uOttawa

L'Université canadienne
Canada's university

FACULTÉ DES ÉTUDES SUPÉRIEURES ET POSTOCTORALES

uOttawa

L'Université canadienne
Canada's university

FACULTY OF GRADUATE AND POSDOCTORAL STUDIES

Güneş Karabulut

GAUTEUR DE LA THÈSE / AUTHOR OF THESIS

Ph.D. (Electrical Engineering)

GRADE / DEGREE

School of Information Technology and Engineering

FACULTÉ, ÉCOLE, DÉPARTEMENT / FACULTY, SCHOOL, DEPARTMENT

Matching Pursuit Algorithms with Applications in Wireless Communication Systems

TITRE DE LA THÈSE / TITLE OF THESIS

Abbas Yongacoglu

DIRECTEUR (DIRECTRICE) DE LA THÈSE / THESIS SUPERVISOR

CO-DIRECTEUR (CO-DIRECTRICE) DE LA THÈSE / THESIS CO-SUPERVISOR

EXAMINATEURS (EXAMINATRICES) DE LA THÈSE / THESIS EXAMINERS

Martin Bouchard

Lucia Moura

Khalid Sayood

Halim Yanikomeroglu

Gary W. Slater

LE DOYEN DE LA FACULTÉ DES ÉTUDES SUPÉRIEURES ET POSTDOCTORALES /
DEAN OF THE FACULTY OF GRADUATE AND POSTDOCORAL STUDIES

University of Ottawa

Matching Pursuit Algorithms with Applications in Wireless Communication Systems

Güneş Karabulut

Ottawa-Carleton Institute for Electrical and Computer Engineering

School of Information Technology and Engineering

Faculty of Engineering

University of Ottawa

07.04.2006

Supervisor: Dr. Abbas Yongaçoğlu

Your file Votre référence
ISBN: 978-0-494-18588-9
Our file Notre référence
ISBN: 978-0-494-18588-9

Abstract

Signal decomposition finds its applications in a wide range of areas. The problem of basis selection for signal decomposition consists of determining a small, possibly smallest, subset of vectors chosen from a large redundant set of vectors to match the given data. However, it is shown that finding the sparsest solution to linear systems is a nondeterministic polynomial time (NP) hard problem and requires combinatorial search. For a simpler solution, greedy heuristic algorithms, namely matching pursuit (MP) algorithms are proposed. These algorithms require sequential selection of basis vectors from a set of vectors, termed as dictionary. This dictionary can be undercomplete, overcomplete or complete, depending on the number of basis vectors it contains.

The contributions of this dissertation are composed of two complementary parts. In the first part, the limitations of the MP based algorithms are considered. Combination of orthogonal matching pursuit (OMP) algorithm with tree-based combinatorial search techniques is proposed in order to overcome the limitations of the greedy heuristic algorithms. A novel flexible tree-search based OMP (FTB-OMP) algorithm is introduced for sparse signal representations. The algorithm provides some design parameters, giving flexibility to establish a tradeoff between performance and complexity. The efficiency is achieved by using a correlation based pruning in the search tree, and reducing the number of leaves as the depth of nodes increases. It must be noted that this algorithm can be applied to any kind of detection

problem that can be modeled as a set of linear equations.

In the second part of the dissertation, parameter estimation problems related to wireless communications are considered. These problems are channel estimation, direction of arrival (DOA) estimation and multi-user detection (MUD). These problems are modeled with different types of dictionaries due to their physical characteristics. The channel estimation problem is modeled with undercomplete dictionary. DOA estimation and MUD have overcomplete and complete dictionary models, respectively. It is shown that the family of MP algorithms, especially FTB-OMP, provides a low complexity solution for parameter detection problems with different types of dictionaries. These algorithms provide a practical solution for detection and estimation problems that are widely encountered in wireless communications.

Acknowledgments

I would like to begin by extending my most sincere thanks to my advisor Dr. Abbas Yongaçoğlu for his help, guidance and patience throughout this work. I am also very grateful to Dr. Lucia Moura and Dr. Daniel Panario for their invaluable support throughout the discussion of heuristic algorithm and implementation procedures. This thesis would not be completed without our discussions for several long hours.

I would also like to express my gratitude to Dr. John H. Husøy for introducing me to matching pursuit algorithms during the Multiresolution Signal Decomposition: Analysis and Applications course he offered at the University of Ottawa. His encouraging comments and ideas about integration of the algorithm to different problems helped me a lot to determine my thesis topic.

I would to like to thank Dr. Martin Bouchard, Dr. Khalid Sayood and Dr. Halim Yanıkömeroğlu for spending time by serving on my examination committee. Their valuable comments helped me shape up the final version of my thesis.

I would like to acknowledge Dr. John Litva from TenXc Wireless for his guidance about engineering and for his understanding throughout the time period that I worked for long hours on my thesis.

I owe many thanks to my family, for understanding and supporting me in all my decisions. Special thanks goes to my grandmother for calling me every single day from Istanbul for more than 5 years and for all her prayers. I also need to thank my sister Deniz Karabulut for spreading joy around, especially in the times that I needed it the most.

Finally, I would like to express my deepest appreciation and gratitude to my husband Tolga Kurt for his extraordinary support, patience, love and knowledge. Without him I could not have completed this work.

Gunes Karabulut

Ottawa, April 2006.

to my family...

Contents

List of Figures

List of Tables

Acronyms

ABMP	Adaptive Basic Matching Pursuit
AOMP	Adaptive Orthogonal Matching Pursuit
AWGN	Additive White Gaussian Noise
BF	Block Fading
BOOMP	Backward Optimized Orthogonal Matching Pursuit
BMP	Basic Matching Pursuit
BS	Basis Pursuit
BPSK	Binary Phase Shift Keying
CDMA	Code Division Multiple Access
CE	Channel Estimation
CRB	Cramér-Rao Bound
DFE	Decision Feedback Equalizer
DOA	Direction of Arrival
EDAMP	Estimation of Directions of Arrival by Matching Pursuit
ESPRIT	Estimation of Signal Parameters via Rotational Invariance Techniques
QBF	Quasi-Block Fading
FTB-OMP	Flexible Tree-search Based Orthogonal Matching Pursuit
I/O	Input/Output
ISI	Inter-Symbol Interference
KLT	Karhunen-Loève Transform
LMS	Least Mean Square
LS	Least Squares
MA	Multiple Access
MIMO	Multiple Input Multiple Output
MISO	Multiple Input Single Output

MMSE	Minimum Mean Square Error
MSE	Mean Square Error
MSIE	Mean Squared Identification Error
MUD	Multi-User Detector
MUSIC	Multiple Signal Classification
MVDR	Minimum variance distortionless response
MVUE	Minimum Variance Unbiased Estimator
NWAF	Normalized Windowed Autocorrelation Function
O-CDMA	Optical Code Division Multiple Access
OMP	Orthogonal Matching Pursuit
ORMP	Order Recursive Matching Pursuit
OOMP	Optimized Orthogonal Matching Pursuit
RMSE	Root Mean Square Error
RLS	Recursive Least Squares
SBS	Sequential Basis Selection
SIC	Successive Interference Canceler
SIMO	Single Input Multiple Output
TB-MP	Tree-search Based Matching Pursuit
TV	Time Variant
TI	Time Invariant
ULA	Uniform Linear Array
WOMP	Weak Orthogonal Matching Pursuit

Notation

$\mathbf{A}$	Dictionary matrix
$\mathbb{C}$	Set of complex numbers
$\mathcal{B}$	Orthonormal basis set
B	Block length for channel estimation applications
$\mathcal{D}$	Dictionary for sequential basis selection algorithm
$\mathbb{H}$	Hilbert space
$\mathbf{e}_M$	Residue vector at M^{th} iteration
$\mathbb{N}$	Set of natural numbers
ξ	Correlation threshold for tree search based sequential basis selection algorithms
ϵ	Error threshold defining a stopping criterion for sequential basis selection algorithms
φ_λ	Dictionary atom (vector)
Λ	Index set of approximant signal
$\{c_\lambda\}$	Set of representation coefficients
Γ	Index set of the dictionary
K	Number of orthonormal bases
M	Sparsity index, the number of atoms selected to represent the input signal
P	Number of distinct columns of dictionary $\mathcal{D}$
Q	Number of basis functions in basis expansion model of time variant channels
σ^2	Variance of AWGN
N	Size of input signal $\mathbf{x}$
$\mathbf{n}$	AWGN component
$\mathbf{x}$	Input signal to be approximated or resolved
$\mathbf{x}_M$	Approximation to input signal with M components

Chapter 1

Introduction and Motivations

With the advent of technology, the speed and robustness requirement of transmission and signal processing increased. This led to the search of more efficient signal representations, bringing the necessity of signal decomposition. Signal decomposition finds its applications in a wide range of areas.

The traditional signal decomposition techniques use linear expansion techniques such as Fourier transform, wavelet transform and discrete cosine transform (DCT). These techniques are very powerful for specific symbol types. For example, Fourier transform is efficient when signals are well localized in frequency. However, it provides a poor representation when signals are well localized in time. This shows us the necessity for flexible representations for non-stationary signals.

The flexibility can be achieved by using redundant vectors, i.e., more vectors than necessary, to span the space of interest in the signal expansion. The set of these vectors is termed as an overcomplete dictionary. The time-frequency and time-scale communities have developed a large number of overcomplete waveform dictionaries such as stationary wavelets, wavelet packets, cosine packets, chirplets, and warplets that constitute the frame theory. A

detailed overview of the frame theory can be found in [17].

The problem of basis selection for signal decomposition consists of determining a small, possibly the smallest, subset of vectors chosen from a large redundant set of vectors to match the given data. In the literature, the basis selection algorithms are treated with the specific application that they are used for. These applications include time/frequency representations [73], speech coding [64], spectral estimation [14], video coding [4,77–79,111], and channel estimation [25]. The main motivation for considering the basis selection problem is this vast possibility of application areas as shown throughout the dissertation.

The sparse linear approximation problem is a nonlinear optimization problem for which there are no universal methods for a globally optimum solution [16]. Since, finding an optimal solution using an exhaustive search is infeasible, suboptimal methods that require lower computational complexity are proposed. A common approach that yields a unique solution is to find the minimum 2-norm solution [40]. The goal of the 2-norm solution is to obtain many non-zero entries, which is contradictory to the sparsity of basis selection methods [2, 40]. Other methods include the method of frames (MOF) [28] and for special dictionaries, the best orthogonal basis (BOB) [19].

As an efficient solution, the basic matching pursuit (BMP) algorithm was proposed by Mallat [73]. Later on, several different versions of the BMP algorithm were proposed. The different algorithms are referred to as matching pursuit (MP) algorithms throughout this dissertation.

Although MP algorithms are suboptimal, they give practical results in detection and estimation problems. These algorithms reduce the computational complexity by employing a greedy strategy[1]. Due to this strategy in MP algorithms, vectors are selected one by one from the dictionary and the approximation is optimized at each step. The selection criteria

[1] Greedy algorithms construct a set of objects from the smallest possible constituent parts recursively [65].

varies between the MP algorithms forming different variations such as, orthogonal matching pursuit (OMP) [29,82], order recursive matching pursuit (ORMP) [13,76], optimized orthogonal matching pursuit (OOMP) [85] and backward optimized orthogonal matching pursuit (BOOMP) [5].

A slight variation of the optimization problem results in a linear programming algorithm, that is named the basis pursuit (BP) algorithm [15]. Although an optimum representation in terms of absolute value norm can be obtained by the BP algorithm, the complexity increases enormously. There are also parallel basis selection algorithms that contradict the greedy approach [40]. Although parallel basis selection algorithms give better performance results, their high computational complexity makes them unattractive for applications where processing time is a concern especially for the cases where the input dictionary is complete or overcomplete [21].

In this dissertation, the main focus is on the OMP algorithm for parameter detection problems. The motivations behind this choice can be summarized as follows.

Sparseness The OMP algorithm, along with the other MP algorithms, guarantees that only few coefficients are evaluated for signal approximation. This is not the case for other algorithms such as least-squares techniques or method of frames.

Speed The fast convergence of the OMP algorithm relative to BMP algorithm is a well known result [73] with a slight increase in the computational complexity. Although, BMP is the most popular MP algorithm in the literature, it is empirically shown that it has inferior performance than OMP algorithm.

Solid theory Although the sparse approximation problem has been dealt with for over fifty years, there have been limited theoretical results for the proposed algorithms. Over these years, researchers have published a wide variety of heuristic algorithms including the MP algorithms, however, there is no proof that these methods can fully recover the

input signal. One exception to these algorithms is the OMP algorithm. Recently, Tropp showed some convergence properties for specific dictionary types for OMP algorithm [104], proving that the optimum representation of a signal can be obtained with OMP algorithm for specific dictionary types.

In the following section, mathematical notation is explained along with the definition of the sparse approximation problem. The main contributions areas of this dissertation are highlighted in Section 1.2.

1.1 Problem Statement

Consider the space generated by signals of size N. Let $\mathcal{D}= \{\varphi_\lambda\}_{\lambda\in\Gamma}$ be a set of P redundant vectors where $P > N$ and with N linearly independent vectors that define $\mathbb{C}^N$ of length N signals [2]. Since the application of the sparse approximation problem will focus on physical parameter estimation problems, it is more convenient to concentrate on finite dimensional inner-product spaces. In the notation, Γ is the index set and $|\Gamma| = |\mathcal{D}| = P$, where $|\cdot|$ returns the cardinality of a set.

The sparse approximation problem is the problem of approximating a given input signal, $\mathbf{x}$, with a linear combination of M elements that are selected from a highly redundant dictionary, $\mathcal{D}$. The basis vectors of a dictionary are referred to as atoms. The goal is to construct an M-term approximation to $\mathbf{x}$, denoted by $\mathbf{x}_M$ with $M < N$, that gives a solution to the following minimization problem[3]

$$\min_{|\Lambda|=M} \left\{ \min_{\{c_\lambda\}} \|\mathbf{x} - \sum_{\lambda\in\Lambda} c_\lambda \varphi_\lambda\| \right\}, \tag{1.1}$$

[2] Although finite dimensional vectors are used in this notation, it is possible to convert the theory to infinite dimensional Hilbert spaces.

[3] Throughout the text $\|\cdot\|$ represents the Euclidean norm unless otherwise stated.

where the set of complex coefficients for approximation are denoted by $\{c_\lambda\}$, and the index set $\Lambda \in \Gamma$.

It is known that the minimization over the coefficients can be accomplished by least squares techniques [37]. The more challenging part of this problem is the selection of the index set Λ. It is shown that finding the sparsest solution to an underdetermined linear system is a nondeterministic polynomial time (NP) hard problem [40, 76].

The approximant signal $\mathbf{x}_M$ is a linear combination of M atoms

$$\mathbf{x}_M = \sum_{\lambda \in \Lambda} c_\lambda \varphi_\lambda. \tag{1.2}$$

Another name for M that is used throughout this dissertation is the sparsity index [15]. Recovering the signal corresponds to identifying the atoms and corresponding coefficients that compose the actual input signal.

1.2 Contributions

In this dissertation, we have addressed the basis selection problem. The importance of this problem is demonstrated for several different applications in wireless communication systems. These applications areas included the channel estimation problem, direction of arrival detection problem and the multi-user detection problem.

The main contributions of this dissertation are:

- **Improving sequential basis selection algorithms**

  Due to the greedy structure of the sequential basis selection algorithms, an erroneous basis selection at an iteration will impact the resulting signal approximation. The selection of a basis vector not leading to the sparsest solution is defined as the error propagation problem.

In this research, combinations of orthogonal matching pursuit (OMP) algorithm with tree-search based techniques are proposed in order to reduce error propagation problems. Although the tree-search based MP algorithms were previously proposed in [24], their advantages were left uncommented. The error propagation problem is considered as the research topic in Chapter 3 where an efficient tree-search based OMP algorithm for sparse signal representations is proposed. The proposed algorithm is named the flexible tree-search based orthogonal matching pursuit (FTB-OMP) algorithm.

The algorithm provides some design parameters that give flexibility to establish a trade-off between performance and running time. The efficiency is achieved by using a correlation based pruning in the search tree, and reducing the number of children as the depth of nodes increase.

- **Application of sequential basis selection algorithms in wireless communication problems**

  Sparse solution requirement for an observed data vector is a frequently encountered problem in wireless communication applications. After reducing the error propagation problem in OMP algorithm or evaluating the limitations of the OMP algorithm, several parameter detection problems can be tackled. Three major areas that are considered for application of OMP algorithm are channel estimation, direction of arrival detection and multi-user detection problems. Of all three applications, only the channel estimation problem has been previously treated using the BMP algorithm in the literature [25]. The two other proposed areas are introduced for the first time in this dissertation. These problems are carefully selected in order to demonstrate the effectiveness of OMP and FTB-OMP algorithms for three different types of dictionaries, namely; undercomplete, overcomplete, and complete dictionaries. Details of these estimation problems are summarized below.

1. **Channel Estimation - Undercomplete dictionary example**

   Sparse channels are encountered in several communication applications. Exploiting the sparsity, a channel estimate can be obtained by using matching pursuit (MP) algorithms. Previously, it was demonstrated that the BMP based channel estimation outperforms the conventional least squares (LS) estimation algorithm for sparse channels [21].

   In this dissertation, OMP and FTB-OMP algorithms are applied to channel estimation problem considering both block fading and quasi-block fading channels, showing cases where the detection performance converges to minimum variance unbiased estimator.

2. **Direction of arrival estimation - Overcomplete dictionary example**

   In recent years, the impact of adaptive antennas and array processing to the system performance of wireless communication systems has gained intense attention.

   One of the most important problems for adaptive antenna systems in order to perform well is to have reliable reference inputs. These references include array element positions and characteristics, directions of arrivals, planar properties and dimensionality of the incoming signals.

   The direction of arrival (DOA) estimation problem is defined for overcomplete dictionary structures, and we show shown that MP based algorithms are very effective in DOA estimation. Furthermore, we verify with simulation results that FTB-OMP algorithm can provide super-resolution property that is necessary for accurate estimation of DOA.

3. **Multi-user detection - Complete dictionary example**

   Code division multiple access (CDMA) has been selected as the multiple access scheme in 3^{rd} generation (3G) mobile communication systems. In CDMA systems, all users share the same frequency band at all times. In order to differentiate

between different users, each user is assigned a code. Multiple access interference is the major constraint on the performance of CDMA systems. Multi-user detection (MUD) has been used in order to increase the system capacity through mitigation of multiple access interference.

We show the equivalence between BMP algorithm and the successive interference cancelation for MUD. Using these relations, a low complexity near-far resistant blind multi-user detector implementation is detailed. Then, we introduce an OMP based detector by using the structural advantages of OMP over BMP.

As we were studying the MUD problem in wireless communications, we have applied the proposed algorithms and procedures to optical CDMA systems. Since this does not fit in line with the topic of this dissertation, this work on optical CDMA is included in Appendix B.

1.3 Outline of Dissertation

The rest of this dissertation is organized as follows. In Chapter 2, literature review on the matching pursuit (MP) algorithms are given. The basic matching pursuit (BMP) and orthogonal matching pursuit (OMP) algorithms are explained.

In Chapter 3, tree-search based MP algorithms are reviewed. A novel algorithm entitled flexible tree-search based orthogonal matching pursuit (FTB-OMP) is introduced. The proposed FTB-OMP algorithm uses correlation based pruning in tree search structures for lower computational complexity. Experimental results are displayed proving the effectiveness of the proposed algorithms.

Channel estimation problem is considered in Chapters 4 and 5. In Chapter 4, block fading channels are of concern. OMP and FTB-OMP algorithms are applied for channel detection.

A training sequence design method is introduced. Simulation results are presented showing the detection and estimation performances of the channel taps.

In Chapter 5, MP based algorithms are applied to quasi-block fading channels. Adaptive orthogonal matching algorithm is proposed for detection of time varying channel taps. Proposed algorithms are compared with polynomial based channel expansion models.

Application of MP algorithms for direction of arrival problem is given in Chapter 6. Problem statement is formulated so that possible directions are included in the designed overcomplete dictionary. BMP, OMP and FTB-OMP algorithms are compared with the well known existing techniques in terms of simulations.

Wireless CDMA multi-user detection problem is considered in Chapter 7. Existing algorithms are reviewed. Relations between interference cancelation procedures and MP algorithms are introduced. Several detector structures are proposed and simulation results are presented.

In Appendix A, Hilbert space definition is reviewed for completeness. Extension of wireless CDMA multi-user detection work to optical systems is given in Appendix B. List of publications related to this dissertation is given in Appendix C.

Chapter 2

Matching Pursuit Algorithms

This chapter introduces the signal approximation concept and concentrates on the sparse signal approximation problem. The required background information about matching pursuit (MP) algorithms are presented. MP algorithms of interest; the basic matching pursuit (BMP) and the orthogonal matching pursuit (OMP) algorithms are summarized. Their approximation performances are compared. Convergence conditions for the OMP algorithm to the sparsest representation are reviewed.

2.1 Signal Approximations

The signal approximation techniques can be categorized into two major classes:

1. Linear signal approximations
2. Non-linear signal approximations

In Section 2.1.1 the linear signal approximation technique is summarized. In Section 2.1.2, the outlines of the non-linear signal approximation problem is given, along with the adaptive representation.

2.1.1 Linear Signal Approximations

Let $\mathcal{B}$= $\{\varphi_m\}_{m\in\mathbb{N}}$ be an orthonormal basis set of a Hilbert space $\mathbb{H}$[1]. The best linear approximation of $\mathbf{x} \in \mathbb{H}$, computed with the first M vectors of $\mathcal{B}$ can be obtained as

$$\mathbf{x}_M = \sum_{m=0}^{M-1} \langle \mathbf{x}, \varphi_m \rangle \varphi_m. \tag{2.1}$$

The expression given in (2.1) corresponds to the orthogonal projection of $\mathbf{x}$ over $\mathbf{V}_M = \{\text{span}\{\varphi_m\}\}_{0\leq m<M}$. The residue vector can be evaluated as

$$\mathbf{e}_M = \mathbf{x} - \mathbf{x}_M = \sum_{m=M}^{\infty} \langle \mathbf{x}, \varphi_m \rangle \varphi_m. \tag{2.2}$$

The approximation error can be calculated from the sum of remaining coefficients

$$\|\mathbf{e}_M\|^2 = \|\mathbf{x} - \mathbf{x}_M\|^2 = \sum_{m=M}^{\infty} |\langle \mathbf{x}, \varphi_m \rangle|^2. \tag{2.3}$$

Since $\|\mathbf{x}\|^2 = \sum_{m=0}^{\infty} |\langle \mathbf{x}, \varphi_m \rangle|^2 < +\infty$, the error should decay to zero, i.e.

$$\lim_{M\to+\infty} \|\mathbf{x} - \mathbf{x}_M\|^2 = 0. \tag{2.4}$$

This approximation is efficient if and only if norm of the residue vector decays rapidly when M increases. The convergence depends on the properties of the input signal $\mathbf{x}$ and the basis set $\mathcal{B}$. More details about non-adaptive signal approximation can be found in [37, 90].

It is known that the Karhunen-Loève transform (KLT) is the best basis projection algorithm that produces the smallest average quadratic error [90]. Although it is possible to construct

[1]Properties of an Hilbert are summarized in Appendix A for completeness.

optimal bases such as KLT, the reconstruction method is not adaptive, making the system a bad one when signal structures are not static [41]. In order to have flexible representations adaptive approximations are required. as explained in the following section.

2.1.2 Non-Linear Signal Approximations

Linear approximations project the input signal $\mathbf{x}$ onto M vectors that are selected a priori. However, selection of these M vectors according to the input signal's properties can provide a more precise approximation. Selection of basis vectors according to $\mathbf{x}$ is referred to as non-linear approximation due to the fact that the elements used in the approximation are not pre-selected.

A signal $\mathbf{x} \in \mathbb{H}$ can be approximated with only M vectors that are adaptively selected from the orthonormal basis $\mathcal{B}= \{\varphi_m\}_{m\in\mathbb{N}}$ of $\mathbb{H}$. Let $\mathbf{x}_M$ be the approximation of $\mathbf{x}$ by projecting it over M vectors with indices from Λ, where $|\Lambda| = M$. The approximant vector $\mathbf{x}_M$ can be expressed as

$$\mathbf{x}_M = \sum_{m\in\Lambda} \langle \mathbf{x}, \varphi_m \rangle \varphi_m. \tag{2.5}$$

and the approximation error can be evaluated as

$$\|\mathbf{e}_M\|^2 = \|\mathbf{x} - \mathbf{x}_M\|^2 = \sum_{m\notin\Lambda} |\langle \mathbf{x}, \varphi_m \rangle|^2. \tag{2.6}$$

In order to minimize this error, the indices in Λ must be picked from the vectors that have a high correlation with the input signal $\mathbf{x}$. Since the index set changes according to the input signal, these approximation techniques are adaptive with respect to $\mathbf{x}$.

Non-linear algorithms outperform linear projections by approximating the signal of interest with adaptive vector selection. This adaptive selection introduces a degree of freedom depending on the properties of the signal of interest.

Adaptive Non-linear Signal Approximations

In order to optimize non-linear signal approximations shown in (2.6), the index set Λ must be picked depending on the signal. As the number of the dictionary components increases, more flexibility can be obtained in terms of basis selection. One method is to use structured overcomplete dictionaries that contain redundant basis vectors such as wavelet packet or local cosine dictionaries [72]. However, these include $P = N \log_2 N$ different vectors. The set of orthogonal bases that can be constructed by choosing N linearly independent vectors from P is much smaller than the set of non-orthogonal bases. In order to improve the approximation performance, generally non-orthogonal signal decompositions are studied [15,27,28,30–32,73].

The adaptivity of the non-linear signal expansion can be improved by keeping the number of atoms, i.e. the number of basis vectors, that construct the dictionary larger than the length of the input signal. Consider the space generated by signals of size N. Let $\mathcal{D}= \{\varphi_\lambda\}_{\lambda\in\Gamma}$ be a set of redundant vectors with $|\Gamma| = P$ distinct atoms where $P > N$ and with N linearly independent vectors. For any $M \geq 1$, an approximation to $\mathbf{x}$, $\mathbf{x}_M$ can be evaluated as

$$\mathbf{x}_M = \sum_{\lambda\in\Lambda} c_\lambda \varphi_\lambda, \tag{2.7}$$

where, $|\Lambda| = M$. The freedom of choice in the equation above leads to a combinatorial explosion. Selecting M vectors that represent the data with minimum error would require searching over $\binom{P}{M}$ possible ways in which the basis can be formed to obtain the best approximation. The cost of such a search is prohibitive. For example for $M = 3$ and $P = 200$ this corresponds to a search of 1313400 linear combinations.

For $P > N$, the computation of $\mathbf{x}_M$ that minimizes $\|\mathbf{x} - \mathbf{x}_M\|$ is shown to be an NP hard problem [40, 76]. That is, there is no known polynomial time algorithm to solve this non-linear optimization problem. In order to find practical approximations, some sub-optimal basis selection algorithms are proposed. Those are explained in the following section.

2.2 Matching Pursuit Algorithms

Matching pursuit (MP) algorithms are adaptive approximations that select the approximation vector with no orthogonality constraint. These greedy algorithms are sufficiently good to build compact representations for signals such as speech, music or image data [72]. The greedy method is used when the set of approximating functions is a linear combination of the basis functions. Initially only the first term is optimized. Optimization corresponds to minimizing the discrepancy between the input signal that can be observed as the training data and the current model. This term is kept fixed and then the next term is optimized. This process continues until all M terms are evaluated. This approach is termed as greedy since at any point only a single term is added to the model in order to get a closer approximation. In the neural network literature greedy algorithms are known as network growing algorithms or constructive procedures [16]. Greedy algorithms are frequently used in many statistical methods.

For MP algorithms, since the problem is pursuing the goal of determining a small subset of vectors in the dictionary $\mathcal{D}$, that best match the vector $\mathbf{x}$, the algorithms proposed for solution are termed as matching pursuit algorithms.

Dictionary $\mathcal{D}$ can be classified in three groups.

1. Undercomplete dictionary ($N > P$)

   Encountered in channel estimation problem.

2. Complete dictionary ($N = P$)

   Encountered in multi-user detection problem.

3. Overcomplete dictionary ($N < P$)

   Encountered in direction of arrival estimation problem.

MP algorithms are adaptive due to the fact that the basis functions are selected adaptively for best matching a given data from a fixed dictionary.

In the following sections, the basic matching pursuit (BMP) and the orthogonal matching pursuit (OMP) algorithms are summarized.

2.2.1 Basic Matching Pursuit Algorithm

The basic matching pursuit (BMP) algorithm was proposed in [73]. The algorithm is closely related to projection pursuit algorithm [33] that is used frequently by statisticians, and the shape-gain vector quantizer [35].

Consider the space generated by signals of size N. Let $\mathcal{D}= \{\varphi_\lambda\}_{\lambda\in\Gamma}$ be a set of redundant vectors with P number of vectors where $P > N$ and with N linearly independent vectors that define $\mathbb{C}^N$ of signals length N. We can also assume that $\|\varphi_\lambda\| = 1$ without loss of generality. In MP algorithms, basis selection is performed sequentially, i.e. one at a time. This forces an index requirement for the residual and selected vectors at each iteration.

The BMP algorithm projects $\mathbf{x}$ onto a vector $\varphi_{\lambda_0} \in \mathcal{D}$ and computes the residue vector $\mathbf{e}_1$ as

$$\mathbf{x} = \langle \mathbf{x}, \varphi_{\lambda_0} \rangle + \mathbf{e}_1, \tag{2.8}$$

where $\mathbf{e}_1$ is orthogonal to φ_{λ_0} and

$$\|\mathbf{x}\|^2 = |\langle \mathbf{x}, \varphi_{\lambda_0} \rangle|^2 + \|\mathbf{e}_1\|^2. \tag{2.9}$$

In order to minimize $\|\mathbf{e}_1\|$, the term $|\langle \mathbf{x}, \varphi_{\lambda_0} \rangle|$ can be maximized. So the vector φ_{λ_0} can be picked as

$$|\langle \mathbf{x}, \varphi_{\lambda_0} \rangle| \geq \sup_{\lambda\in\Gamma} |\langle \mathbf{x}, \varphi_\lambda \rangle|. \tag{2.10}$$

The procedure continues the iterations by further sub-decomposing the residue vector. For completeness of iterations, the residue vector at the initial iteration is set as $\mathbf{e}_0 = \mathbf{x}$. Assume that the residue vector at $(m-1)^{th}$ iteration $\mathbf{e}_m$ is computed for $m \geq 1$. The m^{th} iteration selects $\varphi_{\lambda_m} \in \mathcal{D}$ such that

$$|\langle \mathbf{e}_m, \varphi_{\lambda_m} \rangle| \geq \sup_{\lambda \in \Gamma} |\langle \mathbf{e}_m, \varphi_\lambda \rangle|. \tag{2.11}$$

It should be noted that in (2.11), there is no constraint on the selection of an index in terms of the previously selected indices. Hence in the BMP algorithm it is possible to select a previously selected atom. This problem is referred to as the reselection problem. It may be seen that placing a constraint on the previously selected atoms can be a solution. This is not true since such a constraint would prevent the convergence of the algorithm.

The residue vector $\mathbf{e}_m$ is projected on φ_{λ_m} as

$$\mathbf{e}_m = \langle \mathbf{e}_m, \varphi_{\lambda_m} \rangle \varphi_{\lambda_m} + \mathbf{e}_{m+1}, \tag{2.12}$$

and hence the new residue vector is calculated as

$$\mathbf{e}_{m+1} = \mathbf{e}_m - \langle \mathbf{e}_m, \varphi_{\lambda_m} \rangle \varphi_{\lambda_m}. \tag{2.13}$$

The orthogonality of $\mathbf{e}_{m+1}$ and φ_{λ_m} implies

$$\|\mathbf{e}_m\|^2 = |\langle \mathbf{e}_m, \varphi_{\lambda_m} \rangle|^2 + \|\mathbf{e}_{m+1}\|^2. \tag{2.14}$$

Summing (2.12) for $m = 0, \ldots, M-1$ we obtain the M term approximation vector as

$$\mathbf{x}_M = \sum_{m=0}^{M-1} \langle \mathbf{e}_m, \varphi_{\lambda_m} \rangle \varphi_{\lambda_m} \tag{2.15}$$

$$= \sum_{m=0}^{M-1} c_{\lambda_m} \varphi_{\lambda_m}, \tag{2.16}$$

where the coefficients are evaluated as $c_{\lambda_m} = \langle \mathbf{e}_m, \varphi_{\lambda_m} \rangle$ for $m = 0, 1, \ldots M-1$. The input signal becomes

$$\mathbf{x} = \sum_{m=0}^{M-1} c_{\lambda_m} \varphi_{\lambda_m} + \mathbf{e}_M. \tag{2.17}$$

When we apply a similar summation to (2.14) we get

$$\|\mathbf{x}\|^2 = \sum_{m=0}^{M-1} |c_{\lambda_m}|^2 + \|\mathbf{e}_M\|^2. \tag{2.18}$$

An attractive property of the BMP algorithm is the energy conservation stated in (2.18). Due to (2.18), convergence of the algorithm is guaranteed for $M \to \infty$ [73], i.e. for infinite number of iterations for overcomplete dictionaries. However, since in each iteration the optimization is performed over all vectors in the dictionary, it is possible to re-select a previously selected vector, slowing the convergence [72]. As mentioned above, if a constraint is placed for eliminating the reselection problem, the convergence of the algorithm would not be guaranteed.

The algorithm can be terminated when either sufficient number of terms are picked ($m = M$) or when the norm of the residue vector falls below some predetermined threshold value, $\|\mathbf{e}_m\| \leq \epsilon$.

This iterative procedure can implemented with a fast algorithm by exploiting the recursive structure as follows.

1. **Initialization**

   Set $m = 0$, $\mathbf{e}_0 = \mathbf{x}$ and compute $\langle \mathbf{e}_0, \varphi_\lambda \rangle_{\lambda \in \Gamma}$.

2. **Best Match**

   Find $\varphi_{\lambda_m} \in \mathcal{D}$ such that

$$|\langle \mathbf{e}_m, \varphi_{\lambda_m} \rangle| \geq \sup_{\lambda \in \Gamma} |\langle \mathbf{e}_m, \varphi_\lambda \rangle|. \tag{2.19}$$

3. **Update**

   Update the new residue vector as

$$\mathbf{e}_{m+1} = \mathbf{e}_m - \langle \mathbf{e}_m, \varphi_{\lambda_m} \rangle \varphi_{\lambda_m}. \tag{2.20}$$

4. **Terminate**

   If

$$\|\mathbf{x} - \mathbf{e}_{m+1}\| < \epsilon, \tag{2.21}$$

   or $m = M$ stop. Otherwise $m = m + 1$, and go to step 2.

This approximation is suboptimal due to the fact that it is not performed with all dictionary components but one at a time. In a straightforward implementation, best match step would require P inner product calculations of $N \times 1$ vectors and ordering of P outcomes. If the loop of the BMP algorithm is executed M times, the computational cost of the algorithm is at most $O(PMN)$ [21, 42, 105][2]. Different implementations of the algorithms such as variable dictionary vs. stored dictionary would require different resources in terms of memory and number of operations [21]. A fast implementation method for BMP algorithm is given in [2].

It should also be noted that the output of MP algorithms are not only the coefficient set $\{c_{\lambda_0}, c_{\lambda_1}, \ldots, c_{\lambda_{M-1}}\}$ but also the index set $\{\lambda_0, \lambda_1, \ldots \lambda_{M-1}\}$ which indicate the selected atoms from the dictionary. For demonstration purpose, simple examples are given below.

Examples

As a demonstrative example, consider the case $N = M = 2$, and $P = 3$. Let the dictionary be composed of $\varphi_0 = [1, 0]^T$, $\varphi_1 = [\frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}}]^T$, and $\varphi_2 = [0, 1]^T$, where $(\cdot)^T$ represents the

[2]The O notation determines an upper bound. For example, $f(n) = O(g(n))$ means that there exists a constant c such that $f(n)$ is always less than or equal to $cg(n)$ for large enough n. Details about $O(\cdot)$ notation can be found in [96]. In this context it is used as a measure of the number of floating point operations.

transpose of a vector.

For $\mathbf{x}$= $[4, -5]^T$, the MP iterations are:

- Iteration - 1:

  Initialization, $\mathbf{e}_0 = [4, -5]^T$,

  $|\langle \mathbf{e}_0, \varphi_0 \rangle| = 4$, $|\langle \mathbf{e}_0, \varphi_1 \rangle| = \frac{1}{\sqrt{2}}$, $|\langle \mathbf{e}_0, \varphi_2 \rangle| = 5$.

  Using (2.19), set $\lambda_0 = 2$.

  The residue vector is $\mathbf{e}_1 = \mathbf{e}_0 - \langle \mathbf{e}_0, \varphi_2 \rangle \varphi_2 = [4, -5]^T - (-5)[0, 1]^T = [4, 0]^T$.

  The graphical representation of the first iteration is shown in Fig 2.1.

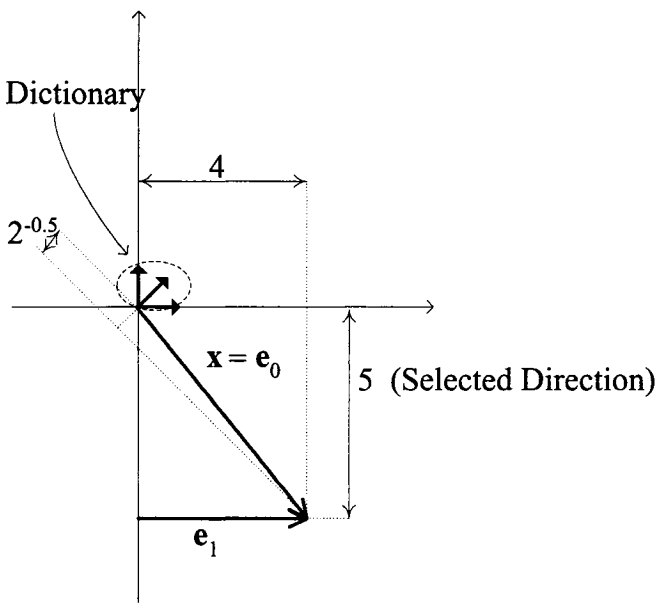


Figure 2.1: First iteration of MP algorithm for $\mathbf{x} = [4, -5]^T$

- Iteration - 2:

  $|\langle \mathbf{e}_1, \varphi_0 \rangle| = 4$, $|\langle \mathbf{e}_1, \varphi_1 \rangle| = 4/\sqrt{2}$, $|\langle \mathbf{e}_1, \varphi_2 \rangle| = 0$.

  Using (2.19), set $\lambda_1 = 0$.

Residue vector is $\mathbf{e}_2 = \mathbf{e}_1 - \langle \mathbf{e}_1, \varphi_0 \rangle \varphi_0 = [4,0]^T - (4)[1,0]^T = [0,0]^T$.

Since $\mathbf{e}_2 = \mathbf{0}$, perfect representation is obtained. This is also the sparsest representation for the given dictionary.

The signal decomposition obtained by using MP is

$$\begin{bmatrix} 4 \\ -5 \end{bmatrix} = -5 \begin{bmatrix} 0 \\ 1 \end{bmatrix} + 4 \begin{bmatrix} 1 \\ 0 \end{bmatrix}. \tag{2.22}$$

The BMP algorithm does not always give the sparsest representation. An example of such a case appears for $\mathbf{x} = [4,5]^T$. The iterations are performed as follows.

- Iteration - 1:

  Initialization, $\mathbf{e}_0 = [4,5]^T$,

  $|\langle \mathbf{e}_0, \varphi_0 \rangle| = 4$, $|\langle \mathbf{e}_0, \varphi_1 \rangle| = \frac{9}{\sqrt{2}} \approx 6.36$, $|\langle \mathbf{e}_0, \varphi_2 \rangle| = 5$.

  Using (2.19), set $\lambda_0 = 1$.

  The residue vector is $\mathbf{e}_1 = [4,-5]^T - \frac{9}{\sqrt{2}}[\frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}}]^T = [-\frac{1}{2}, \frac{1}{2}]^T$.

  Instead of selecting φ_2, the algorithm selects φ_1 due to high values of correlation between the dictionary elements, causing the algorithm to a selection that leads to a non-sparse solution. The rest of the iterations try to cancel out this erroneous selection. This problem is referred to as error propagation throughout the dissertation.

- Iteration - 2:

  $|\langle \mathbf{e}_1, \varphi_0 \rangle| = \frac{1}{2}$, $|\langle \mathbf{e}_1, \varphi_1 \rangle| = 0$, $|\langle \mathbf{e}_1, \varphi_2 \rangle| = \frac{1}{2}$.

  Since both φ_0 and φ_3 result in the same inner product, randomly set $\lambda_1 = 0$[3].

  Residue vector is $\mathbf{e}_1 = [-\frac{1}{2}, \frac{1}{2}]^T - (-\frac{1}{2})[1,0]^T = [0, \frac{1}{2}]^T$.

- Iteration - 3:

  $|\langle \mathbf{e}_2, \varphi_0 \rangle = 0$, $|\langle \mathbf{e}_2, \varphi_1 \rangle| = 0$, $|\langle \mathbf{e}_2, \varphi_2 \rangle| = \frac{1}{2}$.

[3]This is a symmetrical choice. Selecting $\lambda_1 = 2$, does not change the number of required iterations.

Using (2.19), set $\lambda_2 = 2$.

Residue vector is $\mathbf{e}_2 = [0, 0]^T$. However, this is not the sparsest representation, since 3 components are used to span a 2-D space.

The signal decomposition obtained by using BMP is

$$\begin{bmatrix} 4 \\ 5 \end{bmatrix} = \frac{9}{\sqrt{2}} \begin{bmatrix} \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} \end{bmatrix} - \frac{1}{2} \begin{bmatrix} 1 \\ 0 \end{bmatrix} + \frac{1}{2} \begin{bmatrix} 0 \\ 1 \end{bmatrix}. \tag{2.23}$$

2.2.2 Orthogonal Matching Pursuit Algorithm

The orthogonal matching pursuit (OMP) algorithm was proposed in [82] and in [29] independently. OMP is also called the modified matching pursuit algorithm in the literature [2]. Similar to BMP, the aim of OMP is to obtain an approximation to the input signal $\mathbf{x}$, by sequentially selecting vectors from the dictionary. However, the OMP algorithm gives a better approximation performance by orthogonalizing the directions of the projection. This guarantees the convergence of OMP with a finite number of iterations, allowing a maximum of N iterations. In BMP the convergence was guaranteed with infinite number of iterations. With OMP, the computational cost of the algorithm is increased due to the employed Gram-Schmidt orthogonalization procedure.

The indices of the m selected atoms are stored in the index vector $\Lambda_m = [\lambda_0, \lambda_1, \ldots, \lambda_{m-1}]$ with $\Lambda_0 = [\,]$. The OMP algorithm selects the next atom φ_{λ_m} by finding the vector best aligned with the residual obtained by projecting $\mathbf{e}_m$ onto the dictionary components, that is

$$\lambda_m = \arg\max_{\lambda_l \in \Gamma} |\langle \varphi_{\lambda_l}, \mathbf{e}_m \rangle|, \quad l \notin \Lambda_m. \tag{2.24}$$

Then the selected vector component φ_{λ_m} is orthogonalized by the Gram-Schmidt algorithm as

$$u_m = \varphi_{\lambda_m} - \sum_{l=0}^{m-1} \frac{\langle \varphi_{\lambda_m}, u_l \rangle}{\|u_l\|^2} u_l, \tag{2.25}$$

with $u_0 = \varphi_{\lambda_0}$.The residue vector $\mathbf{e}_m$ is updated as

$$\mathbf{e}_{m+1} = \mathbf{e}_m - \frac{\langle \mathbf{e}_m, u_m \rangle}{\|u_m\|^2} u_m. \tag{2.26}$$

The coefficient set $\{c_{\lambda_0}, c_{\lambda_1}, \ldots, c_{\lambda_{m-1}}\}$ change with each iteration and can be evaluated by taking the orthogonal projection of $\mathbf{x}$ onto the selected atoms. Similar to the BMP algorithm, OMP terminates when either $m = M$, or $\|\mathbf{e}_m\| \leq \epsilon$. The computational cost of the algorithm is at most $O(NM(M+P))$ [105]. Similar to BMP, different implementations of the OMP would require different resources in terms of memory and number of operations, which are detailed in [21]. A fast implementation method for OMP algorithm is given in [2].

Examples

In this section, we demonstrate the OMP algorithm, using the simple examples presented previously. The dictionary components are $\varphi_0 = [1, 0]^T$, $\varphi_1 = [\frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}}]^T$, and $\varphi_2 = [0, 1]^T$.

For $\mathbf{x} = [4, -5]^T$, the BMP and OMP iterations are identical. The signal decomposition obtained by using OMP is the same decomposition that is obtained using MP in (2.22).

