMMSE (Correlation based Approximate MMSE) estimate.

In calculating the AMMSE estimate of the channel from Eq. (6.37) we proceed by finding the a posteriori probability of the support vectors $p(\underline{\boldsymbol{\alpha}}|\mathbf{y})$ alongwith the corresponding estimate of the amplitudes for each of the support vectors in $\aleph_\star$. Each estimate of the amplitude is weighted by the corresponding a posteriori probability of its support and normalized over all the support vectors in $\aleph_\star$. In the following we present the AMMSE estimation for each of the three cases of the channel amplitudes:

1. Non-Gaussian Amplitudes

   When the amplitudes are non-Gaussian $\mathbb{E}(\mathbf{a}_{\underline{\boldsymbol{\alpha}}}|\mathbf{y})$ is approximated with the Least-Squares Estimation (LSE).

   $$\mathbb{E}(\mathbf{a}_{\underline{\boldsymbol{\alpha}}}|\mathbf{y}) = (\boldsymbol{\Psi}_{\underline{\boldsymbol{\alpha}}}^H \boldsymbol{\Psi}_{\underline{\boldsymbol{\alpha}}})^{-1} \boldsymbol{\Psi}_{\underline{\boldsymbol{\alpha}}}^H \mathbf{y} \qquad (6.38)$$

The posterior density function $p(\underline{\boldsymbol{\alpha}}|\mathbf{y})$ is given according to the Baye's Rule and is approximated by normalizing over the set $\aleph_\star$.

$$p(\underline{\boldsymbol{\alpha}}|\mathbf{y}) = \frac{p(\mathbf{y}|\underline{\boldsymbol{\alpha}})p(\underline{\boldsymbol{\alpha}})}{\sum_{\underline{\boldsymbol{\alpha}}\in\aleph_\star} p(\mathbf{y}|\underline{\boldsymbol{\alpha}})p(\underline{\boldsymbol{\alpha}})} \tag{6.39}$$

$$= \frac{1}{Z} \exp\left(-\frac{1}{2N_0}\|\Pi^\perp_{\boldsymbol{\Psi}_{\underline{\boldsymbol{\alpha}}}}\mathbf{y}\|_2^2\right) \mathrm{p}_b^{\|\underline{\boldsymbol{\alpha}}\|_0} (1-\mathrm{p}_b)^{N-\|\underline{\boldsymbol{\alpha}}\|_0} \tag{6.40}$$

where $Z = \sum_{\underline{\boldsymbol{\alpha}}\in\aleph_\star} p(\mathbf{y}|\underline{\boldsymbol{\alpha}})p(\underline{\boldsymbol{\alpha}})$ is the normalizing factor. Thus the AMMSE channel estimate is given as,

$$\hat{\boldsymbol{\alpha}}_{AMMSE} = \frac{1}{Z} \sum_{\underline{\boldsymbol{\alpha}}\in\aleph_\star} \exp\left(-\frac{1}{2N_0}\|\Pi^\perp_{\boldsymbol{\Psi}_{\underline{\boldsymbol{\alpha}}}}\mathbf{y}\|_2^2\right) \mathrm{p}_b^{\|\underline{\boldsymbol{\alpha}}\|_0} (1-\mathrm{p}_b)^{N-\|\underline{\boldsymbol{\alpha}}\|_0} (\boldsymbol{\Psi}_{\underline{\boldsymbol{\alpha}}}^H\boldsymbol{\Psi}_{\underline{\boldsymbol{\alpha}}})^{-1}\boldsymbol{\Psi}_{\underline{\boldsymbol{\alpha}}}^H\mathbf{y}$$

$$\tag{6.41}$$

2. Non-Gaussian Amplitudes with Known 2nd Order Statistics

When the amplitudes are non-Gaussian with known second order statistics, the second order statistics are effectively absorbed in the model as described earlier. Therefore, $\mathbb{E}(\mathbf{a}_{\underline{\boldsymbol{\alpha}}}|\mathbf{y})$ is approximated with the Least-Squares Estimation (LSE).

$$\mathbb{E}(\mathbf{a}_{\underline{\boldsymbol{\alpha}}}|\mathbf{y}) = (\mathbf{H}_{\underline{\boldsymbol{\alpha}}}^H\mathbf{H}_{\underline{\boldsymbol{\alpha}}})^{-1}\mathbf{H}_{\underline{\boldsymbol{\alpha}}}^H\mathbf{y} \tag{6.42}$$

The posterior density function $p(\underline{\boldsymbol{\alpha}}|\mathbf{y})$ is given according to the Baye's Rule

and is approximated by normalizing over the set $\aleph_\star$.

$$p(\underline{\boldsymbol{\alpha}}|\mathbf{y}) = \frac{p(\mathbf{y}|\underline{\boldsymbol{\alpha}})p(\underline{\boldsymbol{\alpha}})}{\sum_{\underline{\boldsymbol{\alpha}}\in\aleph_\star} p(\mathbf{y}|\underline{\boldsymbol{\alpha}})p(\underline{\boldsymbol{\alpha}})} \tag{6.43}$$

$$= \frac{1}{Z} \exp\left(-\frac{1}{2N_0}\|\Pi_{\mathbf{H}_{\underline{\boldsymbol{\alpha}}}}^\perp \mathbf{y}\|_2^2\right) \mathrm{p}_b^{\|\underline{\boldsymbol{\alpha}}\|_0} (1-\mathrm{p}_b)^{N-\|\underline{\boldsymbol{\alpha}}\|_0} \tag{6.44}$$

where $Z = \sum_{\underline{\boldsymbol{\alpha}}\in\aleph_\star} p(\mathbf{y}|\underline{\boldsymbol{\alpha}})p(\underline{\boldsymbol{\alpha}})$ is the normalizing factor. Thus the AMMSE

channel estimate is given as,

$$\hat{\underline{\boldsymbol{\alpha}}}_{AMMSE} = \frac{1}{Z} \sum_{\underline{\boldsymbol{\alpha}}\in\aleph_\star} \exp\left(-\frac{1}{2N_0}\|\Pi_{\mathbf{H}_{\underline{\boldsymbol{\alpha}}}}^\perp \mathbf{y}\|_2^2\right) \mathrm{p}_b^{\|\underline{\boldsymbol{\alpha}}\|_0} (1-\mathrm{p}_b)^{N-\|\underline{\boldsymbol{\alpha}}\|_0} (\mathbf{H}_{\underline{\boldsymbol{\alpha}}}^H \mathbf{H}_{\underline{\boldsymbol{\alpha}}})^{-1} \mathbf{H}_{\underline{\boldsymbol{\alpha}}}^H \mathbf{y}$$

$$\tag{6.45}$$

3. Gaussian Amplitudes

   When the amplitudes are Gaussian we obtain the exact $\mathbb{E}(\mathbf{a}_{\underline{\boldsymbol{\alpha}}}|\mathbf{y})$ as,

   $$\mathbb{E}(\mathbf{a}_{\underline{\boldsymbol{\alpha}}}|\mathbf{y}) = \mathbf{D}_{\underline{\boldsymbol{\alpha}}} \boldsymbol{\Psi}_{\underline{\boldsymbol{\alpha}}}^H \boldsymbol{\Sigma}_{\underline{\boldsymbol{\alpha}}}^{-1} \mathbf{a}_{\underline{\boldsymbol{\alpha}}} \tag{6.46}$$

   The posterior density function $p(\underline{\boldsymbol{\alpha}}|\mathbf{y})$ is given according to the Baye's Rule

   and is approximated by normalizing over the set $\aleph_\star$

   $$p(\underline{\boldsymbol{\alpha}}|\mathbf{y}) = \frac{p(\mathbf{y}|\underline{\boldsymbol{\alpha}})p(\underline{\boldsymbol{\alpha}})}{\sum_{\underline{\boldsymbol{\alpha}}\in\aleph_\star} p(\mathbf{y}|\underline{\boldsymbol{\alpha}})p(\underline{\boldsymbol{\alpha}})} \tag{6.47}$$

   $$= \frac{1}{Z} \frac{\exp\left(-\frac{1}{2}\mathbf{y}^H \boldsymbol{\Sigma}_{\underline{\boldsymbol{\alpha}}}^{-1} \mathbf{y}\right)}{\sqrt{\det(\boldsymbol{\Sigma}_{\underline{\boldsymbol{\alpha}}})}} \mathrm{p}_b^{\|\underline{\boldsymbol{\alpha}}\|_0} (1-\mathrm{p}_b)^{N-\|\underline{\boldsymbol{\alpha}}\|_0} \tag{6.48}$$

   where $Z = \sum_{\underline{\boldsymbol{\alpha}}\in\aleph_\star} p(\mathbf{y}|\underline{\boldsymbol{\alpha}})p(\underline{\boldsymbol{\alpha}})$ is the normalizing factor. Thus the AMMSE

channel estimate is given as,

$$\hat{\boldsymbol{\alpha}}_{AMMSE} = \frac{1}{Z} \sum_{\underline{\boldsymbol{\alpha}} \in \aleph_\star} \frac{\exp\left(-\frac{1}{2}\mathbf{y}^H \boldsymbol{\Sigma}_{\underline{\boldsymbol{\alpha}}}^{-1} \mathbf{y}\right)}{\sqrt{\det(\boldsymbol{\Sigma}_{\underline{\boldsymbol{\alpha}}})}} \mathrm{p}_b^{\|\underline{\boldsymbol{\alpha}}\|_0} \left(1 - \mathrm{p}_b\right)^{N - \|\underline{\boldsymbol{\alpha}}\|_0} \mathbf{D}_{\underline{\boldsymbol{\alpha}}} \boldsymbol{\Psi}_{\underline{\boldsymbol{\alpha}}}^H \boldsymbol{\Sigma}_{\underline{\boldsymbol{\alpha}}}^{-1} \mathbf{a}_{\underline{\boldsymbol{\alpha}}}$$

$$(6.49)$$

6.5 Low-Complexity MMSE Channel Estimation

6.5.1 Orthogonal Clustering

We leverage the structure of $\boldsymbol{\Psi}$ to develop the Low-Complexity MMSE (LCMMSE) estimator for the channel. As discussed earlier, $\boldsymbol{\Psi}$ can be divided into nearly orthogonal clusters of fixed width ν. Let there be C such clusters denoted by $\boldsymbol{\Psi}_{\Theta_r}$ for $r = 1, 2, \ldots, C$.