Consider the second example; $\mathbf{x} = [4, 5]^T$. The OMP iterations are:

- Iteration - 1:

  Initialization, $\mathbf{e}_0 = [4, 5]^T$,

  $|\langle \mathbf{e}_0, \varphi_0 \rangle| = 4$, $|\langle \mathbf{e}_0, \varphi_1 \rangle| = \frac{9}{\sqrt{2}} \approx 6.36$, $|\langle \mathbf{e}_0, \varphi_2 \rangle| = 5$.

  Using (2.24), set $\lambda_0 = 1$.

  Since $u_0 = \varphi_{\lambda_0} = \varphi_1$, the residue vector is $\mathbf{e}_1 = \mathbf{e}_0 - \langle \mathbf{e}_0, \varphi_1 \rangle \varphi_1 = [4, -5]^T - \frac{9}{\sqrt{2}}[\frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}}]^T = [-\frac{1}{2}, \frac{1}{2}]^T$.

- Iteration - 2:

  $|\langle \mathbf{e}_1, \varphi_0 \rangle| = \frac{1}{2}$, $|\langle \mathbf{e}_1, \varphi_2 \rangle| = \frac{1}{2}$.

Since both φ_0, and φ_2, result in the same inner product, randomly set $\lambda_1 = 0$[4]. Following (2.25), and (2.26) explicitly, we get, $u_1 = \varphi_0 - \langle \varphi_0, u_0 \rangle u_0 = [-\frac{1}{2}, \frac{1}{2}]^T$ The residue vector at the end of second iteration is $\mathbf{e}_2 = \mathbf{e}_1 - \frac{\langle \mathbf{e}_1, u_1 \rangle}{u_1} = [0, 0]^T$. This gives us the sparsest solution.

The signal decomposition obtained by using OMP is

$$\begin{bmatrix} 4 \\ 5 \end{bmatrix} = \frac{9}{\sqrt{2}} \begin{bmatrix} \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} \end{bmatrix} + \begin{bmatrix} -\frac{1}{2} \\ \frac{1}{2} \end{bmatrix}. \tag{2.27}$$

As can be seen from (2.27), the sparsest decomposition can be obtained using OMP. For the same example, the MP algorithm required 3 terms as was shown in (2.23).

2.2.3 Weak Orthogonal Matching Pursuit Algorithm

Weak OMP (WOMP) algorithm is very similar to the OMP algorithm with a slight suboptimality in the selection of the optimal atom at each iteration [73]. The index λ_m is picked as

$$|\langle \mathbf{e}_m, \varphi_{\lambda_m} \rangle| \geq \alpha \sup_{\lambda \in \Gamma} |\langle \mathbf{e}_m, \varphi_\lambda \rangle|, \tag{2.28}$$

where, $\alpha \in (0, 1]$ is defined as the weakness parameter. Once the new index is selected, the calculations continue as the OMP algorithm. The aim in the introduction of WOMP is to decrease the computational complexity of the OMP algorithm slightly. In [104] some conditions for WOMP are presented showing that it can also converge to the exact solution for sparse approximation problems, given that the input signal is also sparse.

In a real Hilbert space $\mathbb{H}$, the angle between two nonzero vectors $r_1, r_2 \in \mathbb{H}$ is defined to

[4]Similar to the BMP case, selecting $\lambda_1 = 2$, does not change the number of required iterations.

be the number $\theta \in [0, \pi]$ that satisfies

$$\cos\theta = \frac{\langle r_1, r_2 \rangle}{\|r_1\|\|r_2\|}. \tag{2.29}$$

This angle, θ is 0^o (or 180^o) only when the vectors are aligned, resulting in a correlation value of $\cos(0^o) = 1$. When the vectors are orthogonal, by definition $\cos\theta = 90$. In the following chapter, this angle concept is used to add pruning method to the search trees that are composed for OMP algorithms.

More details about geometric interpretation of vectors in real vector spaces can be found in [74]. WOMP is summarized in the text in order to mention the similarity in the tree pruning strategy that is explained in Chapter 3. A similar technique is used to reduce the computational complexity of tree-search based matching pursuit structure. The geometrical interpretation is given here to clarify the notation.

2.3 Dictionary Analysis and Convergence Proofs for Orthogonal Matching Pursuit Algorithm

Recently some convergence properties for specific dictionary types for OMP algorithm are proved [104, 105]. In this section, these results are summarized. The properties associated with a dictionary $\mathcal{D}$ are listed in Section 2.3.1, and the convergence properties are given in Section 2.3.2.

2.3.1 Dictionary Properties

The optimality of the OMP algorithm for sparse signals is proved for dictionaries with special properties. Hence, in order to detect the optimality of OMP, first the corresponding dictionary

must be analyzed. This analysis can be done by using two parameters; coherence, and cumulative coherence.

Coherence

In order to quantify how much atoms look alike, the coherence parameter μ is defined as [73]

$$\mu = \max_{i \neq j} |\langle \varphi_{\lambda_i}, \varphi_{\lambda_j} \rangle|. \tag{2.30}$$

Informally, a dictionary is said to be incoherent when this parameter has a small value.

First results about the coherence parameter are given by Donoho and Huo in [32] where they also explore the behavior of μ for some specific and randomly selected dictionaries.

Cumulative Coherence

The coherence parameter extracts only the most extreme correlation case between two distinct atoms. Hence for a dictionary where there are few high correlation values between dictionary atoms, it does not reflect a realistic view of the dictionary. In order to obtain slightly more information from the dictionary the cumulative coherence is defined as [104]

$$\mu_1(m) = \max_{|\Lambda|=m} \left\{ \max_{\psi} \sum_{\lambda \in \Lambda} |\langle \psi, \varphi_\lambda \rangle| \right\}. \tag{2.31}$$

where m is a positive integer, and the atom ψ is an atom with an index from the set $\Gamma \setminus \Lambda$ For convention $\mu_1(0) = 0$. It can be shown from (2.31) that $\mu_1(1) = \mu$ and $\mu_1(m) \leq m\mu$.

Informally, a dictionary is said to be quasi-incoherent when its cumulative coherence function grows slowly. A similar generalized concept for coherence is defined in [32], however is not developed sufficiently for the OMP convergence proofs.

2.3.2 Convergence of Orthogonal Matching Pursuit Algorithm

The convergence of the OMP and the Basis Pursuit (BP) algorithms for the exact sparse problem is introduced in [104,105]. Here we elaborate on the results about the OMP algorithm since it is shown to be much simpler than the BP algorithm with the same performance results.

The exact sparse problem is considered in these derivations since it is shown that solving or even approximating the sparse approximation problem given in (1.1) is NP-hard for unrestricted dictionaries [40, 76]. However some results showing that OMP is an efficient algorithm for the approximation problem are known [104].

Here, first the problem definition for the exact sparse problem is given. Then the conditions for OMP to recover the exact signal are summarized. Then, the results on good approximations of OMP are listed.

Assume that an input signal $\mathbf{x}$ can be written as a linear combination of at least M atoms from the index set $\Lambda_o = \{\lambda_0, \lambda_1, \ldots \lambda_{M-1}\}$, where $\Lambda_o \in \Gamma$. That is,

$$\mathbf{x} = \sum_{\lambda \in \Lambda_o} c_\lambda \varphi_\lambda. \tag{2.32}$$

It can be assumed that the atoms in Λ_o are linearly independent and the coefficient set $\{c_\lambda\}$ is composed of nonzero components, without loss of generality.

Define $\mathbf{A}$ as the $N \times M$ matrix whose columns are the atoms with indices in Λ_o as

$$\mathbf{A} = \begin{bmatrix} \varphi_{\lambda_0} & \varphi_{\lambda_1} & \cdots \varphi_{\lambda_{M-1}} \end{bmatrix}, \tag{2.33}$$

and the coefficient vector $\mathbf{c}$ with M complex coefficients as

$$\mathbf{c} = \begin{bmatrix} c_{\lambda_0} \\ c_{\lambda_1} \\ \vdots \\ c_{\lambda_{M-1}} \end{bmatrix}. \tag{2.34}$$

Then the input signal can be written as

$$\mathbf{x} = \mathbf{A}\mathbf{c}. \tag{2.35}$$

Let the MoorePenrose generalized inverse of $\mathbf{A}$ be denoted by $\mathbf{A}^{\dagger}$. The matrix $\mathbf{A}$ has full row rank due to the fact that all optimal atoms are linearly independent, and hence,

$$\mathbf{A}^{\dagger} = \left(\mathbf{A}^{H}\mathbf{A}\right)^{-1}\mathbf{A}^{H}, \tag{2.36}$$

where $(\cdot)^{H}$ represents the Hermitian transpose of a matrix. The condition for exact recovery of $\mathbf{x}$ is given by the following theorem.

Theorem 1 *[104] A sufficient condition for the OMP algorithm to recover the sparsest representation of the input signal is given by*

$$\max_{\psi} \|\mathbf{A}^{\dagger}\psi\|_1 < 1, \tag{2.37}$$

where, ψ is an atom indexed by the set $\Gamma \setminus \Lambda_o$.

In the representation $\|\cdot\|_1$ represents the l_1 norm[5]. The condition given in (2.37) is known as the exact recovery condition.

A result about the WOMP algorithm about a recovery condition can be deduced from the theorem above as follows.

Corollary 1 *[104] A sufficient condition for the WOMP algorithm with the weakness parameter α to recover the sparsest representation of the input signal is given by*

$$\max_{\psi} \|\mathbf{A}^{\dagger}\psi\|_1 < \alpha, \tag{2.38}$$

where, ψ is an atom indexed by the set $\Gamma \setminus \Lambda_o$.

[5]The l_1 norm of a vector $v = [v_0, v_1, \ldots, v_{m-1}]$ is defined as $\|v\|_1 = \sum_{i=0}^{m-1} |v_i|$.

The relation between dictionary coherence parameters and the exact recovery condition stated in (2.37) are shown as follows.

Theorem 2 *[104] The exact recovery condition holds for a dictionary with parameters μ and $\mu_1(M)$ for a signal representation with M terms when*

$$M < \frac{1}{2}\left(\frac{1}{\mu} + 1\right), \tag{2.39}$$

or when

$$\mu_1(M) + \mu_1(M-1) < 1. \tag{2.40}$$

The condition given in (2.40) is a more general restriction due to the properties of the coherence parameters [105].

The results above prove that OMP is a correct algorithm over quasi-incoherent dictionaries for the case where $\mathbf{x}$ is a sparse signal, that is $M < N$. However it is shown in [104] that a generic signal does not have a sparse structure.

In this dissertation we consider the problem where the input signal has a sparse nature due to its construction mechanism or physical properties but corrupted by noise. That is, representing the noise component as $\mathbf{n}$ and the M term sparse signal as $\mathbf{x}$, we have

$$\mathbf{r} = \mathbf{x} + \mathbf{n}, \tag{2.41}$$

where $\mathbf{r}$ is treated as the received signal, and consider the conditions given in (2.39) and (2.40) for cases where noise is negligible and hence $\mathbf{r} \approx \mathbf{x}$. Details about this approximation are explained associated with the specific application. These applications are introduced throughout the rest of the dissertation.

Chapter 3

Tree Based Matching Pursuit Algorithms: Integration of Combinatorial Search Methods to Matching Pursuit

For dictionaries that are not incoherent or quasi-incoherent it has been shown that the OMP is not the optimal algorithm [104,105]. We later show that wireless communication applications; channel estimation, direction of arrival detection and multi-user detection dictionaries depend heavily on the problem definition and hence cannot be designed as incoherent or quasi-incoherent. Therefore, OMP may not be the best solution for the problems we will be tackling. In this chapter we elaborate on the combinatorial search techniques that improve the performance of the MP algorithms. The integration method of tree-based search techniques are reviewed along with examples. The flexible tree-search based OMP (FTB-OMP) algorithm is introduced. Experimental results are used to show the efficiency of the FTB-OMP algorithm

for overcomplete dictionaries. A portion of the research work presented in this chapter is published in [54, 55].

3.1 Tree-search Based Matching Pursuit Algorithms

The main problem of MP algorithms is that, an error in the initial stages of the iterations cannot be recovered. These lead to suboptimum representations as shown via examples. In order to overcome such error propagation problems, integration of tree-based search techniques are introduced for MP algorithms [21, 24]. These algorithms, referred here as tree-search based matching pursuit (TB-MP) algorithms, greatly improve the approximation and detection performance of the MP algorithms at the expense of increased computational cost. Implementation of a complete search tree is impractical due to the enormous memory and computational complexity requirements. However, trees of reduced sizes can be implemented and the resulting algorithms give a substantial increase in the performance.

Let $\mathcal{D}=\{\varphi_\lambda\}_{\lambda\in\Gamma}$ be a dictionary for the signal space generated by vectors of size N with $|\Gamma| = P$. In TB-MP, the MP algorithm of choice is extended to a tree structure by keeping L vectors that are maximizing the selection criteria. The parameter L can be set to any value between 1 and P. Each of these vectors represents one alternative to be explored in each of the branches for the current partial solution. This algorithm follows the same basic iterations as MP, but explores L choices for the next vector selected at each iteration. At the end of M iterations, the search grows exponentially to a tree with L^M leaves as shown in Fig. 3.1. Each node in the search tree represents the corresponding residue vector at that particular iteration. Leaves denote the final iterations.

In TB-MP algorithm, the residue vector at m^{th} iteration and j^{th} tree branch is denoted by $\mathbf{e}_m^{(j)}$. The best matching vector indices, $\{\lambda_{m,j}^{(1)}, \lambda_{m,j}^{(2)}, \cdots, \lambda_{m,j}^{(L)}\}$ at the m^{th} iteration coming

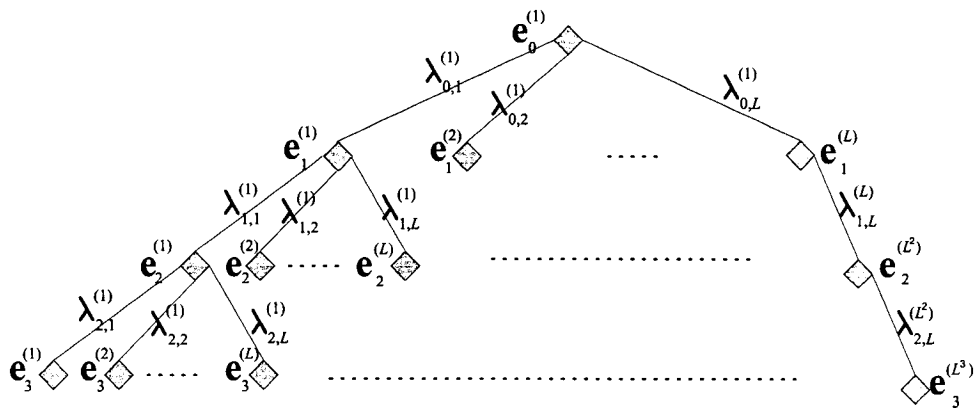


Figure 3.1: Search tree with branching factor L.

from the j^{th} tree branch are selected according to

$$\lambda_{m,j}^{(i)} = \arg\max_{\gamma} F(\varphi_\lambda, \mathbf{e}_m^{(j)}),$$
$$\gamma \in \Gamma \setminus \{\lambda_{m,j}^{(1)}, \lambda_{m,j}^{(2)}, \cdots, \lambda_{m,j}^{(i-1)}\}, \qquad i = 1, \cdots, L, \qquad j = 1, \cdots, L^m. \tag{3.1}$$

where, $F(\cdot)$ is the selection criteria of a particular MP algorithm. Similar to the MP algorithms, TB-MP can be terminated when either sufficient number of terms are picked ($m = M$) or when the norm of the residue vector falls below some predetermined threshold value, $\|\mathbf{e}_m^{(j)}\| \leq \epsilon$. The leaf corresponding to the smallest residual error vector yields the solution.

3.1.1 Example

Consider the previously given example in Chapter 2, with $\varphi_0 = [1, 0]^T$, $\varphi_1 = [\frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}}]^T$, $\varphi_2 = [0, 1]^T$, and $\mathbf{x} = [4, 5]^T$. As was shown in (2.23), the MP algorithm resulted in 3 components. In this section, we show that adding a tree search with $L = 2$ to the MP algorithm, the sparsest solution can be obtained. The iterations can be summarized as follows.

- Iteration - 1:

  Initialization, $\mathbf{e}_0^{(1)} = [4,5]^T$,

  $|\langle \mathbf{e}_0^{(1)}, \varphi_0 \rangle| = 4$, $|\langle \mathbf{e}_0^{(1)}, \varphi_1 \rangle| = \frac{9}{\sqrt{2}} \approx 6.36$, $|\langle \mathbf{e}_0^{(1)} \varphi_2 \rangle| = 5$.

  Using (2.19) as the selection criteria in (3.1), we get $\lambda_{0,1}^{(1)} = 1, \lambda_{0,2}^{(1)} = 2$.

  The residual error vectors are

  $\mathbf{e}_1^{(1)} = \mathbf{e}_0^{(1)} - \langle \mathbf{e}_0^{(1)}, \varphi_1 \rangle \varphi_1 = [4,5]^T - \frac{9}{\sqrt{2}}[\frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}}]^T = [-\frac{1}{2}, \frac{1}{2}]^T$

  for selecting $\lambda_{0,1}^{(1)} = 1$ and

  $\mathbf{e}_1^{(2)} = \mathbf{e}_0^{(1)} - \langle \mathbf{e}_0^{(1)}, \varphi_2 \rangle \varphi_2 = [4,-5]^T - (-5)[0,1]^T = [4,0]^T$,

  for selecting $\lambda_{0,2}^{(1)} = 2$.

- Iteration - 2:

  $|\langle \mathbf{e}_1^{(1)}, \varphi_0 \rangle| = \frac{1}{2}$, $|\langle \mathbf{e}_1^{(1)}, \varphi_1 \rangle| = 0$, $|\langle \mathbf{e}_1^{(1)}, \varphi_2 \rangle| = \frac{1}{2}$.

  The selected indices are $\lambda_{1,1}^{(1)} = 0$, and $\lambda_{1,2}^{(1)} = 2$. The corresponding residue vectors can be shown to be $\mathbf{e}_2^{(1)} = [0, \frac{1}{2}]^T$, and $\mathbf{e}_2^{(2)} = [\frac{1}{2}, 0]^T$. Coming from the second branch of the first iteration, we have $|\langle \mathbf{e}_1^{(2)}, \varphi_0 \rangle| = 5$, $|\langle \mathbf{e}_1^{(2)}, \varphi_1 \rangle| = 4/\sqrt{2}$, $|\langle \mathbf{e}_1^{(2)}, \varphi_2 \rangle| = 0$.

  The selected indices are $\lambda_{1,1}^{(2)} = 0$, and $\lambda_{1,2}^{(2)} = 1$. The corresponding residual error vectors can be shown to be $\mathbf{e}_1^{(3)} = [0,0]^T$, and $\mathbf{e}_1^{(4)} = [2,-2]^T$. Since the zero vector is obtained as residual error vector, the perfect representation is obtained with the index set $\{2,0\}$.

The signal decomposition obtained by using tree based MP, with $L = 2$ is

$$\begin{bmatrix} 4 \\ 5 \end{bmatrix} = 5 \begin{bmatrix} 0 \\ 1 \end{bmatrix} + 4 \begin{bmatrix} 1 \\ 0 \end{bmatrix}. \tag{3.2}$$

This is the same decomposition obtained using OMP algorithm, as was shown in (2.27). The search tree corresponding to the iterations is shown in Fig. 3.2. The sparsest decomposition is represented by red lines. As can be seen from the example, addition of the search tree improves the performance of the algorithm by recovering the error at the first iteration.

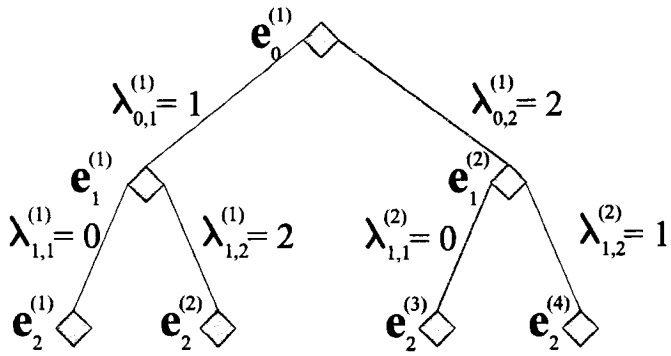


Figure 3.2: $L = 2$-branch search tree of BMP algorithm for $\mathbf{x} = [4, 5]^T$. The path leading to the sparsest solution is shown by red lines.

3.1.2 Implementation of Combinatorial Search

The TB-MP algorithms can be implemented by addition of backtracking to the MP algorithms. The pseudo-code for TB-MP algorithms is given in Table 3.1 that is implemented with backtracking structure presented in [65].

Table 3.1: Pseudo-code for TB-MP algorithms

$TB\text{-}MP(m, M, L, \epsilon, \mathcal{D})$

Global $\Lambda = [\lambda_0, \lambda_1, \ldots]$, Best_res, Best_$\lambda$

Calculate $\mathbf{e}_{m+1}$

If $\|\mathbf{e}_{m+1}\|$ <Best_res

Best_λ=$[\lambda_0, \ldots, \lambda_m]$

Best_res$\leftarrow$ $\|\mathbf{e}_{m+1}\|$

end

If $m > M$ or $\|\mathbf{e}_{m+1}\| < \epsilon$, then **return**

Calculate $\{\lambda_m^{(1)}, \lambda_m^{(2)}, \ldots, \lambda_m^{(L)}\}$

For each $i = 1$ to L **do**

$\lambda_m = \lambda_m^{(i)}$

$TB\text{-}MP(m+1, M, L, \epsilon, \mathcal{D})$

end

end

3.2 Flexible Tree-Search Based Orthogonal Matching Pursuit

Application of tree-based searches are introduced in [24] and [97]. Among the algorithms considered by Cotter and Rao [24], the one that gives the best compromise between performance and complexity is OMP. Moreover, as summarized in the previous chapter, Tropp showed that for some special dictionary classes the OMP algorithm can give the sparsest representation [104]. For these reasons, we focus here on OMP. Even though our proposed algorithm is based on OMP, the tree-search structure introduced here can be applied to all types of matching pursuit algorithms.

The TB-MP algorithm dramatically improves the approximation performance of OMP, since it explores several choices at each iteration. However, the algorithm's computational complexity increases exponentially, making the algorithm impractical for applications where speed is a concern. A similar search tree structure is proposed for the order recursive matching pursuit algorithm [97], however, the authors do not specify the details of the partial search technique.

In this section we introduce an efficient tree-search based OMP algorithm with branch pruning: the flexible tree-search based orthogonal matching pursuit (FTB-OMP) algorithm The flexibility of the algorithm is achieved by three major design parameters which influence the approximation performance and computational complexity. The algorithm has recursive structure.

Let us first introduce the decaying parameter $d \geq 1$, which affects the tree structure by reducing the branching factor at each iteration. A maximum of L branches are searched at each partial solution. In the initial iteration, the branching factor is set to L; at the i^{th} iteration the branching factor is set to $\lceil L/d^i \rceil$, where $\lceil \cdot \rceil$ represents the ceiling function. The

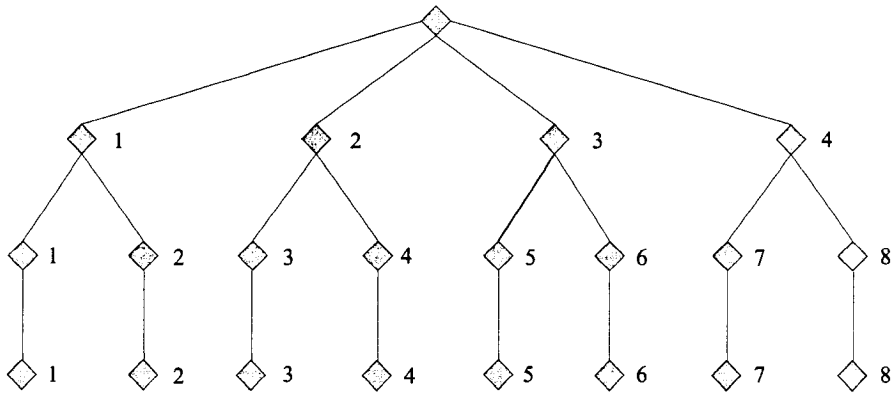


Figure 3.3: Tree structure with exponential decay for $L = 4$, $d = 2$ search tree for $r = 3$.

idea in this algorithm is to start the search with a large number of branches at the initial iteration where an erroneous selection is more likely to appear, and to reduce the branching factor as the number of iterations increases. A search tree for $L = 4$, $d = 2$ is shown in Fig. 3.3. For the special case $d = 1$, the algorithm keeps L as the branching factor.

Further reduction on the tree size is achieved by the correlation threshold $0 \leq \xi \leq 1$. Our objective is to prune the tree branches that are heuristically believed to be unnecessary. Our heuristic is to keep only the branches among $\lambda_m^{(1)}, \lambda_m^{(2)}, \ldots, \lambda_m^{(L)}$ which are closely "aligned" with OMP's first choice branch $\lambda_m^{(1)}$. We measure this alignment by the correlation between vectors which is defined as

$$\rho_j = \frac{\langle \mathbf{e}_m, \varphi_j \rangle}{\|\mathbf{e}_m\| \, \|\varphi_j\|}, \tag{3.3}$$

where, $\mathbf{e}_m$ is the residue signal to be approximated. A branch is assumed to be unnecessary when the candidate vector is not aligned with $\lambda_m^{(1)}$, that is when $|\rho_{\lambda_m^{(1)}, \lambda_m^{(i)}}| < \xi$. A geometrical description for real dictionary components are shown in Fig. 3.4 as an example. Searching the atoms that are only ξ close to the residue signal is inspired from (2.28) of the WOMP algorithm summarized in the previous chapter.

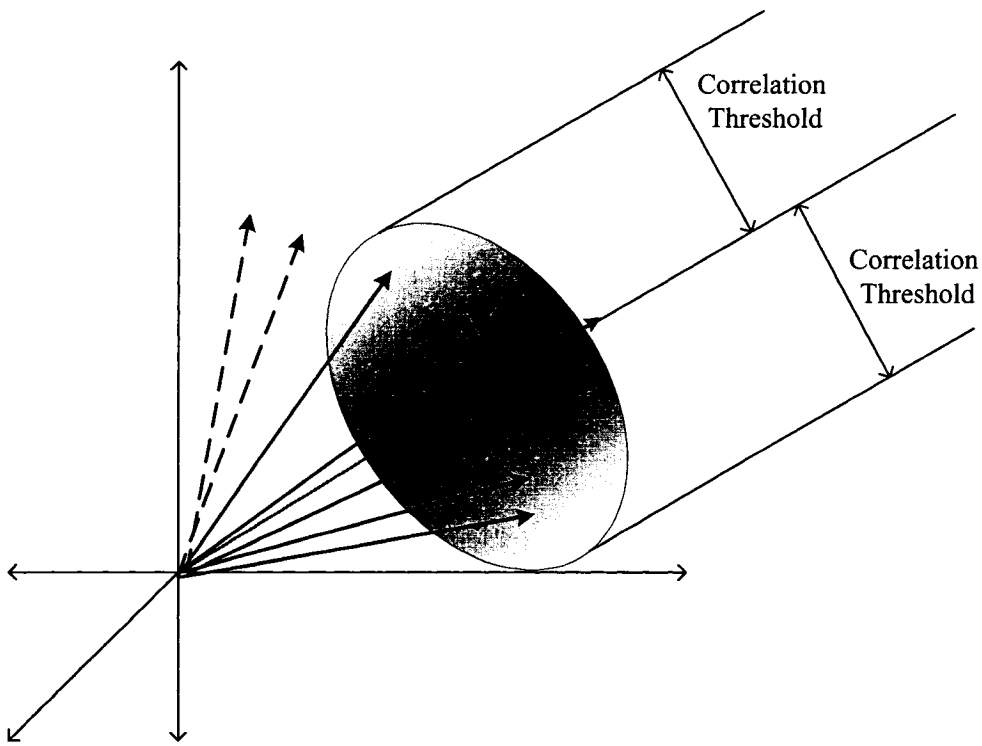


Figure 3.4: Geometrical description for thresholding of real atoms with respect to the residue signal **e**. Atoms depicted with solid lines are within the threshold limits hence they are selected in the tree search. Atoms depicted with dashed lines are believed to be distant to the input vector and are not considered in the search.

Table 3.2: Pseudo-code for FTB-OMP

FTB-OMP$(d, m, M, L, \xi, \epsilon, \mathcal{D})$

Global $\Lambda = [\lambda_0, \lambda_1, \ldots]$, Best_res, Best_$\lambda$

Calculate $\mathbf{e}_{m+1}$ for OMP

If $\|\mathbf{e}_{m+1}\|$ <Best_res

Best_λ=$[\lambda_0, \ldots, \lambda_m]$

Best_res$\leftarrow$ $\|\mathbf{e}_{m+1}\|$

end

If $m > M$ or $\|\mathbf{e}_{m+1}\| < \epsilon$, then **return**

Calculate$\{\lambda_m^{(1)}, \lambda_m^{(2)}, \ldots, \lambda_m^{(L)}\}$ as in (3.1)

For each $i = 1$ to L **do**

If i=1 **or** $|\rho_{\lambda_m^{(i)}}| \geq \xi$

$\lambda_m = \lambda_m^{(i)}$

FTB-OMP$(d, m+1, M, \lceil L/d \rceil, \xi, \epsilon, \mathcal{D})$

end

end

A pseudocode for FTB-OMP is given in Table 3.2. Note that FTB-OMP is a generalization of both OMP and TB-OMP algorithms. By choosing $\xi = 1$, we require full alignment so that only $\lambda_p^{(1)}$ is kept, reproducing OMP. By choosing $\xi = 0$, and $d = 1$, we place no restriction on alignment, reproducing TB-OMP. A value $0 < \xi < 1$ represents a compromise between the number of tree nodes for OMP ($M + 1$ nodes), and for TB-OMP ($\frac{L^{M+1}-1}{L-1}$ nodes).

It should be noted that the exponentially decaying tree-search strategy is proposed in this research work for the first time. This general tree-search strategy could effectively be employed in other problems where the choice on the initial branches are crucial.

The total number of leaves in FTB-OMP is upper bounded by

$$L\left\lceil\frac{L}{d}\right\rceil\left\lceil\frac{L}{d^2}\right\rceil\left\lceil\frac{L}{d^3}\right\rceil\cdots\left\lceil\frac{L}{d^{M-1}}\right\rceil=\prod_{i=0}^{M-1}\left\lceil\frac{L}{d^i}\right\rceil, \tag{3.4}$$

and the number of nodes is upper bounded by

$$1+L+L\left\lceil\frac{L}{d}\right\rceil+\cdots+L\left\lceil\frac{L}{d}\right\rceil\left\lceil\frac{L}{d^2}\right\rceil\left\lceil\frac{L}{d^3}\right\rceil\cdots\left\lceil\frac{L}{d^{M-1}}\right\rceil=\sum_{k=0}^{M}\prod_{i=0}^{k-1}\left\lceil\frac{L}{d^i}\right\rceil. \tag{3.5}$$

This bound for the number of leaves is also a rough complexity bound for FTB-OMP in terms of OMP with the same dictionary and the number of iterations. Threshold based tree pruning is added to this algorithm in order to decrease complexity.

In the algorithms proposed, a low ξ must be selected for a detailed search with less time constraint. For a faster search, a higher ξ value is more suitable, but this gives the optimum expansion with lower probability. In the extreme cases when $\xi = 0$, the algorithm converges to a complete tree-search, and when $\xi = 1$, it becomes the OMP algorithm.

3.3 Comparison of Approximation and Detection Performance of Algorithms

In order to compare the approximation performance and the search tree sizes of the proposed algorithms, various parameter combinations are considered. We run a large selection of tests varying L and ξ values. The experiments are run using MATLAB on Windows XP environment of 2.4 GHZ CPU, and 1024 MB of RAM. We use the data set given in [24] and in [15] where TB-OMP is introduced. In order to compare the algorithm complexities, we considered the number of search nodes and computer running times averaged over 1000 runs. In the results, it can be observed that both quantities are proportional. We should also note that the computer running times are indicated for comparison of relative performance ranking and should not be taken as absolute values. Under a different environment a faster implementation is possible by employing faster machines and lower level programming languages. However, it is expected that the relative ranking of the algorithm execution times is maintained.

3.3.1 Component Detection Experiment

In this experiment, we consider a component detection problem as designed in [24]. A dictionary $\mathbf{D}$, is created as a 20×30 matrix whose components are independent Gaussian random variables with mean 0 and variance 1, so that the correlation among the dictionary elements is low. A sparse solution $\mathbf{x}_M$ is created by adding 7 randomly selected dictionary components (i.e. columns) with random amplitudes that are uniformly distributed in $[0, 1]$. The sparse solution $\mathbf{x}_M$ is then normalized in order to provide a fair comparison for the approximation performances of the algorithms. Gaussian noise is then added to $\mathbf{x}_M$ in order to construct the input signal $\mathbf{x}$. The input signal can be decomposed as

$$\mathbf{x} = \mathbf{D}\mathbf{x}_M + \mathbf{n}, \tag{3.6}$$

where $\mathbf{n}$ is the additive white Gaussian noise (AWGN). Each component of $\mathbf{n}$ is 0 mean and variance σ^2, and signal to noise ratio (SNR) is adjusted as

$$\sigma^2 = \frac{1}{N} 10^{-SNR/10}. \tag{3.7}$$

In the simulations, σ^2 is adjusted so that SNR is 40 dB. The parameter ϵ is selected as $E\{\|n\|\} = 10^{-2}$ assuming that noise level is known at the receiver. The algorithms are run for 1000 distinct input vectors, and the detected number of components is counted. The average running times of the algorithms and average number of searched nodes are evaluated.

The experiment results are given in Tables 3.3 to 3.5. In these tables, NCD-i represent the average percentage that exactly i components are identified correctly. The average number of searched nodes and the average number of leaves in the search tree are denoted by N_{avg} and lv_{avg}, respectively. The maximum number of nodes and leaves that can be searched are represented by N_{max} and lv_{max}, respectively. They can be evaluated using (3.5) and (3.4). These upper bounds can be reached when the stopping criterion ϵ is very small for example 10^{-10}. For a relatively large value of ϵ, algorithms may stop once they have a small enough error vector without reaching the maximum number of allowed iterations M. The average computer running time, t_{avg}, is given in seconds. In the results, it can be observed that N_{avg} and t_{avg} are proportional. The average residual approximation error after the 7^{th} iteration is shown as $\|e_7\|_{avg}$.

In Table 3.3, a wide range of L and d values are investigated with no correlation based pruning ($\xi = 0$). We can observe that the average number of leaves is upper bounded by lv_{max}, and the number of nodes is upper bounded by N_{max} .

From the table, we can see that the OMP algorithm has poor detection capability, since only 60.9% of the time, all components are correctly detected. However, OMP requires low computer running times since for 7 iterations, a maximum of 8 nodes are investigated [1].

[1] 7 iterations correspond to 8 error vectors, and hence demonstrated in 8 nodes.

Table 3.3: Component detection experiment with various L and d values for $\xi = 0$, $M = 7$, $N = 20$, $P = 30$, $\epsilon = 10^{-2}$ and SNR=40 dB. NCD-i represents the percentage of the number of components correctly detected at the i^{th} iteration. N_{avg} and N_{max} denote the average number of searched nodes and the maximum number of nodes that is determined by the input parameters M, L and d, respectively. lv_{avg} and lv_{max} represent the average number of searched leaves and the maximum number of leaves. Average computer running times are shown by t_{avg}. $\|\mathbf{e}_7\|_{avg}$ is the average residual approximation error after 7^{th} iteration.

Algorithm	TB-OMP	FTB-OMP							OMP
L	3	32	12	18	25	16	9	4	1
d	1	4	2	3	5	4	3	2	1
ξ	0	0	0	0	0	0	0	0	0
NCD-0	1.2	1.1	1.2	1.1	1.1	1.3	1.6	3.3	9.7
NCD-1	1.4	1.5	1.6	1.5	1.5	1.7	2.3	4	7.1
NCD-2	1.4	1.4	1.6	1.6	1.8	2	2.3	3.3	6.4
NCD-3	1.7	1.6	1.8	2.1	1.8	2.2	3	3.6	5.1
NCD-4	1.6	1.3	1.7	1.5	1.6	2	2.2	3	4.3
NCD-5	2.2	2.1	1.9	2.3	2.1	2.2	2.4	2.8	3.2
NCD-6	1.9	1.9	1.9	2	1.8	2	2.7	3.6	3.3
NCD-7	88.6	89.1	88.3	87.9	88.3	86.6	83.5	76.4	60.9
N_{avg}	390.478	286.439	217.542	149.5	101.589	64.763	38.499	19.449	7.928
N_{max}	3280	2849	2029	1207	776	401	172	53	8
lv_{avg}	256.591	51.115	45.732	26.404	16.133	10.12	5.853	2.796	1
lv_{max}	2187	512	432	216	125	64	27	8	1
t_{avg}	0.1151	0.0979	0.0732	0.0511	0.037	0.0227	0.0134	0.0068	0.0028
$\|e_7\|_{avg}$	0.0084	0.008	0.0085	0.0088	0.0085	0.0099	0.0118	0.0197	0.0445

The FTB-OMP algorithm performance with $L = 3$, $d = 1$, and $\xi = 0$ is also tabulated. We should note that with the selected set of parameters, this algorithm is equivalent to TB-OMP with $L = 3$. Due to the high computational complexity, $L > 3$ and $d = 1$ cases are not reported. The TB-OMP algorithm with $L = 3$ detects all components correctly at a rate of 88.6%, by searching 390 nodes. From the results we can see that although a good detection performance is achieved with TB-OMP, the tree size is very large. The average approximation error is 0.0084.

In Table 3.3, the executions of the FTB-OMP algorithm are ordered so that an increase in the detection and approximation performances can be observed from right to left while the average number of searched nodes increases. In the table, the algorithm with the best detection performance is $L = 32$, $d = 4$ case. In this case 286 nodes are searched on the average and a detection performance slightly better than TB-OMP is observed. Note that TB-OMP with $L = 3$, $d = 1$ searched through a larger tree structure, taking roughly longer average running time. Although the required time for $L = 32$, $d = 4$ case is lower than for TB-OMP, the average residual approximation error is 4.8% lower than that of TB-OMP. As it can be observed from the results, a better approximation performance can be obtained by keeping L large at the initial iterations.

This conclusion can also be justified by comparing the algorithms $L = 25$, $d = 5$, and $L = 12$, $d = 2$. From their performance evaluations, we can see that $L = 25$, $d = 5$ case can detect more components correctly than the other, while searching only 47% of the nodes of $L = 12$, $d = 2$ case.

From the results in Table 3.3, we can also conclude that detection performance increases as the tree size increases. A tradeoff can be achieved by changing the algorithm parameters of FTB-OMP.

In order to observe the effect of correlation threshold-based pruning we considered $L = 25$,

Table 3.4: Component detection experiment with various ξ values for $L = 25$, $d = 5$, $M = 7$, $N = 20$, $P = 30$, $\epsilon = 10^{-2}$ and SNR=40 dB.

Algorithm	FTB-OMP								
L	25	25	25	25	25	25	25	25	25
d	5	5	5	5	5	5	5	5	5
ξ	0	0.05	0.1	0.15	0.2	0.3	0.4	0.5	0.8
NCD-0	1.1	1.1	1.1	1.3	1.3	1.7	2.5	5.1	9.7
NCD-1	1.5	1.5	1.5	1.6	1.6	2.1	2.9	4.3	7.1
NCD-2	1.8	1.8	1.9	2	2.1	2.1	2.6	3.9	6.4
NCD-3	1.8	1.9	1.9	2.2	2.3	2.5	3.2	3.9	5.1
NCD-4	1.6	1.6	1.6	1.8	1.9	2	2.4	2.9	4.3
NCD-5	2.1	2.1	2.2	2.2	2.2	2.2	2.3	3.4	3.2
NCD-6	1.8	1.8	1.7	1.9	2	2.3	2.9	3.8	3.3
NCD-7	88.3	88.2	88.1	87	86.6	85.1	81.2	72.7	60.9
N_{avg}	101.589	100.621	91.425	79.345	66.359	46.697	27.475	14.723	7.942
lv_{avg}	16.133	15.977	14.494	12.547	10.453	7.279	4.149	2.071	1.002
t_{avg}	0.037	0.0349	0.0319	0.0286	0.0248	0.0181	0.0114	0.0074	0.005
$\|e_7\|_{avg}$	0.0085	0.0085	0.0088	0.0095	0.0099	0.0109	0.0145	0.0264	0.0445

$d = 5$ and $L = 3$, $d = 1$ as two distinct parameter sets. The results are given in Tables 3.4 and 3.5, respectively. In Table 3.4, component detection experiment results are given with various ξ values for $L = 25$, $d = 5$. From the results, we observe that computer running time and average number of nodes decrease as ξ increases, as expected. For the case where $\xi \geq 0.8$, the algorithm almost converges to OMP with single branching factor.

The threshold value 0.4 gives a good component detection performance with NCD-7= 88.5%. The algorithm searches through only 132 nodes by taking approximately 0.05 seconds on the average. This gives a reduction of 57% in running time with respect to the case where $\xi = 0$. We can observe that correlation-based pruning reduces the tree size and computer running times of the algorithm, and gives comparable detection and approximation performance. The effect of the proposed pruning technique becomes more apparent for smaller

Table 3.5: Component detection experiment with various ξ values for $L = 3$, $d = 1$, $M = 7$, $N = 20$, $P = 30$, $\epsilon = 10^{-2}$ and SNR=40 dB.

Algorithm	TB-OMP	FTB-OMP							
L	3	3	3	3	3	3	3	3	3
d	1	1	1	1	1	1	1	1	1
ξ	0	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8
NCD-0	1.2	1.2	1.2	1.2	1.2	3.1	7.4	9.6	9.7
NCD-1	1.4	1.4	1.4	1.4	1.3	2.9	5.8	6.7	7.1
NCD-2	1.4	1.4	1.4	1.4	1.5	2.4	4.9	6.3	6.4
NCD-3	1.7	1.7	1.7	1.7	1.8	2.3	4.7	5	5.1
NCD-4	1.6	1.6	1.6	1.6	1.6	2.2	3.7	4.1	4.3
NCD-5	2.2	2.2	2.2	2.2	2.2	2.7	3	3.2	3.2
NCD-6	1.9	1.9	1.9	1.9	1.9	2.2	3	3.2	3.3
NCD-7	88.6	88.6	88.6	88.6	88.5	82.2	67.5	61.9	60.9
N_{avg}	390.478	390.478	390.301	351.624	132.174	23.625	9.931	8.08	7.942
lv_{avg}	256.591	256.591	256.438	225.426	69.49	6.333	1.4	1.025	1.002
t_{avg}	0.1151	0.1153	0.1150	0.1072	0.0497	0.0126	0.0061	0.005	0.005
$\|e_7\|_{avg}$	0.0084	0.0084	0.0084	0.0084	0.0085	0.0182	0.0368	0.0434	0.0445

values of ϵ as demonstrated in the following section. The reason behind this is that pruning becomes the dominant limit for the stopping criterion since the minimum error limit cannot be achieved in a straightforward search, going through a larger number of nodes.

The effect of correlation based pruning in TB-OMP algorithm for $L = 3$ and $d = 1$ case is investigated in Table 3.5. Similarly to Table 3.4, the TB-OMP algorithm converges to OMP for $\xi \geq 0.8$. By changing the correlation threshold value between 0 and 0.8, a transition is observed between the TB-OMP and OMP algorithms. For $\xi \leq 0.2$ it can be observed that the same tree structures are searched through. The efficiency of tree based pruning is apparent for $\xi = 0.3$. The same detection and approximation performance is achieved while reducing the number of searched nodes by 10%. As ξ increases, the average approximation error increases, while the detection performance, the average number of searched node and

the average computer running times decrease.

Comparing Tables 3.4 and 3.5, we can observe clear advantages of FTB-OMP with $L = 25$ and $d = 5$ over $L = 3$ and $d = 1$. For FTB-OMP with $L = 25$ $d = 5$ and $\xi = 0.4$, the detection values are NCD-7= 81.2%, NCD-6= 2.9% for high number of correctly detected components by searching 27 nodes, with running time of approximately 0.01 seconds. For the FTB-OMP with $L = 3$, $d = 1$ and $\xi = 0.4$, a slightly lower detection performance is observed with NCD-7= 82.2%, NCD-6= 2.2% by searching 23.6 nodes, with running time of approximately 0.012 seconds. Furthermore, the second set of parameters gives a residual error of 0.0182, while the first set of parameters gives 0.0145, which is 20% smaller. These observations show that, by keeping the branching factor large at initial iterations, a more efficient search tree can be obtained in terms of the approximation performance.

The conclusion is not true in terms of detection performance since all components are equally important regardless of the amplitude. Keeping the branching factor high in initial iterations is efficient since the selected coefficients in these iterations have larger amplitudes. As the amplitude values decrease with the increasing number of iterations, their effect on the approximation error decrease. In terms of detection performance all selected atoms have the same importance hence keeping more branches at initial iterations may not be the optimum strategy.

From the experiments, we can conclude that keeping the branching factor high at initial iterations, and decreasing it exponentially gives a good approximation performance with time complexity lower than TB-OMP algorithm for $\xi = 0$. Further pruning via correlation threshold increases the efficiency of the search even more, with a modest decrease of the algorithm's performance. However it is not trivial to optimize all parameters for a given dictionary and signal structure.

3.3.2 Experiments with Various Signals

In this section, we consider the signals that are given in [15]. The approximation performance and relative complexity of FTB-OMP algorithm is investigated with various choices for L, d, and ξ values.

Carbon Signal

In this experiment, the selected dictionary is of size 128×1024, and generated by Symmlets, a class of wavelets. The input signal is called "Carbon" [15], and consists of a linear combination of 6 elements from the dictionary. The signal is shown in Fig. 3.5 and its norm is 2.7316.

The algorithm outputs are represented by the time-frequency (TF) phase plots adapted from Coifman and Wickerhauser [19], and used in [15]. In the TF plane, each rectangle is associated with the corresponding basis vector. When a signal is a superposition of several basis vectors, like the Carbon signal, each basis that appear in the superposition is indicated by shading the corresponding rectangles in the TF plane. The darker regions represents coefficients that are larger in magnitude.

The TF planes of FTB-OMP algorithm with a selected range of parameters for $M = 10$ are given in Fig. 3.6 and 3.9. The number of searched tree nodes (N), number of leaves in the search tree (lv), computer running time in seconds (t), and residual approximation error after the 10^{th} iteration ($\|\mathbf{e}_{10}\|$) are also indicated in the figures. The stopping criterion ϵ is set as 10^{-10} in order to get a very close approximation to the input signal.

In Fig. 3.6 (a), the TF plane of the FTB-OMP algorithm with $L = 3$, $d = 1$, and $\xi = 0$ is shown. The resulting TF plane is the sparsest possible representation of the Carbon signal. In the TF frame, the horizontal line along time axis with constant frequency 0.5 corresponds to a sine wave, and the vertical line along frequency axis with constant time at 0.5 denotes

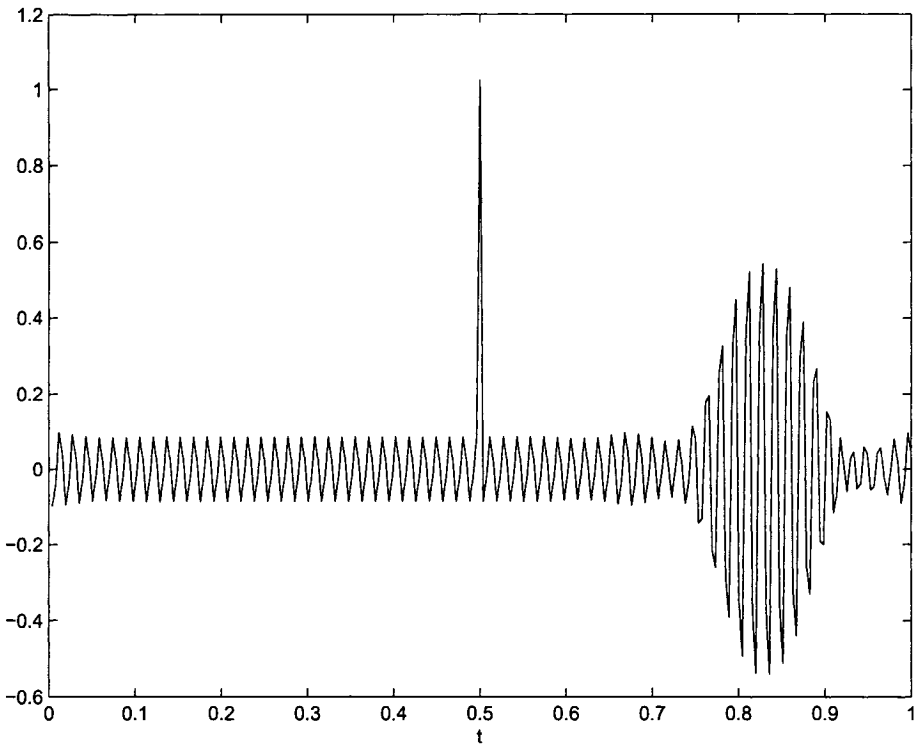
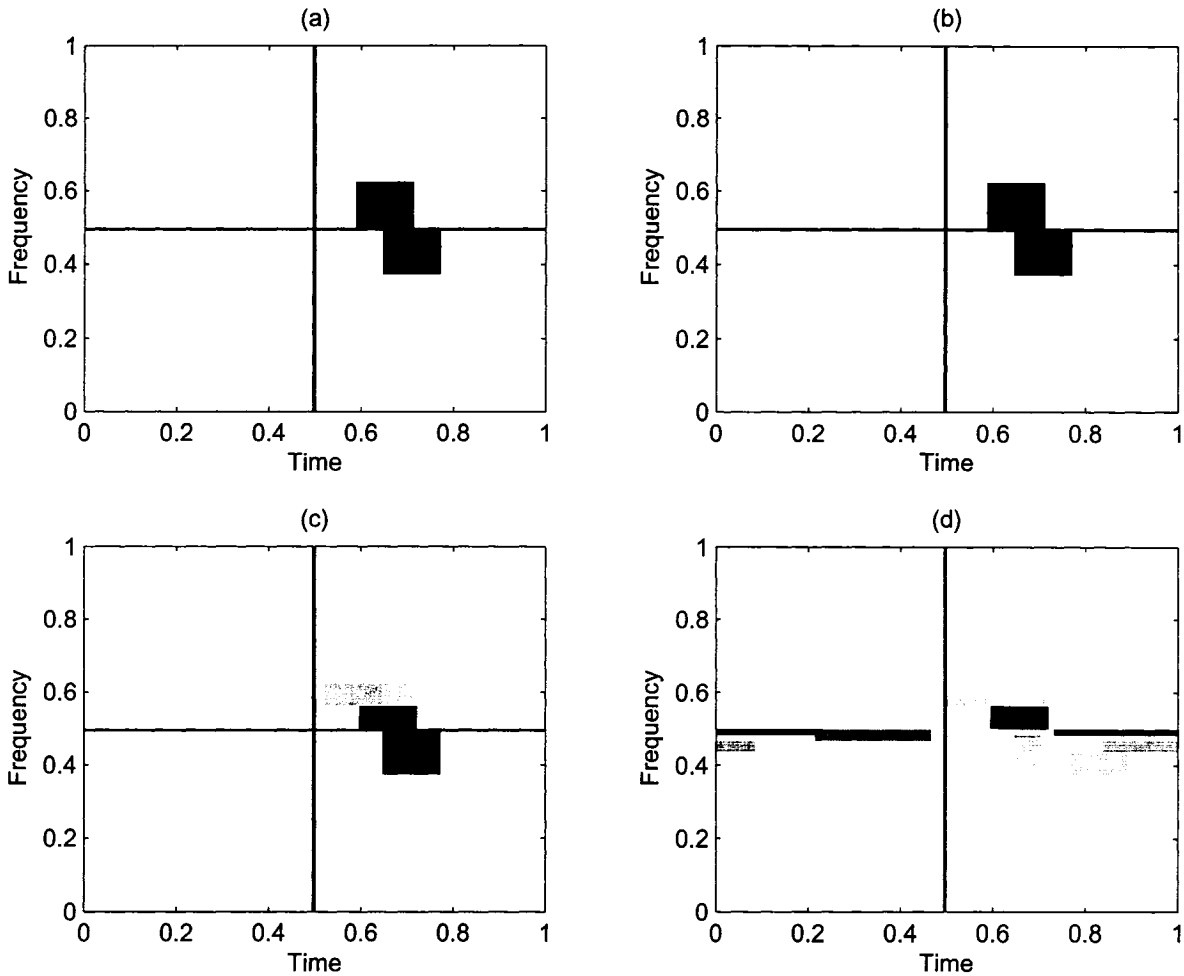


Figure 3.5: Time domain representation of the Carbon signal.

an impulse response at $t = 0.5$. This representation is evaluated by searching through 52736 nodes, which takes approximately 53 seconds. As the correlation threshold increases up to a breaking point of 0.45 (Fig. 3.6 (b)), this sparsest representation can be achieved with a much more efficient search; by searching through 14912 nodes, taking roughly 16 seconds. As the correlation threshold increases, the performance degrades. For $0.45 < \xi \leq 1$ the sparsest representation cannot be obtained as shown in the TF plane given in Fig. 3.6 (c). In Fig. 3.6 (d), the performance of the OMP algorithm is shown. This algorithm cannot even find a good representation in 10 iterations, and can only capture $(2.7316 - 0.40891)/2.7316 \times 100 \approx 85\%$ of the input signal's norm. The approximation obtained with OMP algorithm is shown in Fig. 3.7. The error between the original Carbon signal and the OMP algorithm approximation is



	L	d	ξ	N	lv	t(s)	$\|e_{10}\|$
(a)	3	1	0	52736	35155	52.859	0
(b)	3	1	0.45	14912	9593	15.781	0
(c)	3	1	0.46	14414	9076	14.531	0.031695
(d)	1	1	0	11	1	0.031	0.40891

Figure 3.6: Time/frequency plane of Carbon signal for $M = 10$, $\epsilon = 10^{-10}$ for various parameter sets.

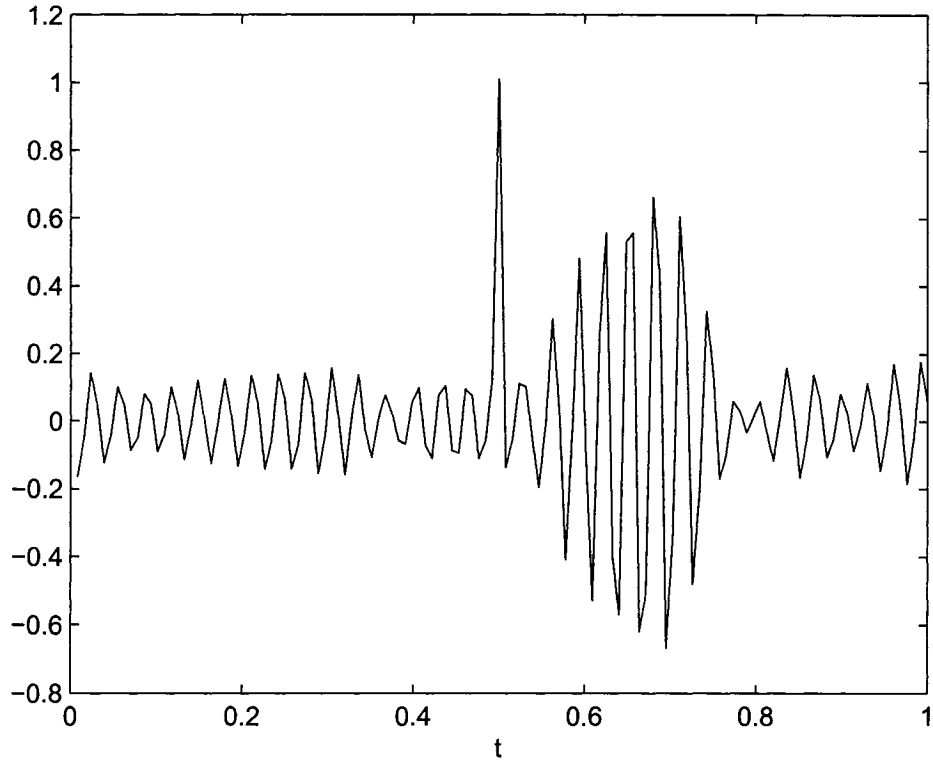


Figure 3.7: Time domain representation of the OMP algorithm approximation of the Carbon signal.

shown in Fig. 3.8.

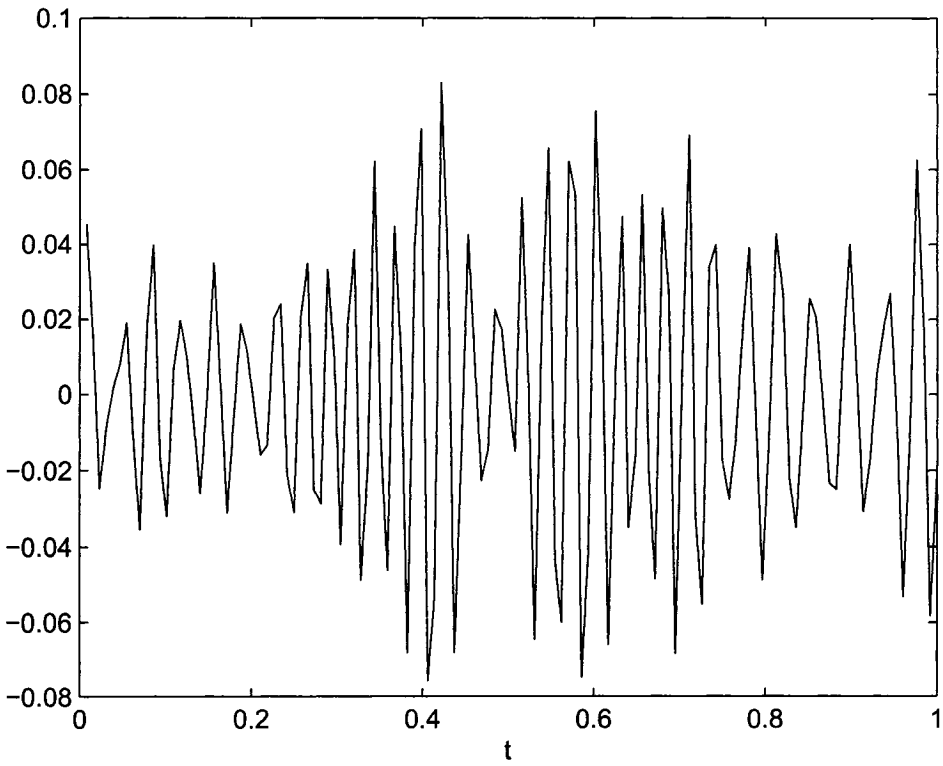
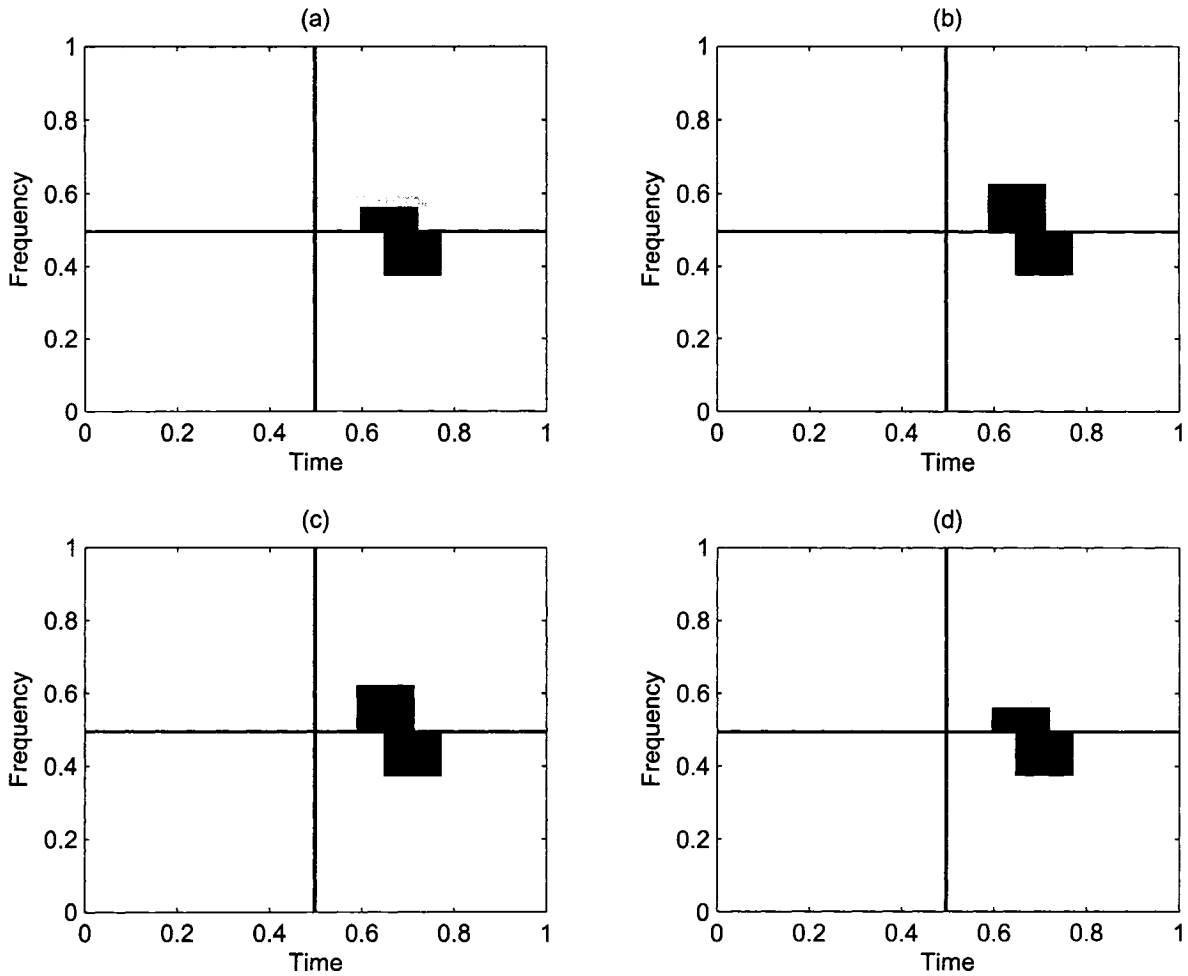


Figure 3.8: Time domain representation of the difference between OMP algorithm approximation of the Carbon signal and the Carbon signal.

In Fig. 3.9, larger values for L are considered with $d > 1$. The TF plane for $L = 8$, $d = 2$, and $\xi = 0$ is shown in Fig. 3.9 (a). This is not the sparsest representation, although it is a better approximation than the OMP algorithm (Fig. 3.6 (d)). As L increases, the approximation performance improves. For $L = 16$, $d = 2$, $\xi = 0$, the sparsest representation can be obtained by searching through 724 nodes (Fig. 3.9 (b)). As we increase the correlation threshold value, and hence the tree pruning, the tree-size decreases. For the case where $L = 16$, $d = 2$, $\xi = 0.47$, the sparsest representation can be reached by searching through only 476 nodes, taking only about 0.8 seconds (Fig. 3.9 (c)). As we increase the correlation threshold beyond this value, the performance degrades. The TF plane with $L = 16$, $d = 2$, $\xi = 0.48$ is shown in Fig. 3.9 (d).

From the experimental results, we can see that the most efficient search is observed for a large L value. For the case with $L = 16$, $d = 2$ and $\xi = 0.47$ only 675 nodes are searched, and the sparsest solution is obtained.



	L	d	ξ	N	lv	t(s)	$\|e_{10}\|$
(a)	8	2	0	553	64	0.922	0.068656
(b)	16	2	0	724	95	0.906	0
(c)	16	2	0.47	675	86	0.843	0
(d)	16	2	0.48	421	53	0.531	0.068656

Figure 3.9: Time/frequency plane of Carbon signal for $M = 10$, $\epsilon = 10^{-10}$ for various parameter sets with $\xi = 0$.

Twinsine Signal

In this experiment, the selected dictionary is a 4-fold discrete cosine transform (DCT) dictionary of size 256×1024 . The input signal is named as "Twinsine" [15], and consists of a linear combination of 2 cosines slightly outside the dictionary. The frequencies of cosines are closer than the Rayleigh distance, $2\pi/m$, where $m = 256$ for this example. The signal is plotted in Fig 3.10, and its norm is 1.812. The algorithm outputs are represented by the normalized frequency plots adapted from [15]. The actual normalized frequencies that form the superposition are represented by dashed lines on the plots, located at the normalized frequency values $124.2/1024$, and $126.2/1024$. The normalized frequencies that can be represented by the dictionary vectors are $k/1024$, for $k = 0, \ldots, 1023$.

The basic OMP cannot resolve these closely spaced cosines components due to its heuristic greedy structure. The normalized frequency plots of FTB-OMP algorithm with a selected range of parameters for $M = 10$ and $\epsilon = 10^{-10}$ are given in Fig. 3.11. In the figure, the number of searched nodes (N), the number of leaves in the search tree (lv), the computer running time in seconds (t), and the residual approximation error after 10^{th} iteration ($\|\mathbf{e}_{10}\|$) are reported. The dashed lines in the figures represent the actual locations of the signal components.

In Fig. 3.11 (a), the normalized frequency plot of the FTB-OMP algorithm with $L = 3$, $d = 1$ and $\xi = 0$ is shown. We remark that this corresponds to TB-OMP with $L = 3$. The resulting TF plane is the sparsest possible representation of the Twinsine signal with two peaks signal. This representation is evaluated by searching through 1093 nodes in approximately 3 seconds.

Similar to the Carbon signal case, as the threshold increases, this sparsest representation can be achieved with a more efficient search. For the correlation threshold value of 0.76, the normalized frequency plot is given in Fig. 3.11 (b). We can see from the figure that the

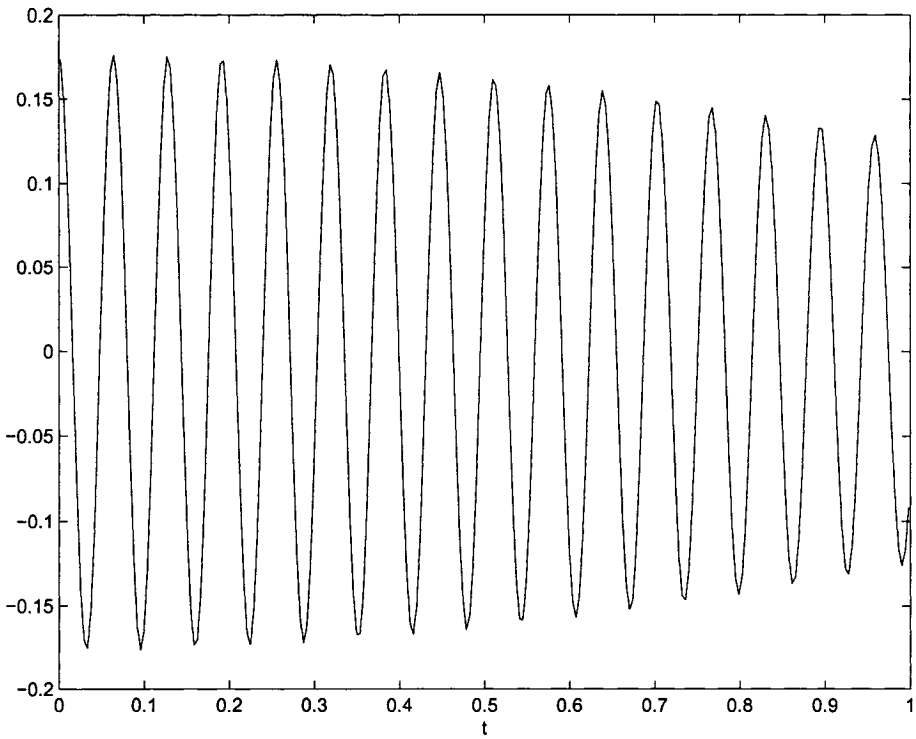
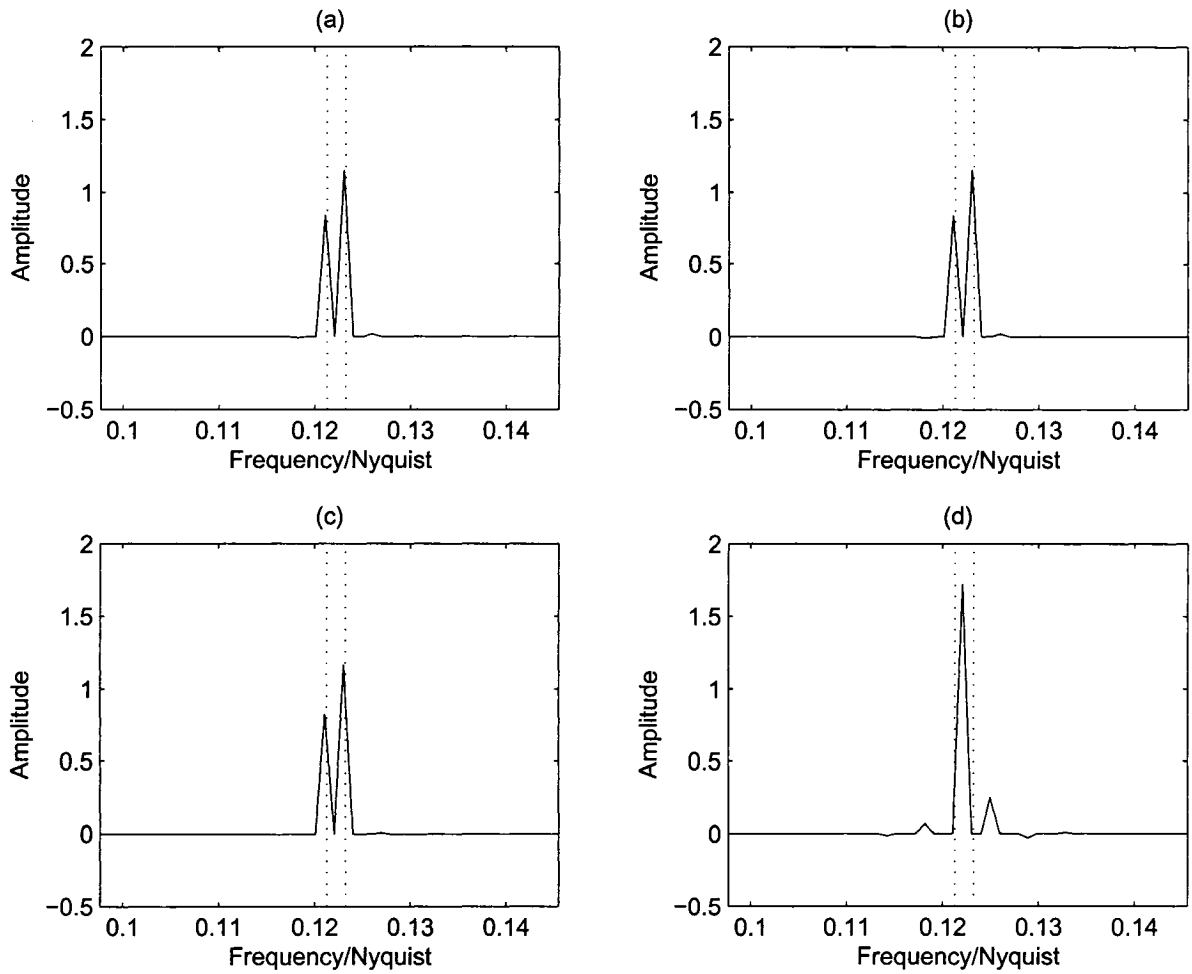


Figure 3.10: Time domain representation of the Twinsine signal.

components can be resolved by searching through 63 nodes, in roughly 0.36 seconds. As the values of L and d increase, a more efficient tree search can be performed. For the parameters $L = 8$, $d = 2$ and $\xi = 0.76$ the actual components can be resolved by searching through only 56 nodes, taking less than 0.1 seconds as shown in Fig. 3.11 (c).

In Fig. 3.11 (d), the performance of the OMP algorithm is shown. As can be seen from the figure, although 99.5% of the input signal's norm is captured, the two components are not resolved after 10 iterations with OMP. We conclude that it is advantageous to increase the values of L and d. Moreover, the parameter ξ can be tuned according to the dictionary for higher computational efficiency.



	L	d	ξ	N	lv	t(s)	$\|e_{10}\|$
(a)	3	1	0	1093	729	2.984	0.00041522
(b)	3	1	0.76	63	13	0.3594	0.001541
(c)	8	2	0.76	56	11	0.1712	0.001541
(d)	1	1	0	7	1	0.0625	0.0092791

Figure 3.11: Normalized frequency plot of Twinsine signal for $M = 10$, $\epsilon = 10^{-10}$ for various parameter sets.

3.4 Conclusions

In this chapter, we proposed a novel and efficient tree-search based orthogonal matching pursuit algorithm for sparse signal representations, called FTB-OMP. The algorithm provides some design parameters, giving flexibility to choose between higher approximation performance and lower complexity. The efficiency is achieved by using a correlation based pruning in the search tree, and by changing the number of children at each tree node. It is verified in our experiments that with respect to TB-OMP, performance improvements can be achieved with much lower computer running times.

The proposed FTB-OMP algorithm is a suitable technique to solve problems where sparse signal representations are required. From the simulation results, we can conclude that FTB-OMP algorithm with high L values and $d > 1$ is more effective than TB-OMP for this problem. The novel exponential decaying structure of the search tree makes the algorithm more computationally efficient and increases the approximation and detection performance. The threshold parameter ξ can be used for further computational efficiency. Results of component detection experiments show that, by tuning parameters of FTB-OMP, a good approximation and detection performance with time complexity lower than TB-OMP algorithm can be obtained. This is also confirmed on various signals where FTB-OMP algorithm can obtain the sparsest representation.

It is known that solving a set of linear equations is a frequently encountered problem in communications and signal processing areas. In the following chapters, it is shown that FTB-OMP is a suitable detection algorithm that can be applied to channel estimation, direction of arrival detection and multi-user detection problems.

Chapter 4

Sparse Channel Estimation by Matching Pursuit Algorithms - Block Fading Channels

Sparse channels are frequently encountered in communication applications [92,98]. Exploiting the sparsity property, the channel estimation problem can be modeled as the basis selection problem. A channel estimate can be obtained by using the matching pursuit (MP) algorithms with undercomplete dictionaries.

In the literature channel estimation techniques are divided into three major classes. These are summarized below.

- Training sequence based channel estimation techniques
- Semi-blind channel estimation techniques
- Blind channel estimation techniques

Traditional training sequence based techniques transmit a known information sequence, referred to as the training sequence, and try to estimate the channel impulse response from the received sequence. Although these methods are known to give the best estimation performance, transmitting a training sequence reduces the communication efficiency [83]. In order to overcome this problem, blind estimation techniques are introduced. A good review of blind estimation methods can be found in [101]. These methods try to estimate the channel by avoiding the training sequence. One popular criterion for blind channel estimation is to use subspace identification methods. Although these methods provide increased communication efficiency by avoiding the overhead data introduced through the training sequence, their performances are inferior to that of training sequence based methods. They also suffer from limitations that restrict their use to some specific applications [101]. In semi-blind channel estimation techniques, a combination of training based and blind channel estimation techniques are used.

In this chapter, we elaborate on training sequence based channel estimation, due to its inherent advantages in terms of detection performance. The associated disadvantage of low efficiency is also tackled and necessary conditions for minimizing the training sequence length are stated.

Application of BMP algorithm to training sequence based channel estimation problem is proposed in [21]. It is shown that, the BMP based channel estimation is more accurate and computationally simpler than the least squares (LS) estimation [23, 70]. Furthermore, a shorter training sequence is required to form an accurate channel estimate leading to greater information throughput [25].

In this chapter, we propose to use the orthogonal matching pursuit (OMP) algorithm [56] and the FTB-OMP algorithm for estimating channel impulse responses of block fading channels. In the BMP algorithm, since each iteration optimization is performed over all vectors in the dictionary, it is possible to re-select a previously selected vector. This process

slows down the convergence speed to a sparse solution [72]. In OMP and FTB-OMP, the re-selection problem is avoided by using the stored dictionary at each iteration. In these algorithms by eliminating the re-selection problem a faster convergence to a sparse solution is obtained. It is shown via simulations that more accurate channel estimates can be acquired by applying the OMP and FTB-OMP algorithm.

Also in this chapter, we consider optimum training sequences design for the OMP algorithm for channel estimation problem. The necessary and sufficient conditions for the OMP to detect the exact channel taps are derived. These conditions allow designing short training sequences that can resolve all existing channel taps for sparse channels, which reduces the overhead by minimizing length of the training sequence and hence increases communication efficiency.

This chapter is organized as follows. In Section 4.1, the system model for block fading channels is given. Least squares channel estimation and popular estimation techniques that are related with basis selection are reviewed in Section 4.2. In Section 4.3, training sequence design method with exhaustive search is introduced for OMP in order to increase efficiency. Simulation results that compare the accuracy of the block fading channel estimates are presented for BMP, OMP and FTB-OMP algorithms in Section 4.4. A portion of the research work presented in this chapter is published in [56, 58].

4.1 Problem Statement

A continuous time model for a digital communication system with a time-varying fading channel is shown in Fig. 4.1. The notation and the channel model is based on [36]. In this model $w_c(t)$ is the complex additive white Gaussian noise with mean 0 and variance σ^2, T_s represents the symbol period. The sequence of complex input symbols tranmitted in time interval $[nT_s, (n+1)T_s)$ is represented by $s(n)$. Time variant (TV) channel impulse response

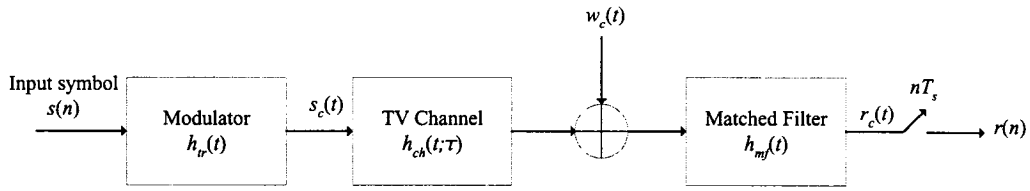


Figure 4.1: Continuous time model of time-varying communication system.

is denoted by $h_{ch}(t;\tau)$. The filtered transmitted pulse and matched filter impulse response are represented by $h_{tr}(t)$ and $h_{mf}(t)$, respectively. The input/output (I/O) relation of this model for the continuous received symbol can be written as

$$r_c(t) = s(n) \star h_{tr}(t) \star h_{ch}(t;\tau) \star h_{mf}(t) + w_c(t) \star h_{mf}(t), \tag{4.1}$$

where $\star$ denotes the convolution operation. Defining $h_t(t,\tau) = h_{tr}(t) \star h_{ch}(t;\tau) \star h_{mf}(t)$ and $\nu_c(t) = w_c(t) \star h_{mf}(t)$, the I/O relation can be simplified as

$$r_c(t) = \sum_{n=-\infty}^{+\infty} s(n) h_t(t; \tau - nT_s) + \nu_c(t). \tag{4.2}$$

In this work, we restrict ourselves to estimating $h_{ch}(t;\tau)$, the channel impulse response only by neglecting the filters at the transmitter and receiver. However the discussion can be extended to include $h_{tr}(t)$ and $h_{mf}(t)$.

After sampling the received signal at a frequency of $1/T_s$, the output for a block of B symbols can be written as

$$r(n) = \sum_{l=0}^{P-1} h(n;l) s(n-l) + \nu(n), \quad n = 0, 1, 2, \ldots, B-1, \tag{4.3}$$

where $h(n;l) = h_{ch}(nT_s; lT_s)$ is the sampled channel impulse response truncated to a span of P symbol periods. This assumption is very common in communications applications [36]. The discrete time system model for TV channel model is shown in Fig. 4.2. The sampled channel impulse response $h(n;l)$ can also be viewed as the l^{th} tap of the channel at time n.

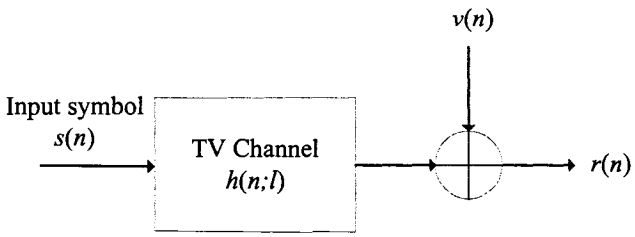


Figure 4.2: Discrete time model of time-varying communication system.

Accurate knowledge about the channel is necessary for many receiver structures including maximum likelihood sequence detectors, decision feedback equalizers and multi-user detectors [83]. Hence detection of $h(n;l)$ is an important problem for wireless communication systems. Length of the training sequence is denoted by N. This sequence of known symbols is transmitted in a fraction of the data packet of B symbols with $N < B$. The channel impulse response, $h(n;l)$, is estimated from the received samples. The estimates are then used to detect the remaining symbols in the data packet. The training sequence can be placed at the end, beginning or the middle of the data packet.

Throughout the dissertation, we investigate this problem considering two channel types; block fading channels in Chapter 4, and quasi-block fading channels in Chapter 5.

4.1.1 Block Fading Channels

Assuming that the channel is slowly varying and it is time invariant (TI) during transmission of B symbols, (4.3) becomes

$$r(n) = \sum_{l=0}^{P-1} h(l)s(n-l) + \nu(n), \quad n = 0, 1, 2, \ldots, B-1. \tag{4.4}$$

This case is referred to as block fading channel model, and is considered in detail in Section 4.1.1. Furthermore for $P = 1$, (4.4) becomes $r(n) = s(n)h(n) + \nu(n)$ where the channel is

not frequency selective but time selective.

The sparse channel estimation problem for block fading channels can be stated as follows. Let $d(n)$ be the complex training sequence for $n = 0, 1, \cdots, N-1$, that is transmitted through a stationary channel. The training sequence symbols $d(n)$ for $n < 0$ can be obtained from the previous estimates or for the first arriving frame they are assumed to be zero.

Considering only the training sequence, the received signal samples can be modeled as

$$r(n) = \sum_{l=0}^{P-1} h(l)d(n-l) + \nu(n) \quad n = 0, 1, 2, \ldots, N-1. \tag{4.5}$$

We assume that the channel is sparse, i.e. $h(l) \neq 0$ for only a few values of l.

Without loss of generality we can assume that the training sequence is the first N symbols transmitted. For channel estimation problem, N must be greater or equal to P for reliable channel detection. This leads to an undercomplete dictionary. The received symbol can be written in a matrix form as [25]

$$\begin{bmatrix} r(0) \\ r(1) \\ \vdots \\ r(N-1) \end{bmatrix} = \begin{bmatrix} d(0) & d(-1) & \cdots & d(-P+1) \\ d(1) & d(0) & \cdots & d(-P+2) \\ \vdots & \vdots & \cdots & \vdots \\ d(N-1) & d(N-2) & \cdots & d(N-P) \end{bmatrix} \begin{bmatrix} h(0) \\ h(1) \\ \vdots \\ h(P-1) \end{bmatrix} + \begin{bmatrix} \nu(0) \\ \nu(1) \\ \vdots \\ \nu(N-1) \end{bmatrix}. \tag{4.6}$$

The equation above can be abbreviated as

$$\mathbf{r} = \mathbf{D}\mathbf{h} + \mathbf{n}. \tag{4.7}$$

The $N \times P$ matrix $\mathbf{D}$ that is composed of the training sequence can be represented in terms of its column vectors as

$$\mathbf{D} = [\, \mathbf{d}_0 \, \mathbf{d}_1 \, \mathbf{d}_2 \, \cdots \, \mathbf{d}_{P-1} \,]. \tag{4.8}$$

Assume that there are at most $M < P$ nonzero taps. Let the M nonzero channel tap values be denoted by the $M \times 1$ vector $\mathbf{h}_s$, and the collection of columns vectors $\mathbf{d}_\lambda$ with indices from $\lambda \in \{0, 1, 2 \ldots, N-1\}$ be denoted by the $N \times M$ matrix $\mathbf{D}_s$. Then the received sequence can be written as

$$\mathbf{r} = \mathbf{D}_s \mathbf{h}_s + \mathbf{n}. \tag{4.9}$$

for the noise vector $\mathbf{n}$.

4.2 Channel Estimation Methods

Estimation of $\mathbf{h}$ from (4.7) is a well-known problem in the literature. There are several methods for detecting the channel taps. Here, we consider the least squares estimation and the matching pursuit (MP) based estimation techniques.

4.2.1 Least Squares Channel Estimation

For block fading channels, the classical channel estimation technique is the least squares (LS) method. In this method the estimates are evaluated as

$$\hat{\mathbf{h}} = \mathbf{D}^\dagger \mathbf{r} = (\mathbf{D}^H \mathbf{D})^{-1} \mathbf{D}^H \mathbf{r}, \tag{4.10}$$

where $(\cdot)^H$ denotes the Hermitian transpose of a matrix, and $(\cdot)^\dagger$ denotes the Moore-Penrose matrix inverse. Further details about application of LS to channel estimation problem and equalization can be found in [37].

Although this solution is the best linear unbiased estimator, it does not exploit the knowledge that many of the channel taps are non-zero. A thresholding method for the channel taps with small amplitudes are proposed by Cotter in [21] in order to exploit the sparsity of the

channel. For further performance improvement, the thresholded coefficients are re-estimated using LS by exploiting the location information.

We must note that although the matrix inversion process in LS based estimation for high values of N is computationally intensive, it can be calculated before estimation process and stored in the processor. Throughout the rest of the text, this channel estimation method is referred to as LS-CE.

4.2.2 Basic Matching Pursuit Based Channel Estimation

Since the channel is sparse, most components of the channel impulse response are zero. Through this knowledge, a sparse solution to $\mathbf{r} \approx \mathbf{Dh}$ is achieved by applying an MP algorithm. The problem is to approximate $\mathbf{r}$ as a linear combination of a small number of columns of $\mathbf{D}$.

BMP can be applied to detect the locations and the amplitudes of non-zero taps by setting $\mathbf{r} = \mathbf{x}$ in the algorithm. The output of the BMP algorithm is the index set $\Lambda_{BMP} = \{\lambda_m\}_{m=0}^{M-1}$ that denotes the location of the non-zero channel taps, and the coefficient set $\mathcal{C}_{BMP} = \{c_{\lambda_m}\}_{m=0}^{M-1}$. The dictionary is the normalized set of column vectors that form $\mathbf{D}$.

Application of BMP to channel estimation problem of block fading channels is proposed in [23]. It is shown that, the BMP based channel estimation (BMP-CE) is more accurate and computationally simpler than the LS-CE [21]. Furthermore, a shorter training sequence is required to form accurate channel estimates leading to greater information throughput [25]. However, its performance severely degrades when the sparseness condition does not hold [70].

4.2.3 Least Square Matching Pursuit Based Channel Estimation

A method for improving the performance of the BMP-CE is a two step technique that involves BMP algorithm and the LS solution. This method first detects the locations of the non-zero taps by BMP algorithm. That is the index set $\Lambda_{BMP} = \{\lambda_m\}_{m=0}^{M-1}$ is evaluated. Then LS estimate is obtained for the non-zero channel taps. Using M nonzero channel tap values, $\mathbf{h}_s$, and the collection of columns vectors $\mathbf{D}_s$, the received sequence can be written as (4.9). The non-zero channel estimates are evaluated as

$$\hat{\mathbf{h}}_s = \mathbf{D}_s^{\dagger} r = (\mathbf{D}_s^H \mathbf{D}_s)^{-1} \mathbf{D}_s^H \mathbf{r}. \tag{4.11}$$

Hence the inversion of the $P \times P$ matrix in (4.10) is reduced to the inversion of an $M \times M$ matrix.

This method, referred here as BMP-LS-CE has been applied to quasi-block fading channels by Liu and Borah [70, 71]. The advantage of this method is that for channels that are not sparse, the estimates are more accurate than the estimates that are solely obtained from the BMP output [70], however the method introduces extra computational complexity when compared to BMP-CE.

4.2.4 Orthogonal Matching Pursuit Based Channel Estimation

A more straightforward solution for improving the estimation performance is the application of OMP algorithm to the channel estimation problem as we recently proposed [56]. In the BMP algorithm, since at each iteration, optimization is performed over all vectors in the dictionary, it is possible to re-select a previously selected vector. This process slows down the convergence to a sparse solution [72].

The output of the OMP algorithm is the index set $\Lambda_{OMP} = \{\lambda_m\}_{m=0}^{M-1}$ that denotes the tap locations and the coefficient set $\mathcal{C}_{OMP} = \{c_{\lambda_m}\}_{m=0}^{M-1}$ that is obtained by projecting the received vector $\mathbf{r}$ onto the orthogonalized selections from the dictionary $\mathcal{D}$ that is formed by normalizing the columns of $\mathbf{D}$. The OMP algorithm, orthogonalizes the selection vector at each iteration guaranteeing the convergence of the algorithm with a finite number of iterations. This method is referred to as OMP-CE throughout the text. In OMP-CE, the re-selection problem is avoided by using the stored dictionary at each iteration hence a faster convergence to a sparse solution can be obtained. In Section 4.4, it is experimentally verified that the OMP based channel estimates are more accurate than BMP based estimates for sufficiently high values of SNR.

Another advantage of the OMP algorithm is that due to the dictionary analysis introduced by Tropp in [104], it is possible to design a training sequence that guarantees the convergence of the OMP algorithm for a specific value of M. Hence unlike the BMP algorithm, the accurate channel estimates are detected by using OMP. Details about the training sequence design is given in Section 4.3.

4.2.5 Channel Estimation by Flexible Tree-Search Based Orthogonal Matching Pursuit

It is shown that the OMP algorithm gives the sparsest representation for quasi-incoherent dictionaries [104]. However, for some channels where M does not satisfy the incoherence requirement of the dictionary, a more detailed search for the channel taps would give more accurate estimates.