Due to the sparsity of the the received vector $\mathbf{y}$, not all the C orthogonal clusters are active, i.e., $\mathbf{y}$ is not composed of columns belonging to all the C othogonal clusters. Therefore, $\mathbf{y}$ is first correlated with all the columns of $\boldsymbol{\Psi}$ and the orthogonal clusters containing the significantly correlated columns are identified and included into the set $\mathcal{S}$. We select χ as the significance level of

correlation between $\mathbf{y}$ and columns of $\boldsymbol{\Psi}$, then for $r = 1, 2, \ldots, C$

$$\Theta_r \in \mathcal{S} \quad \begin{cases} \mathbf{y}^H \boldsymbol{\psi}_i \geq \chi \\ \\ \boldsymbol{\psi}_i \in \boldsymbol{\Psi}_{\Theta_r} \end{cases} \tag{6.50}$$

where $\boldsymbol{\psi}_i$ denotes the ith column of $\boldsymbol{\Psi}$ for $i = 1, 2, \ldots, N-1$.

6.5.2 Non-Gaussian Amplitudes

When the vector of the channel amplitudes $\mathbf{a}_{\underline{\alpha}}$ is non-Gaussian then, conditioned on the support vector $\underline{\alpha}$,

$$\mathbf{y} = \boldsymbol{\Psi}_{\underline{\alpha}} \mathbf{a}_{\underline{\alpha}} + \boldsymbol{\omega} \tag{6.51}$$

The orthogonal projection of $\mathbf{y}$ onto the orthogonal complement of $\boldsymbol{\Psi}_{\underline{\alpha}}$ is Gaussian. Therefore, the likelihood $p(\mathbf{y}|\underline{\alpha})$ can be approximated by,

$$p(\mathbf{y}|\underline{\alpha}) \propto \exp\left(-\frac{1}{2N_0} \|\Pi_{\boldsymbol{\Psi}_{\underline{\alpha}}}^{\perp} \mathbf{y}\|_2^2 \right) \tag{6.52}$$

where,

$$\Pi_{\boldsymbol{\Psi}_{\underline{\alpha}}}^{\perp} = \mathbf{I} - \boldsymbol{\Psi}_{\underline{\alpha}} \left[\boldsymbol{\Psi}_{\underline{\alpha}}^H \boldsymbol{\Psi}_{\underline{\alpha}} \right]^{-1} \boldsymbol{\Psi}_{\underline{\alpha}}^H \tag{6.53}$$

Since there are C orthogonal clusters, then we can express $\mathbf{\Psi_{\underline{\alpha}}}$ in a block matrix form as,

$$\mathbf{\Psi_{\underline{\alpha}}} = [\mathbf{\Psi_{\Theta_1}} \quad \mathbf{\Psi_{\Theta_2}} \ldots \mathbf{\Psi_{\Theta_C}}] \tag{6.54}$$

where $\mathbf{\Psi_{\Theta_i}}$ is the matrix formed by collecting the columns of $\mathbf{\Psi}$ belonging to the ith orthogonal cluster. Now the inverse term that appears in Eq. (6.53) becomes the inverse of the block diagonal matrix,

$$\left[\mathbf{\Psi_{\underline{\alpha}}^H \Psi_{\underline{\alpha}}}\right]^{-1} = \begin{bmatrix} (\mathbf{\Psi_{\Theta_1}^H \Psi_{\Theta_1}})^{-1} & 0 & \ldots & 0 \\ 0 & (\mathbf{\Psi_{\Theta_2}^H \Psi_{\Theta_2}})^{-1} & \ldots & \vdots \\ \vdots & 0 & \ddots & 0 \\ 0 & \vdots & \ldots & (\mathbf{\Psi_{\Theta_C}^H \Psi_{\Theta_C}})^{-1} \end{bmatrix} \tag{6.55}$$

Furthermore, it is also easy to show that, for $r = 1, \ldots, C$;

$$(\mathbf{\Psi_{\Theta_r}^H \Psi_{\Theta_r}})^{-1} = (\mathbf{\Psi_{\Theta_1}^H \Psi_{\Theta_1}})^{-1} \tag{6.56}$$

Since the clusters are orhtogonal, the projection matrix of Eq. (6.7) can be expressed in terms of the sum of individual projection matrix of each cluster as follows,

$$\Pi^{\perp}_{\mathbf{\Psi_{\underline{\alpha}}}} = -C\mathbf{I} + \mathbf{I} + \Pi^{\perp}_{\mathbf{\Psi_{\Theta_1}}} + \cdots + \Pi^{\perp}_{\mathbf{\Psi_{\Theta_C}}} \tag{6.57}$$

Therefore, we can write the likelihood as,

$$p(\mathbf{y}|\underline{\alpha}) \propto \exp\left(\frac{1}{2N_0}\left(C\|\mathbf{y}\|^2 - \|\mathbf{y}\|^2 - \|\Pi^{\perp}_{\mathbf{\Psi_{\Theta_1}}}\mathbf{y}\|^2 - \cdots - \|\Pi^{\perp}_{\mathbf{\Psi_{\Theta_C}}}\mathbf{y}\|^2\right)\right) \tag{6.58}$$

Thus upto an irrelevant proportionality constant the likelihood is given by,

$$p(\mathbf{y}|\underline{\boldsymbol{\alpha}}) = \exp\left[\frac{1}{2N_0}\left((C-1)\|\mathbf{y}\|^2 - \|\Pi^{\perp}_{\boldsymbol{\Psi}_{\Theta_1}}\mathbf{y}\|^2 - \cdots - \|\Pi^{\perp}_{\boldsymbol{\Psi}_{\Theta_C}}\mathbf{y}\|^2\right)\right] \qquad (6.59)$$

$$= \exp\left[\frac{(C-1)}{2N_0}\|\mathbf{y}\|^2 - \frac{1}{2N_0}\sum_{r=1}^{C}\|\Pi^{\perp}_{\boldsymbol{\Psi}_{\Theta_r}}\mathbf{y}\|^2\right] \qquad (6.60)$$

In calculating the above likelihood we need to compute the l_2-norm of the projection of $\mathbf{y}$ onto each orthogonal cluster. These norms can be expressed as,

$$\|\Pi^{\perp}_{\boldsymbol{\Psi}_{\Theta_r}}\mathbf{y}\|^2 = \mathbf{y}^H\boldsymbol{\Psi}_{\Theta_r}\left[\boldsymbol{\Psi}^H_{\Theta_r}\boldsymbol{\Psi}_{\Theta_r}\right]^{-1}\boldsymbol{\Psi}^H_{\Theta_r}\mathbf{y} \qquad (6.61)$$

$$= \mathbf{y}^H_{\Theta_r}\overline{\boldsymbol{\Psi}}_{\Theta_r}\left[\boldsymbol{\Psi}^H_{\Theta_1}\boldsymbol{\Psi}_{\Theta_1}\right]^{-1}\overline{\boldsymbol{\Psi}}^H_{\Theta_r}\mathbf{y}_{\Theta_r} \qquad (6.62)$$

$$= \left(\overline{\boldsymbol{\Psi}}^H_{\Theta_r}\mathbf{y}_{\Theta_r}\right)^H\mathbf{W}\left(\overline{\boldsymbol{\Psi}}^H_{\Theta_r}\mathbf{y}_{\Theta_r}\right) \qquad (6.63)$$

$$= \|\overline{\boldsymbol{\Psi}}^H_{\Theta_r}\mathbf{y}_{\Theta_r}\|^2_{\mathbf{W}} \qquad (6.64)$$

where we have used Eq. (6.56) to obtain Eq. (6.62) and $\mathbf{W} = [\boldsymbol{\Psi}^H_{\Theta_1}\boldsymbol{\Psi}_{\Theta_1}]^{-1}$. The vector $\mathbf{y}_{\Theta_r}$ is $\nu \times 1$ and is the masked version of $\mathbf{y}$ which includes only those elements of $\mathbf{y}$ which correspond to the rth cluster and $\overline{\boldsymbol{\Psi}}_{\Theta_r}$ is the $\nu \times \nu$ sub-matrix of $\boldsymbol{\Psi}_{\Theta_r}$ composed of the non-zero portion of the rth cluster. We note here that from the Toeplitz structure,

$$\overline{\boldsymbol{\Psi}}_{\Theta_r} = \overline{\boldsymbol{\Psi}}_{\Theta_1} \qquad (6.65)$$

for $r = 1, 2, \ldots, C$. Now Eq. (6.60) is expressed as follows,

$$p(\mathbf{y}|\underline{\boldsymbol{\alpha}}) = \exp\left[\frac{(C-1)}{2N_0}\|\mathbf{y}\|^2 - \frac{1}{2N_0}\sum_{r=1}^{C}\|\overline{\boldsymbol{\Psi}}^H_{\Theta_1}\mathbf{y}_{\Theta_r}\|^2_{\mathbf{W}}\right] \qquad (6.66)$$

97

Therefore, we need to perform the matrix inversion only for the first cluster to find $\mathbf{W}$ and obtain the projection matrix $\overline{\boldsymbol{\Psi}}_{\Theta_1}^H$. When a vector $\mathbf{y}$ is received, we need to calculate its norm, mask it to get $\mathbf{y}_{\Theta_r}$ for $r = 1, 2, \ldots, C$ and find the weighted norm of the projections $\overline{\boldsymbol{\Psi}}_{\Theta_1}^H \mathbf{y}_{\Theta_r}$ for all the masked vectors $\mathbf{y}_{\Theta_r}$ to calculate the likelihood of Eq. (6.66).