In order to further improve the estimation performance of the selected MP algorithm, the flexible tree-search based OMP (FTB-OMP) algorithm can be applied to obtain more accurate channel estimates. For cases where OMP does not give the sparsest representation,

the FTB-OMP gives a more accurate channel estimate with higher computational complexity that is associated with the branching factor L, the decay parameter d, and the correlation threshold ξ.

The output of the FTB-OMP algorithm is the index set $\Lambda_{FTB-OMP(L,d,\xi)} = \{\lambda_m\}_{m=0}^{M-1}$ that denotes the tap locations and the coefficient set $\mathcal{C}_{FTB-OMP(L,d,\xi)} = \{c_{\lambda_m}\}_{m=0}^{M-1}$. Similar to the OMP case, $\mathcal{C}_{FTB-OMP(L,d,\xi)}$ is obtained by projecting the received vector $\mathbf{r}$ onto the orthogonalized vectors from the final selected index set from the dictionary $\mathcal{D}$.

The application of FTB-OMP to channel estimation is referred to throughout the text as FTB-CE. The advantages of the OMP-CE are kept with FTB-CE. Furthermore more accurate estimates are obtained for coherent dictionaries where OMP-CE does not converge to the sparsest solution. However, the computational complexity increases and the parameters L, d and ξ must be adjusted according to system limitations.

4.3 Training Sequence Design for Channel Estimation using OMP Algorithm

It is stated above that the least square estimate for the channel taps given (4.10) is known to be the best estimator when there is no knowledge about the sparseness of the channel. An optimum training sequence for LS-CE is the sequence that gives a perfectly diagonal covariance matrix. In [26], some sequence examples are given as a result of exhaustive computer search.

In Section 4.4, it is shown that using the sparseness information that is obtained by MP algorithms, a more accurate estimate than LS-CE can be obtained. Especially OMP algorithm gives a good tradeoff between performance and complexity. In this section, the constraints for designing an optimum training sequence for OMP-CE is introduced. Using the proposed

conditions, the shortest possible training sequence can be picked depending on the channel parameters, reducing the overhead due to training sequence.

The conditions that needed to be satisfied by the dictionary for the convergence of the OMP algorithm were stated in (2.39) and (2.40). Having these conditions for exact recovery of OMP, we can design the training sequence so that it can minimize the value of N for given P and M values, and estimate all taps of the channel when there is negligible amount of noise [58].

First, we define the normalized windowed autocorrelation function (NWAF) as

$$R_w(i,j) = \frac{\sum_{k=0}^{N} d(k+i)d(k+j)^*}{\sqrt{\sum_{k=0}^{N} d(k+i)d(k+i)^*}\sqrt{\sum_{k=0}^{N} d(k+j)d(k+j)^*}}, \tag{4.12}$$

with $d(l) = 0$ for $l \geq N$ or $l < 0$. The denominator appears due to the normalization requirement of the dictionary $\mathcal{D}$ for MP algorithms.

It can be shown that the dictionary coherence is

$$\mu = \max_{\substack{i,j \\ i \neq j}} |R_w(i,j)|, \tag{4.13}$$

and cumulative coherence is

$$\mu_1(m) = \max_{|\Lambda|=m} \left\{ \max_i \sum_{\lambda \in \Lambda} |R_w(i,\lambda)| \right\}. \tag{4.14}$$

for i an integer from the set $\{0,1,2,\ldots,N-1\} \setminus \Lambda$. Relating the NWAF with the conditions for exact recovery of OMP stated in (2.39) and (2.40), a training sequence for detection of M channel taps can be designed so that the dictionary is quasi-incoherent. That is the NWAF, hence the training sequence, must satisfy one of the conditions below for the convergence of OMP to the exact representation, and get the exact channel taps when the noise is negligible.

Corollary 2 *The exact recovery condition holds for a training sequence* $\{d(n)\}_{n=0}^{N-1}$ *for a*

channel with M nonzero taps when

$$M < \frac{1}{2}\left(\frac{1}{\max_{\substack{i,j \\ i\neq j}} |R_w(i,j)|} + 1\right), \tag{4.15}$$

or when

$$\max_{|\Lambda|=M}\left\{\max_{i\in\{0,\ldots,N-1\}\setminus\Lambda}\sum_{\lambda\in\Lambda}|R_w(i,\lambda)|\right\} + \max_{|\Lambda|=M-1}\left\{\max_{i\in\{0,\ldots,N-1\}\setminus\Lambda}\sum_{\lambda\in\Lambda}|R_w(i,\lambda)|\right\} < 1. \tag{4.16}$$

The OMP algorithm can be applied to $\mathbf{r}$ for detection of the channel taps by using the normalized columns of $\mathbf{D}$ as dictionary $\mathcal{D}$. The algorithm can resolve all M nonzero taps of an P tap channel by using a training sequence that satisfies (4.15) and (4.16) through NWAF. Hence these relations can be used to determine the possible shortest training sequence for a multipath channel.

Training Sequence Design for Binary Phase Shift Keying

For BPSK signaling the minimum length of training sequence $\{d(n)\}_{n=0}^{N-1}$ for an M tap channel is evaluated by an exhaustive search. The results are summarized in Table 4.1. Two training sequences per case are presented in the table, however two more can be given by using antipodals of the given sequences. These sequences are tabulated here in order to demonstrate that they contain no special features.

Two sequence examples are tabulated for each (M, P) parameter set. Multiplication of the sequence by -1 also gives an optimal training sequence for specified N, M and P values due to symmetry properties. The table concentrates on a small channel span and comparable values of nonzero channel taps. This exhaustive search can simply be extended to sequences with more levels. It should also be noted that a training sequence designed for fixed value of P and M can also be used when the number of non-zero channel taps value is less than M. As an example, the sequence designed for $P = 5$ and $M = 4$ can also detect channels with 3 or less non-zero taps.

For cases where P is large and M is relatively small, setting N on the order of P and random selection of sequence symbols are very likely to give the optimal sequence, eliminating the need for an exhaustive search.

Table 4.1: Minimum length of training sequence examples for estimating M tap channel for BPSK signaling. Legend # represents the number of distinct binary sequences that satisfy (4.15) and (4.16).

P	M	N	#	2 Example Sequences
5	2	8	4	[1, −1, −1, −1, 1, −1, −1, 1]
				[1, 1, −1, 1, 1, 1, −1, −1]
5	3	9	8	[1, −1, −1, −1, −1, 1, −1, −1, −1]
				[1, 1, −1, 1, −1, −1, −1, 1, −1]
5	4	10	20	[1, 1, −1, 1, −1, −1, −1, 1, −1, −1]
				[1, 1, −1, 1, 1, 1, −1, 1, −1, −1]
5	5	10	20	[1, 1, −1, 1, −1, −1, −1, 1, −1, −1]
				[1, 1, −1, 1, 1, 1, −1, 1, −1, −1]
10	2	14	8	[1, −1, −1, −1, 1, −1, −1, 1, −1, 1, 1, −1, −1, −1]
				[1, 1, −1, −1, 1, −1, 1, 1, 1, 1, 1, 1, −1, −1]
10	3	17	4	[1, 1, −1, −1, 1, −1, 1, 1, 1, 1, −1, 1, 1, 1, −1, −1, −1]
				[1, −1, −1, 1, 1, 1, 1, −1, 1, −1, −1, −1, 1, −1, −1, 1, −1]
10	4	18	4	[−1, 1, −1, −1, −1, 1, 1, 1, 1, −1, 1, 1, −1, 1, 1, 1, −1, −1]
				[1, 1, 1, −1, 1, 1, −1, 1, −1, −1, −1, 1, 1, 1, −1, 1, 1, −1]
10	5	19	8	[−1, 1, 1, 1, −1, −1, 1, −1, 1, 1, 1, 1, 1, −1, 1, 1, −1, −1, −1]
				[−1, 1, −1, −1, −1, 1, 1, 1, 1, −1, 1, 1, −1, 1, 1, 1, −1, −1, −1]

4.4 Performance Comparison of Channel Estimation Algorithms for Block Fading Channels

In this section, the channel estimation algorithms explained above are compared in terms of estimation accuracy, and equalization performance based on channel estimates.

4.4.1 Detection of Nonzero Channel Taps

The performances of the MP algorithms are compared in the detection of the non-zero tap locations. This idea is closely related with the component detection experiment. Hence the same conclusions are expected from the experiment results. The re-evaluation is due to drawing some guidelines for specific channel conditions.

First sparse channels with short delay spreads are considered. The simulation parameters are set as $P = 15$ and $N = 20$. The channel is assumed to have at most M nonzero coefficients. BPSK signaling is used. In practice a priori information about the exact number of nonzero taps is not available at the receiver. However, in order to compare the accuracy of the tap locations detected by an MP algorithm, the maximum number of iterations is set to M. In order to model the finite precision arithmetic of digital signal processors, stopping error is selected as $\epsilon = 10^{-3}$.

The tap location detection performances of the MP-CE, OMP-CE, and FTB-CE methods are investigated with various parameter sets. Channel tap values are assumed to have a normalized exponentially decaying structure. Simulations are run for 1000 times and results are averaged. Channel tap locations and values are randomly selected for each run. Training sequences are designed so that they satisfy (4.15) or (4.16). The training sequence symbols $d(n)$ for $n < 0$ are randomly set to -1 or $+1$.

For the FTB-OMP algorithm parameters are selected as $L = 10$, $d = 5$ and $\xi = 0.05$. Hence using (3.4), the computational complexity associated with the tree structure is expected to be upper bounded roughly by $10 \times \lceil 10/5 \rceil \times 1 = 20$ times that of OMP in terms of the number of leaves when there is no threshold based pruning in the search tree.

The histograms for the correctly detected channel tap locations are given in Fig. 4.3 for SNR values of 0, 10, 20 and 30 dB, respectively. From the histograms we can see that OMP-CE has slightly better estimation accuracy than BMP-CE, and FTB-CE has the best accuracy among the compared algorithms for SNR values of 10 dB or more. For a low SNR value, the performance of the BMP algorithm is better than the rest. The reason for this result is that the OMP based algorithms tend to minimize the residual error. Hence for a low SNR case they select different paths resulting in the wrong tap locations. With increasing SNR, their performance increase significantly, exceeding the detection performance of the BMP-CE. FTB-CE can further be improved by increasing L and decreasing d, and ξ values however, the complexity associated with the algorithm also increases.

In terms of computational complexity the average number of tree nodes, leaves and the average running times are plotted in Figs. 4.4 and 4.5. The average number of nodes associated with the OMP and BMP algorithms are $M + 1$, and the leaves are always 1. The running times are increasing linearly for OMP and BMP. For FTB-OMP, it increases exponentially as shown in Fig. 4.5. This is consistent with the increasing number of nodes in the search tree.

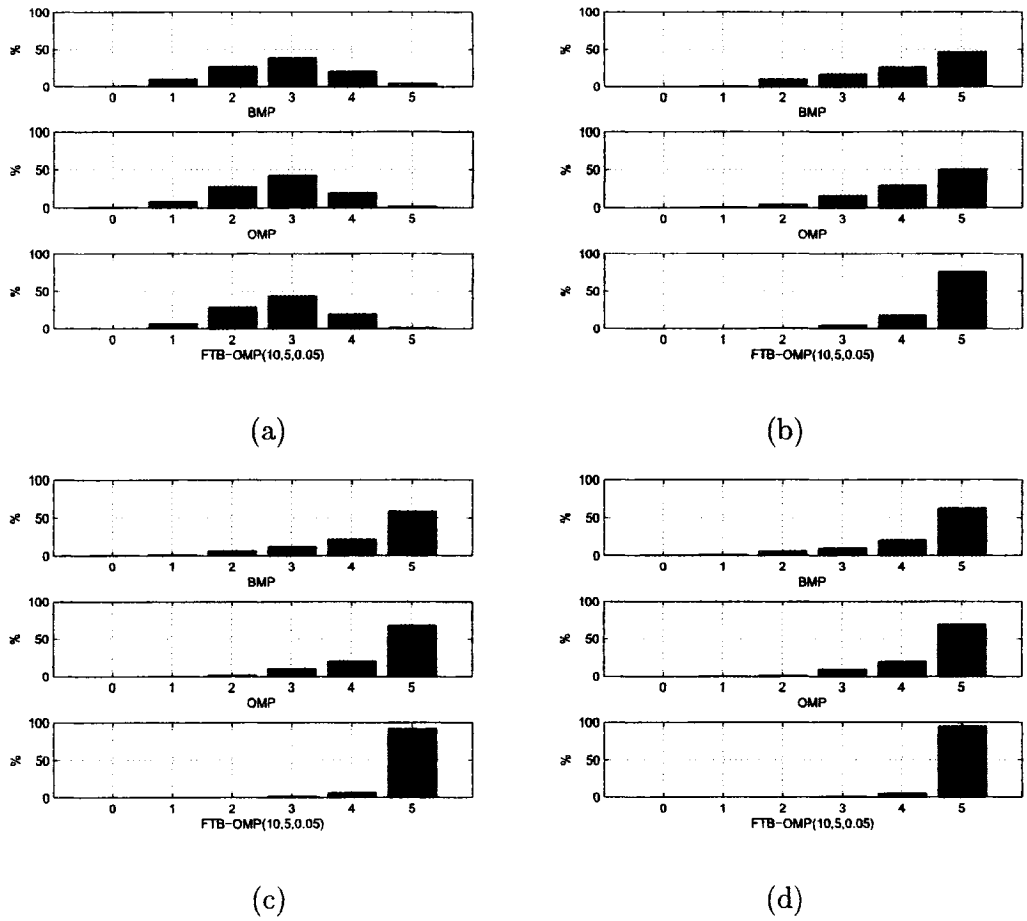


Figure 4.3: Histograms of number of correctly detected components by using BMP, OMP and FTB-OMP$(10, 5, 0.05)$ algorithms, $M = 5$, $N = 20$, $P = 15$, $\epsilon = 10^{-3}$ for (a) SNR=0 dB, (b) SNR=10 dB, (c) SNR=20 dB, (d) SNR=30 dB.

After comparing the detection performances of the algorithms with respect to a fixed M value, we evaluate the algorithms' performances for various M values. For the simulations we considered $M = 1, 2, \ldots 10$, and obtained the percentages when all channel taps are correctly detected. The results are shown in Fig. 4.6 for SNR values of 0, 10, 20 and 30 dB. For SNR=0 the performances of all algorithms are poor. For SNR=10 dB the BMP-CE has a good performance. As mentioned earlier, the reason behind this is, for high SNR, the OMP based estimation techniques try to minimize the average residual error. This results in deviations from actual tap locations when noise variance σ^2 is large. OMP-CE gives more accurate estimates than BMP-CE for SNR values of 20 and 30 dB. For all cases the detection accuracy of the FTB-CE is higher than OMP-CE and BMP-CE.

4.4.2 Estimation of Nonzero Channel Taps

The accuracy of the LS-CE, BMP-CE, LS-BMP-CE, OMP-CE and FTB-CE techniques are compared in terms of the mean-squared error in the channel estimates, through a normalized channel. The channel length is set as $P = 30$ and the training sequence is $N = 50$ bits. The channel has at most 5 nonzero taps with uniformly distributed indices and exponentially decaying amplitudes. The stopping criteria for the MP algorithms is set to 6 terms assuming that the number of nonzero channel coefficients is not available at the receiver. The stopping parameter is $\epsilon = 10^{-3}$ assuming that knowledge about SNR is not available at the receiver.

It is mentioned above that the LS-CE is the best estimate when there is no knowledge about the sparseness of the channel taps. Let us assume that it is certain that at most M taps are nonzero. Without loss of generality it can be assumed that these taps are located as the first M coefficients, that is

$$\mathbf{r} = \mathbf{D}\mathbf{h} + \mathbf{n} = [\mathbf{D_1}\ \mathbf{D_2}][\mathbf{h_1}\ \mathbf{h_2}]' + \mathbf{n}, \tag{4.17}$$

where $(\cdot)'$ denotes transpose operation, $\mathbf{D}_1 \in \mathbb{C}^{N\times M}$, $\mathbf{D}_2 \in \mathbb{C}^{N\times(P-M)}$, $\mathbf{h}_1 \in \mathbb{C}^{M}$, $\mathbf{h}_2 \in$

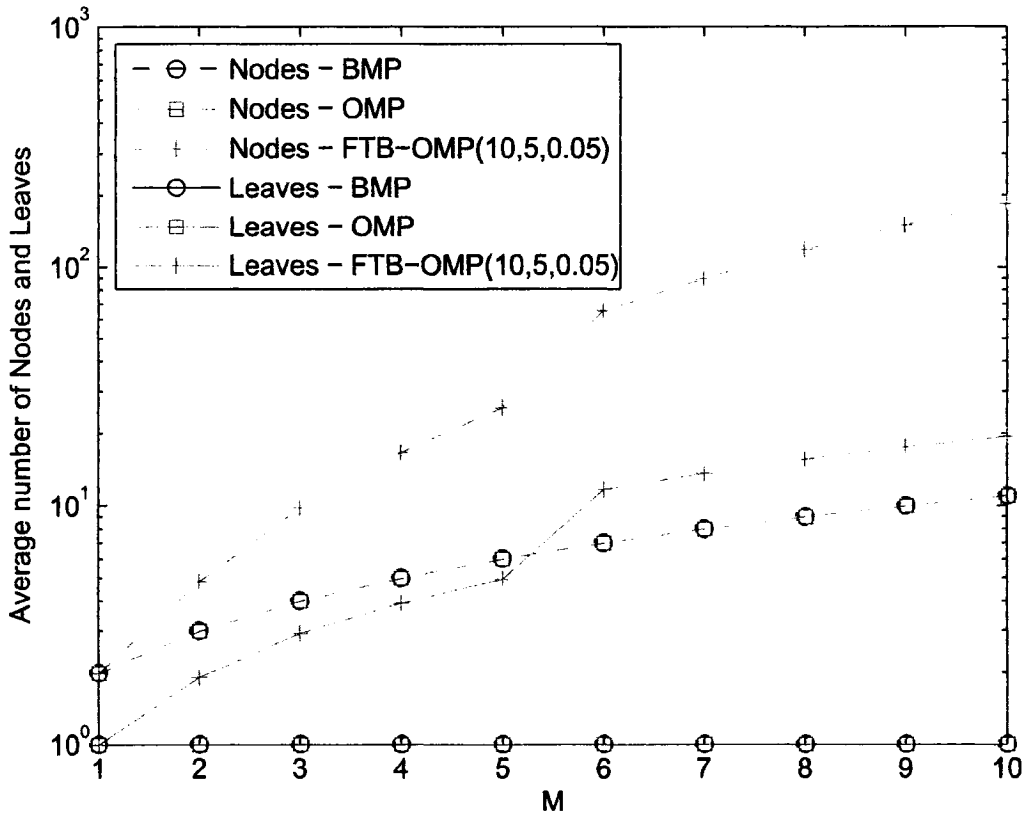


Figure 4.4: Average number of tree nodes and leaves of BMP, OMP and FTB-OMP$(10, 5, 0.05)$ algorithms for various M values, $N = 20$, $P = 15$, $\epsilon = 10^{-3}$, SNR=15 dB.

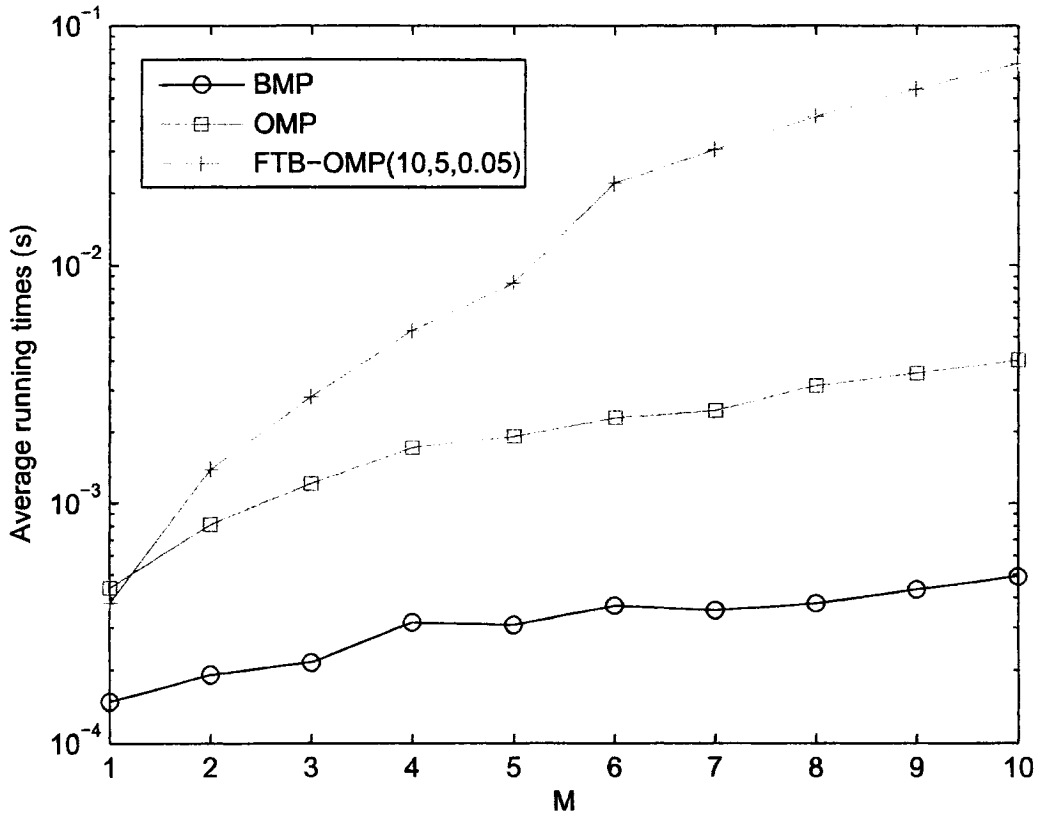


Figure 4.5: Average running times of BMP, OMP and FTB-OMP$(10, 5, 0.05)$ algorithms for various M values, $N = 20$, $P = 15$, $\epsilon = 10^{-3}$, SNR=15 dB.

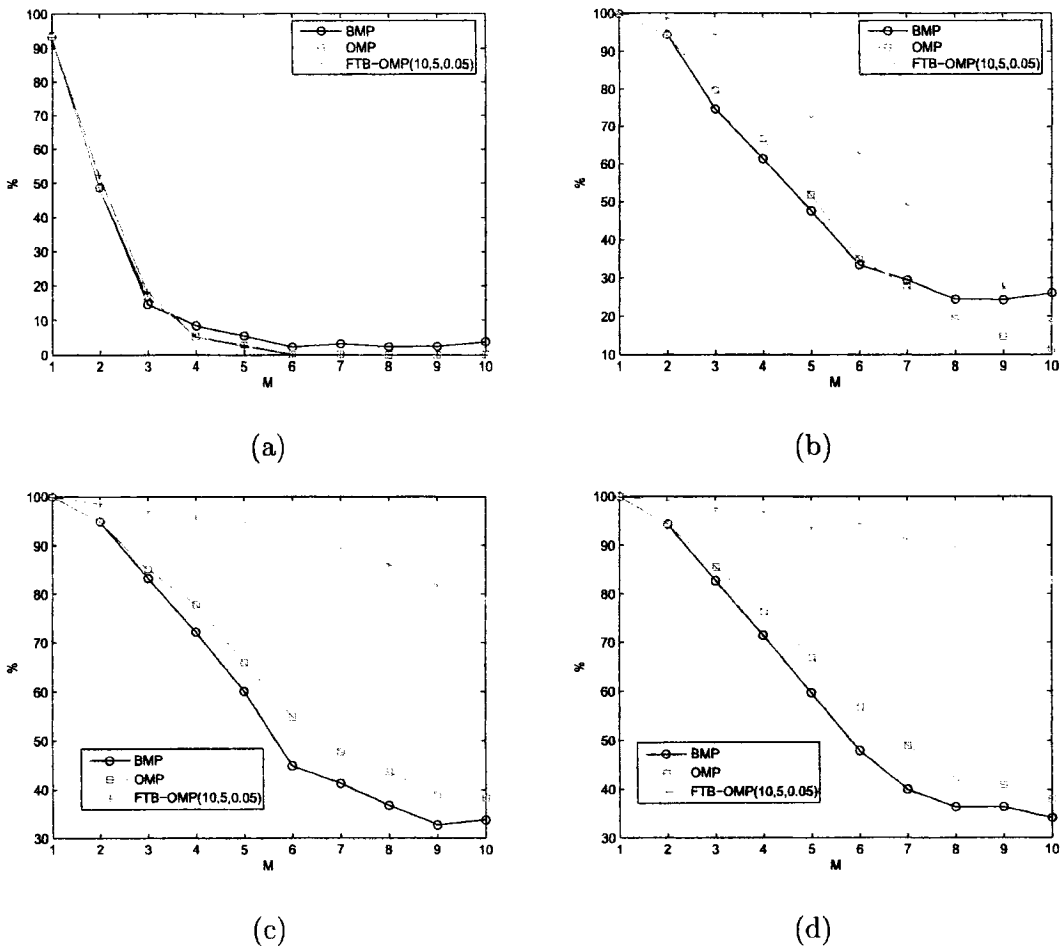


Figure 4.6: Probability of detecting all components correctly by using BMP, OMP and FTB-OMP$(10, 5, 0.05)$ algorithms, $N = 20$, $P = 15$, $\epsilon = 10^{-3}$ for (a) SNR=0 dB, (b) SNR=10 dB, (c) SNR=20 dB, (d) SNR=30 dB.

$\mathbb{C}^{(P-M)}$. Knowing that the elements of $\mathbf{h}_2$ are 0, minimum variance unbiased estimate for $\mathbf{h}_1$ can be shown to be [66]

$$\hat{\mathbf{h}} = \mathbf{D}^{\dagger}\mathbf{r} = (\mathbf{D_1}^{H}\mathbf{D_1})^{-1}\mathbf{D_1}^{H}\mathbf{r}. \tag{4.18}$$

Simulation results are compared in terms of the mean squared identification error (MSIE) identified as [69]

$$MSIE = E\left[\sum_{l=0}^{P-1}|h(l) - \hat{h}(l)|^2\right], \tag{4.19}$$

where $E[\cdot]$ represents the expected value function and $\hat{h}(l)$ represents the channel estimates. MSIE values from the minimum variance unbiased estimator given in (4.18) are shown with the legend MVUE.

From Fig. 4.7 it can be observed that FTB-CE gives the most accurate estimates converging to MVUE for SNR values larger than 10 dB. The OMP-CE outperforms the BMP-CE and LS-BMP-CE and the LS-CE. For SNR values larger than 10 dB FTB-CE also converges to MVUE, providing the most accurate estimates possible. LS-BMP-CE outperforms LS-CE up to 20 dB SNR however for higher SNR values LS-CE gives better estimates and it is of lower computational complexity.

The performances of the algorithms with respect to the sparseness is examined in Fig. 4.8 for SNR 15 dB. In the figure it is observed that the LS-CE, OMP-CE and FTB-CE are robust to such channels, however performances of the BMP-CE and LS-BMP-CE degrade severely as the number of channel components increase. The most accurate estimates are evaluated through FTB-CE.

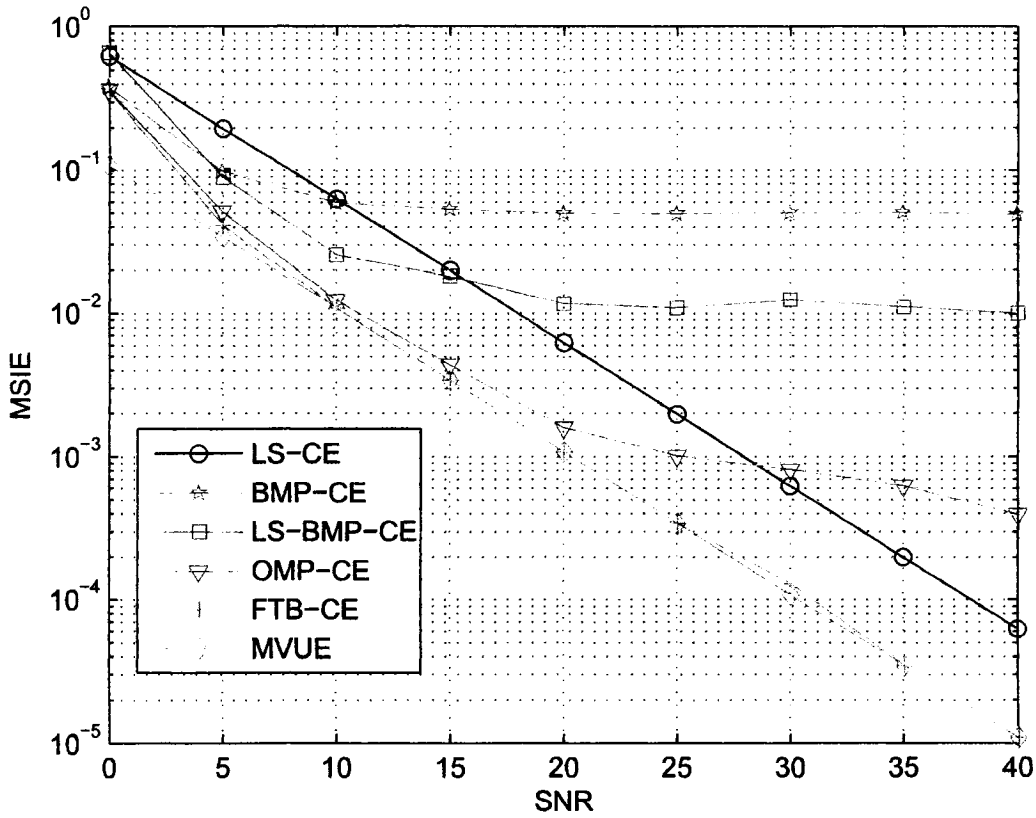


Figure 4.7: Mean squared identification error of the channel estimates over a range of values of SNR for LS-CE, BMP-CE, BMP-LS-CE, OMP-CE and FTB-CE with $(10, 5, 0.05)$, $M = 5$, $N = 50$, $P = 30$, $\epsilon = 10^{-3}$.

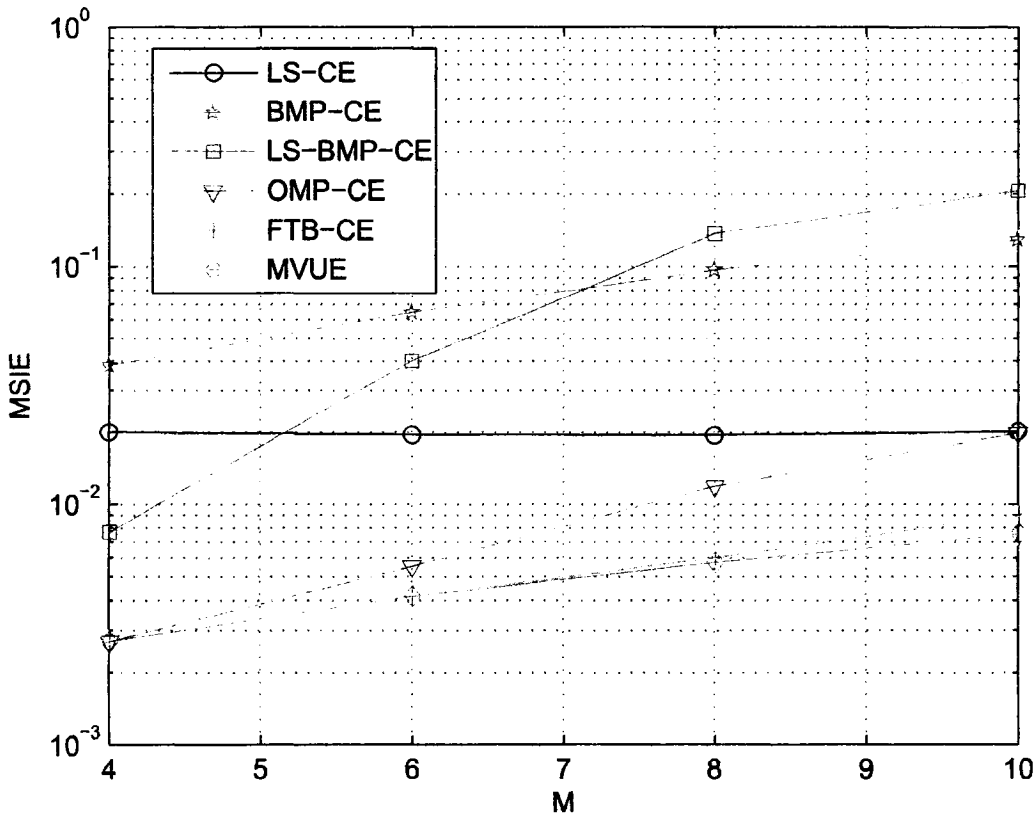


Figure 4.8: Mean squared identification error of the channel estimates over a range of values of M for LS-CE, BMP-CE, BMP-LS-CE, OMP-CE and FTB-CE with $(10, 5, 0.05)$ for SNR=15 dB, $M = 5$, $N = 50$, $P = 20$, $\epsilon = 10^{-3}$.

4.5 Conclusions

In this chapter, we proposed to use the OMP algorithm and the FTB-OMP algorithm for channel estimation. In the BMP algorithm, since each iteration optimization is performed over all vectors in the dictionary, it is possible to re-select a previously selected vector, slowing the convergence [72]. This re-selection problem is avoided in the OMP and FTB-OMP with the stored indices. It is experimentally verified that by avoiding the re-selection problem, more accurate channel estimates can be obtained by using the OMP based algorithms. It is shown by simulation results that best possible estimates can be obtained by using FTB-CE for low SNR values. For higher SNR values OMP-CE, or a FTB-CE with a search tree of reduced dimension are good options.

Furthermore, we elaborated on the training sequence design technique for the OMP algorithm for channel estimation. We investigated the limitations on the ability of OMP for detecting the channel taps. We introduced relations between correlation properties of the training sequence and algorithm's performance. The proposed relations allow designing short training sequences that can resolve all existing channel taps for sparse channels.

Chapter 5

Sparse Channel Estimation by Matching Pursuit Algorithms - Quasi-Block Fading Channels

In the previous chapter, block fading channels were considered where the dictionary $\mathcal{D}$ has a time invariant structure. In this chapter, we consider the channel estimation problem for quasi-block fading (QBF) channels where $\mathcal{D}$ is changing with time. Adaptive matching pursuit based methods and channel estimation methods with basis expansion models are considered. Simulation results are presented comparing estimation accuracy of the mentioned estimation techniques for quasi-block fading channels. A portion of the research work presented in this chapter is published in [57].

5.1 Problem Statement

Fading channels with a coherence time less than a symbol period are called quasi-block fading channels [84]. Estimation of such channels form a challenging problem. It is possible to update the channel tap coefficients hence the equalizer taps by directly using an adaptive algorithm such as recursive least squares (RLS) or least mean square (LMS) algorithms [83]. However, it is shown that tracking the channel and using these estimates for the equalizer tap evaluation yields superior performance than sole adaptation of the equalizer taps by RLS or LMS [95]. Hence here we concentrate on the channel estimation methods that track the channel.

There are several methods for detecting the time variant (TV) channel taps. Similar to the block fading case, here we elaborate on the least squares estimation and the matching pursuit (MP) based estimation techniques for TV channels.

5.2 Adaptive Matching Pursuit Based Channel Estimation

In the previous chapter, it was shown that a sampled received symbol for a block of B symbols through a TV channel can be modeled as

$$r(n) = \sum_{l=0}^{P-1} h(n;l)s(n-l) + \nu(n), \quad n = 0, 1, 2, \ldots, B-1, \tag{5.1}$$

where $h(n;l) = h_t(nT_s; lT_s)$ is the sampled channel impulse response truncated to a span of P symbol periods, $s(n)$ is the transmitted symbol and $\nu(n)$ is the noise component with variance σ^2. The sampled channel impulse response $h(n;l)$ can also be viewed as the l^{th} tap of the channel at time n.

The received symbol can be written in a matrix form as

$$\begin{bmatrix} r(0) \\ r(1) \\ \vdots \\ r(N-1) \end{bmatrix} = \begin{bmatrix} s(0) & s(-1) & \cdots & s(-P+1) \\ s(1) & s(0) & \cdots & s(-P+2) \\ \vdots & \vdots & \cdots & \vdots \\ s(N-1) & s(N-2) & \cdots & s(N-P) \end{bmatrix} \begin{bmatrix} h(0;0) \\ h(0;1) \\ \vdots \\ h(0;P-1) \end{bmatrix} + \begin{bmatrix} \nu(0) \\ \nu(1) \\ \vdots \\ \nu(N-1) \end{bmatrix}, \tag{5.2}$$

For causality purpose, $s(n) = 0$ for $n < 0$. Assuming that the estimate of $s(N)$ is correct, the received block starting with the symbol $r(1)$ can be written as

$$\begin{bmatrix} r(1) \\ r(2) \\ \vdots \\ r(N) \end{bmatrix} = \begin{bmatrix} s(1) & s(0) & \cdots & s(-P+2) \\ s(2) & s(1) & \cdots & s(-P+3) \\ \vdots & \vdots & \cdots & \vdots \\ s(N) & s(N-1) & \cdots & s(N-P+1) \end{bmatrix} \begin{bmatrix} h(1;0) \\ h(1;1) \\ \vdots \\ h(1;P-1) \end{bmatrix} + \begin{bmatrix} \nu(1) \\ \nu(2) \\ \vdots \\ \nu(N) \end{bmatrix}. \tag{5.3}$$

Let us consider the special case, channel estimation process with a training sequence and let $d(n)$ be the complex training sequence for $n = 0, 1, \ldots, N-1$, that is transmitted through a TV channel where $N \geq P$. Let $d(n)$ be placed at the front of the transmitted block. Hence we have $s(n) = d(n)$ for $n = 0, 1, \ldots, N-1$. Equation (5.2) can be written as

$$\begin{bmatrix} r(0) \\ r(1) \\ \vdots \\ r(N-1) \end{bmatrix} = \begin{bmatrix} d(0) & d(-1) & \cdots & d(-P+1) \\ d(1) & d(0) & \cdots & d(-P+2) \\ \vdots & \vdots & \cdots & \vdots \\ d(N-1) & d(N-2) & \cdots & d(N-P) \end{bmatrix} \begin{bmatrix} h(0;0) \\ h(0;1) \\ \vdots \\ h(0;P-1) \end{bmatrix} + \begin{bmatrix} \nu(0) \\ \nu(1) \\ \vdots \\ \nu(N-1) \end{bmatrix}. \tag{5.4}$$

Since $d(n)$ is known, the channel taps at time 0, $[h(0;0), h(0;1), \ldots h(0;P-1)]]'$ can be obtained by using the methods given in Chapter 4.

In the next time interval, the channel taps can be evaluated by using

$$\begin{bmatrix} r(1) \\ r(2) \\ \vdots \\ r(N) \end{bmatrix} = \begin{bmatrix} d(1) & d(0) & \cdots & d-P+2) \\ d(2) & d(1) & \cdots & d(-P+3) \\ \vdots & \vdots & \cdots & \vdots \\ \hat{s}(N) & d(N-1) & \cdots & d(N-P+1) \end{bmatrix} \begin{bmatrix} h(1;0) \\ h(1;1) \\ \vdots \\ h(1;P-1) \end{bmatrix} + \begin{bmatrix} \nu(1) \\ \nu(2) \\ \vdots \\ \nu(N) \end{bmatrix}, \tag{5.5}$$

where $\hat{s}(N)$ is the estimate of the received symbol at time N.

Equations (5.4) and (5.5) above can be abbreviated and generalized as

$$\mathbf{r} = \mathbf{D}_i \mathbf{h}_i + \mathbf{n}_i, \tag{5.6}$$

where $i = 0, 1, 2, \ldots, B - N - 1$ represents the time index.

5.2.1 Adaptive Basic Matching Pursuit

One method to estimate time varying channel coefficients is to apply the MP algorithms to (5.6) consecutively for $i = 0, 1, \ldots, B-N-1$. This corresponds to $B-N$ executions of the MP algorithm of choice. This recursive implementation is proposed by Cotter and Rao for basic matching pursuit (BMP) algorithm, and named here as adaptive BMP (ABMP) [22,24]. The channel estimation technique with ABMP is referred to as ABMP-CE throughout the text. It is shown in [21] that the ABMP-CE can adapt the changes in the channel impulse response more quickly than the LMS and the RLS, and gives more accurate channel estimates.

5.2.2 Adaptive Orthogonal Matching Pursuit

We have shown in Chapter 4 that the OMP based channel estimates are more accurate than the BMP based estimates. We now propose the adaptive OMP algorithm that uses the same

principles as ABMP. Avoiding the re-selection problem that is inherent to the BMP algorithm, the AOMP gives more accurate channel estimates as shown via simulation results in Section 5.4. This technique is referred to as AOMP-CE.

Using the fast implementation techniques proposed in [2], similar computer running times are observed for both ABMP and AOMP algorithms. However, application of FTB-OMP algorithms repetitively is not very practical due to increased computational complexity that depends on the selection of L, d and ξ parameters.

A similar improvement for the ABMP structure is proposed by an MP variation called parallel matching pursuit by Borah [6]. However, the algorithm requires weight optimization for each scenario and has no known convergence proofs.

5.3 Channel Estimation with Basis Expansion Models

A different technique to estimate quasi-block fading channel taps is to use adaptive basis expansions models for the channel. These models include a KLT based channel expansion [112], complex exponential based channel expansion [36, 106, 107], and polynomial based channel expansion [7]. They may be integrated with the MP algorithms to obtain good channel tracking performances. In the literature, the BMP-CE, and BMP-LS-CE methods are used with the polynomial based channel expansion [70, 71]. In this work, we also consider the same model in order to retain consistent results with the existing literature. It should be noted that it is straightforward to integrate any other basis expansion model to the considered channel estimation methods.

5.3.1 Polynomial Based Channel Expansion Model

It is known that the time varying channel taps $h(n;l)$ can be modeled as a basis expansion model with a priori known basis functions $f_q(n)$ and unknown coefficients $c_q(l)$ as

$$h(n;l) = \sum_{q=0}^{\infty} c_q(l) f_q(n), \quad 0 \leq n \leq B-1. \tag{5.7}$$

A finite term approximation can be written as

$$h(n;l) \approx \sum_{q=0}^{Q-1} c_q(l) f_q(n), \quad 0 \leq n \leq B-1, \tag{5.8}$$

where Q is the number of basis functions that is used in the model. As mentioned earlier, there are several sets of basis functions that model quasi-block fading channels. In the polynomial basis expansion model $f_q(n)$ is selected as the q^{th} order polynomial function and are evaluated

by orthogonalization of the set of vectors $\{\mathbf{g}_0, \mathbf{g}_1, \cdots, \mathbf{g}_{Q-1},\}$, where the n^{th} element of $\mathbf{g}_q$ is evaluated using $(n - B/2)^q$ [7].

Substituting (5.8) into (5.1), the received symbol can be expressed as

$$r(n) = \sum_{l=0}^{P-1} \sum_{q=0}^{Q-1} c_q(l) f_q(n) s(n-l) + \nu(n), \quad n = 0, 1, 2, \ldots, B-1, \tag{5.9}$$

and can be converted to vector matrix notation as [71]

$$\begin{bmatrix} r(0) \\ r(1) \\ \vdots \\ r(B-1) \end{bmatrix} = \begin{bmatrix} a_{0,0} & a_{0,1} & \cdots & a_{0,QP-1} \\ a_{1,0} & a_{1,1} & \cdots & a_{1,QP-1} \\ \vdots & \vdots & & \vdots \\ a_{B-1,0} & a_{B-1,1} & \cdots & a_{B-1,QP-1} \end{bmatrix} \begin{bmatrix} c_0(0) \\ c_1(0) \\ \vdots \\ c_{Q-1}(0) \\ c_0(1) \\ c_1(1) \\ \vdots \\ c_{Q-1}(1) \\ \vdots \\ c_0(P-1) \\ c_1(P-1) \\ \vdots \\ c_{Q-1}(P-1) \end{bmatrix} + \begin{bmatrix} \nu(0) \\ \nu(1) \\ \vdots \\ \nu(B-1) \end{bmatrix}, \tag{5.10}$$

with the coefficient $a_{i,j} = s(i-m) f_q(i)$ for $j = mQ + q$, $m = 0, 1, 2, \ldots, P-1$, $q = 0, 1, \ldots, Q$ and can be simplified as

$$\mathbf{r} = \mathbf{A}\mathbf{c} + \mathbf{n}. \tag{5.11}$$

Until now we assumed that the training sequence is the first N symbols transmitted. For model based channel estimation techniques we place no restriction on the location of

the training sequence since the location affects the performance of the channel estimation techniques extensively in quasi-block fading channels [1, 11].

Let $\mathbf{U}$ be defined as the training sample selection matrix from the received symbol vector $\mathbf{r}$. Hence we have

$$\begin{aligned} \mathbf{r}_t &= \mathbf{U}\mathbf{r} \\ &= \mathbf{U}\mathbf{A}\mathbf{c} + \mathbf{U}\mathbf{n} \\ &= \mathbf{Q}\mathbf{c} + \mathbf{U}\mathbf{n}, \end{aligned} \tag{5.12}$$

where $\mathbf{r}_t$ represents the vector composed of the received symbols from the transmitted training sequence that is used to evaluate the channel estimates as explained in the following sections.