To further reduce the computations, the likelihood $p(\mathbf{y}|\underline{\boldsymbol{\alpha}})$ is computed over only those orthogonal clusters which are included in the set $\mathcal{S}$ of Eq. (6.50) and used in Eq. (6.41) to obtain the Low-Complexity AMMSE (LC-AMMSE) channel estimate for the case of non-Gaussian amplitudes.

Figure 6.3 shows the NRMSE in the estimation of the support using AMMSE estimation and Figure 6.4 shows the correponding performance in terms of the energy capture.

6.5.3   Non-Gaussian Amplitudes with Known 2nd Order Statistics

When the vector of the channel amplitudes $\mathbf{a}_{\underline{\alpha}}$ is non-Gaussian but its Second-Order Statistics is known, this information incorporated into the modified sensing matrix $\mathbf{H}$. Then conditioned on the support vector $\underline{\boldsymbol{\alpha}}$,

$$\mathbf{y} = \mathbf{H}_{\underline{\alpha}} \mathbf{a}_{\underline{\alpha}} + \boldsymbol{\omega} \tag{6.67}$$

The orthogonal projection of $\mathbf{y}$ onto the orthogonal complement of $\mathbf{H}_{\underline{\alpha}}$ is Gaussian. Therefore, the likelihood $p(\mathbf{y}|\underline{\alpha})$ can be approximated by,

$$p(\mathbf{y}|\underline{\alpha}) \propto \exp\left(-\frac{1}{2N_0}\|\Pi_{\mathbf{H}_{\underline{\alpha}}}^{\perp}\mathbf{y}\|_2^2\right) \tag{6.68}$$

where,

$$\Pi_{\mathbf{H}_{\underline{\alpha}}}^{\perp} = \mathbf{I} - \mathbf{H}_{\underline{\alpha}}\left[\mathbf{H}_{\underline{\alpha}}^{H}\mathbf{H}_{\underline{\alpha}}\right]^{-1}\mathbf{H}_{\underline{\alpha}}^{H} \tag{6.69}$$

Similar to the previous section, if there are C orthogonal clusters, we can express $\mathbf{H}_{\underline{\alpha}}$ in a block matrix form as,

$$\mathbf{H}_{\underline{\alpha}} = [\mathbf{H}_{\Theta_1} \quad \mathbf{H}_{\Theta_2} \ldots \mathbf{H}_{\Theta_C}] \tag{6.70}$$

where $\mathbf{H}_{\Theta_i}$ is the matrix formed by collecting the columns of $\mathbf{H}$ belonging to the ith orthogonal cluster. We observe that,

$$\mathbf{H}_{\Theta_r}^{H}\mathbf{H}_{\Theta_r} = \gamma_r(\mathbf{H}_{\Theta_1}^{H}\mathbf{H}_{\Theta_1}) \tag{6.71}$$

where,

$$\gamma_r = \frac{\mathbb{E}(a_r^2)}{\mathbb{E}(a_1^2)} \tag{6.72}$$

$$= \exp\left(-r\frac{\nu}{\Gamma}\right) \tag{6.73}$$

Γ is the decay constant, $\mathbb{E}(a_r^2)$ is the variance of the first element of the rth orthogonal cluster for $r = 1, \ldots, C$; ν is as defined earlier and Eq. (6.73) follows from the exponential nature of the APDP. The value of the factors γ_r depend on the location of the rth orthogonal cluster.

The inverse term that appears in Eq. (6.69) becomes the inverse of the block diagonal matrix,

$$[\mathbf{H}_{\underline{\alpha}}{}^H \mathbf{H}_{\underline{\alpha}}]^{-1} = \begin{bmatrix} (\mathbf{H}_{\Theta_1}^H \mathbf{H}_{\Theta_1})^{-1} & 0 & \cdots & 0 \\ 0 & (\mathbf{H}_{\Theta_2}^H \mathbf{H}_{\Theta_2})^{-1} & \cdots & \vdots \\ \vdots & 0 & \ddots & 0 \\ 0 & \vdots & \cdots & (\mathbf{H}_{\Theta_C}^H \mathbf{H}_{\Theta_C})^{-1} \end{bmatrix} \tag{6.74}$$

It is also easy to show that, for $r = 1, \ldots, C$;

$$(\mathbf{H}_{\Theta_r}^H \mathbf{H}_{\Theta_r})^{-1} = \frac{1}{\gamma_r}(\mathbf{H}_{\Theta_1}^H \mathbf{H}_{\Theta_1})^{-1} \tag{6.75}$$

Since the clusters are orhtogonal, the projection matrix of Eq. (6.69) can be expressed in terms of the sum of individual projection matrix of each cluster as,

$$\Pi_{\mathbf{H}_{\underline{\alpha}}}^{\perp} = -C\mathbf{I} + \mathbf{I} + \Pi_{\mathbf{H}_{\Theta_1}}^{\perp} + \cdots + \Pi_{\mathbf{H}_{\Theta_C}}^{\perp} \tag{6.76}$$

Therefore, we can write the likelihood as,

$$p(\mathbf{y}|\underline{\boldsymbol{\alpha}}) \propto \exp\left(\frac{1}{2N_0}\left(C\|\mathbf{y}\|^2 - \|\mathbf{y}\|^2 - \|\Pi_{\mathbf{H}_{\Theta_1}}^{\perp}\mathbf{y}\|^2 - \cdots - \|\Pi_{\mathbf{H}_{\Theta_C}}^{\perp}\mathbf{y}\|^2\right)\right) \tag{6.77}$$

Thus upto an irrelevant proportionality constant the likelihood is given by,

$$p(\mathbf{y}|\underline{\boldsymbol{\alpha}}) = \exp\left[\frac{1}{2N_0}\left((C-1)\|\mathbf{y}\|^2 - \|\Pi^{\perp}_{\mathbf{H}_{\Theta_1}}\mathbf{y}\|^2 - \cdots - \|\Pi^{\perp}_{\mathbf{H}_{\Theta_C}}\mathbf{y}\|^2\right)\right] \qquad (6.78)$$

$$= \exp\left[\frac{(C-1)}{2N_0}\|\mathbf{y}\|^2 - \frac{1}{2N_0}\sum_{r=1}^{C}\|\Pi^{\perp}_{\mathbf{H}_{\Theta_r}}\mathbf{y}\|^2\right] \qquad (6.79)$$

We set the $\nu \times 1$ vector $\mathbf{y}_{\Theta_r}$ as the masked version of $\mathbf{y}$ which includes only those elements of $\mathbf{y}$ which correspond to the rth cluster and $\overline{\mathbf{H}}_{\Theta_r}$ is the $\nu \times \nu$ sub-matrix of $\mathbf{H}_{\Theta_r}$ composed of the non-zero portion of the rth cluster. We note here that from the Toeplitz structure,

$$\overline{\mathbf{H}}_{\Theta_r} = \sqrt{\gamma_r}\,\overline{\mathbf{H}}_{\Theta_1} \qquad (6.80)$$

for $r = 1, 2, \ldots, C$. In calculating the likelihood of Eq. (6.79) we need to compute the l_2-norm of the projections of $\mathbf{y}$ onto each orthogonal cluster. These norms can be expressed as,

$$\|\Pi^{\perp}_{\mathbf{H}_{\Theta_r}}\mathbf{y}\|^2 = \mathbf{y}^H\mathbf{H}_{\Theta_r}\left[\mathbf{H}^H_{\Theta_r}\mathbf{H}_{\Theta_r}\right]^{-1}\mathbf{H}^H_{\Theta_r}\mathbf{y} \qquad (6.81)$$

$$= \frac{1}{\gamma_r}\mathbf{y}^H_{\Theta_r}\overline{\mathbf{H}}_{\Theta_r}\left[\mathbf{H}^H_{\Theta_1}\mathbf{H}_{\Theta_1}\right]^{-1}\overline{\mathbf{H}}^H_{\Theta_r}\mathbf{y}_{\Theta_r} \qquad (6.82)$$

$$= \mathbf{y}^H_{\Theta_r}\overline{\mathbf{H}}_{\Theta_1}\left[\mathbf{H}^H_{\Theta_1}\mathbf{H}_{\Theta_1}\right]^{-1}\overline{\mathbf{H}}^H_{\Theta_1}\mathbf{y}_{\Theta_r} \qquad (6.83)$$

$$= \left(\overline{\mathbf{H}}^H_{\Theta_1}\mathbf{y}_{\Theta_1}\right)^H\mathbf{W}'\left(\overline{\mathbf{H}}^H_{\Theta_1}\mathbf{y}_{\Theta_r}\right) \qquad (6.84)$$

$$= \|\overline{\mathbf{H}}^H_{\Theta_1}\mathbf{y}_{\Theta_r}\|^2_{\mathbf{W}'} \qquad (6.85)$$

where we have used Eq. (6.75) to obtain Eq. (6.82) and $\mathbf{W}' = [\mathbf{H}^H_{\Theta_1}\mathbf{H}_{\Theta_1}]^{-1}$. Now

Eq. (6.79) is expressed as follows,

$$p(\mathbf{y}|\underline{\boldsymbol{\alpha}}) = \exp\left[\frac{(C-1)}{2N_0}\|\mathbf{y}\|^2 - \frac{1}{2N_0}\sum_{r=1}^{C}\|\overline{\mathbf{H}}_{\Theta_1}^{H}\mathbf{y}_{\Theta_r}\|_{\mathbf{W}}^2\right] \qquad (6.86)$$

Therefore, we need to perform the matrix inversion only for the first cluster to find $\mathbf{W}'$ and obtain the projection matrix $\overline{\mathbf{H}}_{\Theta_1}^{H}$. When a vector $\mathbf{y}$ is received, we need to calculate its norm, mask it to get $\mathbf{y}_{\Theta_r}$ for $r = 1, 2, \ldots, C$ and find the weighted norm of the projections $\overline{\mathbf{H}}_{\Theta_1}^{H}\mathbf{y}_{\Theta_r}$ for all the masked vectors $\mathbf{y}_{\Theta_r}$ to calculate the likelihood of Eq. (6.86).

To further reduce the computations, the likelihood $p(\mathbf{y}|\underline{\boldsymbol{\alpha}})$ is computed over only those orthogonal clusters which are included in the set $\mathcal{S}$ of Eq. (6.50) and used in Eq. (6.45) and we obtain the Low-Complexity AMMSE (LC-AMMSE) channel estimate for the case of non-Gaussian amplitudes with known second order statistics.

Figure 6.3 shows the NRMSE in the estimation of the support using AMMSE estimation and Figure 6.4 shows the correponding performance in terms of the energy capture.

6.5.4 Gaussian Amplitudes

When the vector of the channel amplitudes $\mathbf{a}_{\underline{\alpha}}$ is Gaussian then, conditioned on the support vector $\underline{\alpha}$,

$$\mathbf{y} = \mathbf{\Psi}_{\underline{\alpha}}\mathbf{a}_{\underline{\alpha}} + \boldsymbol{\omega} \tag{6.87}$$

The likelihood is given by,

$$p(\mathbf{y}|\underline{\alpha}) = \frac{\exp\left(-\frac{1}{2}\mathbf{y}^H\mathbf{\Sigma}_{\underline{\alpha}}^{-1}\mathbf{y}\right)}{\sqrt{\det(\mathbf{\Sigma}_{\underline{\alpha}})}} \tag{6.88}$$

where $\mathbf{\Sigma}_{\underline{\alpha}} = N_0\mathbf{I} + \mathbf{\Psi}_{\underline{\alpha}}\mathbf{D}_{\underline{\alpha}}\mathbf{\Psi}_{\underline{\alpha}}^H$ is the co-variance matrix.

Similar to previous section, if there are C orthogonal clusters, then the overall likelihood becomes the product of the individual likelihoods of each orthogonal cluster, as follows,

$$p(\mathbf{y}|\underline{\alpha}) = \frac{\exp\left(-\frac{1}{2}\mathbf{y}^H\mathbf{\Sigma}_{\Theta_1}^{-1}\mathbf{y}\right)}{\det(\mathbf{\Sigma}_{\Theta_1})} \frac{\exp\left(-\frac{1}{2}\mathbf{y}^H\mathbf{\Sigma}_{\Theta_2}^{-1}\mathbf{y}\right)}{\det(\mathbf{\Sigma}_{\Theta_2})} \cdots \frac{\exp\left(-\frac{1}{2}\mathbf{y}^H\mathbf{\Sigma}_{\Theta_C}^{-1}\mathbf{y}\right)}{\det(\mathbf{\Sigma}_{\Theta_C})} \tag{6.89}$$

$$= \frac{\exp\left[-\frac{1}{2}\sum_{r=1}^{C}\left(\mathbf{y}^H\mathbf{\Sigma}_{\Theta_r}^{-1}\mathbf{y}\right)\right]}{\prod_{r=1}^{C}\det(\mathbf{\Sigma}_{\Theta_r})} \tag{6.90}$$

where $\boldsymbol{\Sigma}_{\Theta_r}$ is the $\nu \times \nu$ co-variance matrix when only the rth cluster is active. $\boldsymbol{\Sigma}_{\Theta_r}$ for $r = 1, 2, \ldots, C$ is given by,

$$\boldsymbol{\Sigma}_{\Theta_r} = \mathbb{E}[\mathbf{yy}^H | \Theta_r] \tag{6.91}$$

$$= N_0 \mathbf{I} + \boldsymbol{\Psi}_{\Theta_r} \mathbf{D}_{\Theta_r} \boldsymbol{\Psi}_{\Theta_r}^H \tag{6.92}$$

where $\mathbf{D}_{\Theta_r}$ is the $\nu \times \nu$ sub-matrix of $\mathbf{D}$ given in Eq. (2.12) corresponding to the columns of $\boldsymbol{\Psi}_{\Theta_r}$. Thus the likelihood for the rth cluster is given as,

$$p(\mathbf{y}|\Theta_r) = \frac{\exp\left(-\frac{1}{2}\mathbf{y}^H \boldsymbol{\Sigma}_{\Theta_r}^{-1} \mathbf{y}\right)}{\sqrt{\det(\boldsymbol{\Sigma}_{\Theta_r})}} \tag{6.93}$$

We note that for the rth cluster the calculation of the inverse of $\boldsymbol{\Sigma}_{\Theta_r}$ is the computationally intensive part. We proceed with finding the inverse by splitting the matrix $\mathbf{D}_{\Theta_r}$ as a product of its square root diagonal matrices and applying the matrix inversion lemma, as follows,

$$\boldsymbol{\Sigma}_{\Theta_r}^{-1} = \left(N_0 \mathbf{I} + \boldsymbol{\Psi}_{\Theta_r} \mathbf{D}_{\Theta_r} \boldsymbol{\Psi}_{\Theta_r}^H\right)^{-1} \tag{6.94}$$

$$= \left(N_0 \mathbf{I} + \boldsymbol{\Psi}_{\Theta_r} (\mathbf{D}_{\Theta_r}^{1/2})(\mathbf{D}_{\Theta_r}^{1/2}) \boldsymbol{\Psi}_{\Theta_r}^H\right)^{-1} \tag{6.95}$$

$$= \left(N_0 \mathbf{I} + (\boldsymbol{\Psi}_{\Theta_r} \mathbf{D}_{\Theta_r}^{1/2})(\boldsymbol{\Psi}_{\Theta_r} \mathbf{D}_{\Theta_r}^{1/2})^H\right)^{-1} \tag{6.96}$$

We set $\boldsymbol{\Psi}_{\Theta_r}\mathbf{D}_{\Theta_r}^{1/2} = \boldsymbol{\Omega}_{\Theta_r}$ and use the matrix inversion lemma to obatin,

$$\boldsymbol{\Sigma}_{\Theta_r}^{-1} = \left(N_0\mathbf{I} + \boldsymbol{\Omega}_{\Theta_r}\boldsymbol{\Omega}_{\Theta_r}^{H}\right)^{-1} \tag{6.97}$$

$$= \frac{1}{N_0}\left(\mathbf{I} + \frac{1}{N_0}\boldsymbol{\Omega}_{\Theta_r}\boldsymbol{\Omega}_{\Theta_r}^{H}\right)^{-1} \tag{6.98}$$

$$= \frac{1}{N_0}\left(\mathbf{I} - \frac{1}{N_0}\boldsymbol{\Omega}_{\Theta_r}\left(N_0\mathbf{I} + \boldsymbol{\Omega}_{\Theta_r}^{H}\boldsymbol{\Omega}_{\Theta_r}\right)^{-1}\boldsymbol{\Omega}_{\Theta_r}^{H}\right) \tag{6.99}$$

Now we only consider the inverse part, where we decompose the symmetric matrix $\boldsymbol{\Omega}_{\Theta_r}^{H}\boldsymbol{\Omega}_{\Theta_r}$ using the eigenvlaue decomposition,

$$\left(N_0\mathbf{I} + \boldsymbol{\Omega}_{\Theta_r}^{H}\boldsymbol{\Omega}_{\Theta_r}\right)^{-1} = \left(N_0\mathbf{I} + \mathbf{Q}\boldsymbol{\Lambda}_{\Theta_r}\mathbf{Q}^{H}\right)^{-1} \tag{6.100}$$

$$= \left(N_0\mathbf{Q}\mathbf{I}\mathbf{Q}^{H} + \mathbf{Q}\boldsymbol{\Lambda}_{\Theta_r}\mathbf{Q}^{H}\right)^{-1} \tag{6.101}$$

$$= \mathbf{Q}\left(N_0\mathbf{I} + \boldsymbol{\Lambda}_{\Theta_r}\right)^{-1}\mathbf{Q}^{H} \tag{6.102}$$

where $\boldsymbol{\Lambda}_{\Theta_r}$ is the diagonal matrix of the eigenvalues and $\mathbf{Q}$ contains the corresponding eigenvectors. For the first cluster, we can write,

$$\boldsymbol{\Sigma}_{\Theta_1}^{-1} = \frac{1}{N_0}\left(\mathbf{I} - \frac{1}{N_0}\boldsymbol{\Omega}_{\Theta_1}\mathbf{Q}\left(N_0\mathbf{I} + \boldsymbol{\Lambda}_{\Theta_1}\right)^{-1}\mathbf{Q}^{H}\boldsymbol{\Omega}_{\Theta_1}^{H}\right) \tag{6.103}$$

$$= \frac{1}{N_0}\mathbf{I} - \frac{1}{N_0^2}(\boldsymbol{\Omega}_{\Theta_1}\mathbf{Q})\left(N_0\mathbf{I} + \boldsymbol{\Lambda}_{\Theta_1}\right)^{-1}(\mathbf{Q}^{H}\boldsymbol{\Omega}_{\Theta_1}^{H}) \tag{6.104}$$

The computation for the matrix inversion is reduced as we can see from Eq. (6.104) that the inverse operation is simply the inverse of a diagonal matrix. We note that, the eigenvectors for all the C orthogonal clusters are the same and hence the matrix $\mathbf{Q}$ remains the same for all the clusters. If we denote by $\overline{\boldsymbol{\Omega}}_{\Theta_r}$

105

the $\nu \times \nu$ non-zero sub-matrix of $\mathbf{\Omega}_{\Theta_r}$, then due to the Toeplitz structure of $\mathbf{\Psi}$, we can write,

$$\overline{\mathbf{\Omega}}_{\Theta_r} = \sqrt{\gamma_r}\,\overline{\mathbf{\Omega}}_{\Theta_1} \tag{6.105}$$

Thus, the eigenvalue matrix for the rth cluster $\mathbf{\Lambda}_{\Theta_r}$ can be easily calculated from the eigenvalue decomposition of the first cluster as,

$$\mathbf{\Lambda}_{\Theta_r} = \sqrt{\gamma_r}\,\mathbf{\Lambda}_{\Theta_1} \tag{6.106}$$

Therefore, we only need to perform the Eigenvalue decomposition for the first cluster to obtain $\mathbf{Q}$ and $\mathbf{\Lambda}_{\Theta_1}$ and based on the location we compute the eigenvalues for all the remaining orthogonal clusters (i.e. $\mathbf{\Lambda}_{\Theta_r}$ for $r = 1, \ldots, C$). Now all the matrix inversions in Eq. (6.89) become simple inversions of diagonal matrices.