5.3.2 Least Squares with Channel Model

The coefficients of the polynomial channel model in the least squares (LS) sense can be obtained as

$$\hat{\mathbf{c}}_s = \mathbf{Q}^{\dagger}\mathbf{r}_t. \tag{5.13}$$

Finally the channel is estimated by placing the values of $\hat{\mathbf{c}}_s$ into (5.8). This channel estimation method is referred to as P-LS-CE throughout the text. For cases when the product of the channel span and the polynomial order is larger than the length of the training sequence (i.e. $P \times Q > N$) the $\mathbf{Q}^{\dagger}$ cannot be evaluated since the matrix evaluated by $\mathbf{Q}^H\mathbf{Q}$ becomes singular. Hence model order must be kept low for P-LS-CE. This is not a limitation for the rest of the model based algorithms that are summarized in the following sections.

5.3.3 Basic Matching Pursuit with Channel Model

Application of BMP algorithm to (5.12) is proposed by Liu and Borah [70, 71]. Since $\mathbf{Q}$ is known a priori, BMP can be applied to $\mathbf{r}_t$. The dictionary is composed by using the matrix $\mathbf{Q}$ of size $N \times PK$. This method is very similar to BMP-CE however, the computational complexity is higher since for the block fading case the dictionary is of size $N \times P$. This method is named as P-BMP-CE.

5.3.4 Least Squares Matching Pursuit with Channel Model

Application of LS-BMP-CE to time varying channels is possible by applying the P-BMP-CE first, and then projecting the received sequence onto the columns of $\mathbf{Q}$ that are selected by the BMP algorithm. As mentioned earlier this method, referred here as P-LS-BMP-CE, is applied to quasi-block fading channels by Liu and Borah [70, 71].

The advantage of this method is that for channels that are not sparse, the estimates are more accurate than the estimates that are solely obtained from the BMP output [70], however the method introduces extra computational complexity when compared to P-BMP-CE.

5.3.5 Orthogonal Matching Pursuit with Channel Model

Again, a straightforward solution in order to enhance the channel estimation performance of P-BMP-CE and P-LS-BMP-CE is to consider the OMP algorithm instead of BMP when finding the channel estimates from (5.12). This method is referred to as P-OMP-CE and proposed here for the first time.

It is expected that with the faster convergence of the OMP algorithm P-OMP-CE gives more accurate channel tracking ability to the system. However in the simulation results it is

Table 5.1: Summary of the time-varying channel estimation methods

ABMP-CE	Adaptive basic matching pursuit
AOMP-CE	Adaptive orthogonal matching pursuit
P-LS-CE	Least squares based on polynomial channel model
P-BMP-CE	Basic matching pursuit based on polynomial channel model
P-LS-BMP-CE	Least squares matching pursuit on polynomial channel model
P-OMP-CE	Orthogonal matching pursuit on polynomial channel model

shown that the performance of the algorithm is fairly close to the rest of the model based estimation methods that are mentioned above. The reason for this is that the tracking ability is dominated by the channel model used but not the parameter estimation method.

5.4 Performance Comparison of Channel Estimation Algorithms for Quasi-Block Fading Channels

A summary of the channel estimation methods is given in Table 5.1. These methods are applied to fading channels changing with different fading rates. Similar to Chapter 4, simulation results are evaluated in terms of the mean squared identification error (MSIE) identified as [69]

$$MSIE = E\left[\sum_{l=0}^{P-1} |h(n;l) - \hat{h}(n;l)|^2\right], \tag{5.14}$$

where $E[\cdot]$ represents the expected value function and $\hat{h}(n;l)$ represents the channel estimates.

For channel model based estimation techniques, the training sequences are placed into the data packets as shown in Fig. 5.1 where $B = 500$, $N = 100$ with 30 symbols are placed at

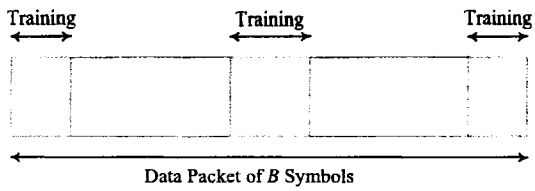


Figure 5.1: Data packet.

Table 5.2: Power profile for hilly-terrain.

Delay (Symbol)	0	1	7	8	10
Average Power (dB)	0	-1.4	-7.0	-14.8	-16.8

the beginning and the end of data packets, 40 of them are placed in the middle. For channel model based algorithms, the channel order Q is taken as 2 unless otherwise stated.

Simulation results are obtained with Monte-Carlo analysis by transmitting 1000 data blocks. For adaptive MP based estimation techniques, the training sequence is placed in the preamble of the data packet. BPSK signaling is used as the modulation technique. We assume that the channel has a maximum delay spread of $P = 11$ symbol periods. Hilly terrain (HT) type tap power profile is considered as shown in Table 5.2 [62]. In the simulation results the best possible estimates are represented by MVUE legend to act as a reference.

In the simulations the received symbol estimate at time i, $\hat{s}(i)$, is estimated directly from the received signal in order to observe the performance of the algorithms solely. Error floors in the simulation results can be reduced by applying an equalizer such as a decision feedback equalizer [86] to the system.

The MSIE values for the channel estimation techniques for a range of SNR values are shown in Fig. 5.2. Each tap amplitude is time varying with $f_dT = 0.0003$, where f_d is the Doppler frequency, and T is the bit period. From the figure it is observed that the P-LS-CE

gives the most accurate channel estimations. However, it should be noted that the P-LS-CE technique has limitations due to the channel model as mentioned earlier. Another problem for the P-LS-CE is the low sparsity properties. For channels with small number of nonzero coefficients, the LS technique cannot detect the exact tap locations hence the estimate is not sparse. A solution to this problem is thresholding the channel taps that have low powers. In the simulations, the channel taps with absolute value of the amplitudes less than 10^{-2} are assumed to be 0.

Another point that we should note is with the thresholding method, the P-LS-CE outperforms all of the MP algorithms that uses channel model. In [70, 71] the opposite claim was expressed however, the thresholding is not considered in the simulations. We should emphasize that the channel model based algorithms result in fairly close MSIE curves. The main reason behind that is the channel model used in the system. The performances of the algorithms mainly depend on the channel model since it is an approximation to the actual channel, and they have a lower dependence to the parameter estimation technique of choice.

In order to verify the tracking abilities of the estimation techniques, the fading rate is increased to $f_dT = 0.003$. The MSIE values for the algorithms of interest are plotted in Fig. 5.3. From this figure, we can see that the tracking ability of the ABMP-CE and AOMP-CE methods are superior to that of the model based techniques for high SNR values. The adaptive algorithms outperform the channel based algorithms for SNR values of 15 dB and higher. We should also note that the performance of the adaptive algorithms are almost independent from the channel fading speed. The main reason behind this is the repetitive application of the OMP or the BMP algorithm for each received symbol and hence for each time instant of the fading channel. As a result, one can apply the OMP and BMP with a lower frequency and can decrease the computational complexity of the algorithm substantially. Then, their performance would change with the fading speed. For example, consider a channel where taps are changing with every 10 symbols. OMP and BMP algorithms can be applied every

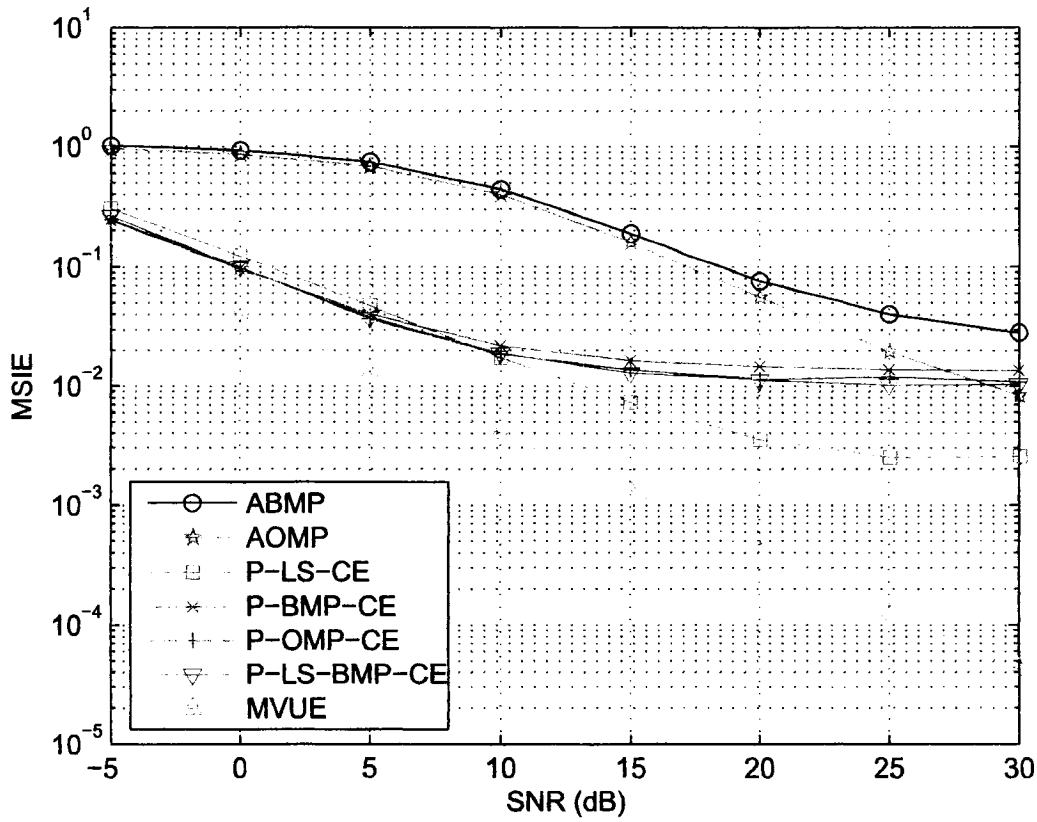


Figure 5.2: Mean squared identification error for a range of SNR values for $f_dT = 0.0003$, $M = 4$, $N = 100$, $P = 11$, $\epsilon = 10^{-3}$, $B = 500$.

10 received symbols instead of for all received symbols and channel can be tracked. This would reduce the overall complexity roughly to $1/10^{th}$ of the AOMP/ABMP cases. However if the fading rate of the channel is faster, then channel cannot be tracked by the simplified estimation techniques.

For the fading rate $f_dT = 0.003$, the channel model based algorithms give quite similar performance results and converge to an error floor for high SNR. However, for low SNR values they outperform the adaptive algorithms. This performance difference of the adaptive algorithms is due to the assumption that the sign of the received symbol is correct. For low SNR values this assumption barely holds, resulting in a poor channel tracking. The estimation performance of the adaptive algorithms can be improved by an integrated equalizer in order to increase the probability that the bit estimate is correct. However this introduces extra computational complexity.

When the ABMP-CE and AOMP-CE are compared, we can see that the latter gives more accurate channel estimates due to the inherent advantages of the OMP algorithm.

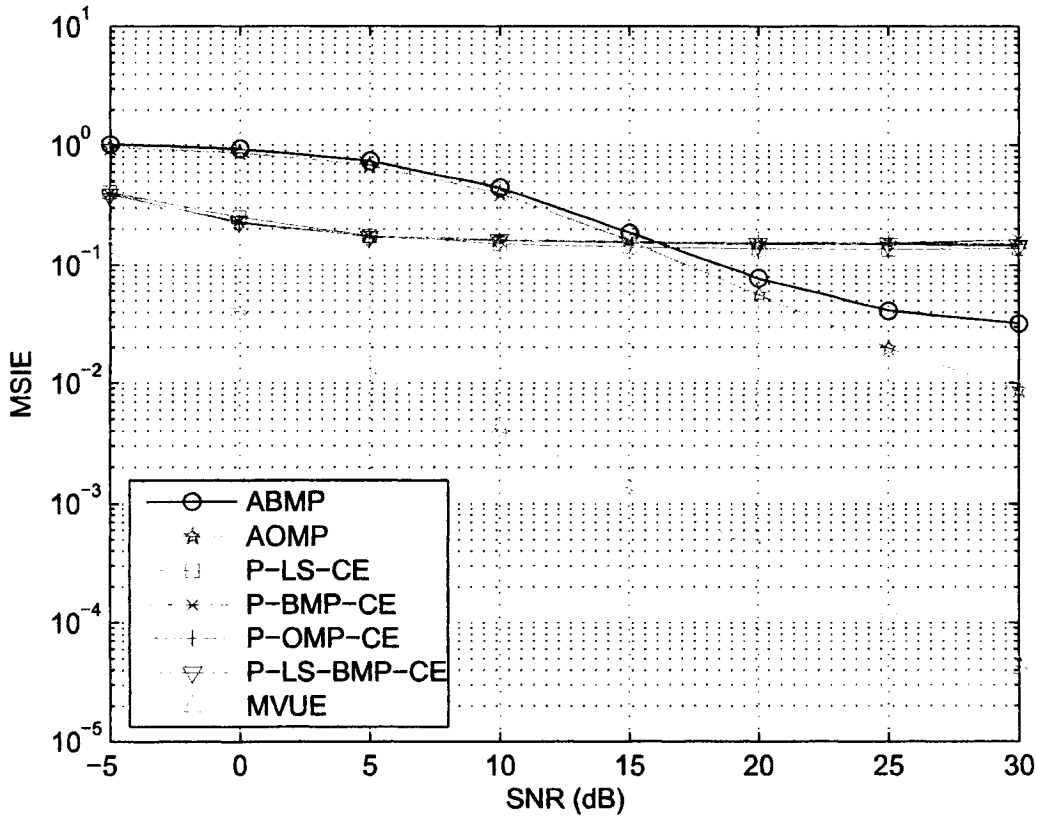


Figure 5.3: Mean squared identification error for a range of SNR values for $f_dT = 0.003$, $M = 4$, $N = 100$, $P = 11$, $\epsilon = 10^{-3}$, $B = 500$.

In order to observe the effect of channel mobility (f_dT) in more detail, simulations have been performed for varying f_dT values for a fixed SNR value of 20 dB. The performances of the channel estimation algorithms are shown in Fig. 5.4. From the figure we can see that the channel model based estimation algorithms give a detection performance that is very dependent on the channel fading speed by giving quite accurate estimates for slowly fading channels at the order of 10^{-4}. However, as the speed of the channel increases, the accuracy of the channel estimates decrease considerably. As mentioned previously, the reason for that is the dependency of the algorithms to the channel model. For the adaptive algorithm ABMP-CE and AOMP-CE, the performance is almost independent from the fading speed. This is due to the fact that both of these algorithms solve a set of equations for each time instant. If these algorithms were to apply BMP or OMP once every $H > 1$ symbols, then their performances would be affected by the fading rate. However, then they would have lower computational complexity. The error floors of both adaptive algorithms are appearing because of the existing AWGN in the system.

As mentioned above, the performances of the channel model based algorithms mainly depend on the channel model and hence the channel order Q. The effect of the channel order in the algorithms' estimation performance is summarized in terms of the MSIE Figures 5.5 and 5.6 for $f_dT = 0.0003$ and $f_dT = 0.003$, respectively. In these figures, SNR is kept at 20 dB. The MSIE of the adaptive algorithms are also plotted in order to act as a reference. It is obvious that their performances are independent from the channel model order. From the figures we can see a drawback of the model based algorithms. The channel model order must be determined with respect to the channel fading rate. Hence neither increasing the channel order to the order that the computational complexity allows nor increasing it to the order when the $P \times Q > N$ is the solution for the best estimation accuracy. Increasing Q more than necessary also causes in performance degradation. We should also note that the dependency on Q decreases as the fading speed increases.

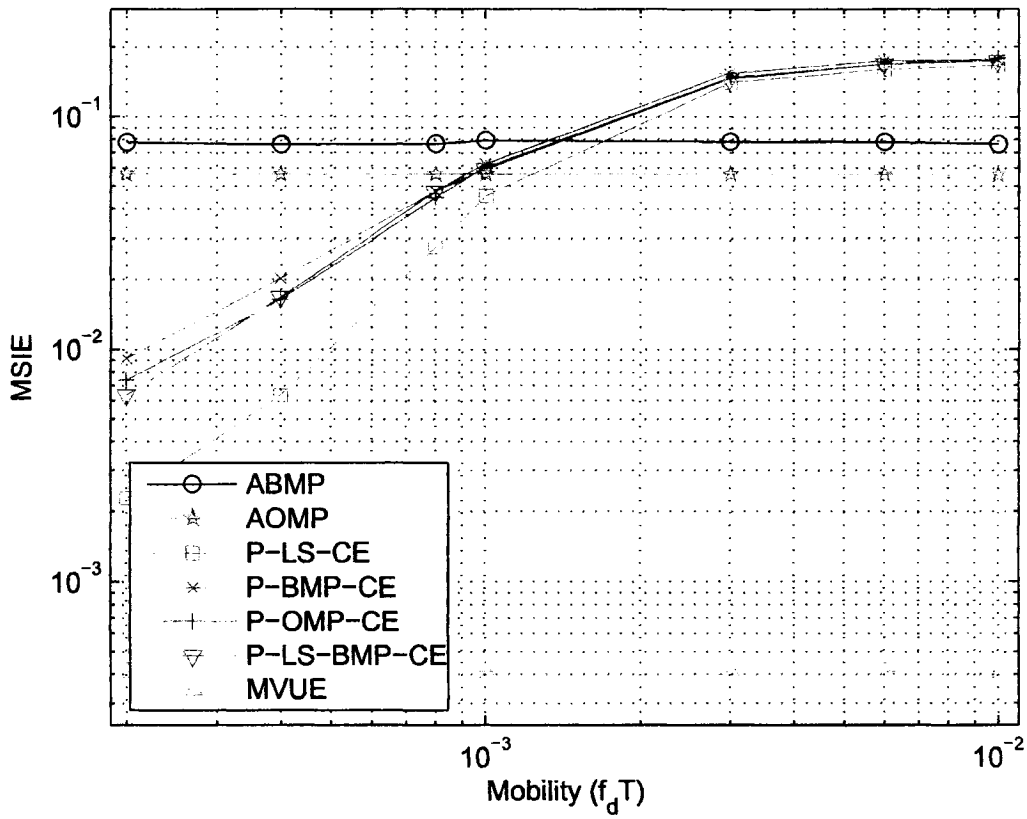


Figure 5.4: Mean squared identification error for a range of fading rates for SNR= 20 dB, $M = 4$, $N = 100$, $P = 11$, $\epsilon = 10^{-3}$, $B = 500$.

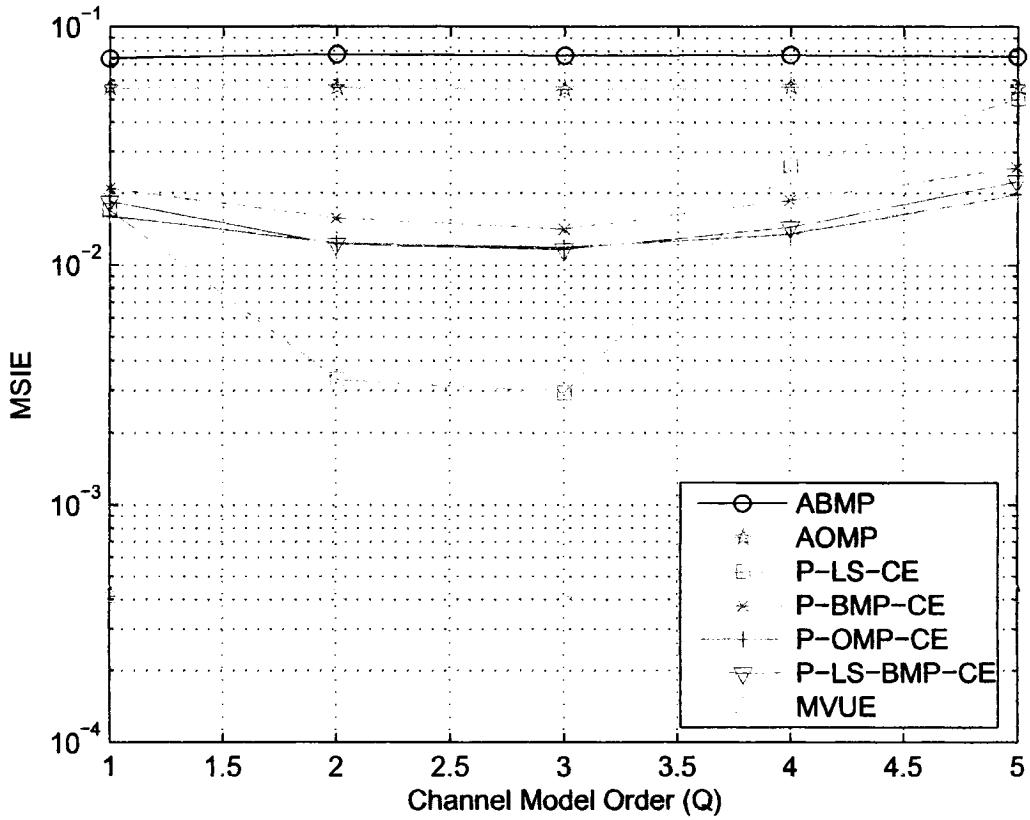


Figure 5.5: Mean squared identification error with various channel order values for SNR= 20 dB $f_dT = 0.0003$, $M = 4$, $N = 100$, $P = 11$, $\epsilon = 10^{-3}$, $B = 500$.

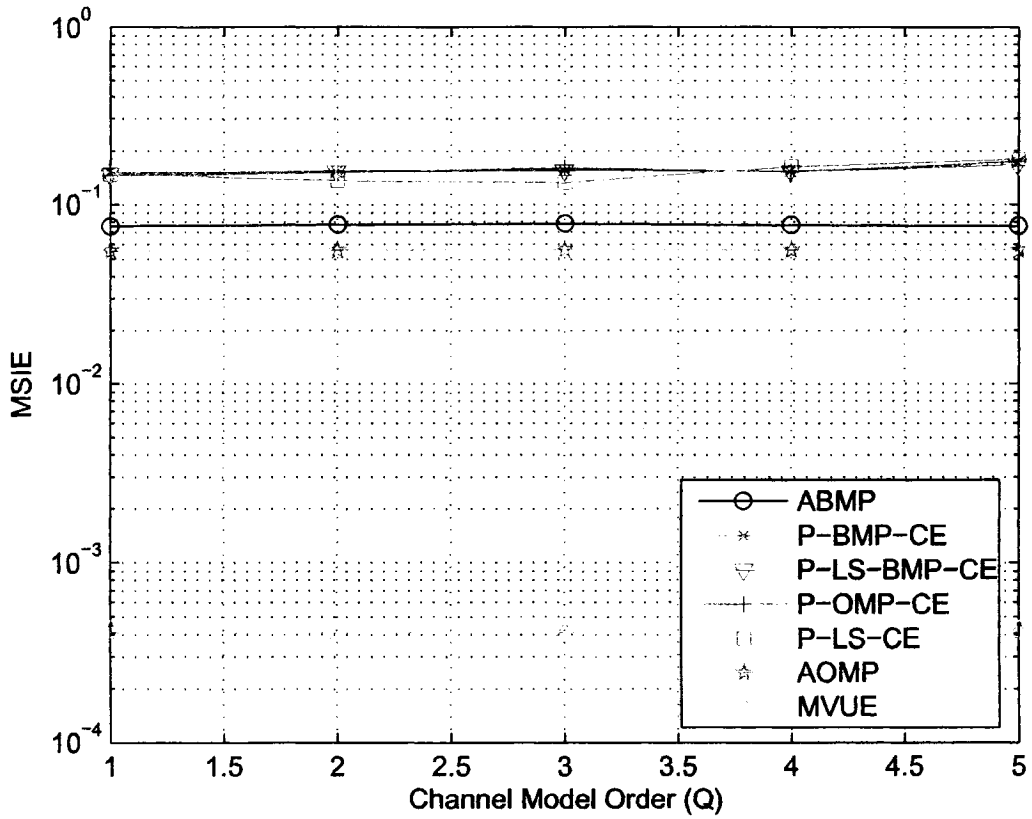


Figure 5.6: Mean squared identification error with various channel order values for SNR= 20 dB $f_dT = 0.003$, $M = 4$, $N = 100$, $P = 11$, $\epsilon = 10^{-3}$, $B = 500$.

5.5 Conclusions

In this chapter, MP based and basis expansion model based tracking methods have been considered for quasi-block fading channels. Adaptive orthogonal matching pursuit (AOMP) algorithm is introduced. Two novel methods, AOMP-CE and P-OMP-CE are described, and their performances are evaluated. It is shown that the ABMP-CE and AOMP-CE algorithms can track channels quite accurately, with the latter giving more precise estimates due to algorithmic properties. We should also note that in [70, 71] it was stated that the ABMP-CE alone is shown to perform poorly for quasi-block fading channels . However we have shown that this is not necessarily the case. Their results were based on a fading rate on the order of 10^{-4} where the channel model based techniques are shown to track the channel more accurately. However for faster fading channels this is not the case, as shown in the previous section via simulation results. This result is observed due to the fact that the accuracy of the ABMP-CE and AOMP-CE uses the MP algorithm of choice repetitively for each received symbol. Hence the process can be visualized as the solution of a set of linear equations with changing coefficients. Once the basis elements are determined, the coefficients can be determined accurately by the MP algorithms.

Their performance difference lies in the fact that the OMP algorithm can detect the basis elements more accurately as was shown in the component detection experiment results. As mentioned above, their performance can be greatly improved by removing the ISI by an equalizer. In [21] a decision feedback equalizer was used for this purpose for the ABMP-CE.

The channel model based algorithms are good for low SNR regions with not "so fast" fading channels. Comparing these algorithms, the performances of P-LS-CE, P-BMP-CE, P-LS-BMP-CE and P-OMP-CE methods are almost indistinguishable. Hence the least complex one, P-BMP-CE is a reasonable choice.

Chapter 6

Directions of Arrival Estimation by MP Algorithms

In this chapter, we introduce a novel application area for MP algorithms; the directions of arrival (DOA) estimation. Although the DOA estimation has been investigated briefly by using parallel basis selection algorithm FOCUSS [40], here it is introduced for the first time for MP algorithms with overcomplete dictionaries.

In adaptive antenna applications, there are two different reasons for the estimation of the DOA. The first one is to identify the DOA so that the appropriate beam can be utilized for each source. This selection of correct beam from a group of beams is typically encountered in switched beam antennas [103]. For this problem, the position of the DOA within a beam is not important, rather the selection of the beam is important. Hence, we refer to this problem as medium resolution problem.

The second case is typically encountered in beam-forming antennas and the coverage beam is steered to the direction of the user and a null is formed in the angle where the interferer resides. However, in most of the cases, the null is very steep and a minor estimation

error in the interferer angle may degrade the interference cancelation of the adaptive antenna. Likewise, although typically not as critical as the previous case [103], the gain for the desired source is maximum at the boresight angle and any error in the estimation of desired source DOA, decreases the boresight gain experienced by the corresponding source. Hence, in this case, the estimation accuracy is much more critical than the former case. We refer to the latter case as the high resolution problem.

Due to the above described importance of beamwidth in the context of adaptive antennas, most of the resolution results presented are normalized by the null-to-null beamwidth of the antennas.

We start the chapter by introducing the problem of DOA estimation in Section 6.1. In Section 6.2, we provide the overview of well known DOA estimators available in the literature, as well as an introduction to DOA estimation with MP algorithms.

In Section 6.3, the medium resolution DOA is investigated. Various scenarios are investigated such as uncorrelated vs. correlated sources, AWGN channels vs. Rayleigh fading channels. Also a complexity analysis is provided. In Section 6.4, the high resolution DOA is investigated under the similar scenarios of 6.3.

In the estimation problems, three different MP based algorithms are employed: BMP, OMP, and FTB-OMP. In the literature, we named the application of the FTB-OMP as EDAMP [51], which are used throughout the chapter. A portion of the research work presented in this chapter is published in [49, 51, 52, 59].

6.1 Problem Statement

In our system model for DOA estimation, we consider a narrow band adaptive antenna array of N elements. We consider the narrow-band, far-field estimation problem, and hence we

assume that the information sources are point sources, and the incoming waves are plane waves. Coherent detection is also assumed. The amplitude response of the l^{th} sensor to the i^{th} source is represented by $b_{i,l}(t)$. The output at the l^{th} sensor can then be written as [103]

$$x_l(t) = \sum_{i=1}^{M} b_{i,l}(t) e^{jw_0\tau(\theta_i)} + n(t), \quad l = 1, 2, \ldots, N, \tag{6.1}$$

where j is the complex exponent, M is the number of distinct point sources, $n(t)$ is the AWGN component with mean 0 and variance σ^2. The center frequency is w_0, and $\tau(\theta_i)$ is the time delay between the reference sensor (first sensor) and the l^{th} sensor for i^{th} source. The DOA of i^{th} source is denoted by θ_i, where $-\pi \leq \theta_i \leq 0$.

For a uniform linear array, we can construct the corresponding dictionary $\mathcal{D}$ as

$$\mathcal{D} = \frac{1}{\sqrt{N}} \begin{Bmatrix} 1 & 1 & \cdots & 1 \\ e^{j\psi_1} & e^{j\psi_2} & \cdots & e^{j\psi_P} \\ \vdots & \vdots & \vdots & \vdots \\ e^{j(N-1)\psi_1} & e^{j(N-1)\psi_2} & \cdots & e^{j(N-1)\psi_P} \end{Bmatrix}, \tag{6.2}$$

where ψ_i is the phase difference between elements of the antenna array when the signal arrives from angle θ_i, and hence $\psi_i = \frac{2\pi l}{\lambda}\cos(\theta_i)$, where $l \leq \frac{\lambda}{2}$ is the separation between antenna elements [103].

As explained above, there is a particular phase shift pattern for distinct DOAs in the antenna array output. In practice, only a few angles are received with substantial amplitudes. In the DOA estimation problem, we exploit two facts: the bijective relation between phases and DOAs, and the sparsity of the DOAs. For the case in (6.2), the possible range of DOAs is divided into P parts forming the dictionary $\mathcal{D}$. In (6.2), P is a design parameter related to minimum possible resolution. Since the number of antenna elements is limited in real life, in DOA problem $P > N$ resulting in overcomplete dictionary.

Representing the i^{th} column of $\mathcal{D}$, by $\mathbf{d}_i$, the vector matrix model for this system can be

simplified to

$$\mathbf{x} = \sum_{i=1}^{M} c_i \mathbf{d}_{k_i} + \mathbf{n}, \quad k_i \in \{1, 2, \ldots, P\}, \tag{6.3}$$

where c_i is the received signal amplitude from arriving angle θ_i, k_i is the column number of $\mathcal{D}$ corresponding to the i^{th} direction and $\mathbf{n}$ is the AWGN vector. Assuming $M \leq N$ we have

$$\mathbf{x} \approx \sum_{i=1}^{M} c_i \mathbf{d}_{k_i}. \tag{6.4}$$

Depending on the DOA, the received signal vector of size $N \times 1$ is a linear combination of the columns of $\mathcal{D}$ plus noise. Hence, detecting the DOA problem is reduced to finding correct linear combination of the columns of $\mathcal{D}$.

In the literature, different methods for achieving this goal are presented:

- The first one is the maximum likelihood (ML) approach [99]. Although it is the best one in terms of performance, it has formidable complexity. As a result, in practice, sub-optimum algorithms are proposed which generally converge to ML performance at high SNR.

- The second approach is finding the array response in the spectral domain for different angles, and recovering the local maximas as DOA [39, 108].

- The third one is the eigenstructure method. In this method the space spanned by the eigenvectors is partitioned into signal subspace and noise subspace, hence they are referred to as subspace algorithms. After partitioning, signal subspace is investigated to recover DOA. The most popular subspace algorithms are ESPRIT [86] and MUSIC [91]. These algorithms are more complex than spectral domain algorithms since they require eigenvalue decomposition. However they have performances in between ML algorithm and spectral domain algorithms. On the other hand, they have poor performances in the low SNR regions [39, 103].

Many different techniques, including independent component analysis [89], and many modified versions of these algorithms have been proposed in addition to the main ones mentioned above [10, 12, 39, 114].

6.2 Directions of Arrival Estimation Methods

In this section, frequently used DOA methods are briefly reviewed [39, 103].

6.2.1 Bartlett Beam former

Bartlett is the most basic and earliest method for DOA estimation. By steering the array in the corresponding direction, the mean power is estimated in each direction. A set of steering vectors is defined initially, and the output power from each direction is computed. It is similar to mechanically steering the array in different directions and calculating the power in those directions.

However since there are sidelobes in the array, the output power is not only a function of the main direction but the other directions also count. Hence the performance of the estimator is usually very low and it is strongly dependent on the relationship between main beamwidth and sidelobe levels.

6.2.2 Minimum Variance Distortionless Response

Minimum variance distortionless response (MVDR) algorithm finds the estimate of the power arriving from a point source in a certain direction assuming all other sources as interferences. It maximizes the output SNR while passing the desired signal undistorted.

The output power level for the MVDR algorithm can be defined as

$$\begin{aligned} P_{MVDR} &= \frac{1}{K}\sum_{k=1}^{K} |\mathbf{w}_{MVDR}^{H}\mathbf{X}_k|^2 \quad -\pi \leq \theta \leq \pi \\ &= \frac{1}{K}\sum_{k=1}^{K} |\frac{\mathbf{v}^H(\theta)\mathbf{C}_x^{-1}}{\mathbf{v}^H(\theta)\mathbf{C}_x^{-1}\mathbf{v}(\theta)}\mathbf{X}_k|^2, \quad -\pi \leq \theta \leq \pi, \end{aligned} \tag{6.5}$$

where K is the number of snapshots, $\mathbf{w}$ is the antenna element weights vector, $\mathbf{X}_k$ is the k^{th} snapshot, $\mathbf{v}$ is the array manifold vector corresponding to the response of the array from a specific arrival angle θ, and $\mathbf{C}_x$ is the snapshot spectral matrix.

In order to find the DOA, the power of the MVDR algorithms is deduced for whole range of angles, and the directions corresponding to the peaks in the power are selected as DOAs.

The method tries to adjust the array weights such that it minimizes the output power (minimum variance) by keeping the power of the desired direction at unity (distortionless). Although MVDR does not provide the resolution properties of more complex algorithms, it performs much better than Bartlett algorithm.

6.2.3 Multiple Signal Classification (MUSIC)

In principle, the eigenstructure-based methods divide the space into two subspaces; noise subspace and signal subspace. They search for the directions in the noise subspace that are orthogonal to the signal subspace.

In order identify the eigen vectors, first the power spectrum of the received signals in the frequency domain is formulated as

$$\mathbf{S}_x = \mathbf{V}(\theta)\mathbf{S}_f\mathbf{V}^H(\theta) + \sigma_w^2\mathbf{I}, \tag{6.6}$$

where $\mathbf{S}_x$, $\mathbf{V}(\theta)$, $\mathbf{S}_f$ correspond to the frequency domain snapshots of the received signal, array manifold vector, source signals, respectively. $\mathbf{I}$ is the identity matrix and σ_w^2 is the noise variance.

Next, eigen distribution of $\mathbf{S}_x$ is expressed as

$$\mathbf{S}_x = \mathbf{\Phi}_i \mathbf{\Lambda} \mathbf{\Phi}_i^H, \tag{6.7}$$

where $\mathbf{\Lambda}$ is the diagonal matrix containing the eigen values and columns of $\mathbf{\Phi}_i$ correspond to eigen vectors. Then the eigen vectors are grouped as D signal eigen vectors,

$$\mathbf{U}_S = [\mathbf{\Phi}_1 \vdots \mathbf{\Phi}_2 \vdots \cdots \vdots \mathbf{\Phi}_D], \tag{6.8}$$

and $N - D$ noise eigen vectors,

$$\mathbf{U}_N = [\mathbf{\Phi}_{D+1} \vdots \mathbf{\Phi}_{D+2} \vdots \cdots \vdots \mathbf{\Phi}_N]. \tag{6.9}$$

First, a weight vector located in the noise subspace is searched for which is orthogonal to the signal subspace. Then, the directions that are orthogonal to this vector are investigated by projecting different DOA to the noise subspace $\mathbf{U}_N$ as

$$\mathbf{Q} = \mathbf{v}^H \mathbf{U}_N \mathbf{U}_N^H \mathbf{v}. \tag{6.10}$$

Next D minimum points of the projection $\mathbf{Q}$ are selected as direction of arrivals. This method requires an estimate on number of distinct DOAs.

Also the opposite of the method can be employed, where the steering vectors that are included in the signal subspace are searched. This time the maximum points of the projection are selected. The selection between two different methods depends on the size of the subspaces, the one with smaller size is generally preferred.

6.2.4 Estimation of Signal Parameters via Rotational Invariance Techniques (ESPRIT)

ESPRIT is both computationally efficient and a very accurate method when compared to MUSIC. It employs two imaginary arrays in the array elements as shown in the Fig. 6.1.

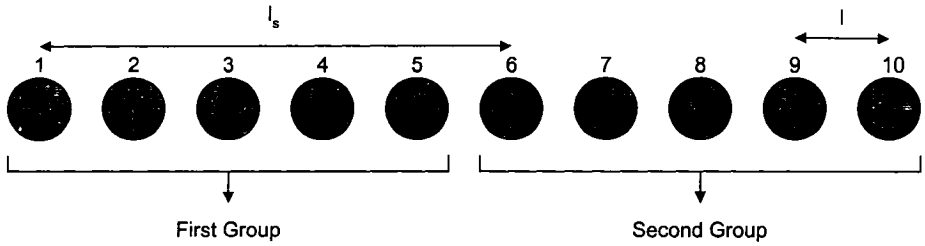


Figure 6.1: Subarrays for ESPRIT: First five elements of the original array form the first subarray, and last five elements of the original array form the second subarray. Here, l is the difference between two adjacent elements and l_s is the difference between the first elements of each group.

The DOA estimations are acquired by investigating the constant phase shift between two arrays. Since the phase shift is a function of the DOA, it is unique for every DOA. The estimation process becomes straightforward and computationally efficient. The main steps of the ESPRIT algorithm can be summarized as follows.

- Find $\mathbf{U}_S$ as in the case of MUSIC.
- Calculate the selection matrices for the two shifts $\mathbf{J}_1 = [\mathbf{I}\vdots\mathbf{0}_{N\times l_s}]$ and $\mathbf{J}_2 = [\mathbf{0}_{N\times l_s}\vdots\mathbf{I}]$.
- Calculate $\mathbf{U}_{S1} = \mathbf{J}_1\mathbf{U}_S$ and $\mathbf{U}_{S2} = \mathbf{J}_2\mathbf{U}_S$.
- Find the least square solution for $\mathbf{\Psi}_{LS}$ that minimizes the value of the expression $\|\mathbf{U}_{S2} - \mathbf{U}_{S1}\mathbf{\Psi}\|$.
- Compute the D different eigen values λ_i of $\mathbf{\Psi}_{LS}$.
- Finally, the angles of arrivals are given by $\theta_i = \frac{1}{l_s}(arg(\lambda_i))$.

6.2.5 Direction of Arrival Estimation by Matching Pursuit Algorithms

When the signal arrives from an individual angle only, the problem is straight-forward and algorithm chooses the column of $\mathcal{D}$, which has the maximum inner product with the received vector $\mathbf{x}$. However when the signal arrives from more than one angle, $\mathbf{x}$ is a linear combination of columns of $\mathcal{D}$ and trying every possible linear combination would give the ML solution. On the other hand, this would bring formidable complexity to the system.

Application of a basis selection algorithm in order to estimate the DOA's modeled in (6.4) would decrease the computational complexity. In this section three variations of this method are introduced. These are

1. Application of BMP algorithm.
2. Application of OMP algorithm.
3. Application of FTB-OMP algorithm (EDAMP).

Among these algorithms a high resolution can be obtained by FTB-OMP algorithm due to its performance as presented in Chapter 3. By employing the FTB-OMP algorithm a heuristic approximation to ML solution is obtained.

FTB-OMP algorithm selects the columns of $\mathcal{D}$ which are estimated to form $\mathbf{x}$, and these columns correspond to the DOA. FTB-OMP also gives the coefficients of these columns, which represent the amplitudes of the corresponding DOA.

There are four main advantages of the application of FTB-OMP:

1. It does not require the number of directions to be estimated. By comparing the amplitude in $\mathbf{x}$ and amplitude of the resolved signals defined by the space spanned by

the columns of $\mathcal{D}$, which have already been chosen by the algorithm, it is capable of deciding whether all the components are resolved or not. Considering that most of the spectral and subspace algorithms require the number of directions as an input, this is a very important advantage.

2. The algorithm allows flexibility between complexity and resolution property. By increasing the search depth, a closer solution to ML can be achieved, by decreasing the search depth algorithm running time can be decreased. But for both cases, it is computationally advantageous to the subspace based algorithms, since it works on spectral domain and does not require eigenvalue decomposition.

3. By comparing the power in $\mathbf{x}$ and power of the resolved signals defined by the space spanned by the columns of $\mathcal{D}$, the algorithm can decide whether all the components are resolved or not. Hence the selection of the threshold value, ϵ, becomes important. By selecting ϵ, receiver decides the percentage of the power that should be resolved. Generally, ϵ is taken much less than one percent.

4. In FTB-OMP, not the signal subspaces but the amplitudes of the received signals are used. As a result, system performance is robust to correlation between the inputs from different angles.

In the following section we support these advantages by simulation results. Due to the advantages of the FTB-OMP algorithm, it can resolve much more closely located DOA when compared to BMP and OMP algorithms. Hence the simulation results are presented in two sections: medium resolution, and high resolution. Here medium resolution corresponds to the case where the DOAs are apart from each other at least with a distance of the size of the main lobe beam. High resolution is the case when they can be inside the same main lobe. Medium resolution may correspond to a practical case of switched beam antenna whereas high resolution may correspond to adaptive beamforming [103].

In the medium resolution part, BMP and OMP algorithms are compared with well known algorithms in the literature in terms of resolution performance. Next in the high resolution section, all of these algorithm are compared with FTB-OMP for very closely positioned DOA. In the literature there exists more complicated versions of MUSIC and ESPRIT which perform better than traditional versions. However, the MP based detection procedures are more computationally efficient when compared to even traditional MUSIC and ESPRIT. Hence we limit ourselves with these versions.

6.3 Medium Resolution Simulation Results

In the simulations we consider a $N = 10$ element uniform linear array (ULA) that has element separation of $\lambda/2$, as shown in Fig. 6.2. The selection of separation of $\lambda/2$ optimal in terms of grating lobes [103]. For the dictionary $\mathcal{D}$, we considered $P = 201$. The selection of P determines maximum resolution for a single beam. Selecting a larger and more dense dictionary would result in better resolution accuracy. Selecting $P = 201$ results in an asymptotic root mean square error (RMSE) performance of -22.5 dB [103]. The asymptotic behavior can be observed in the following RMSE simulation results. Many other dictionary selections can be made, however the selected accuracy is sufficient for almost all practical systems. Furthermore, it is a sufficient number to show the convergence to the Cramér-Rao bound.

The SNR values correspond to the signal to noise ratios at the input of each antenna element and they are assumed to be the same. The noise at each element is assumed to be independent identically distributed additive white Gaussian noise (AWGN). Since we consider the SNR at each antenna element, the overall system SNR is much higher. For the special case of uncorrelated fading at individual antennas, the received total SNR from the antenna system is N times that of the SNR at individual antennas. Hence the low SNR per antenna element cases considered in the simulation results correspond to possible practical overall SNR

cases.

Let us define $u = \cos(\theta)$. Unless stated otherwise, two different signal directions with $u_1 = 0.433$ ($\theta_1 = 64.34$) and $u_2 = -0.433$ ($\theta_1 = 115.66$), 10 times the maximum resolution [103], are considered. The powers in both directions are assumed to be the same. The angles are chosen wide apart, in order to represent the low SNR performance more clearly. Depending on the application, the two signals may represent the signals of a multiple input multiple output (MIMO) channel, a desired signal and an interferer or even signals from two different users. 100 independent snapshots are used for the subspace based algorithms for the convergence of the eigenvalues. The results are averaged over 1000 Monte Carlo simulations.

6.3.1 Cramér-Rao Bound

In our simulations, we also compare the results with Cramér-Rao Bound (CRB). CRB provides an upper-limit on the resolution performance of the DOA estimation [103]. The CRB for the DOA problem, $\mathbf{B}_{CR}$ can be given as

$$\mathbf{B}_{CR} = \frac{\sigma_w^2}{2K}\{\Re\{[\mathbf{S}_f[(\mathbf{I} + \mathbf{D}_a^H\mathbf{D}_a\frac{\mathbf{S}_f}{\sigma_w^2})^{-1}(\mathbf{D}_a^H\mathbf{D}_a\frac{\mathbf{S}_f}{\sigma_w^2})]] \odot \mathbf{G}^T\}\}^{-1}, \tag{6.11}$$

where, $\mathbf{S}_f$ is the spatial spectral response of different DOAs, $\mathbf{I}$ is the identity matrix, $\mathbf{D}_a$ is a matrix formed by combining the columns of $\mathbf{D}$ that correspond to the DOAs, $|.|^H$ corresponds to Hermitian transform, $\odot$ is the Hadamard product, and $\mathbf{G}$ is defined by

$$\mathbf{G} = (\frac{d\mathbf{D}_a}{d\theta})^H[\mathbf{I} - \mathbf{D}_a(\mathbf{D}_a^H\mathbf{D}_a)^{-1}\mathbf{D}_a^H]\frac{d\mathbf{D}_a}{d\theta}, \tag{6.12}$$

where $d\mathbf{D}_a/d\theta$ corresponds to the derivative of $\mathbf{D}_a$ with respect to the DOAs.