For the rth cluster using Eq. (6.99) and the eigenvalue decomposition, we can write,

$$\mathbf{y}^H \mathbf{\Sigma}_{\Theta_r}^{-1} \mathbf{y} = \frac{1}{N_0}\|\mathbf{y}\|^2 - \frac{1}{N_0^2}(\mathbf{y}^H \mathbf{\Omega}_{\Theta_r} \mathbf{Q})\,(N_0 \mathbf{I} + \mathbf{\Lambda}_{\Theta_r})^{-1}\,(\mathbf{Q}^H \mathbf{\Omega}_{\Theta_r}^H \mathbf{y}) \tag{6.107}$$

Using and the fact that $\mathbf{D}_{\Theta_r}^{-1/2}$ si symmetric and $\mathbf{D}_{\Theta_r}^{-1/2} = \sqrt{\gamma_r}\, \mathbf{D}_{\Theta_1}^{-1/2}$, the second term in the above equation can be expressed as,

$$\frac{1}{N_0^2}(\mathbf{y}^H \mathbf{\Omega}_{\Theta_r} \mathbf{Q})\,(N_0 \mathbf{I} + \mathbf{\Lambda}_{\Theta_r})^{-1}\,(\mathbf{Q}^H \mathbf{\Omega}_{\Theta_r}^H \mathbf{y}) \tag{6.108}$$

$$= \frac{1}{N_0^2}(\mathbf{y}^H \mathbf{\Psi}_{\Theta_r} \mathbf{D}_{\Theta_r}^{-1/2} \mathbf{Q})\,(N_0 \mathbf{I} + \sqrt{\gamma_r}\,\mathbf{\Lambda}_{\Theta_r})^{-1}\,(\mathbf{Q}^H \mathbf{D}_{\Theta_r}^{-1/2} \mathbf{\Psi}_{\Theta_r}^H \mathbf{y}) \tag{6.109}$$

$$= \frac{\gamma_r}{N_0^2}(\mathbf{y}_{\Theta_r}^H \mathbf{D}_{\Theta_1}^{-1/2} \mathbf{Q})\,(N_0 \mathbf{I} + \sqrt{\gamma_r}\,\mathbf{\Lambda}_{\Theta_r})^{-1}\,(\mathbf{Q}^H \mathbf{D}_{\Theta_1}^{-1/2} \mathbf{y}_{\Theta_r}^H) \tag{6.110}$$

$$= \frac{\gamma_r}{N_0^2}(\mathbf{Q}^H \mathbf{D}_{\Theta_1}^{-1/2} \mathbf{y}_{\Theta_r}^H)^H\,(N_0 \mathbf{I} + \gamma_r\,\mathbf{\Lambda}_{\Theta_1})^{-1}\,(\mathbf{Q}^H \mathbf{D}_{\Theta_1}^{-1/2} \mathbf{y}_{\Theta_r}^H) \tag{6.111}$$

$$= \frac{\gamma_r}{N_0^2}\|(\mathbf{Q}^H \mathbf{D}_{\Theta_1}^{-1/2} \mathbf{y}_{\Theta_r}^H)\|_{\tilde{\mathbf{W}}}^2 \tag{6.112}$$

where we set $\tilde{\mathbf{W}} = (N_0 \mathbf{I} + \gamma_r\,\mathbf{\Lambda}_{\Theta_1})^{-1}$ and the $\nu \times 1$ vector $\mathbf{y}_{\Theta_r}$ as the masked version of $\mathbf{y}$ which includes only those elements of $\mathbf{y}$ which correspond to the rth cluster, $\mathbf{y}_{\Theta_r} = \overline{\mathbf{\Psi}_{\Theta_r}^H \mathbf{y}}$.

Using the determinant lemma and the fact that $(\mathbf{\Psi}_{\Theta_r}^H \mathbf{\Psi}_{\Theta_r}) = (\mathbf{\Psi}_1^H \mathbf{\Psi}_1)$ due to the Toeplitz structure of $\mathbf{\Psi}$, we can write,

$$\det(\mathbf{\Sigma}_{\Theta_r}) = \det\left(N_0 \mathbf{I} + \mathbf{\Omega}_{\Theta_r} \mathbf{\Omega}_{\Theta_r}^H\right) \tag{6.113}$$

$$= N_0^{N-\nu}\det\left(N_0 \mathbf{I} + \mathbf{\Omega}_{\Theta_r}^H \mathbf{\Omega}_{\Theta_r}\right) \tag{6.114}$$

$$= N_0^{N-\nu}\det\left(N_0 \mathbf{I} + \left[\mathbf{\Psi}_{\Theta_r} \mathbf{D}_{\Theta_r}^{-1/2}\right]^H \left[\mathbf{\Psi}_{\Theta_r} \mathbf{D}_{\Theta_r}^{-1/2}\right]\right) \tag{6.115}$$

$$= N_0^{N-\nu}\det\left(N_0 \mathbf{I} + \gamma_r\,\mathbf{D}_{\Theta_1}^{-1/2}(\mathbf{\Psi}_{\Theta_r}^H \mathbf{\Psi}_{\Theta_r})\mathbf{D}_{\Theta_1}^{-1/2}\right) \tag{6.116}$$

$$= N_0^{N-\nu}\det\left(N_0 \mathbf{I} + \gamma_r\,\mathbf{D}_{\Theta_1}^{-1/2}(\mathbf{\Psi}_{\Theta_1}^H \mathbf{\Psi}_{\Theta_1})\mathbf{D}_{\Theta_1}^{-1/2}\right) \tag{6.117}$$

Thus the likelihood for the rth cluster is computed as,

$$p(\mathbf{y}|\Theta_r) = -\frac{\exp\left(\frac{1}{N_0}\|\mathbf{y}\|^2 + \frac{\gamma_r}{N_0^2}\|(\mathbf{Q}^H \mathbf{D}_{\Theta_1}^{-1/2}\mathbf{y}_{\Theta_r}^H)\|_{\tilde{\mathbf{W}}}^2\right)}{N_0^{\frac{N-\nu}{2}}\sqrt{\det\left(N_0\mathbf{I} + \gamma_r\,\mathbf{D}_{\Theta_1}^{-1/2}(\mathbf{\Psi}_{\Theta_1}^H\mathbf{\Psi}_{\Theta_1})\mathbf{D}_{\Theta_1}^{-1/2}\right)}} \tag{6.118}$$

The overall likelihood is now given as,

$$p(\mathbf{y}|\underline{\boldsymbol{\alpha}}) = \frac{\exp\left[-\frac{1}{2}\sum_{r=1}^{C}\left(\frac{1}{N_0}\|\mathbf{y}\|^2 + \frac{\gamma_r}{N_0^2}\|(\mathbf{Q}^H \mathbf{D}_{\Theta_1}^{-1/2}\mathbf{y}_{\Theta_r}^H)\|_{\tilde{\mathbf{W}}}^2\right)\right]}{N_0^{\frac{C(N-\nu)}{2}}\prod_{r=1}^{C}\sqrt{\det\left(N_0\mathbf{I} + \gamma_r\,\mathbf{D}_{\Theta_1}^{-1/2}(\mathbf{\Psi}_{\Theta_1}^H\mathbf{\Psi}_{\Theta_1})\mathbf{D}_{\Theta_1}^{-1/2}\right)}} \tag{6.119}$$

Now we present the calculation of the log-likelihood function of Eq. (6.93) for a single orthogonal cluster in an order recursive way for the prescence of $1, 2, \ldots, k_c$ active columns where k_c was given earlier We begin with the case of a single active column in a cluster and denote the corresponding covariance matrix as $\mathbf{\Sigma}_1$ where we drop the subscript indicating the cluster for clarity. Similarly for two paths within a custer, the covariance matrix is expressed as $\mathbf{\Sigma}_2$. Now we can write,

$$\mathbf{\Sigma}_{i+1} = \mathbf{\Sigma}_i + \boldsymbol{\psi}_j \mathbb{E}[a_j^2]\boldsymbol{\psi}_j^H \tag{6.120}$$

where $\boldsymbol{\psi}_j$ is the column of $\mathbf{\Psi}$ added when moving from order i to $i+1$. Similarly, we can write the determinant as;

$$\det(\mathbf{\Sigma}_{i+1}) = \det(\mathbf{\Sigma}_i + \boldsymbol{\psi}_j \mathbb{E}[a_j^2]\boldsymbol{\psi}_j^H) \tag{6.121}$$

$$= \det(\mathbf{\Sigma}_i + \boldsymbol{\psi}_j \sigma_j^2 \boldsymbol{\psi}_j^H) \tag{6.122}$$

$$= \det(\mathbf{\Sigma}_i)\det(1 + \sigma_j^2 \boldsymbol{\psi}_j^H (\mathbf{\Sigma}_i)^{-1}\boldsymbol{\psi}_j) \tag{6.123}$$

108

where $\mathbb{E}[a_j^2] = \sigma_j^2 \propto \exp(-j\frac{\nu}{\Gamma})$ and we have used the matrix determinant lemma.