Other than the proposed BMP and OMP algorithms as described in the previous section, Bartlett [108], MUSIC [91] and ESPRIT [86] algorithms are also considered. These algorithms are simulated with the parameters stated above and all of the results presented in this

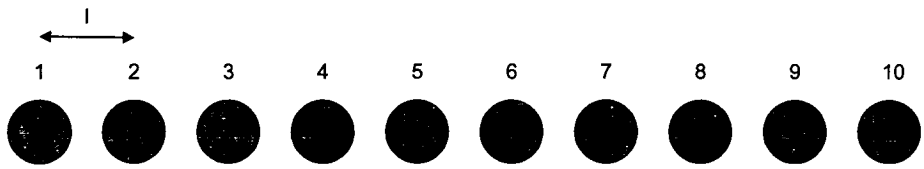


Figure 6.2: Array structure of ULA.

work about these algorithms are completely calibrated with the results on their performances presented in the literature prior to this work [39, 103].

For MUSIC, the parameters described in [91, 103] are employed for 10 antenna elements. For the ESPRIT algorithm, the antenna array is divided into two subarrays, one being shifted version of the other in space. The constant phase shift between two subarrays gives the resolution. For the simulations, 5 element shifted ESPRIT is considered as shown in Fig. 6.1.

6.3.2 Uncorrelated Inputs

We first look at the case when signals arriving from different angles are uncorrelated. In Fig. 6.3, the proposed BMP and OMP algorithms are compared with the algorithms mentioned above, in terms of probability of resolution. We define the probability of resolution as finding the estimates $\hat{u}_1$ and $\hat{u}_2$ such that

$$|\hat{u}_1 - u_1| \leq 0.0433 \text{ and } |\hat{u}_2 - u_2| \leq 0.0433,$$

which is the resolution limit of 10 element antenna array [103].

As it can be seen in Fig. 6.3, the OMP performs well, especially in the low SNR region; and the probability of resolution decreases linearly with decreasing SNR. Also, Fig. 6.3 shows that OMP converges to ML in low SNR regime as well.

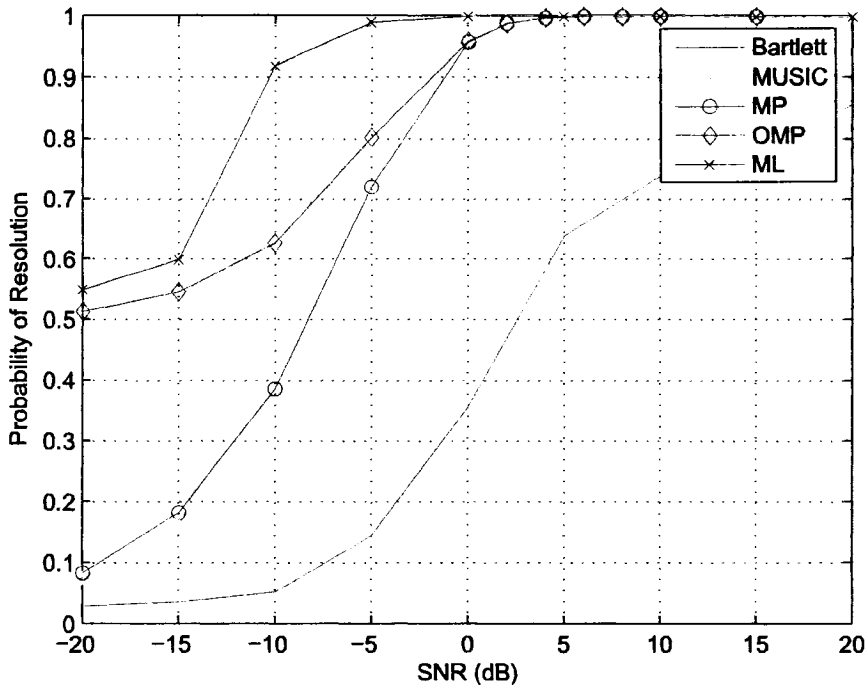


Figure 6.3: Probability of resolution vs. SNR for two uncorrelated inputs that are 50 degrees apart, 100 snapshots, $P = 201$, $M = 2$, $\epsilon = 10^{-10}$.

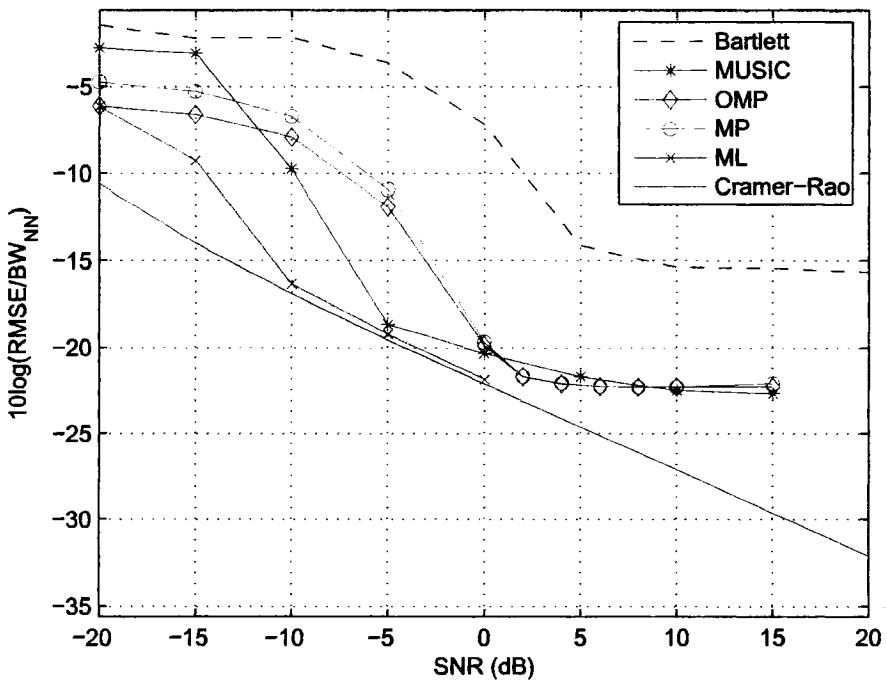


Figure 6.4: RMSE of DOA normalized by null-to-null beamwidth versus SNR per antenna element. Two uncorrelated inputs that are 50 degrees apart, 100 snapshots, $P = 201$, $M = 2$, $\epsilon = 10^{-10}$.

In Fig. 6.4, the RMSE in the estimated angles is shown. RMSE is normalized by the null-to-null beamwidth (BW_{nn}) of the 10 element antenna array. As it is seen in Fig. 6.4, the BMP and OMP algorithms have superior performance especially at low SNR values.The convergence of ML solution to the CRB is also depicted in Fig. 6.4. It is also shown that both MUSIC, BMP and OMP algorithms converge to the CRB, however their asymptotic performance is limited due to element imperfections and angular quantization at the dictionary level, as it is seen in Fig. 6.4.

6.3.3 Correlated Inputs

In this section, we investigate the effect of correlation of the input signals on the system performance. The correlation may be a result of different aspects of the communication channel such as multi-paths from a single source. The performance of subspace algorithms, namely MUSIC and ESPRIT are highly dependent on the correlation between input signals arriving from different angles [39, 86, 91, 103]. Performances of these subspace algorithms deteriorates with increasing correlation values of the inputs. This is a natural outcome of subspace algorithms which make use of eigenspace decomposition in order to separate noise, signal and interference.

The performances of BMP and OMP algorithms are independent of correlation in the signals, since their resolving capability depends solely on the powers received in different directions. This is supported by the results of Fig. 6.5 and 6.6. Even for 90% correlation, the performances of BMP and OMP are the same as its performance with uncorrelated inputs. However, as depicted in Fig. 6.5 and 6.6, the performance of MUSIC is degraded with increased correlation. In this case only BMP and OMP algorithms converge to the CRB. Although not included in this work, simulation results show that BMP and OMP algorithms have the same performance even with 100% correlated inputs.

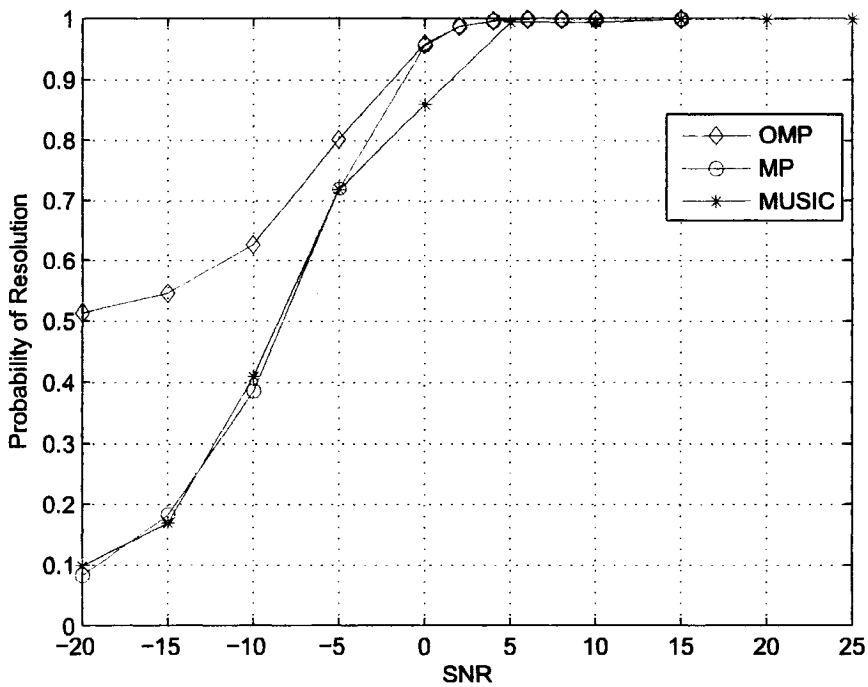


Figure 6.5: Probability of resolution vs. SNR for two 90% correlated inputs that are 50 degrees apart, 100 snapshots, $P = 201$, $M = 2$, $\epsilon = 10^{-10}$.

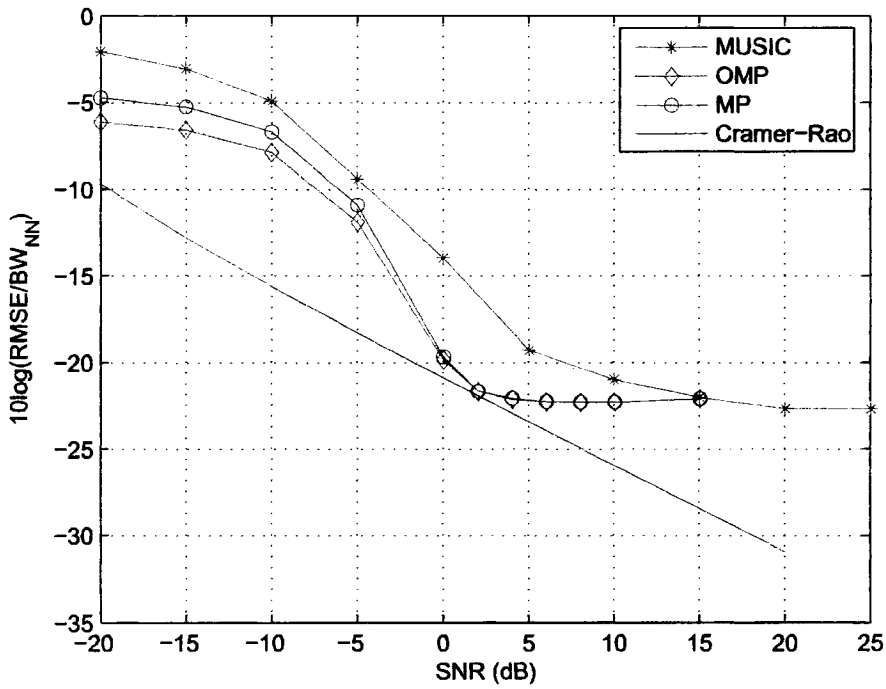


Figure 6.6: RMSE of DOA normalized by null-to-null beamwidth for two 90% correlated inputs that are 50 degrees apart, 100 snapshots, $P = 201$, $M = 2$, $\epsilon = 10^{-10}$.

6.3.4 Comparison with ESPRIT

In order to compare BMP and OMP algorithms with ESPRIT, the angular separation must be reduced due to the resolution limitations of ESPRIT. In ESPRIT the angles to be separated must be in the interval [103]

$$-1/l_s < u < 1/l_s, \tag{6.13}$$

where l_s is the number of antenna elements between the first element of the first group, and the first element of the second group. In our ESPRIT scenario since $l_s = 5$, our limitation becomes $-0.2 < u < 0.2$. Hence, for the simulations from this point on, we update u values as $u_1 = 0.15$ (81.4^o) and $u_2 = -0.15$ (98.6^o).

In Fig. 6.7 it is clearly depicted that both BMP and OMP perform much better than ESPRIT in terms of RMSE. It is seen that for highly correlated signals the BS based resolution performance is superior to subspace algorithms such as MUSIC and ESPRIT.

6.3.5 Performance with Rayleigh Fading

Up to now, we have investigated the performances of the algorithms under the assumption that the average signal strengths from the two directions are equal and constant. Now we investigate the effect of Rayleigh fading channel. We assume that the amplitudes of the signals from two directions are independent and Rayleigh distributed. As shown in Fig. 6.8, although there is a performance loss compared to non-fading case, for fading channels OMP outperforms ESPRIT and MUSIC as well.

Simulation results show that OMP algorithm clearly outperforms MP algorithm for all the scenarios that are investigated. Using the fast implementation techniques proposed in [2], comparable computer running times are observed for both BMP and OMP algorithms. In the BMP algorithm, since each iteration optimization is performed over all vectors in the dic-

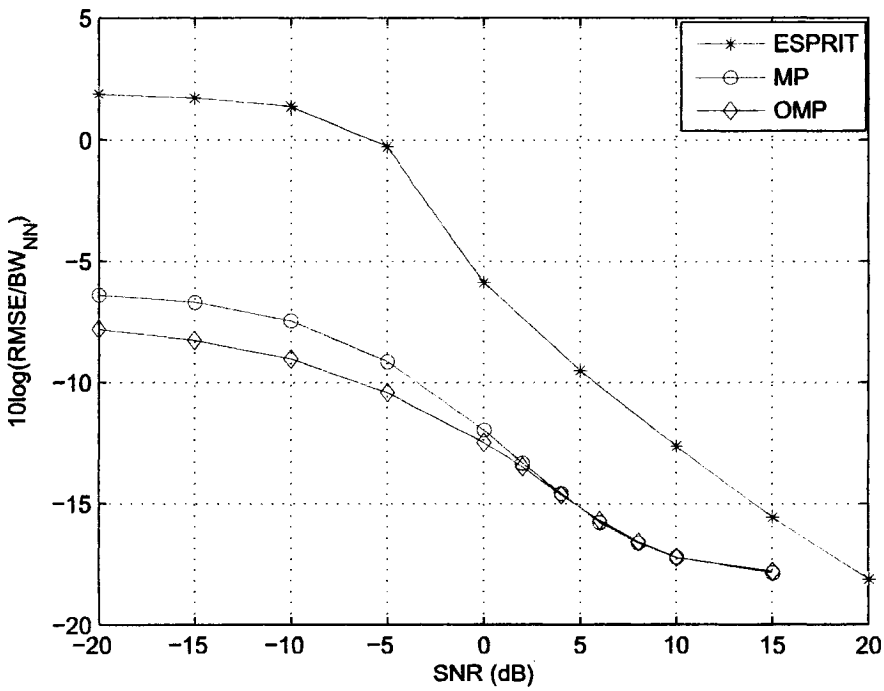


Figure 6.7: RMSE of DOA normalized by null-to-null beamwidth for two 90% correlated inputs that are 17 degrees apart, 100 snapshots, $P = 201$, $M = 2$, $\epsilon = 10^{-10}$.

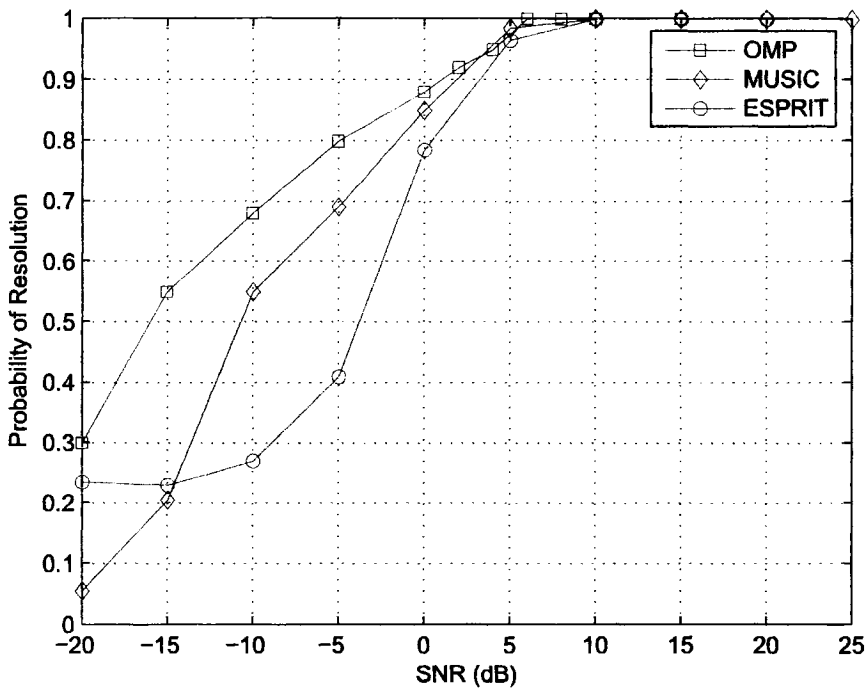


Figure 6.8: Probability of resolution for two 90% correlated inputs for fading channel coefficients, 100 snapshots, $P = 201$, $M = 2$, $\epsilon = 10^{-10}$.

tionary, it is possible to re-select a previously selected vector, slowing the convergence [72]. This re-selection problem is avoided in the OMP algorithm with the stored dictionary. Experimental results show that the OMP algorithm is a better candidate for the DOA estimation problem. However, we should also note that the orthogonalization process causes OMP to have slightly higher computational complexity than BMP.

The resolution properties of ESPRIT depend critically on the number of samples used in the algorithm. Next we investigate the effect of the number of samples taken from the channel to the probability of resolution. As depicted in Fig. 6.9 for SNR=3 dB, the performance of ESPRIT is very low for low number of snapshots, however with OMP even observing a single snapshot from the channel gives reasonable DOA estimates.

Up to this point, we assumed angles that are very well separated from each other. Next we investigate the probability of resolution for various angular separations. As can be seen from Fig. 6.10, after increasing the separation more than a threshold, the performance of the ESPRIT algorithm quickly becomes worse. This is due to the limitations of ESPRIT stated in (6.13). The performance of ESPRIT for closely separated angles surpasses that of OMP, at the cost of increased complexity. Next we compare the complexities of the aforementioned algorithms.

6.3.6 Computational Complexity Analysis

In order compare the complexity of the basis selection algorithms with the ones in the literature, the number of floating point operations (flops) required by each algorithm is investigated. The number of flops are computed at an SNR of 10 dB. Simulation results indicate that for other SNR values of interest, the number of flops are close to the ones in 10 dB within 1%. As it can be seen in Fig. 6.11, both BMP and OMP have marginally higher computational complexities than the simple Bartlett beamformer. However both perform much less

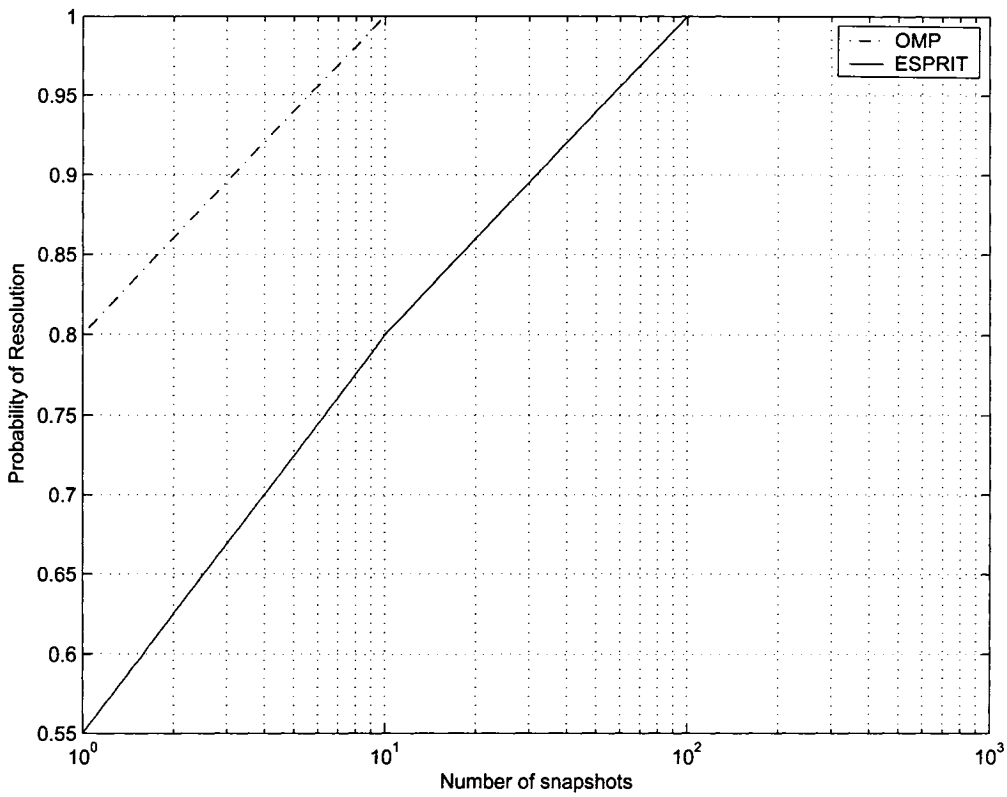


Figure 6.9: Probability of resolution vs. number of iterations for two 90% correlated inputs for SNR=3 dB in Rayleigh fading channel.

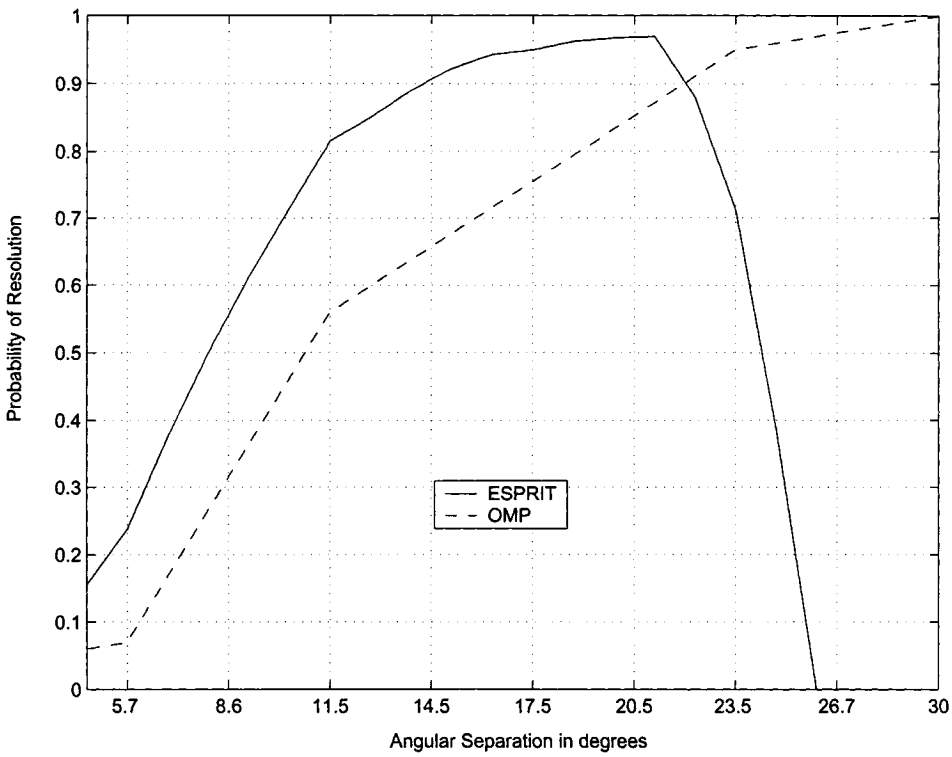


Figure 6.10: Probability of resolution vs. angular separation for 90% correlated inputs for SNR=3 dB, 100 snapshots, $P = 201$, $M = 2$, $\epsilon = 10^{-10}$ in Rayleigh fading channel.

operations when compared to MVDR, MUSIC and ESPRIT. BMP requires approximately one percent more flops than Bartlett, and OMP requires one percent more than BMP. This is a very important result, because the Bartlett beamformer, which is of comparable complexity, has a very insufficient performance for most of the practical systems.

Because of the well known high complexity and graphical presentation purposes, the number of flops for ML case is not included in Fig. 6.11. However simulation results show that the number of flops required by ML is approximately two orders of magnitude higher than MUSIC. These results show that the proposed algorithms are much more computationally efficient than the conventional DOA estimation algorithms.

The presented complexity analysis is important from a different aspect as well. Although sub-space based algorithms lose performance with increased input correlation, there exist methods to improve their performance under correlation. However, these algorithms further increase the complexity, as a result reducing the probability of wide-scale development. On the other hand, BMP and OMP based algorithms do not require additional complexity in terms of protection against the input correlation level.

6.4 High-Resolution Simulation Results

In high-resolution simulations we consider a 10 element uniform linear array (ULA) that has element separation of $\lambda/2$ as shown in Fig. 6.12. The SNR values correspond to the signal to noise ratios at the input of each antenna element and they are assumed to be the same. However the noise at each element is assumed to be independent identically distributed (iid) additive white Gaussian noise (AWGN). The system SNR is much higher than the SNR at each element. Hence, low SNR results presented in this work are of practical interest as well.

Unless stated otherwise, two different signal directions with $u_1 = 0.0433$ and $u_2 =$

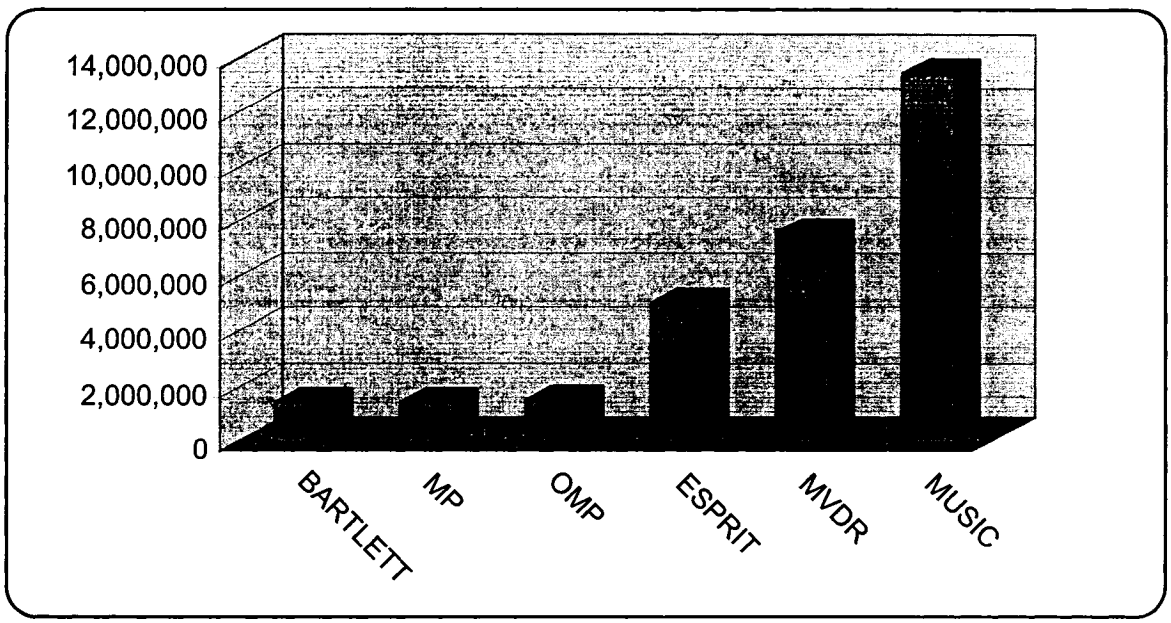


Figure 6.11: Comparison of number of floating point operations required for each algorithm for two 90% correlated inputs, with 201 column dictionary, $M = 2$, $\epsilon = 10^{-10}$.

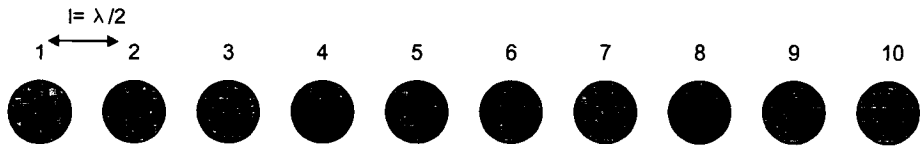


Figure 6.12: Array structure of simulated ULA.

-0.0433 (the minimum distance that can be resolved for a 10 element ULA [103]) are considered. The amplitudes in both directions are assumed to be the same. These u values correspond to 87.52^o and 92.48^o degrees. Hence the minimum resolvable angle is 4.96^o. As shown in Fig. 6.13, the range of estimation is in between 0^o and 180^o .

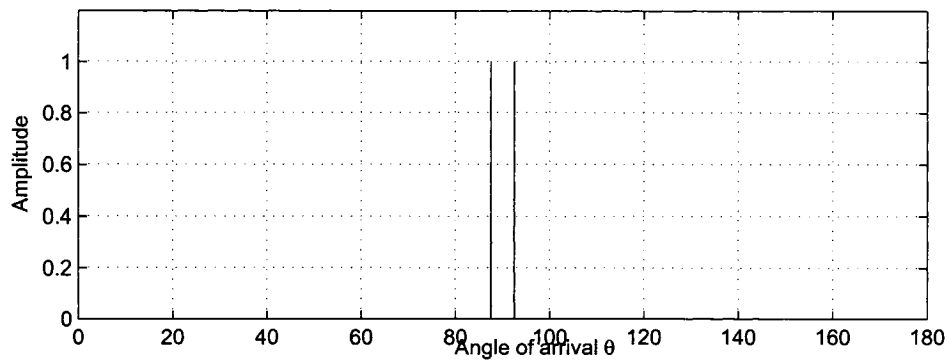


Figure 6.13: Arrival angles $\theta_1 = 87.52^o, \theta_2 = 92.48^o$

In the subspace based algorithms, for the convergence of the eigenvalues, 100 independent snapshots are used. The results are averaged over 1000 Monte Carlo simulations.

Other than the proposed EDAMP algorithm as described in the previous section, Bartlett [108], MVDR [103], MUSIC [91] and ESPRIT [86] algorithms have also been considered. These algorithms have been simulated with the parameters defined above, and all of the results presented in this work about these algorithms have been calibrated with the results on

their performances presented in the literature prior to this work [39, 103].

In Table 6.1, the parameters used for FTB-OMP algorithm employed in the simulations are given. With these parameters, EDAMP requires much less computational time when compared to ESPRIT and MUSIC. In terms of floating point operations in MATLAB simulation platform, EDAMP requires approximately half the number of flops required by ESPRIT, and one fourth the number of flops required by MUSIC.

Table 6.1: Parameters of FTB-OMP algorithm used in EDAMP simulations

Parameter	Value
Tree-pruning (ξ)	0.25
Number of branches (L)	100
Decaying parameter (d)	10
Maximum iteration (r)	3

6.4.1 Uncorrelated Inputs

We first look at the case when the signals arriving from different angles are uncorrelated. In Fig. 6.14, the novel FTB-OMP algorithm is compared with all four algorithms mentioned above. As can be seen in Fig. 6.14, EDAMP performs well especially in the low SNR region and the probability of resolution increases linearly with SNR. For uncorrelated channels at low SNR, EDAMP outperforms every other algorithm, and at high SNR, ESPRIT performs the best.

In Fig. 6.15, RMSE in the estimated angles is shown. RMSE is normalized by the null-to-null beamwidth (BW_{nn}) of the 10 element antenna array. As it is seen in Fig. 6.15, at low SNR EDAMP outperforms ESPRIT and at high SNR, ESPRIT is better in terms of RMSE performance.

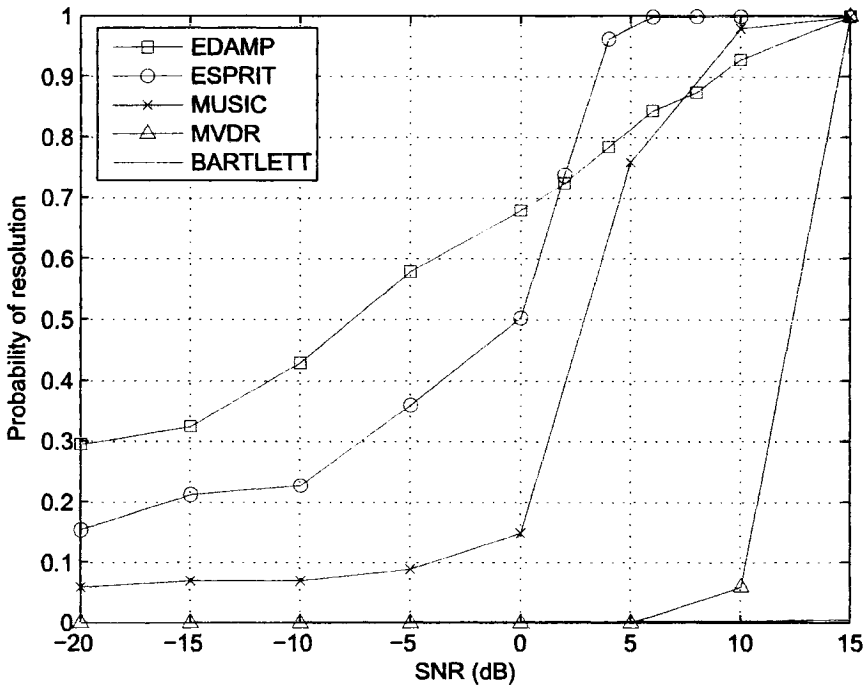


Figure 6.14: Probability of resolution vs. SNR for uncorrelated inputs, 100 snapshots, $P = 201$, $M = 2$, $\epsilon = 10^{-10}$.

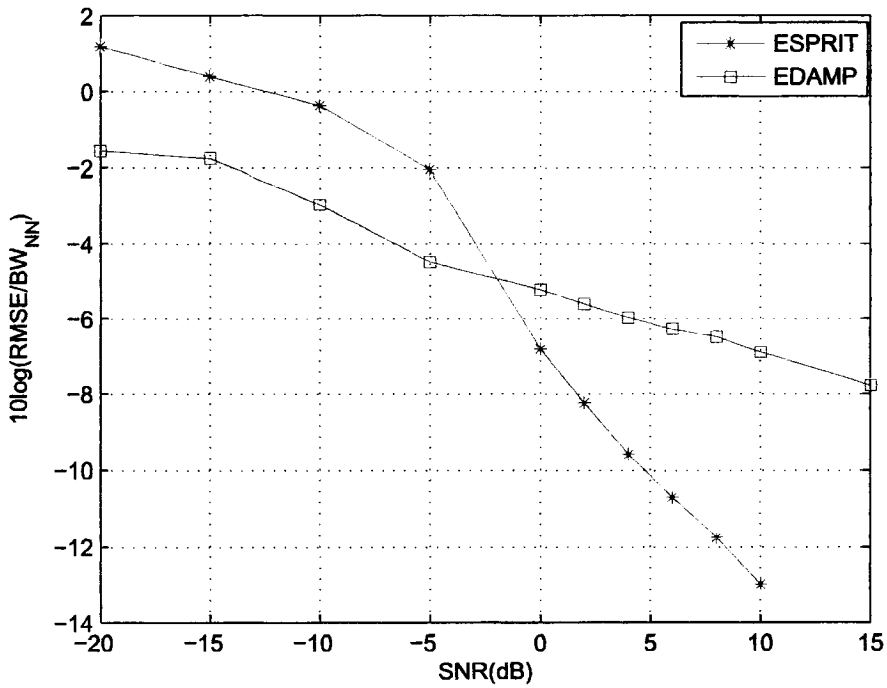


Figure 6.15: RMSE of DOA normalized by null-to-null beamwidth for uncorrelated inputs, 100 snapshots, $P = 201$, $M = 2$, $\epsilon = 10^{-10}$.

Next, the effect of angular separation on the probability of resolution is investigated. In Fig. 6.16, it is depicted that for SNR=3 dB, EDAMP can resolve more closely separated signals when compared to ESPRIT. Also in Fig. 6.16, we can see another limitation of ESPRIT. In ESPRIT algorithm, the antenna array is divided into two symmetric subarrays. The resolution property is highly dependent on the distance between the first element of the first array and first element of the second array, which is denoted by l_s [103]. The ESPRIT scheme that we employ in our simulations is the one with highest resolution available for a 10 element antenna array [103]. However, in ESPRIT algorithm, the resolvable angles are limited by the relation

$$-\frac{1}{l_s} < u < \frac{1}{l_s}. \tag{6.14}$$

For the scheme employed $l_s = 5$. Since

$$-\frac{1}{5} < u < \frac{1}{5}, \tag{6.15}$$

the largest value of Δu, for resolution is $\frac{1}{5} + \frac{1}{5} = 0.4$. It is clearly seen that for $u > 0.4$ which approximately corresponds to 46 degrees, the performance of ESPRIT degrades very fast. On the other hand, EDAMP has no such limitation. One could select an ESPRIT scheme with smaller l_s hence increasing the resolvable range, but this would result in lower probability of resolution and worse RMSE in the resolvable range [86, 103].

6.4.2 Correlated Inputs

Above we considered the case when two signals arriving from different angles were uncorrelated. Here, we investigate the effect of correlation on the system performance. The performance of subspace algorithms, namely MUSIC and ESPRIT are highly dependent on the correlation between input signals arriving from different angles [39, 86, 91, 103]. This is a natural outcome of subspace algorithms making use of eigenspace decomposition in order to separate noise, signal and interference.

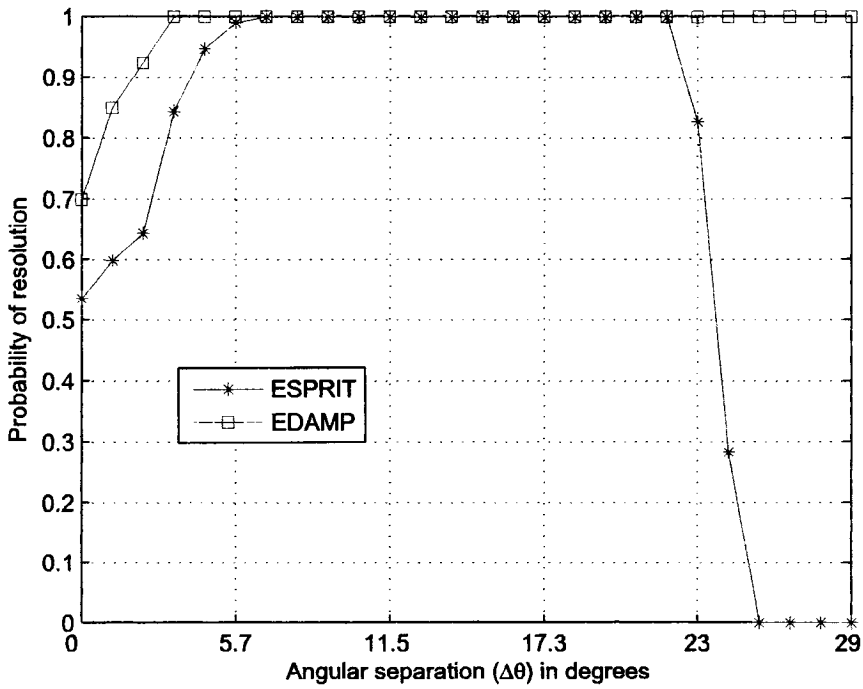


Figure 6.16: Probability of resolution vs. angular separation for 2 uncorrelated inputs for SNR=3 dB , 100 snapshots, $P = 201$, $M = 2$, $\epsilon = 10^{-10}$.

On the other hand, the performance of EDAMP is independent of correlation in the signals, since for a constant dictionary size and angular separation, its resolving power depends solely on the amplitudes in different directions. This is supported by the results of Fig. 6.17 and 6.18. Even for 90% correlation, the performance of EDAMP is the same as its performance with uncorrelated channels. However, as shown in Fig. 6.17 and 6.18, the performances of MUSIC and ESPRIT are severely degraded with increased correlation.

It is seen that for highly correlated signals EDAMP resolution performance is much better than subspace algorithms such as MUSIC and ESPRIT.

6.4.3 Effect of Number of Snapshots

In wireless communications, especially for real-time applications, delays in the system are very critical. In DOA estimation, a number of snapshots is required for the estimation to be accurate [39]. When the number of snapshots increases, the processing delay in the system increases. It is well known that with insufficient number of snapshots, traditional DOA algorithms perform poorly. In EDAMP, snapshots are only utilized for running the algorithm again and averaging the estimations. For known signals, the snapshots can be utilized to increase the SNR by averaging the signals from different snapshots. The number of snapshots, therefore, is not very critical as in the case of subspace algorithms. Here we investigate the effect of number of snapshots by decreasing it from 100 to 10, and the effect of number of snapshots when the SNR is 15 dB.

In Figures 6.19, 6.20 and 6.21, it is clearly depicted that EDAMP performs much better for low number of snapshots. Even at 10 snapshots, EDAMP shows acceptable performance, which makes EDAMP even more valuable for applications requiring short delays.

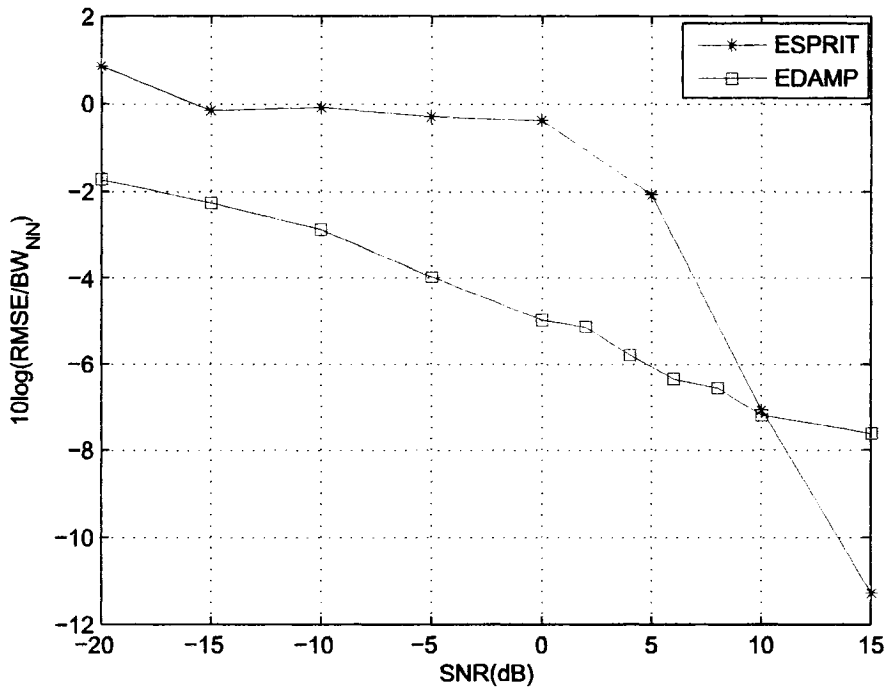


Figure 6.17: RMSE of DOA normalized by null-to-null beamwidth for 90% correlated inputs, 100 snapshots, $P = 201$, $M = 2$, $\epsilon = 10^{-10}$.