Now we can write,

$$\frac{\det(\mathbf{\Sigma}_{i+1})}{\det(\mathbf{\Sigma}_i)} = 1 + \sigma_{a_j}^2 \boldsymbol{\psi}_j^H (\mathbf{\Sigma}_i)^{-1} \boldsymbol{\psi}_j \tag{6.124}$$

$$= c_{i+1} \tag{6.125}$$

Therefore, the inverse of Eq. (6.120) is now expressed as,

$$(\mathbf{\Sigma}_{i+1})^{-1} = (\mathbf{\Sigma}_i + \boldsymbol{\psi}_j \sigma_j^2 \boldsymbol{\psi}_j^H)^{-1} \tag{6.126}$$

$$= (\mathbf{\Sigma}_i)^{-1} - (\mathbf{\Sigma}_i)^{-1} \boldsymbol{\psi}_j \left[\frac{1}{\sigma_j^2} + \boldsymbol{\psi}_j^H (\mathbf{\Sigma}_i)^{-1} \boldsymbol{\psi}_j \right]^{-1} \boldsymbol{\psi}_j^H (\mathbf{\Sigma}_i)^{-1} \tag{6.127}$$

$$= (\mathbf{\Sigma}_i)^{-1} - \sigma_j^2 (\mathbf{\Sigma}_i)^{-1} \boldsymbol{\psi}_j \left[1 + \sigma_j^2 \boldsymbol{\psi}_j^H (\mathbf{\Sigma}_i)^{-1} \boldsymbol{\psi}_j \right]^{-1} \boldsymbol{\psi}_j^H (\mathbf{\Sigma}_i)^{-1} \tag{6.128}$$

$$= (\mathbf{\Sigma}_i)^{-1} - \frac{\sigma_j^2}{c_{i+1}} (\mathbf{\Sigma}_i)^{-1} \boldsymbol{\psi}_j \boldsymbol{\psi}_j^H (\mathbf{\Sigma}_i)^{-1} \tag{6.129}$$

Thus the likelihood function of Eq. (6.118) is computed in an order recursive way for each cluster with the reduced complexity in evaluating the terms as shown above.

To further reduce the computations, the likelihood $p(\mathbf{y}|\underline{\boldsymbol{\alpha}})$ is computed over only those orthogonal clusters which are included in the set $\mathcal{S}$ of Eq. (6.50) and used in Eq. (6.49) to obtain the Low-Complexity AMMSE (LC-AMMSE) channel estimate for the case of Gaussian amplitudes where the co-variance matrices and determinants for the orthogonal clusters are computed in a recursive

way as shown above.

Figure 6.3 shows the NRMSE in the estimation of the support using AMMSE estimation and Figure 6.4 shows the correponding performance in terms of the energy capture.
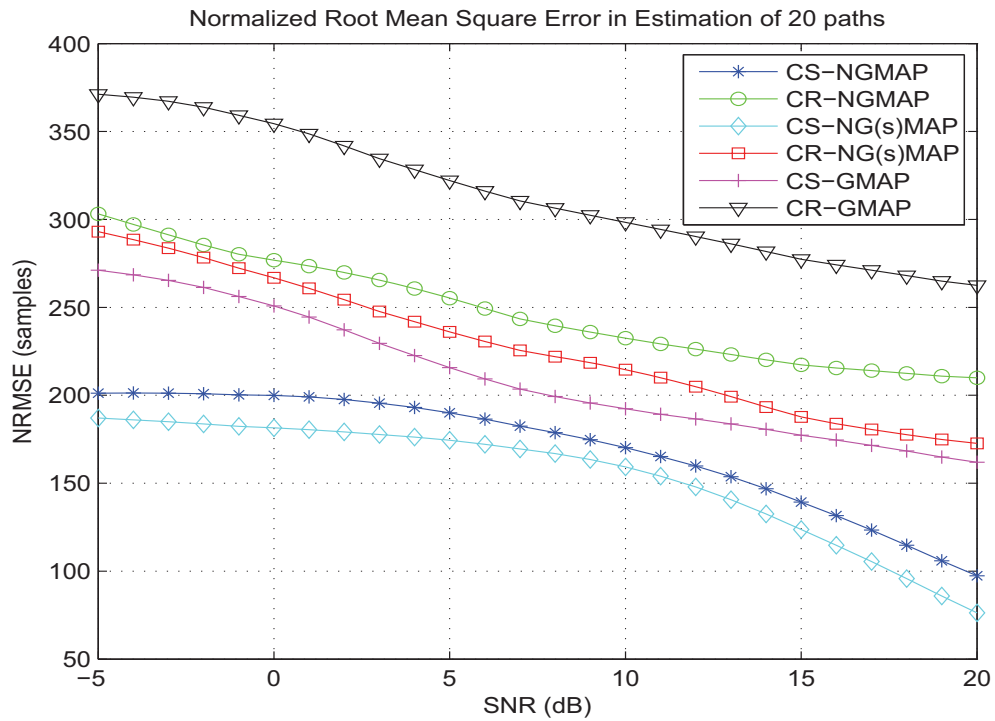
6.6 Results

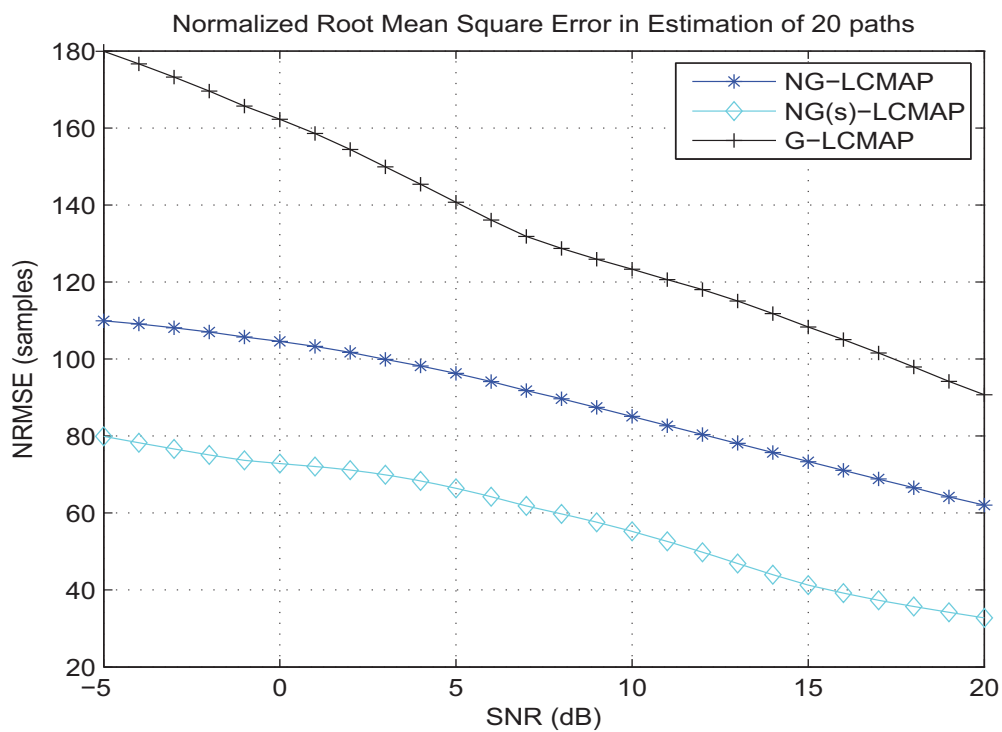Figure 6.1: NRMSE in MPCs Arrival Time Estimation using Decomposed Channel Estimation

Figure 6.2: NRMSE in MPCs Arrival Time Estimation using Decomposed Channel Estimation
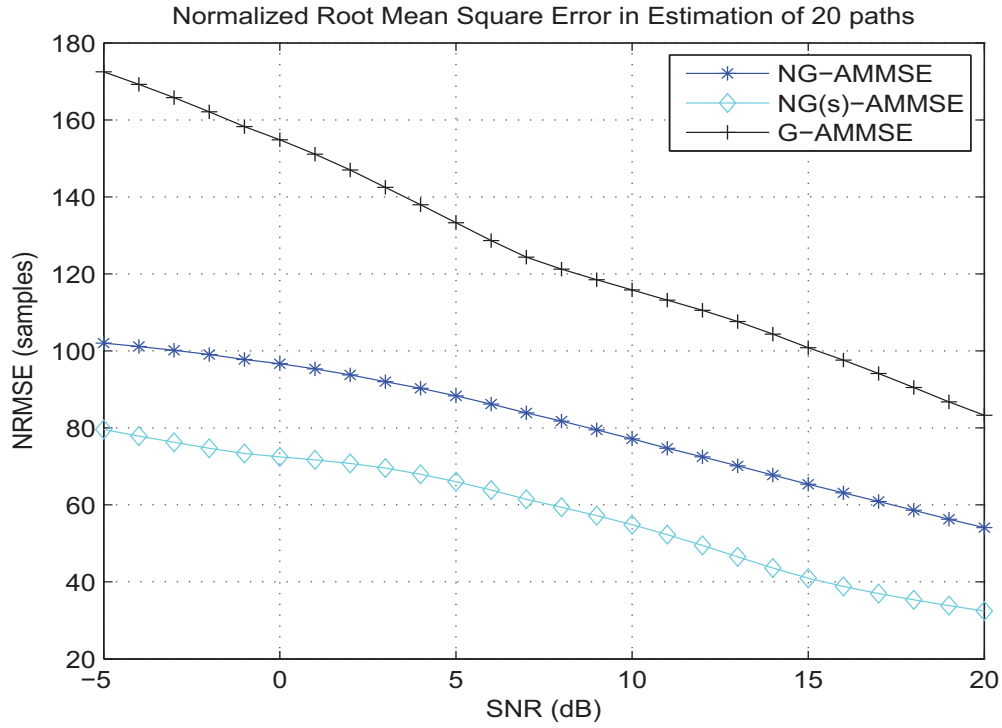
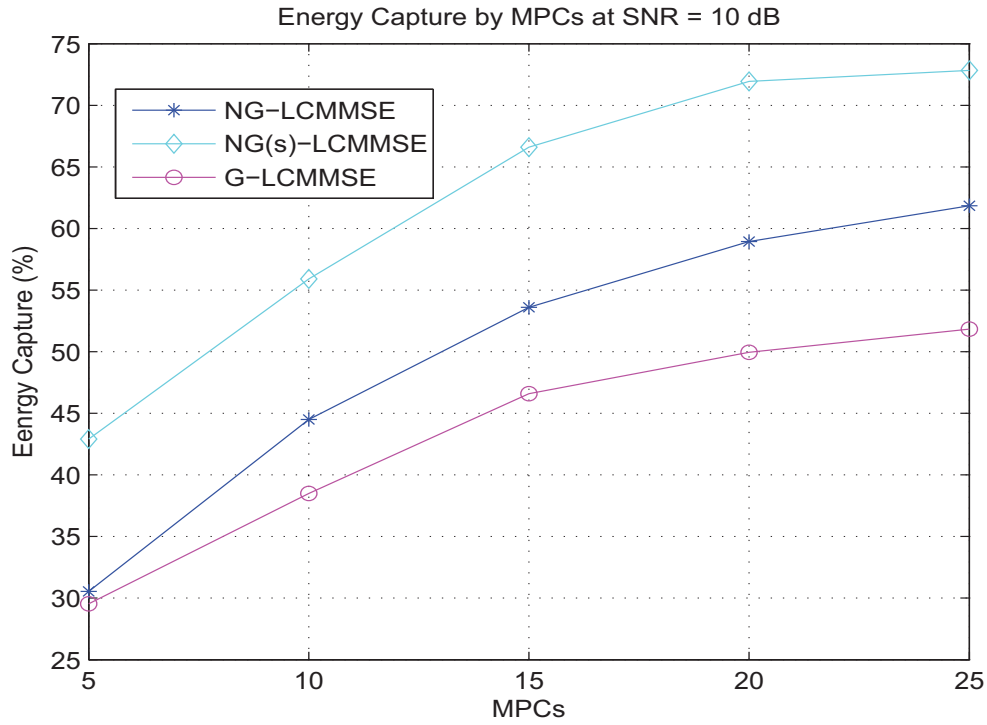Figure 6.3: NRMSE in MPCs Arrival Time Estimation using Joint Channel Estimation

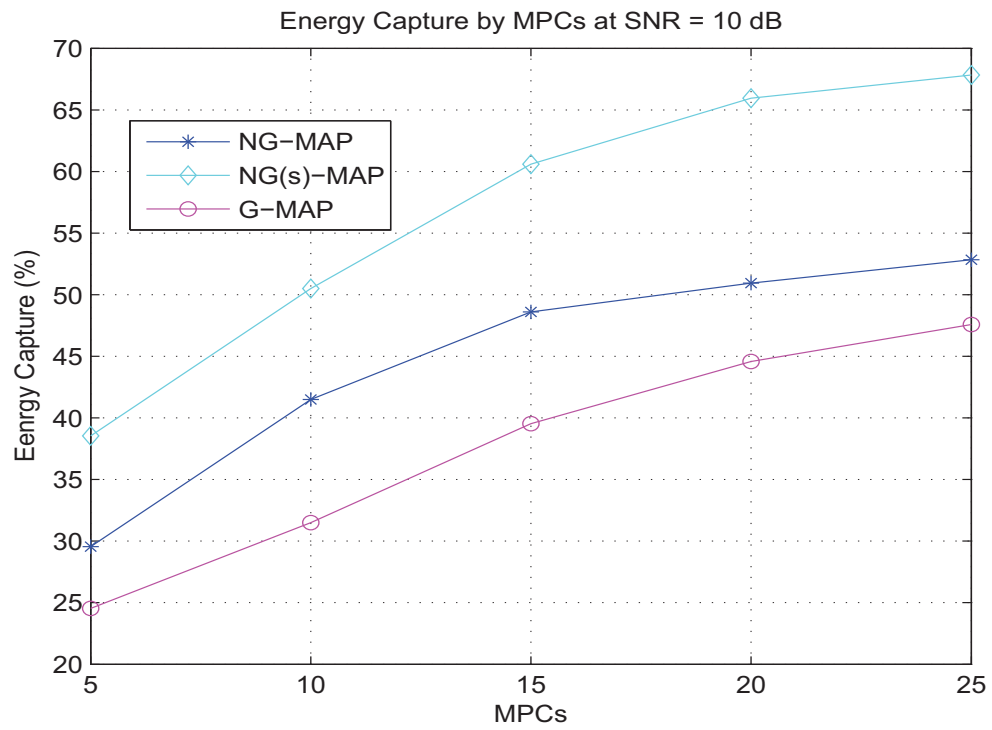Figure 6.4: Energy Capture using LC-AMMSE Estimation

Figure 6.5: Energy Capture using LC-AMMSE Estimation

CHAPTER 7

CONCLUSION

7.1 Conclusion

The Bayesian estimation provided the best estimates of the UWB channel parameters. Among the Bayesian estimates the AMMSE channel estimate for the case of non-Gaussian amplitudes with known 2nd order statistics performed the best. Among the Classical estiamtion methods GA followed by ML provided the best estimates of the channel but it requires a very high computation time primarily because of the GA step. The LC-MAP was reasonably good but had a remarkably reduced computational complexity.

7.2 Future Work

In the future work, we would like to include the effect of UWB channel's frequency selective fading in the model and develop low-complexity channel estimator. Another future work is to address the order recursive step for the case of Gaussian

amplitudes where the non-identical variances inhibit the usefulness of the metrics

from one orthogonal cluster to the next.

REFERENCES

[1] A. Molisch, K. Balakrishnan, C. Chong, S. Emami, A. Fort, J. Karedal, J. Kunisch, H. Schantz, U. Schuster, and K. Siwiak, "IEEE 802.15. 4a channel model-final report," *IEEE P*, vol. 15, pp. 802–1504, 2006.

[2] S. Gezici, Z. Sahinoglu, A. Molisch, H. Kobayashi, and H. Poor, "A two-step time of arrival estimation algorithm for impulse radio ultra wideband systems," *EUSIPCO*, 2005. [Online]. Available: http://arxiv.org/abs/cs.IT/0507006

[3] F. M. Naini, R. Gribonval, L. Jacques, and P. Vandergheynst, "Compressive sampling of pulse trains: Spread the spectrum!" *2009 IEEE International Conference on Acoustics, Speech and Signal Processing*, pp. 2877–2880, Apr. 2009. [Online]. Available: http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=4960224

[4] A. F. Molisch, P. Orlik, Z. Sahinoglu, and J. Zhang, "UWB-based sensor networks and the IEEE 802.15.4a standard - a tutorial," *2006 First International Conference on Communications*

and Networking in China, pp. 1–6, Oct. 2006. [Online]. Available:
http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=4149882

[5] a.F. Molisch, K. Balakrishnan, D. Cassioli, S. Emami, a. Fort, J. Karedal,
J. Kunisch, H. Schantz, and K. Siwiak, "A comprehensive model for ultra-
wideband propagation channels," GLOBECOM '05. IEEE Global Telecom-
munications Conference, 2005., pp. 3648–3653, 2005. [Online]. Available:
http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=1578452

[6] M. Davenport, P. Boufounos, M. Wakin, and R. G. Baraniuk, "Signal Pro-
cessing with Compressive Measurements," IEEE Journal of Selected Topics
in Signal Processing, vol. 4, no. 2, pp. 445–460, 2010.

[7] R. a. Scholtz, D. M. Pozar, and W. Namgoong, "Ultra-
Wideband Radio," EURASIP Journal on Advances in Signal Pro-
cessing, vol. 2005, no. 3, pp. 252–272, 2005. [Online]. Available:
http://www.hindawi.com/journals/asp/2005/758540.abs.html

[8] S. Roy, J. Foerster, V. Somayazulu, and D. Leeper, "Ul-
trawideband Radio Design: The Promise of High-Speed,
Short-Range Wireless Connectivity," Proceedings of the IEEE,
vol. 92, no. 2, pp. 295–311, Feb. 2004. [Online]. Available:
http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=1266915

[9] J. Hu and T. Lv, "Low-complexity frequency-domain algorithm for
UWB timing," Proceedings of the 4th international conference on

mobile technology, applications, and systems and the 1st international symposium on Computer human interaction in mobile technology - Mobility '07, vol. 07, p. 79, 2007. [Online]. Available: http://portal.acm.org/citation.cfm?doid=1378063.1378077

[10] M. Win and R. Scholtz, "Characterization of ultra-wide bandwidth wireless indoor channels: a communication-theoretic view," *IEEE Journal on Selected Areas in Communications*, vol. 20, no. 9, pp. 1613–1627, Dec. 2002.

[11] R. Qiu, "A generalized time domain multipath channel and its application in ultra-wide-band (UWB) wireless optimal receiver design: system performance analysis," *2004 IEEE Wireless Communications and Networking Conference (IEEE Cat. No.04TH8733)*, pp. 901–907, 2004. [Online]. Available: http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=1311306

[12] Z. N. C. Huseyin Arslan and M.-G. D. Benedetto, *Ultra Wideband Wireless Communication*, 2006.

[13] L. Yang and G. Giannakis, "Timing Ultra-Wideband Signals With Dirty Templates," *IEEE Transactions on Communications*, vol. 53, no. 11, pp. 1952–1963, Nov. 2005.