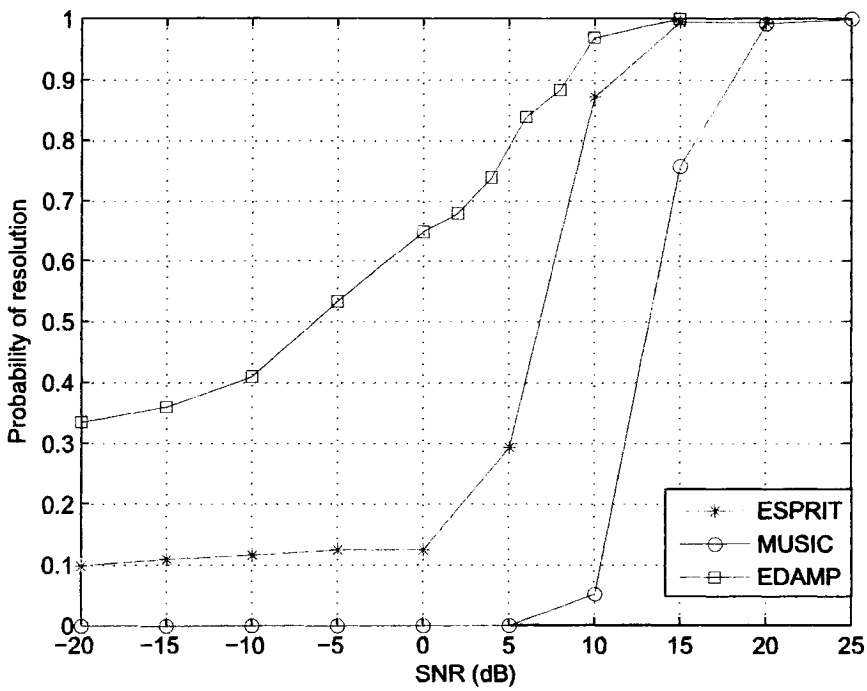


Figure 6.18: Probability of resolution vs. SNR for 90% correlated inputs, 100 snapshots, $P = 201$, $M = 2$, $\epsilon = 10^{-10}$.

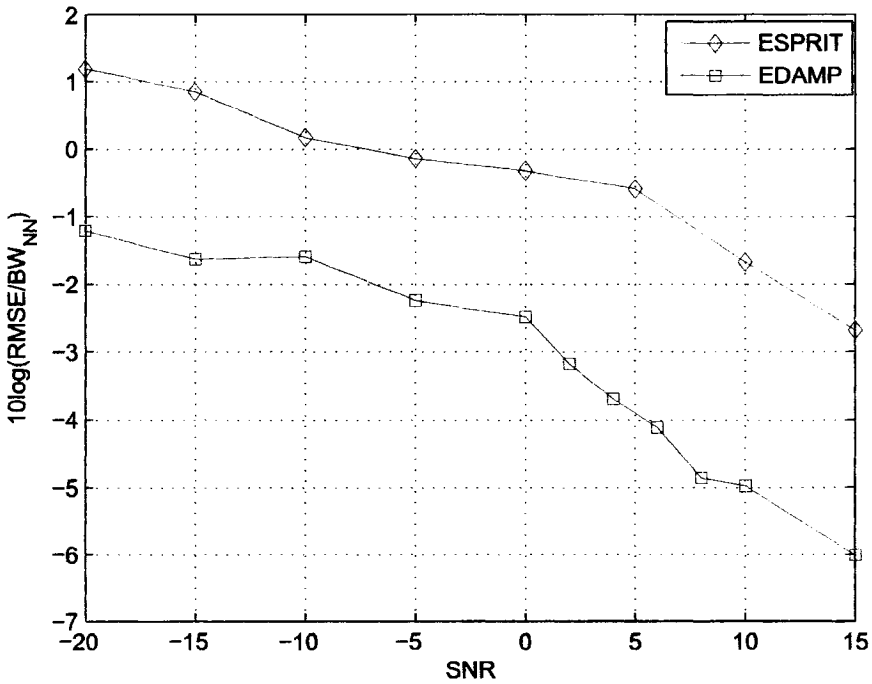


Figure 6.19: RMSE of DOA normalized by null-to-null beamwidth for 90% correlated inputs with 10 snapshots, $P = 201$, $M = 2$, $\epsilon = 10^{-10}$.

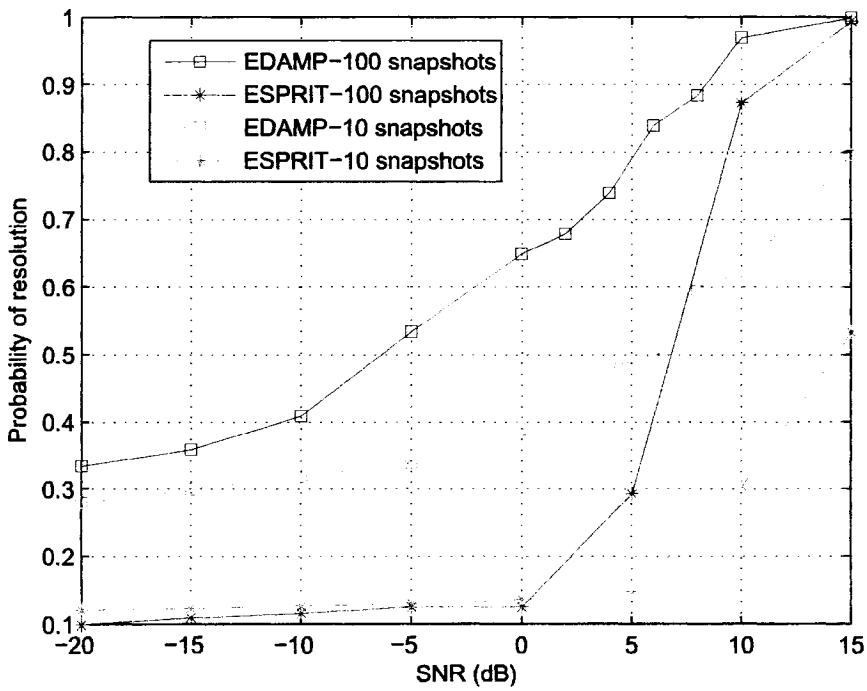


Figure 6.20: Comparison of probabilities of resolution of 90% correlated inputs for 10 and 100 snapshots, $P = 201$, $M = 2$, $\epsilon = 10^{-10}$.

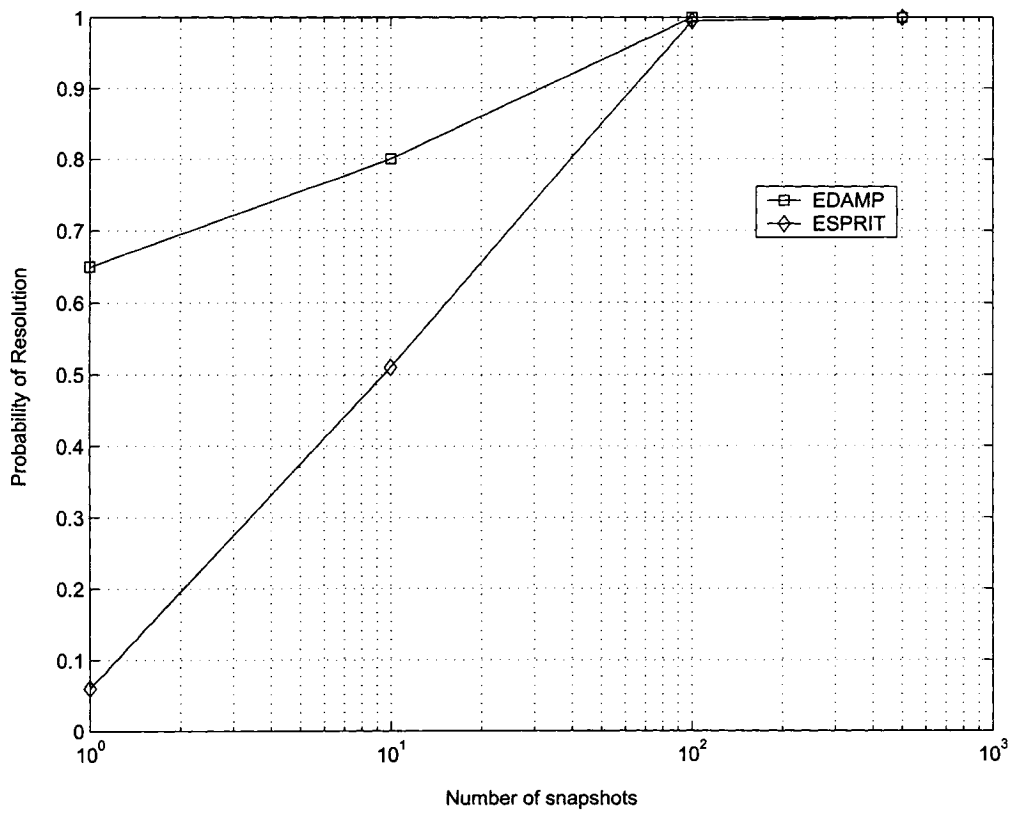


Figure 6.21: Probability of resolution vs. number of snapshots for 90% correlated inputs with SNR=15 dB, $P = 201$, $M = 2$, $\epsilon = 10^{-10}$.

6.5 Conclusions

In this chapter, we modeled DOA estimation problem so that MP algorithms can be applied for detection of the angles. We presented a novel DOA estimator, EDAMP, which employs a basis selection algorithm, namely FTB-OMP. Many advantages of EDAMP when compared to the traditional algorithms are presented, which can be summarized as follows:

The EDAMP algorithm gives directions of arrival and their corresponding amplitudes as output, so it does not require post-processing to detect amplitudes after detecting directions. At the same time, the algorithm does not need pre-processing since it does not require the number of DOA as input.

EDAMP is not affected by the correlations in the signals from different DOA, hence it is expected to perform better in multi-path situations when compared to traditional techniques.

Since it is a heuristic approach to ML solution, it gives good resolution properties even at low SNR situations. It also requires very few snapshots, when compared to subspace algorithms, thus decreasing processing time.

Chapter 7

Multi-User Detection by Matching Pursuit Algorithms

The amount of radio spectrum that can be allocated to a communication system is a limited resource. Using multiple access schemes many users simultaneously share this limited spectrum. For efficient communications, the bandwidth allocation to multiple users must be performed without causing severe degradation in the system performance. There are three major multiple access (MA) schemes, namely frequency division multiple access (FDMA), time division multiple access (TDMA) and code division multiple access (CDMA). Details about MA schemes can be found in [84].

CDMA has been approved as the standard in 3^{rd} generation (3G) mobile communication systems[1]. Although this technology is already in use, a strong demand still exists for improvements to respond to the increasing use of mobile communication devices including the cellular phones and wireless LANs.

In CDMA systems all users share the same frequency band at all times. In order to differ-

[1]3G specifications are available at `www.3gpp.org`.

entiate between different users, each user is assigned a code. The process of differentiating between users is referred to as multi-user detection problem.

In this chapter, we present MP based multi-user detectors for CDMA systems with different problem definitions including complete dictionaries. We point out the similarities between the BMP and basic successive interference canceler (SIC). Using the structural advantages of OMP over BMP, we introduce an OMP based detector. We verify via simulation results that the OMP algorithm actually converges to the performance of linear decorrelator detector. A practical implementation for the successive interference canceler is presented by integrating the basic matching pursuit algorithm to the detector. Furthermore, we present a blind single user detector that combats the degrading effects of the near-far problem. A portion of the research work presented in this chapter is published in [53].

7.1 Problem Statement

In a CDMA system, the binary signals of multiple users are modulated by spreading codes assigned to each user, also named as the signature sequence, and these modulated sequences are transmitted to a base station. The transmitted waveform would be this signature multiplied by the data stream. Considering uplink communications, the receiver is the base station and has to detect all mobile stations' information. In downlink communications the receiver is the mobile station that needs to detect only the related data. The signatures are chosen to minimize their cross correlations in order to keep the interference caused by other users low [110].

Originally, a matched filter receiver was proposed for CDMA detection. This was based on the fact that in the case of a large number of users are active in the system, their interference can be considered to be Gaussian by the central limit theorem, and hence, the optimum

receiver is the single user receiver. Although this matched filter receiver is the simplest in its implementation and only requires the signature of the required user, it suffers a large performance degradation, especially when an interfering user has larger transmitting power than the desired user due to its proximity to the base station. This problem is named as the near-far effect.

The base station receives a mixture of the modulated sequences and some noise. A detector at the base station extracts the original binary signals from the received signals using the knowledge of the users' spreading codes. This process is named as the multi-user detection (MUD). By simultaneously detecting multiple user signals, a multi-user detector suppresses mutual interference and can provide much better detection performance than a matched filter receiver [110].

The received signal in this M user system, can be written as

$$r(t) = \sum_{m=1}^{M} A_m b_m s_m(t) + n(t), \quad t \in [0, T], \tag{7.1}$$

where, T is the symbol period, $n(t)$ is the white Gaussian noise. For the m^{th} user where $m = 1, 2, \ldots, M$, A_m is the received amplitude, $b_m \in \{+1, -1\}$ is the transmitted bit, $s_m(t)$ is the signature waveform with unit energy[2]. For simplicity of analysis we consider a chip synchronous system. Hence it is assumed that, when $t \notin [0, T]$, $s_m(t) = 0$ for eliminating intersymbol interference.

The similarity between signature waveforms affects the performance of the system [110]. This similarity can be quantified by the crosscorrelation between sequences which is defined as

$$\rho_{ij} = \langle s_i, s_j \rangle = \int_0^T s_i(t) s_j(t) dt. \tag{7.2}$$

[2] $\|s_m(t)\| = \int_0^T |s_m(t)|^2 dt = 1$.

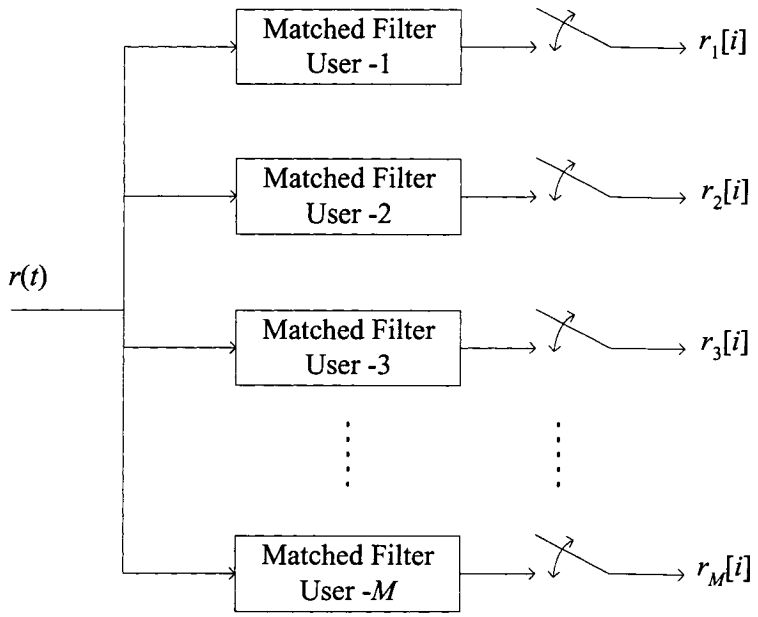


Figure 7.1: M-user matched filter outputs for synchronous system.

For MUD case, where detection of all active users' signals are of interest, a bank of matched filters can be used for converting the received continuous waveform into discrete-time waveform. The model for a synchronous M user CDMA system is shown in Fig. 7.1. Each filter is matched to the signature waveform of the user to be detected. The filter-bank outputs can be represented as

$$\begin{aligned} r_1 &= \int_0^T y(t)s_1(t)dt \\ &\vdots \\ r_M &= \int_0^T y(t)s_M(t)dt. \end{aligned} \tag{7.3}$$

Placing (7.1), and (7.2) in (7.3), we can obtain

$$r_m = A_m b_m + \sum_{j \neq m} A_j b_j \rho_{jm} + n_m, \tag{7.4}$$

where

$$n_m = \int_o^T n(t)s_m(t)dt. \tag{7.5}$$

The sampled noise n_m, is a Gaussian random variable with zero mean and variance σ^2.

Equation (7.4) can be represented in a vector form as[3]

$$\mathbf{r} = \mathbf{RAb} + \mathbf{n}, \tag{7.6}$$

where, $\mathbf{R}$ is the normalized crosscorrelation matrix, with $\{i,j\}^{th}$ component defined as $[\mathbf{R}]_{i,j} = \{\rho_{ij}\}$, and

$$\begin{aligned} \mathbf{r} &= [r_1, r_2, \ldots, r_M]^T \\ \mathbf{b} &= [b_1, b_2, \ldots, b_M]^T \\ \mathbf{n} &= [n_1, n_2, \ldots, n_M]^T \\ \mathbf{A} &= \text{diag}\{A_1, A_2, \ldots, A_M\}. \end{aligned} \tag{7.7}$$

We should also note that the additive noise $\mathbf{n}$ is not white due to filtering process and $E[\mathbf{n}\mathbf{n}^T] = \sigma^2\mathbf{R}$, where $E[\cdot]$ represents the expected value function. In the following section frequently used MUD schemes are summarized.

7.2 Literature Review of Existing Detector Structures

As mentioned earlier, the simplest method to demodulate a CDMA signal is to use a filter that is matched to the signature sequence of the user of interest. This structure is referred to as the conventional detector in the literature [110]. Although this is the optimal receiver structure

[3]Details about this notation can be found in [110].

for single user channel, its performance becomes worse when there are multiple active users in the system. In order to enhance detection performance, interference cancelation methods are used in the implementation of multi-user detectors.

Assume that the CDMA system of interest can accommodate a maximum of $P \geq M$ users. The optimum multiuser detector, i.e the maximum likelihood detector can be implemented by selecting the information bit vector $\mathbf{b}$ that maximizes the metric Ω where,

$$\Omega(\mathbf{b}) = 2\mathbf{b}^T\mathbf{A}\mathbf{r} - \mathbf{b}^T\mathbf{A}\mathbf{R}\mathbf{A}\mathbf{b}. \tag{7.8}$$

For systems that have the knowledge of active users, selection of the optimum $\mathbf{b}$ can be done in $O(2^P)$ operations since b_m can be $\{+1, -1\}$ with $O(2^P/P)$ time complexity per bit. For systems without such knowledge, this process can be done in $O(3^P)$ operations since b_m can be $\{+1, -1\}$ or 0 when the m^{th} user is inactive. Hence the time complexity per bit is $O(3^P/P)$. This problem is known to be NP-hard and as a consequence there is no known algorithm that can solve this problem in polynomial time. Hence suboptimal algorithms are frequently used for multi-user detection. In the literature there are several MUD schemes that provide a corresponding performance and computational complexity tradeoff. A good review of these algorithms can be found in [110].

In this section, we review some of the frequently used detection schemes; the decorrelating detector, linear minimum mean square error (MMSE) detector, and the successive interference canceler.

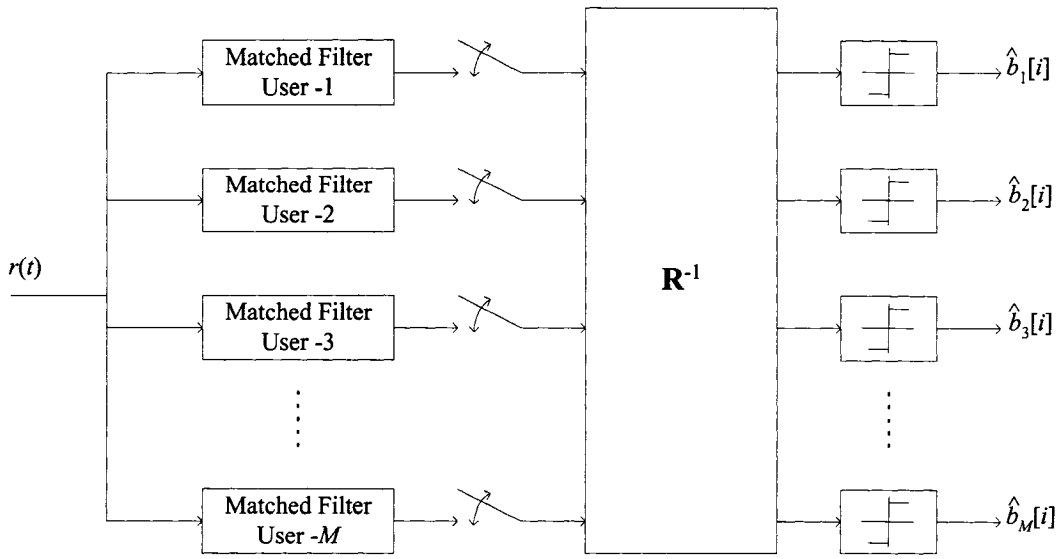


Figure 7.2: M-user decorrelating detector for synchronous channel

7.2.1 Decorrelating Detector

Assuming that the crosscorrelation matrix is invertible, a multi-user detector can be implemented as

$$\begin{aligned} \mathbf{R^{-1}r} &= \mathbf{R^{-1}RAb} + \mathbf{R^{-1}n} \\ &= \mathbf{Ab} + \mathbf{R^{-1}n}. \end{aligned} \tag{7.9}$$

The transmitted data of m^{th} user can be evaluated by using

$$\hat{b}_m = \mathrm{sgn}(\mathbf{R^{-1}r})_m, \tag{7.10}$$

where, $(\cdot)_m$ returns the m^{th} component of the input vector, and $\mathrm{sgn}(\cdot)$ is the signum function. For the cases when the crosscorrelation matrix is singular, a vector formulation can also be stated as shown in [110].

In (7.9), the effect of interference on the m^{th} component is eliminated. However the effect of background noise affects the detection performance. Such a detector is named as decorrelating detector. A decorrelating detector for M-user synchronous channel is shown in Fig. 7.2.

The computational complexity per bit that is required by the direct inversion of the matrix $\mathbf{R}$ is $O(M^3)$. There are other low complexity decorrelator implementation techniques such as integration of adaptive algorithms [93] and decomposition of the correlation matrix [63]. However the first technique introduces convergence problems and the latter requires relatively low crosscorrelation values between signature waveforms.

7.2.2 Linear MMSE detector

A multiuser detector can be constructed by minimizing the mean square error (MSE). A common approach is to use linear transformation technique to obtain the minimum MSE (MMSE) estimates due to its simpler computation process. This detector provides an optimal trade-off between interference cancelation and robustness against noise.

Linear MMSE detector evaluates the transmitted data of m^{th} user as

$$\hat{b}_m = \mathrm{sgn}((\mathbf{R} + \sigma^{\mathbf{2}}\mathbf{A}^{-\mathbf{2}})^{-\mathbf{1}}\mathbf{r})_m. \tag{7.11}$$

Similar to decorrelating detector, the computational complexity per bit that is required by the direct inversion of the matrix $\mathbf{R}$ is $O(M^3)$. Furthermore, noise variance, σ must be available at the receiver.

An MMSE detector for M-user synchronous channel is shown in Fig. 7.3. It should be noted that the MMSE approach does not necessarily lead to minimum bit error rate.

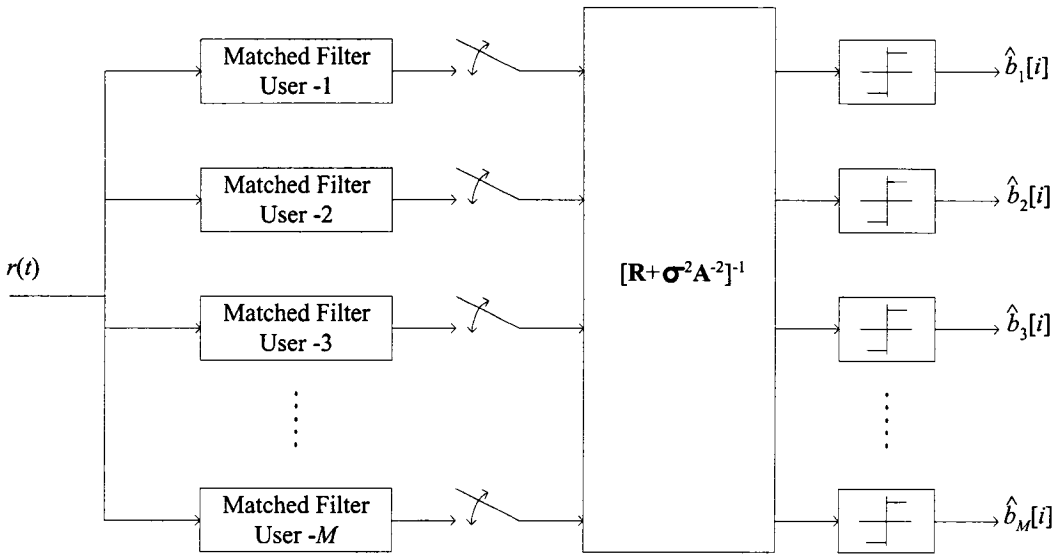


Figure 7.3: M-user MMSE detector for synchronous channel

7.2.3 Successive Interference Canceler

Successive decoding is a receiving strategy that is shown to achieve capacity under certain conditions [34]. A 2-user successive decoder is shown as an example in Fig. 7.4. The decoder detects the first user by treating rest of the users as noise. Then the signal of that user is modulated and subtracted from the received signal. The process is repeated until all users are decoded.

A successive interference canceler (SIC) is established around the successive decoding idea. This nonlinear detector has an adaptive structure due to its decision driven structure.

In the implementation, SIC requires the signature sequences and the amplitudes of the received users' signals. An error in any of these estimates results in error propagation as mentioned earlier due to the unreliability of the tentative decisions. Since the user signals are processed with a decreasing signal strength order, received signals that are weaker than that

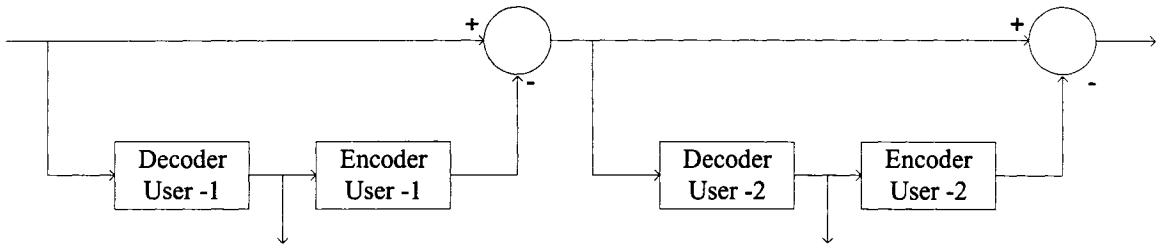


Figure 7.4: 2-user successive decoder.

of the users of interest are not processed. Due to its prioritizing sequential nature, successive cancelation is a detection strategy that is more suited to high SNR channels with power imbalances. This property results in strong mitigation with the near-far effect.

Successive cancelation procedure is a relatively simple technique for MUD. It does not require any algebraic operations other than inner product calculations and subtractions. The time complexity of SIC increases linearly with the number of users however the successive procedure results in time delays as the number of users increase.

The most important issue for the SIC is the cancelation of the strongest signal, that is the signal with the maximum amplitude. This importance is due to two factors. First, strongest signal gives the most reliable estimation. Secondly, it reduces the risks for error propagation. Hence ranking of the received amplitudes is crucial for the performance of the detector. These amplitudes can be obtained through the conventional detectors or from a separate channel estimator. The most frequently used method is based on the correlator outputs.

There are two frequently used implementation methods for the basic SIC [44]. The first type of ranking involves identification of the strongest signal and feeding back its modulated signal estimate. Without loss of generality, assume that the signal strengths are decreasing with the user's index, i.e. 1^{st} user has the strongest signal and M^{th} user has the weakest

signal. In order to detect information of iterations above can also be expressed as

$$\hat{b}_m = \text{sgn}\left(y_m - \sum_{j=m+1}^{M} A_j \hat{b}_j\right), \tag{7.12}$$

where y_m is the output of m^{th} user's matched filter.

Although the iterations summarized above are the simplest method for successive interference cancelation, they are not necessarily the optimum due to nonzero crosscorrelation between signature waveforms. The second type of ranking can be implemented by obtaining the estimates for the m^{th} user as

$$\hat{b}_m = \text{sgn}\left(y_m - \sum_{j=m+1}^{M} A_j \rho_{jm} \hat{b}_j\right). \tag{7.13}$$

The steps of a simple SIC can be summarized as [45] follows.

1. Identification (i.e. estimation) of the strongest signal.
2. Decoding the strongest signal.
3. Estimation of the amplitude of the strongest user from decorrelator output.
4. Regeneration of the user's signal using the corresponding signature waveform.
5. Cancelation of the strongest signal.
6. Repetition of the steps above until desired users are detected.

There are several other types of SICs in the literature such as soft decision based SIC. Details about these can be found in [110]. Since our goal is to point out the similarities with the SIC method mentioned above and the MP algorithms they are outside the scope of this work.

7.3 Greedy Multi-user Detection: A New Perspective for Successive Interference Cancelation

In this section we introduce the "basic matching pursuit" view of the successive interference cancelation procedure. We show that when some users are active, MUD problem can be solved by using an MP algorithm. The similarity between the iterations of the BMP algorithm and the successive interference canceler are very straightforward to grasp. In this section we present the practical implementation consequences of the BMP view of the SIC. Furthermore, we introduce an OMP (FTB-OMP) based near-far resistant multi-user detector that provides a complexity tradeoff between performance and complexity by improving interference cancelation method.

First let us start with the multi-user detection problem definition for matching pursuit algorithms. Assume that M users are active, at time interval i. As stated before, the received signal can be written as

$$r(t) = \sum_{m=1}^{M} A_m b_m s_m(t) + n(t), \quad t \in [0, T]. \tag{7.14}$$

Sampling $r(t)$ at multiples of chip period T_c, the received signal can be written in discrete form as

$$r_c(i) = \sum_{m=1}^{M} A_m b_m s_m(i) + n_c(i), \quad i = 1, \ldots, N, \tag{7.15}$$

where N is the signature sequence length and the processing gain of the CDMA system.

Equation (7.15) can be written in vector matrix notation as

$$\mathbf{r}_c = \mathbf{SAb} + \mathbf{n}_c, \tag{7.16}$$

where

$$\begin{aligned} \mathbf{r}_c &= [r(1), r(2), \ldots, r(N)]^T \\ \mathbf{n}_c &= [n(1), n(2), \ldots, n(N)]^T \\ \mathbf{S} &= [s_1, s_2, \ldots, s_M]. \end{aligned} \tag{7.17}$$

In the equation above, the i^{th} column of the $N \times M$ matrix $\mathbf{S}$ is composed of the signature sequence of the i^{th} user, i.e. s_i.

Equation (7.16) can be rewritten as

$$\mathbf{r}_c = \mathbf{S}\mathbf{c} + \mathbf{n}_c, \tag{7.18}$$

where $\mathbf{c}(m) = A_m b_m$ for $i = 1 \ldots, N$. Assuming that a phase ambiguity can be prevented (i.e. $A_i \geq 0$), m^{th} user's signal can be optimally detected as

$$\hat{b}_m = \operatorname{sgn}(\mathbf{c}(m)), \tag{7.19}$$

and

$$A_m = |\mathbf{c}(m)|. \tag{7.20}$$

Approximating

$$\mathbf{r}_c \approx \mathbf{S}\mathbf{c} \tag{7.21}$$

a matching pursuit algorithm can be used to estimate $\mathbf{c}$. For such a case, $\mathbf{S}$ is the dictionary. The number of columns in the dictionary is $P = M$ and hence the set of equations forms a complete dictionary. Hence the dictionary is not overcomplete. Despite this fact, in the following sections it is shown that MP algorithms can be used as equation solvers for MUD.

7.3.1 Basic Matching Pursuit and Successive Interference Cancelation: Are they the same?

In the previous section the problem definition for application of an MP algorithm is introduced. Here we show the equivalence of the steps of basic matching pursuit (BMP) algorithm and the SIC. Let us rewrite the steps of a simple SIC with the ranking criterion given in (7.13).

The iterations of the SIC are:

1. Estimation of the strongest signal. This corresponds to the detection of maximum correlation value, and can be expressed as

$$m = \arg\left\{\max_j \langle r(t), s_j(t)\rangle\right\} = \arg\left\{\max_j \langle \mathbf{r}_c, s_j\rangle\right\}. \tag{7.22}$$

2. Decoding the strongest signal,

$$\hat{b}_m = \operatorname{sgn}\left(\langle r(t), s_m(t)\rangle\right) = \operatorname{sgn}\left(\langle \mathbf{r}_c, s_m\rangle\right). \tag{7.23}$$

3. Estimation of amplitude of the strongest user, A_m, from the correlator output,

$$A_m = \langle r(t), s_m(t)\rangle = \langle \mathbf{r}_c, s_m\rangle. \tag{7.24}$$

4. Regeneration of the m^{th} user's signal using the corresponding signature waveform as

$$u_m(t) \;=\; A_m \hat{b}_m s_m(t), \tag{7.25}$$

and can be written in vector notation as

$$\mathbf{u}_m \;=\; c(m) s_m, \tag{7.26}$$

where $\hat{b}_m = \operatorname{sgn}\left(\mathbf{c}(m)\right)$.

5. Cancelation of the strongest signal,

$$r(t) = r(t) - u_m(t), \tag{7.27}$$

or equivalently

$$\mathbf{r}_c = \mathbf{r}_c - \mathbf{u}_m. \tag{7.28}$$

6. Go to step 1 until all desired users are detected.

Clearly, iterations summarized above correspond to the BMP iterations that are summarized in Chapter 2. The reselection problem can be avoided in the scenario explained above since it is known that all the columns of the dictionary matrix are a part of the received vector $\mathbf{r}_c$. Hence it can be concluded that SIC is a specific case of BMP for $P = M$ case. Although this result does not provide any computational complexity reduction or performance enhancement to SIC, it is important in the sense that the conclusion provides a broader perspective to SIC as shown in Section 7.5.

7.3.2 Orthogonal Matching Pursuit based Multi-User Detection: A Novel Detector?

Modeling the received signal as a linear combination of signature sequences in (7.21), a detector can be implemented via applying OMP and using $\mathbf{S}$ as the dictionary instead of the BMP. Due to the inherent structural differences with the OMP and BMP, it is expected that the OMP based MUD performs better. This is mainly due to the additional orthogonalization procedure of the OMP. At each iteration the effect of interference is removed along with the aligned components of the interfering users' signature sequences. The improved performance is verified also by simulation results.

However an unexpected result appears when the performances of the OMP based detector and the decorrelator detector are compared. They are equivalent for the case where $M \leq N$, i.e. when the dictionary $\mathbf{S}$ is undercomplete or complete instead of the overcomplete case. For this case, all of the basis components in the dictionary must appear at the OMP output. In such a scenario OMP converges to the least square solution. Although this result is not yet proven analytically, simulation results confirm this expectancy for both AWGN and Rayleigh fading channels as shown in Section 7.4.

The OMP point of view for the decorrelator detector has the advantage of preventing the cases where the correlation matrix $\mathbf{R}$ is singular, and prevent the calculation of the $\mathbf{R}^{-1}$ which has a computational complexity of $O(M^3)$ in case of a direct implementation. Although the process gives estimates for all M users, for estimation of a single user the computational complexity is kept due to the matrix inversion process. Hence the required number of operations by decorrelator detector is bounded by $O(M^3)$ per bit.

The OMP has the complexity of $O(M^2N)$ per implementation. Considering that the iterations can be stopped once they reach the desired user, computational complexity of OMP based detector is less than $O(M^2N)$ per bit.

A naturally arising question is: "Is there a possible performance enhancement by using a broader search". Application of such a search with FTB-OMP algorithm however would not introduce a considerable performance gain. This is due to the fact that the FTB-OMP tries to minimize the Euclidean distance between $\mathbf{r}_c$ and an estimate $\hat{\mathbf{r}}_c$ by searching through linear combinations. This in turn can only try to approximate the performance of the MMSE detector.

7.4 Performance Comparison of Multi-User Detection Schemes

In this section, simulation results are presented for AWGN and flat Rayleigh fading channels. Due to well known resistance of SIC, decorrelator and MMSE detectors to near-far effect [9, 20], only perfect power control case is considered for simulations. Near-far effect is considered in Section 7.5.

In simulations, randomly generated signature sequences are considered. These sequences are regenerated every 100 blocks. Bit error rate (BER) curves are evaluated by counting 10000 errors. Processing gain is $N = 31$. In the simulations, matched filter, SIC and BMP, decorrelator, OMP based detector, and FTB-OMP based detectors are considered. Single user limits are also shown in the figures.

AWGN channels are considered in Figs. 7.5 and 7.6. FTB-OMP based detector is not considered for these channels in the figures due to its very close performance to OMP based detector for a tree-search range.

Bit error rate performance of the MUD schemes of interest for SNR= 8 dB is shown in Fig. 7.5. In the figure, SIC and BMP curves match perfectly as expected. BER curves of OMP based detector and decorrelator detector also match due to least squares solution convergence.

Error performances of MUD schemes for $M = 15$ users are shown in Figs. 7.6. The same results hold for these curves.

Flat, block fading Rayleigh channels are considered in Figs. 7.7, and 7.8. BER curves of the MUD schemes of interest for SNR= 8 dB is shown in Fig. 7.7. Similar to AWGN case, SIC and BMP curves match perfectly, and OMP based detector and decorrelator detector

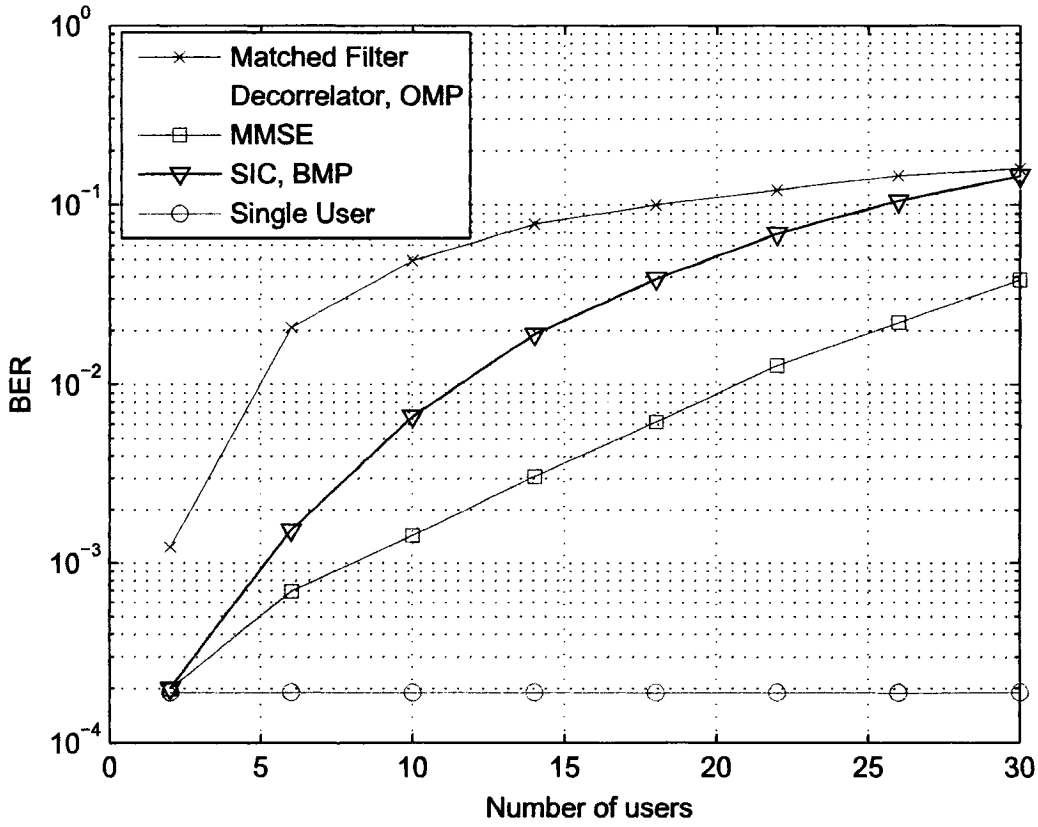


Figure 7.5: BER vs. number of users (capacity) plot for MUD schemes for AWGN channels under perfect power control, $N = 31$ and SNR= 8 dB.

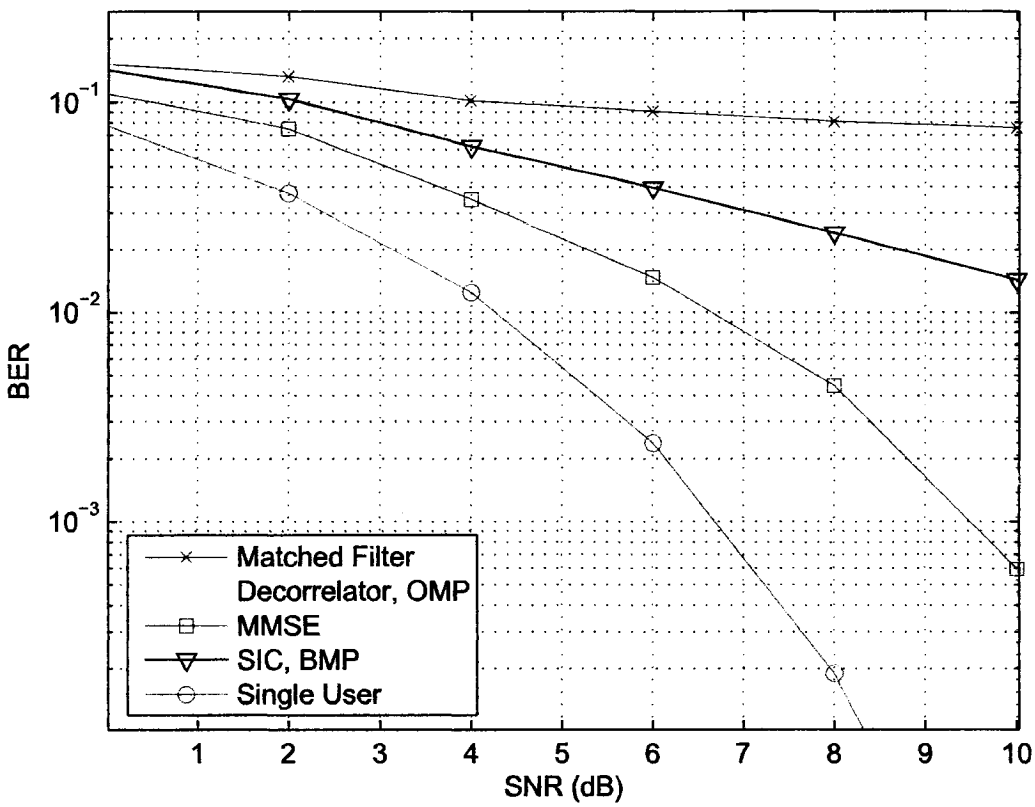


Figure 7.6: BER vs. SNR plot for MUD schemes for AWGN channels under perfect power control, $N = 31$ and $M = 15$.

also give the same error performance. In Fig. 7.8, SNR is varying where the number of users is kept constant, $M = 15$. The same results hold for this case.

In the simulations above we showed that the SIC and BMP based detector give the same error performance. Advantages of the MP viewpoint can be summarized as

1. All users are active in the systems. Due to this fact, index constraint hence, the re-selection problem of BMP is eliminated with SIC viewpoint of the algorithm. However, this directs us to a blind detector design where only some users are active.

2. Perfect power control is not essential due to well known near-far resistance of the SIC.

3. The BMP based detector performs joint channel estimation and multi-user detection, not only MUD. This is also a beneficial fact for blind detector design. Delay tracking is possible with integration of training sequence as mentioned in chapters 4 and 5. Since channel taps are available it is easy to exploit multipath components with RAKE receiver.

4. Angles can be detected just by extending the dictionary. This can be accomplished by including phase shifted versions of the signature sequences according to the delays between the antenna elements.

7.5 A Broader Perspective for Interference Cancelation: Blind Interference Cancelation

Synchronous communication is attainable in CDMA downlink channels. However, the detector must be low in complexity due to the battery requirements of the mobile terminal.

In the previous sections it is assumed that all users in the CDMA system are active, that

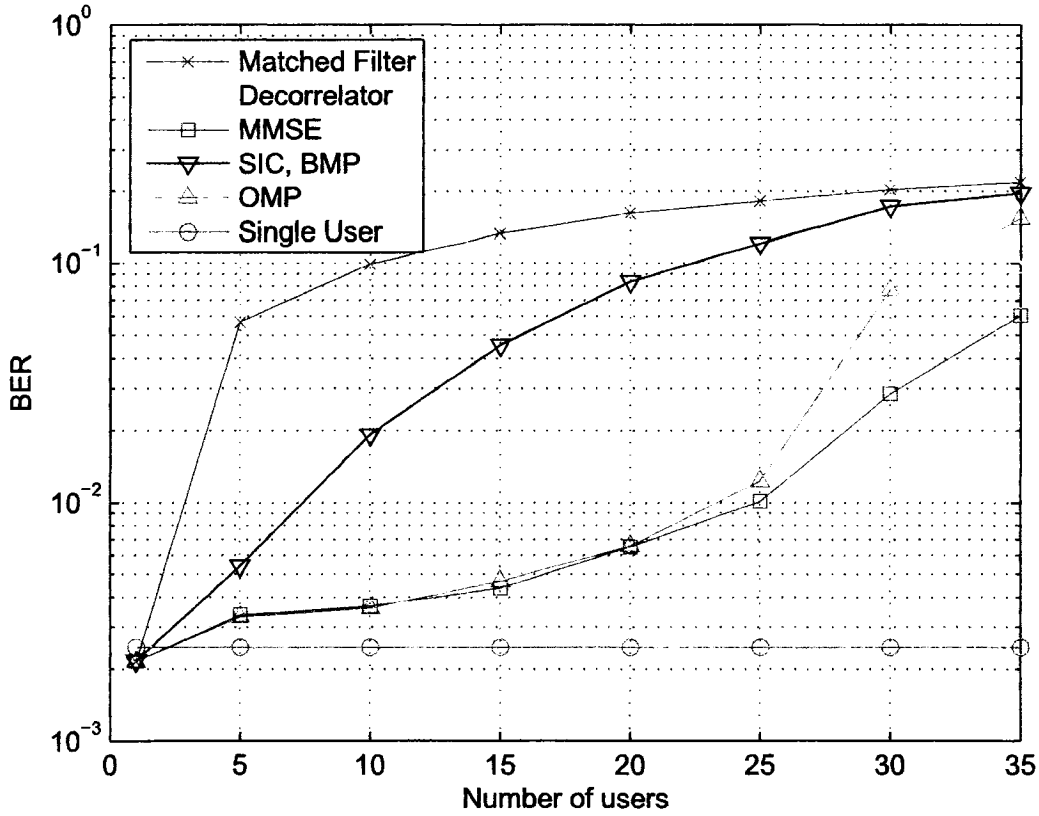


Figure 7.7: BER vs. number of users plot for MUD schemes for Rayleigh fading channels under perfect power control, $N = 31$ and SNR= 8 dB.