[14] a.F. Molisch, D. Cassioli, S. Emami, a. Fort, B. Kannan, J. Karedal, J. Kunisch, H. Schantz, K. Siwiak, and M. Win, "A Comprehensive Standardized Model for Ultrawideband Propagation Channels," *IEEE Transactions on Antennas and Propagation*,

vol. 54, no. 11, pp. 3151–3166, Nov. 2006. [Online]. Available: http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=4012455

[15] W. Malik, D. Edwards, and C. Stevens, "Frequency Dependence of Fading Statistics for Ultrawideband Systems," *IEEE Transactions on Wireless Communications*, vol. 6, no. 3, pp. 800–804, Mar. 2007. [Online]. Available: http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=4133863

[16] T. Blumensath and M. E. Davies, "Stagewise Weak Gradient Pursuits," *IEEE Transactions on Signal Processing*, vol. 57, no. 11, pp. 4333–4346, Nov. 2009. [Online]. Available: http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=5109633

[17] Z. Wang, G. Arce, B. Sadler, J. Paredes, and X. Ma, "Compressed detection for pilot assisted ultra-wideband impulse radio," in *Ultra-Wideband, 2007. ICUWB 2007. IEEE International Conference on.* IEEE, 2007, pp. 393–398. [Online]. Available: http://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=4380976

[18] L.-C. Wang and W.-C. Liu, "Bit error rate analysis in IEEE 802.15.3a UWB channels," *IEEE Transactions on Wireless Communications*, vol. 9, no. 5, pp. 1537–1542, May 2010. [Online]. Available: http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=5463206

[19] D. Yang, H. Li, and G. D. Peterson, "Message passing Bayesian Compressed Sensing for UWB pulse acquisition," *2010 44th Annual Conference on Infor-*

119

mation Sciences and Systems (CISS), pp. 1–5, Mar. 2010. [Online]. Available: http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=5464905

[20] L. Yang, "Ultra-Wideband Communications: From Concept to Reality," Ph.D. dissertation, 2004.

[21] D. M. Malioutov, S. R. Sanghavi, and A. S. Willsky, "Sequential Compressed Sensing," *IEEE Journal of Selected Topics in Signal Processing*, vol. 4, no. 2, pp. 435–444, Apr. 2010. [Online]. Available: http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=5418893

[22] E. G. Larsson and Y. Selen, "Linear Regression With a Sparse Parameter Vector," *IEEE Transactions on Signal Processing*, vol. 55, no. 2, pp. 451–460, Feb. 2007. [Online]. Available: http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=4063557

[23] R. Robucci, J. D. Gray, L. K. Chiu, J. Romberg, and P. Hasler, "Compressive Sensing on a CMOS Separable-Transform Image Sensor," *Proceedings of the IEEE*, vol. 98, no. 6, pp. 1089–1101, Jun. 2010. [Online]. Available: http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=5452994

[24] G. R. Arce, J. L. Paredes, and B. M. Sadler, "Compressed detection for ultra-wideband impulse radio," *2007 IEEE 8th Workshop on Signal Processing Advances in Wireless Communications*, pp. 1–5, Jun. 2007. [Online]. Available: http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=4401384

[25] V. Lottici, "Low-complexity ML timing acquisition for UWB communications in dense multipath channels," *IEEE Transactions on Wireless Communications*, vol. 4, no. 6, pp. 3031–3038, Nov. 2005. [Online]. Available: http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=1545877

[26] V. Lottici, A. D'Andrea, and U. Mengali, "Channel estimation for ultra-wideband communications," *IEEE Journal on Selected Areas in Communications*, vol. 20, no. 9, pp. 1638–1645, Dec. 2002.

[27] M. Win and R. Scholtz, "Impulse radio: how it works," *IEEE Communications Letters*, vol. 2, no. 2, pp. 36–38, 1998. [Online]. Available: http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=660796

[28] J. Foerster, I. A. Labs, and I. Corp, "Ultra-Wideband Technology for Short- or Medium-Range Wireless Communications," pp. 1–11, 2001.

[29] R. Qiu, H. Liu, and X. Shen, "Ultra-wideband for multiple access communications," *IEEE Communications Magazine*, vol. 43, no. 2, pp. 80–87, Feb. 2005. [Online]. Available: http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=1391505

[30] I. Maravic and M. Vetterli, "Low-complexity subspace methods for channel estimation and synchronization in ultra-wideband systems," in *Proc. International Workshop on UWB Systems*, no. 1. Citeseer, 2003.

[31] R. G. Baraniuk, "Compressive sensing," *IEEE Signal Processing Magazine*, vol. 24, no. 4, p. 118, 2007.

[32] J. L. Paredes, G. R. Arce, and Z. Wang, "Ultra-Wideband Compressed Sensing: Channel Estimation," *IEEE Journal of Selected Topics in Signal Processing*, vol. 1, no. 3, pp. 383–395, Oct. 2007.

[33] Z. Wang, G. R. Arce, B. M. Sadler, J. L. Paredes, S. Hoyos, and Z. Yu, "Compressed UWB signal detection with narrowband interference mitigation," *2008 IEEE International Conference on Ultra-Wideband*, vol. 2, pp. 157–160, Sep. 2008. [Online]. Available: http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=4653375

[34] Y. Sel, "Model Selection and Sparse Modeling," Ph.D. dissertation, 2007.

[35] Y. Selen and E. Larsson, "Empirical Bayes linear regression with unknown model order," *Digital Signal Processing*, vol. 18, no. 2, pp. 236–248, Mar. 2008. [Online]. Available: http://linkinghub.elsevier.com/retrieve/pii/S1051200407000383

[36] S. M. Kay, *Fundamentals of Statistical Signal Processing*, 1993.

[37] R. Tibshirani, "Regression shrinkage and selection via the lasso," *Journal of the Royal Statistical Society. Series B (Methodological)*, vol. 58, no. 1, pp. 267–288, 1996. [Online]. Available: http://www.jstor.org/stable/2346178

[38] M. E. Tipping, "Bayesian Inference : An Introduction to Principles and Practice in Machine Learning From Least-Squares to Bayesian Inference," pp. 1–19, 2006.

[39] T. Jia and D. I. Kim, "Analysis of Channel-Averaged SINR for Indoor UWB Rake and Transmitted Reference Systems," *IEEE Transactions on Communications*, vol. 55, no. 9, pp. 1822–1822, Sep. 2007. [Online]. Available: http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=4303343

[40] A. Muqaibel, "Directional modelling of ultra wideband communication channels," *IET Communications*, vol. 4, no. 1, p. 51, 2010. [Online]. Available: http://link.aip.org/link/ICEOCW/v4/i1/p51/s1&Agg=doi

[41] F. O. Karray and C. W. D. Silva, *Soft Computing and Intelligent Systems Design, Theory, Tools and Applications*, May 2004, vol. 17, no. 3. [Online]. Available: http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=1629107

[42] K. Schnass, "Sparsity & Dictionaries - Algorithms & Design," Ph.D. dissertation, 2009.

[43] O. Holtz, "Compressive sensing: a paradigm shift in signal processing," 2008. [Online]. Available: http://arxiv.org/pdf/0812.3137

[44] R. Baraniuk, V. Cevher, M. Duarte, and C. Hegde, "Model-based compressive sensing," *Information Theory, IEEE Transactions on*, vol. 56, no. 4, pp. 1982–2001, 2010. [Online]. Available: http://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=5437428

[45] L. Jacques and P. Vandergheynst, "Compressed Sensing:When sparsity meets sampling," *Signal Processing*, pp. 1–30, 2010. [Online]. Available: http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.162.1020

[46] J. Haupt, W. U. Bajwa, G. Raz, R. Nowak, and A. Background, "Toeplitz Compressed Sensing Matrices with Applications to Sparse Channel Estimation," 2010.

[47] J. A. Tropp and A. C. G. Gilbert, "Signal recovery from Random Measurements Via Orthogonal Matching Pursuit," *IEEE Transactions on Information Theory*, vol. 53, no. 12, pp. 4655–4666, 2007.

[48] S. G. Mallat and Z. Zhang, "Matching Pursuits With Time-Frequency Dictionaries," *IEEE Transactions On Signal Processing*, vol. 41, no. 12, pp. 3397–3415, 1993.

[49] T. Blumensath, "On the difference between orthogonal matching pursuit and orthogonal least squares," *Proceedings of the IEEE*, vol. 44, no. 0, pp. 1–3, 2007. [Online]. Available: http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.63.3258

[50] T. Blumensath and M. E. Davies, "Gradient Pursuits," *IEEE Transactions on Signal Processing*, vol. 56, no. 6, pp. 2370–2382, Jun. 2008. [Online]. Available: http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=4480155

[51] M. Navarro, S. Prior, and M. Najar, "Low Complexity Frequency Domain TOA Estimation for IR-UWB Communications," *IEEE Vehicular Technology Conference*, pp. 1–5, Sep. 2006. [Online]. Available: http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=4109420

[52] L. Stoica, a. Rabbachin, and I. Oppermann, "A low-complexity noncoherent IR-UWB transceiver architecture with TOA estimation," *IEEE Transactions on Microwave Theory and Techniques*, vol. 54, no. 4, pp. 1637–1646, Jun. 2006. [Online]. Available: http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=1618591

[53] I. Guvenc, Z. Sahinoglu, and P. Orlik, "TOA estimation for IR-UWB systems with different transceiver types," *IEEE Transactions on Microwave Theory and Techniques*, vol. 54, no. 4, pp. 1876–1886, Jun. 2006. [Online]. Available: http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=1618617

Vitae

- Name: Syed Faraz Ahmed

- Nationality: Pakistani

- Date of Birth: March 31, 1987

- Email: *faraz107@gmail.com*

- Permenant Address: A-851 Sector 11-B North Karachi, Karachi, Pakistan.

- Received Bachelor of Engineering (B.E.) in Electronic Engineering from N.E.D University of Engineering and Technology, Karachi, Pakistan in December 2007.

- Joined King Fahd University of Petroleum and Minerals, Dhahran, Saudi Arabia as a Research Assistant in October 2008.

- Completed Master of Science (M.S.) in Electrical Engineering in June 2011.