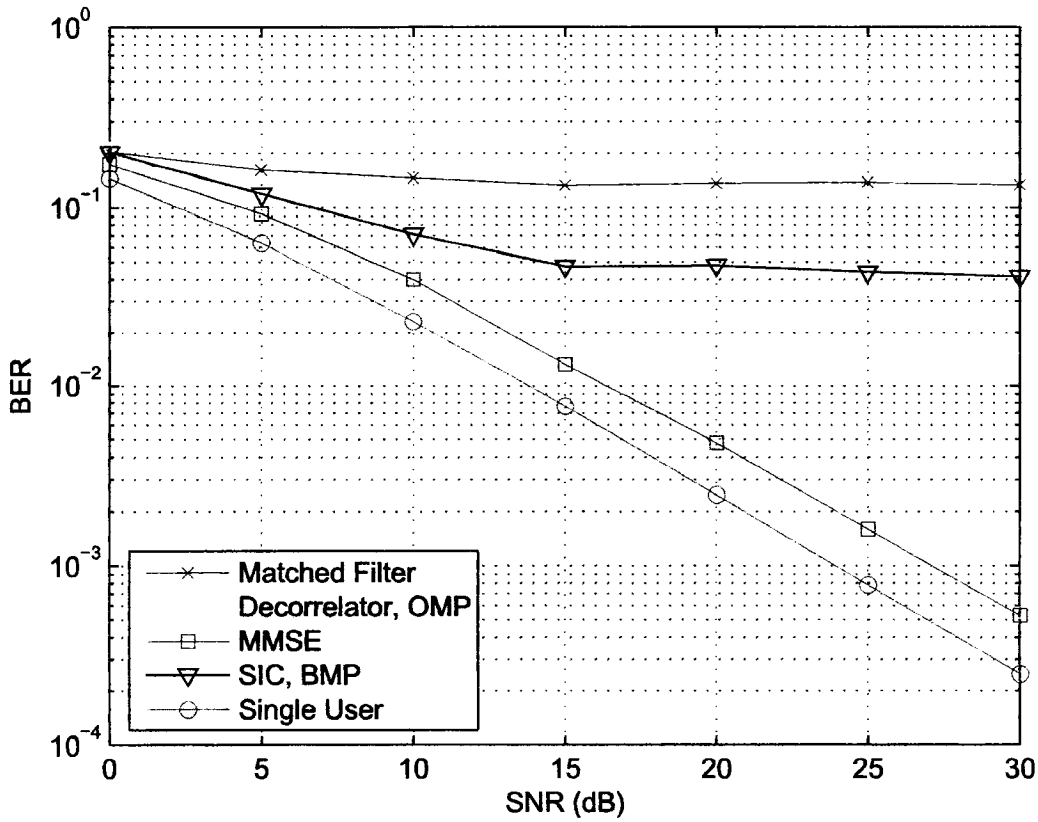


Figure 7.8: BER vs. SNR plot for MUD schemes for Rayleigh fading channels under perfect power control, $N = 31$ and $M = 15$.

is the knowledge about all users are available at the receiver. Although this is possible for uplink communications, it is not necessarily the case for downlink. Using the basis selection approach introduced earlier, in this section blind detector structures are discussed based on MP algorithms. All possible signature sequences are available at the mobile terminal however neither the number of active users nor the knowledge of corresponding signature sequences of these users are available.

Assume that $M \leq P$ users are active, at any given time interval. The case with $M = P$ is discussed in the previous section. For a case, when $M < P$ users are active, some of the $b_k = 0$. Without loss of generality, we can construct an extended bit vector as[4]

$$\mathbf{b}_e = [b_1, b_2, \ldots, b_M, 0, \ldots, 0]^T. \tag{7.29}$$

Using the sparsity of $\mathbf{b}_e$, and having the a priori knowledge about the signature sequences of the active users, the transmitted information can be detected using an MP algorithm.

The received signal at the chip matched filter output can be expressed as

$$\mathbf{r}_c = \mathbf{S}_e \mathbf{A}_e \mathbf{b}_e + \mathbf{n}_c, \tag{7.30}$$

where an extended dictionary $\mathbf{S}_e$ is composed as

$$\mathbf{S} = [s_1, s_2, \ldots, s_P], \tag{7.31}$$

and the amplitudes are

$$\mathbf{A}_e = \text{diag}\{A_1, A_2, \ldots, A_M, , 0, \ldots, 0\}. \tag{7.32}$$

Equation (7.30) can be re-expressed as

$$\mathbf{r}_c = \mathbf{S}_e \mathbf{c}_e + \mathbf{n}_c. \tag{7.33}$$

[4]This ordering is not necessary. However it is mentioned here for demonstration of the sparseness property.

Assuming that the additive noise is negligible, the received vector can be approximated as

$$\mathbf{r}_c \approx \mathbf{S}_e \mathbf{c}_e, \tag{7.34}$$

and an estimate for the extended coefficient vector $\mathbf{c}_e$ can be obtained by using an MP algorithm. The information bits can be extracted in the same method for the case with $M = P$, i.e.

$$\hat{b}_m = \operatorname{sgn}\left(\mathbf{c}_e(m)\right). \tag{7.35}$$

The stopping criterion for the MP algorithms is given in terms of an error threshold ϵ since M is not available at the receiver.

Application of BMP algorithm to (7.34) for blind interference cancelation techniques, referred to as B-BMP, are considered for simulations. The computational complexity is at most $O(MNP)$ per application. We must also mention that the iteration number in the implementation can be limited by a number U where $U < M$, hence only very strong components can be canceled, and a filter matched to the desired user's signature waveform can be applied after strong interferers are canceled. An index constraint in the application of BMP is crucial. Hence the overall complexity is $O(UNP)$.

7.6 Performance Evaluation for Blind Interference Cancelation

In this section, simulation results of B-BMP detector introduced in the previous section are presented. Flat Rayleigh fading channels are considered. Near-far resistance of the proposed detector is investigated.

In simulations, codes are randomly generated every 100 blocks. Even though orthogonal codes are usually employed in the downlink transmission, interference from neighboring cells

typically resemble random codes. It is also necessary to state that changing the codes in the simulations to orthogonal sets would enhance the performance. Hence results presented in this section can be taken as worst case. Bit error rate (BER) curves are evaluated by counting 10000 errors. Processing gain is again $N = 31$. Matched filter and B-BMP detector with $U \in \{1, 3, 5\}$ are considered.

In the labels, SNR_1 represents the signal to noise ratio of the user of interest (first user). The signal to noise ratio of the second user is denoted by SNR_2. The rest of the active users have signal to noise ratios equal to SNR_1. Near-far effect is introduced by the second user's SNR imbalance. U is the number of iterations of BMP algorithm.

In Figs. 7.9 and 7.10, 15 users are active. Perfect power control is assumed where all users have the same power levels. As can be seen from the figures, all detectors have similar performances. Hence adding BMP iterations does not degrade the performance on the contrary it improves the performance slightly.

In Fig. 7.11 and 7.12, again 15 users are active. All users except one have same power levels. One user (an interferer with SNR_2) has a signal 10 dB stronger than the rest. In the figure we can see that BMP considerably cancels the degrading effects of this near-far scenario. Law of diminishing returns is apparent from the results for increasing U. Hence a low value for U is suitable in detector implementation.

In Figs. 7.13, and 7.14 several case of near-far problem is integrated in the system with $SNR_1 = SNR_3 = 5$ dB and $SNR_1 = SNR_3 = 10$ dB, respectively . There are $M = 3$ users and SNR_2 is used for the near-far effect. B-BMP detector with $U = 3$, and $U = 5$ are included in order to demonstrate that a mismatch in U does not cause a severe performance degradation. From the figures, we can see that with this adaptive interference cancelation, B-BMP approached to single user performance.

Furthermore, not all K users have to be active all the time. Having a priori knowledge

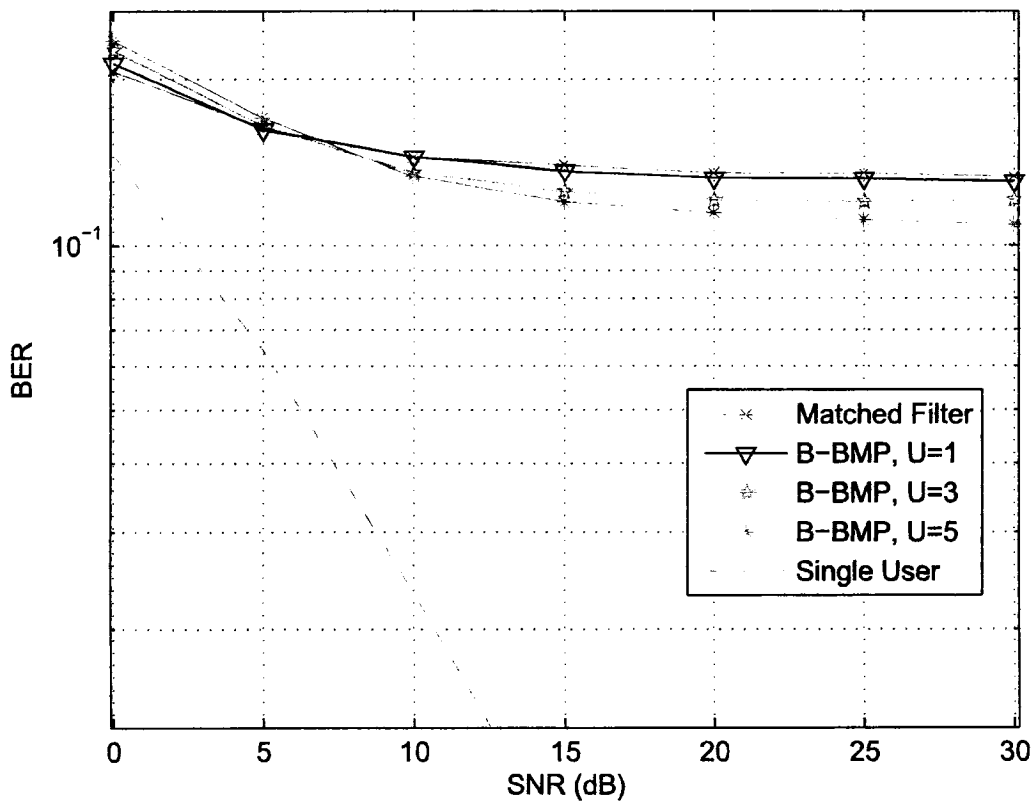


Figure 7.9: BER vs. SNR in 15-user Rayleigh fading channel with $SNR_2 = SNR_1$, $N = 31$, $\epsilon = 10^{-10}$.

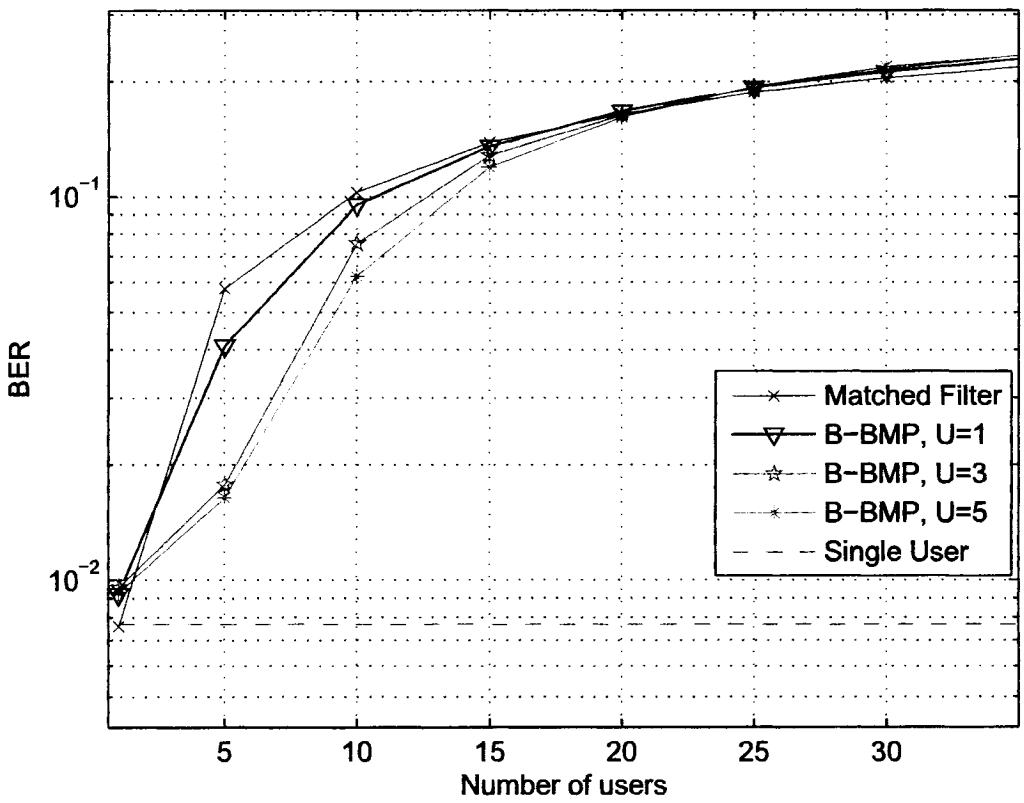


Figure 7.10: BER vs. number of users in Rayleigh fading channel with $SNR_2 = SNR_1 =$ 15 dB, $N = 31$, $\epsilon = 10^{-10}$.

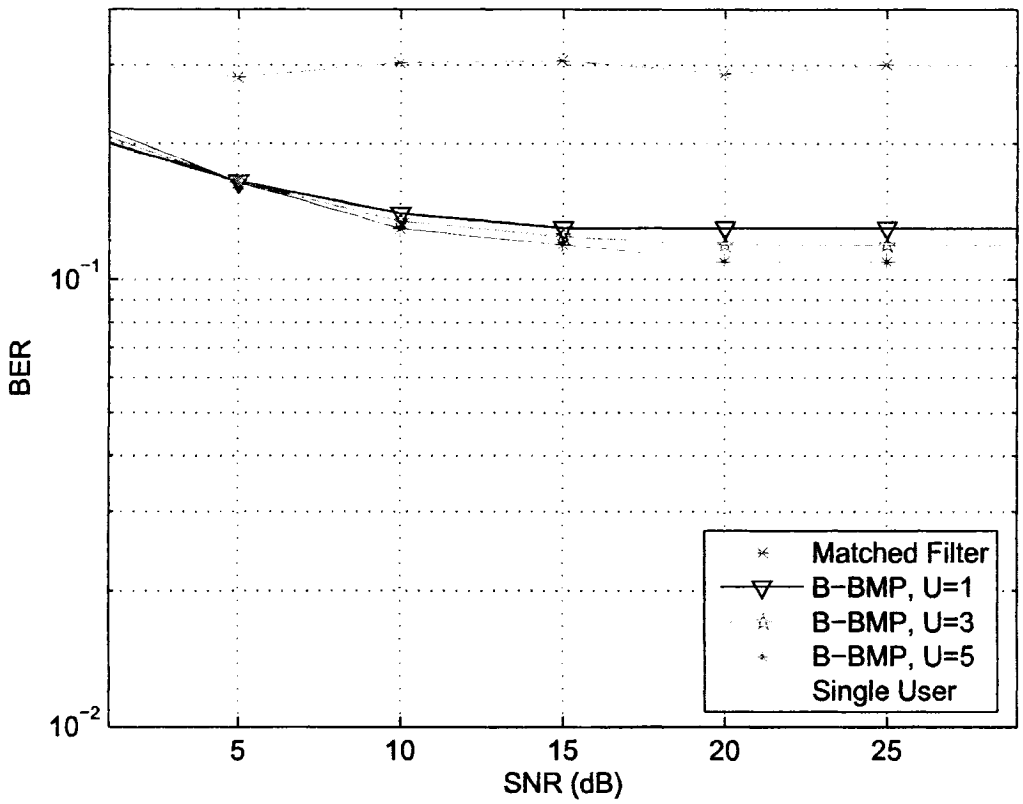


Figure 7.11: BER vs. SNR in 15-user Rayleigh fading channel with $SNR_2 = SNR_1 + 10$ dB, $N = 31$, $\epsilon = 10^{-10}$.

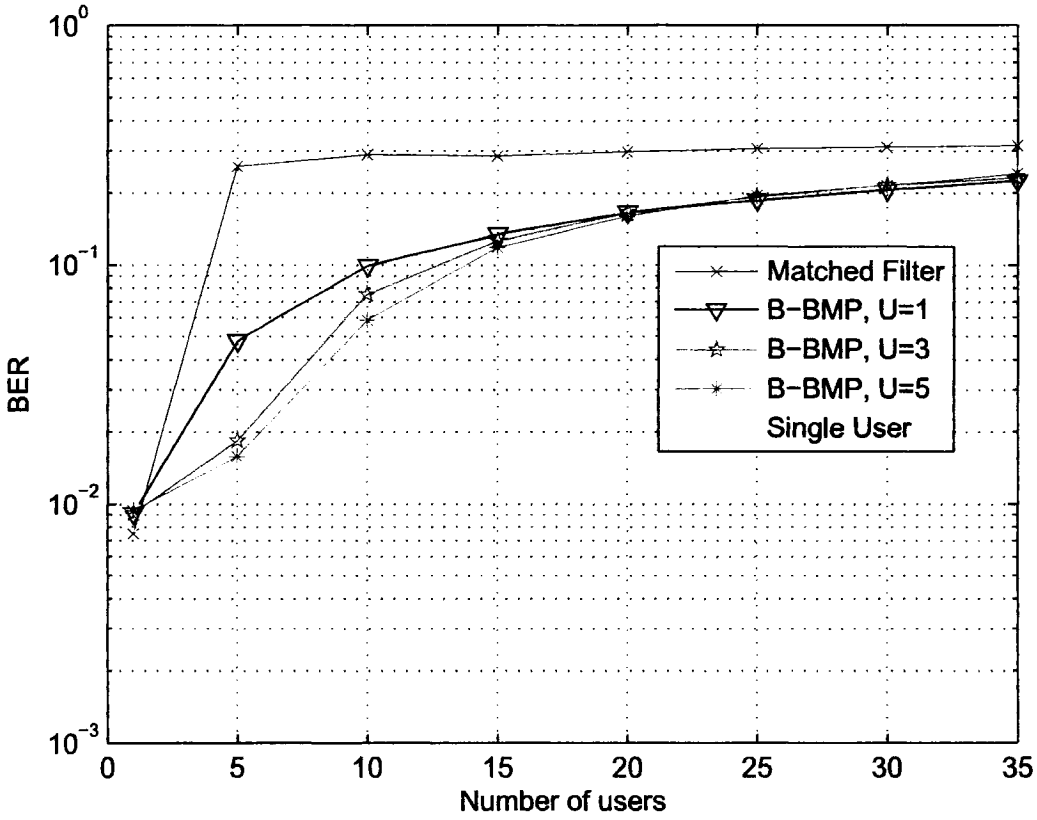


Figure 7.12: BER vs. number of users in Rayleigh fading channel with $SNR_1 = 15$ dB, $SNR_2 = 25$ dB, $N = 31$, $\epsilon = 10^{-10}$.

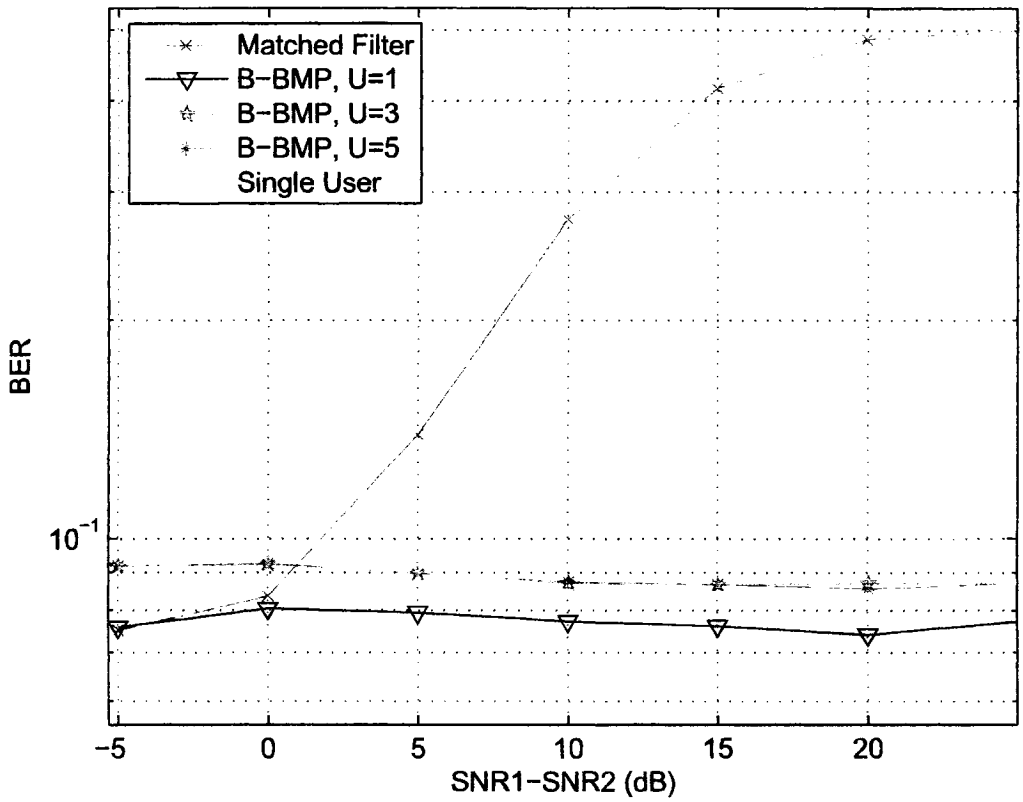


Figure 7.13: BER vs. SNR difference in 3-user Rayleigh fading channel with $SNR_1 = 5$ dB, $N = 31$, $\epsilon = 10^{-10}$.

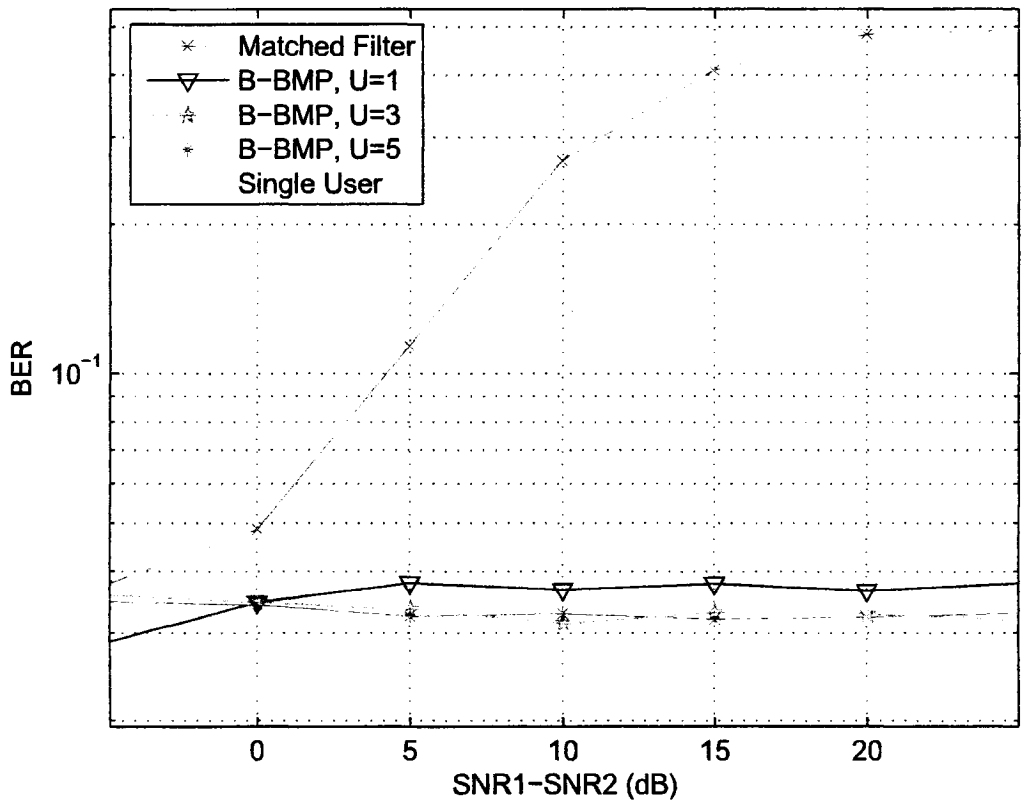


Figure 7.14: BER vs. SNR difference in 3-user Rayleigh fading channel with $SNR_1 = 10$ dB, $N = 31$, $\epsilon = 10^{-10}$.

of all signature waveforms, active users can be detected with the SBS algorithm. We should note that, there is no need to know which users are active, since they are resolved using the algorithm.

7.7 Blind Channel Estimation for CDMA

In this section, we propose a joint channel estimation and MUD scheme for uplink CDMA using OMP. The channel estimation is accomplished in a 'blind' fashion, which is then followed by MUD. Application of MP algorithms to channel estimation problem is considered in chapters 4 and 5. There, the formulation required transmission of training symbols in order to detect the channel taps. However in CDMA systems, using the properties of spreading codes channel estimated can be obtained by the MP algorithms. Formation of channel estimates is implicit in the CDMA detectors explained above. These channel estimates are used in order to cancel interference.

In the literature, blind channel estimation is achieved by the implementation of sub-space based approaches. The details of these methods are described in [3] and [102]. Here, we propose the application of OMP in order to decrease complexity and improve performance of the blind channel estimators. In order to make a fair comparison with the methods that exist in the literature, and isolate the results from delay estimation methods, we assume that the delays of the different users are known to the receiver. However the channel multi-path delays are unknown. Hence the multi-path delay distribution with respect to the mean delay is unknown. We model the channel as tapped delay line with taps at integer multiples of $1/BW$, where BW is the bandwidth of the spread signal. The channel taps are assumed to be Rayleigh fading (no line of sight) and the tap gains are exponentially decaying with arrival time.

The most basic way of joint detection would be including every single chip shift of every user's signature $s_i(t)$ in the signature matrix $\mathbf{S}$. However this would increase its size to $N \times 2M$, and increase the complexity to a formidable level. Instead the following method is employed. For each user a temporary channel estimation dictionary $\mathbf{S}_c^i$ of size $N \times M$ is formed:

$$\mathbf{S}_c^i = [s_i(t), s_i(t+T_c), ..., s_i(t+(N-1)T_c)] \tag{7.36}$$

This dictionary includes all the chip shifted signature sequences for a specific user. From this dictionary, by employing OMP algorithm, successful L_i' candidates are selected and the $L_i \times N$ matrix $\mathbf{S}_l^i$ is formed. Hence at this stage, most likely taps that include multi-path components are determined. Then the new detection matrix including every channel tap of every user $\mathbf{S}_C$ is formed by

$$\mathbf{S}_C = [\mathbf{S}_l^1, \mathbf{S}_l^2 ... \mathbf{S}_l^M] \tag{7.37}$$

The detection is accomplished as in the non-blind case from this point on. However, instead of giving the symbol amplitude values $\mathbf{A}$, the algorithm gives the amplitude values of different channel taps for different users. Then the transmitted symbol is detected by diversity combining. As long as the channel is constant, the dictionary $\mathbf{S}_C$ can be used for MUD. However, it should regularly be updated by using the dictionaries $\mathbf{S}_c^i$. Next, we compare the performance of the described OMP detector with the sub-space methods.

7.7.1 Simulation Results

For the simulations, we consider a CDMA2000 channel with $PG = 32$ and chip duration of $0.26\mu s$. Each user has an independent channel with delay spread of approximately 6 chip durations ($1.5\mu s$), which is a typical case for urban environment [68].

We first compare the mean square error (MSE) per user between the proposed algorithm

and matching pursuit methods. The MSE is defined as:

$$MSE = \sum(||\mathbf{h} - \hat{\mathbf{h}}||^2) \tag{7.38}$$

where $\mathbf{h}$ and $\hat{\mathbf{h}}$ are respectively the channel and its estimate for a specific user . However, the results are averaged for multiple users that exist in the system. We employ the SVD method of [3], which has very similar results to [102]. As shown in Fig. 7.15, both OMP and BMP methods outperform the SVD based method. Especially, for OMP based method, the advantage is very high for high SNR values.

Next we compare the complexities of the models by depicting required floating point operations for different processing gains. For a CDMA2000 system, the processing gain can be adaptively changed between 2 and 128. As shown in Fig. 7.16, the complexity of the SVD blind detector increases exponentially with increasing code length, whereas the OMP method only increases linearly.

The difference in MSE performance directly effects the BER performance of the MUD. Although a less complex method which doesn't require SVD but matrix inversions are proposed recently in [81]. However the performance of [81] is worse than SVD and its complexity is still much higher than the matching pursuit based detectors proposed in this work.

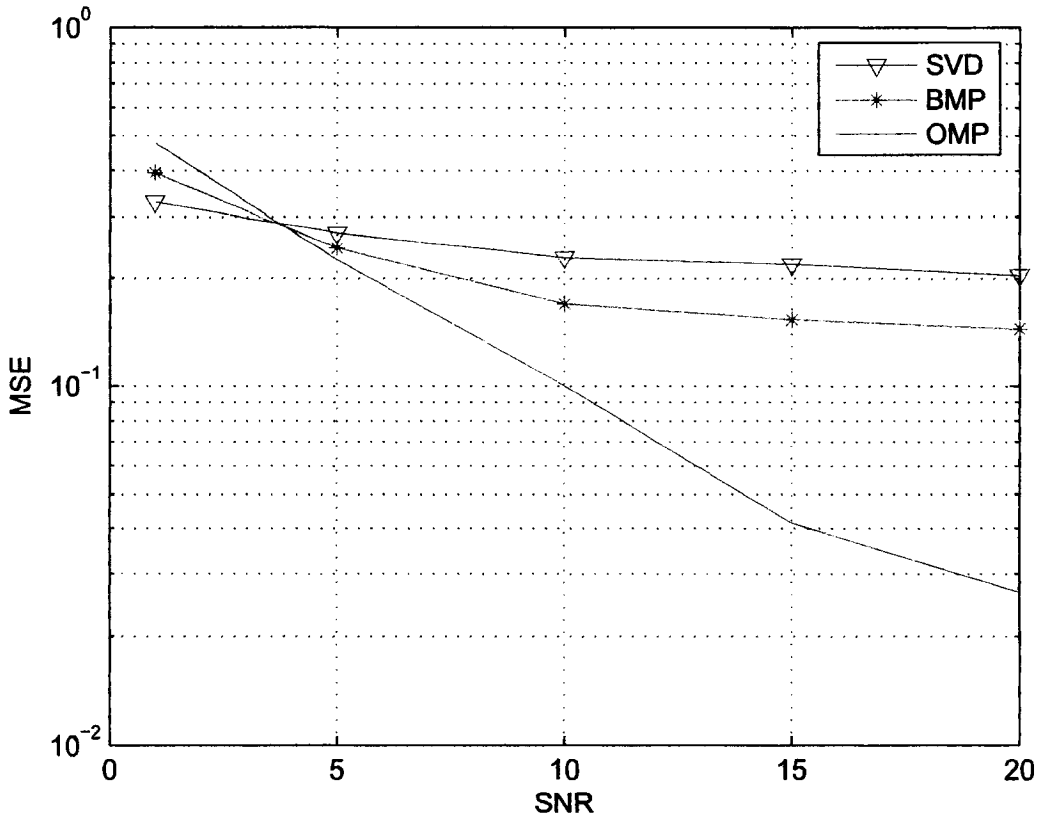


Figure 7.15: MSE vs. SNR plot for different blind estimation methods

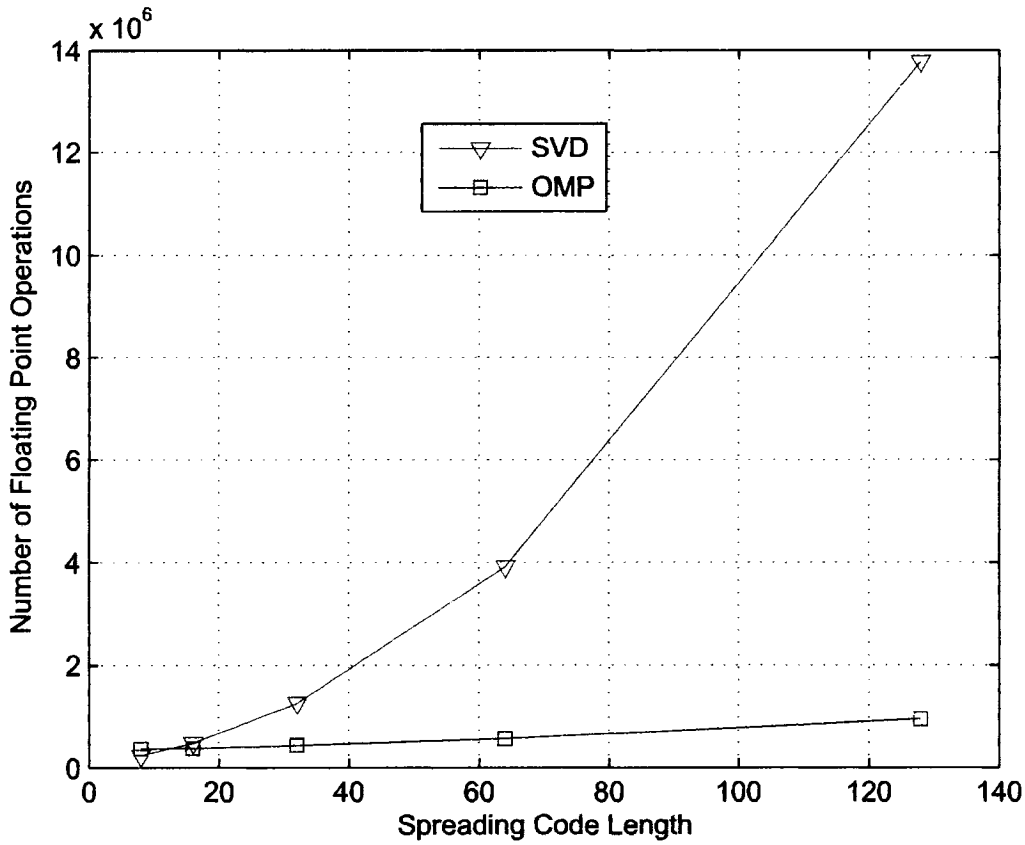


Figure 7.16: Number of floating point operations vs. processing gain

7.8 Conclusions

In this chapter, relations between SIC and BMP based MUD schemes are shown. An OMP based MUD is proposed that gives the same performance with the decorrelator detector with less complexity.

After having shown the relations between interference cancelation and MP algorithms a low complexity near-far resistant blind detector is proposed.

The proposed algorithm is not restricted to cellular CDMA systems, but can be effectively employed for tactical applications such as anti-jamming. Also, the MP algorithms are especially suitable for systems without power control in general or without power control in particular instants such as initial call setup in cellular CDMA sytems.

We also presented a novel multi-user detector and a blind channel estimator for CDMA systems employing BMP and OMP algorithms. We have compared these algorithms' methodology both analytically and by simulation to the traditional MUD algorithms that exist in the literature. We have shown the equivalence between the basic matching pursuit algorithm and basic SIC. Then, using the structural advantages of OMP over BMP, we introduced an OMP based detector. We have presented that the OMP algorithm actually converges to the performance of linear decorrelator detector. A practical implementation for the successive interference cancelation has been presented by integrating the basic matching pursuit algorithm to the detector.

Finally, we have also shown that, the proposed architecture can easily be extended to multi-path case. As a result, a novel joint blind channel estimator and multi-user detector is proposed. This novel blind channel estimator is shown to outperform SVD based methods with lower complexity.

Chapter 8

Conclusions

In this dissertation we have addressed the basis selection problem. The importance of this problem is demonstrated for several different applications in wireless communication systems. These applications areas include the channel estimation problem, direction of arrival detection problem and the multi-user detection problem. These problems contain three distinct dictionary types; undercomplete, overcomplete and complete set of equations, respectively.

Due to the greedy structure of the sequential basis selection algorithms, an erroneous basis selection at an iteration has an impact the resulting signal approximation. The selection of a basis vector not leading to the sparsest solution is referred to as the error propagation problem. An efficient exponentially decaying tree structure is integrated in the efficient sequential basis selection algorithm, OMP and the resulting algorithm is named as flexible tree-search based orthogonal matching pursuit (FTB-OMP). The algorithm provides some design parameters that give flexibility to establish a tradeoff between performance and running time. The efficiency is achieved by using a correlation based pruning in the search tree, and reducing the number of children as the depth of nodes increase. The efficiency of the proposed algorithm is demonstrated with several experiment results. From the experimental results, we can conclude

that FTB-OMP algorithm with high L values and $d > 1$ is more effective than TB-OMP for $\xi = 0$. The novel exponential decaying structure of the search tree makes the algorithm more computationally efficient and increases the approximation and detection performance. The threshold parameter ξ can be used for further computational efficiency.

It is important to emphasize that the proposed FTB-OMP algorithm is not tied to any specific application. The parameter sets for FTB-OMP can be designed so that the resulting search tree is suitable for the application of interest.

In the second part of the dissertation, the family of MP algorithms are applied to several wireless communication problems. Sparse solution requirement for an observed data vector is a frequently encountered problem in wireless communication applications. These problems contain undercomplete set of equations (channel estimation), overcomplete set of equations (direction of arrival estimation) and complete set of equations (special case of multi-user detection). It is shown for all three types of dictionaries OMP based algorithms provide low complexity solutions for these detection problems. Performance of OMP algorithm can further be improved via FTB-OMP with parameters designed for the specific application.

Among the application areas, only channel estimation problem is treated using MP algorithm in the literature [25]. The two other proposed areas are introduced in this dissertation. These problems are carefully selected in order to demonstrate the effectiveness of OMP and FTB-OMP algorithms for undercomplete, overcomplete, and complete dictionaries. It is shown by simulation results that the proposed FTB-OMP algorithm is a good choice in terms of performance and complexity for all these applications with a wide variety of dictionary possibilities.

8.1 Recommendations for Future Research

The proposed FTB-OMP algorithm can be used in a wide range of applications from image processing to data compression. Sensitivity analysis of FTB-OMP (or tree search methods integrated with any selection of MP algorithms) parameters including branching factor, correlation threshold, stopping threshold and search depth would be a valuable research contribution. Also analytical results on the convergence of the FTB-OMP to the sparsest representation would be a valid contribution.

In terms of applications perspective, extension of parameter detection techniques to multiple input multiple input multiple output (MIMO) systems would be a straightforward extension to this work for future research. The DOA work in Chapter 6 is mainly concentrated on single input multiple output systems. The formulation for a joint channel tap detection and AOA estimation process for MIMO systems can easily be generalized from the stated problem definitions. This would also include joint parameter detection estimation problems for MIMO, single input multiple output (SIMO), and multiple input single output (MISO) cases. Another proposal for future work is analytical performance evaluations for the simulated results in detection problems.

In closing, we envisage that the application areas of the basis selection algorithms are not limited to the areas listed above. Most problems can be set as linear equations. Having shown that MP algorithms are effective in undercomplete, overcomplete and complete set of equations. set of equations, it can be stated that these algorithms can be low complexity solutions to several distinct research areas.

Appendix A

Hilbert Space

A Banach space is defined as a vector space with norm definition. A norm should satisfy the following properties.

$$\forall x \in \mathbb{H}, \quad \|x\| \geq 0 \quad \|x\| = 0 \quad \Leftrightarrow x = 0, \tag{A.1}$$

$$\forall \lambda \in \mathbb{C} \quad \|\lambda x\| = |\lambda| \|x\|, \tag{A.2}$$

$$\forall x, y \in \mathbb{H}, \|x + y\| \leq \|x\| + \|y\|. \tag{A.3}$$

A Hilbert space is a Banach space with an inner product definition. Angles and orthogonality are defined by inner products with the properties stated below.

$$\forall \lambda_1, \lambda_2 \in \mathbb{C}, \quad \langle \lambda_1 x_1 + \lambda_2 x_2, y \rangle = \lambda_1 \langle x_1, y \rangle + \lambda_2 \langle x_2, y \rangle \tag{A.4}$$

$$\langle x, y \rangle = \langle y, x \rangle^*. \tag{A.5}$$

Moreover

$$\langle x, x \rangle \geq 0 \text{ and } \Leftrightarrow \langle x, x \rangle \geq 0 \tag{A.6}$$

Norm definition in a Hilbert space is given by

$$\langle x, x \rangle^{1/2} = \|x\|, \tag{A.7}$$

$$\langle x, y \rangle \leq \|x\| \|y\|, \tag{A.8}$$

with equality of and only if x and y are linearly independent.

Appendix B

Optical CDMA Detection with Matching Pursuit Algorithms

In Chapter 7, we have presented the application of BS algorithms to wireless CDMA detection problem. CDMA is being widely employed in wireless communication systems, but it has also been shown to offer some major advantages for optical systems [87, 88]. Here, we investigate the multiuser detection for optical CDMA systems.

Optical code division multiple access (O-CDMA) uses the vast bandwidth of optical communication systems in order to benefit from the flexibility of the CDMA systems. Other than the bandwidth utilization, asynchronous access and secure communication makes O-CDMA systems a promising candidate for optical communication systems. Due to its asynchronous access property, network control and management becomes easier.

O-CDMA systems became a very promising candidate for implementation of an all optical system because of their robustness to network size and complexity [87]- [61]. Also due to its broad bandwidth operation and asynchronous access structure, O-CDMA became a candidate for backbone transmission of radio over fiber systems as well [62]. In terms of

available services, multi-rate O-CDMA systems were proposed to provide different services to different user profiles [47, 48].

The main difference of O-CDMA systems from wireless CDMA is the code structure. Optical systems are mainly intensity modulated; hence the signal space is non-negative. The chips in the CDMA system are alternating '1's and'0's (unipolar symbols), instead of '-1's and '1's (bipolar symbols). A set of good wireless CDMA codes may not be suitable for optical CDMA. For example, a pair of bipolar codes with zero correlation would have very high correlation when they are converted to unipolar codes. In order to find a good set of codes for optical systems, optical orthogonal codes (OOC) have been proposed initially in [18] and recently in [113].

The main challenge in the O-CDMA systems is in the detection process. Recently many different O-CDMA receiver architectures have been proposed, in order to decrease system complexity and cost, and to increase the interference rejection property. The best solution in terms of performance is achieved by maximum likelihood (ML) solution [109]. However the ML solution has a formidable computational complexity and it is known to be NP-hard. Hence some sub-optimal detectors such as correlation receivers, optical hard limiters, and chip level detectors have been proposed. A detailed comparison of these algorithms is given in [115]. An interference canceler has been presented in [8]. Lately, algorithms such as expectation-maximization [75] and serial search [61] as well as two dimensional optical codes [94] have also been proposed. However none of these algorithms are cost-efficient, simple and can provide good and robust performance at the same time. This situation delays the wide range deployment of O-CDMA systems.

Here, we propose the application of BS algorithms, in particular the OMP algorithm for detection of O-CDMA signals. This permits us to have a cost efficient, simple and robust detection of O-CDMA signals.

The chapter is organized as follows: In Section B.1, the system model is given. Some of the popular conventional O-CDMA receivers are briefly reviewed in Section B.2. Simulation results for different O-CDMA detector models are given in Section B.3. The 2D-CDMA codes are investigated in Section B.4 and the conclusions are presented in Section B.5. A portion of the research work presented in this appendix is published in [50, 60, 67].

B.1 System Model

An O-CDMA system with K users is considered as shown in Fig. B.14. In the figure, each user transmits the information sequence $b_i \in \{0,1\}$, for $i = 1, 2, \ldots K$. OOC's consisting of (0,1) sequences are employed. Sticking to the notation of [18], OOC's are defined as $\Phi(N, w, \lambda_a, \lambda_b)$, where N is the code length, w is the code weight (i.e. the number of '1's in a block of N), and they satisfy the correlation properties for any code sequence $\mathbf{c}_i, \mathbf{c}_j \in \Phi$ as follows:

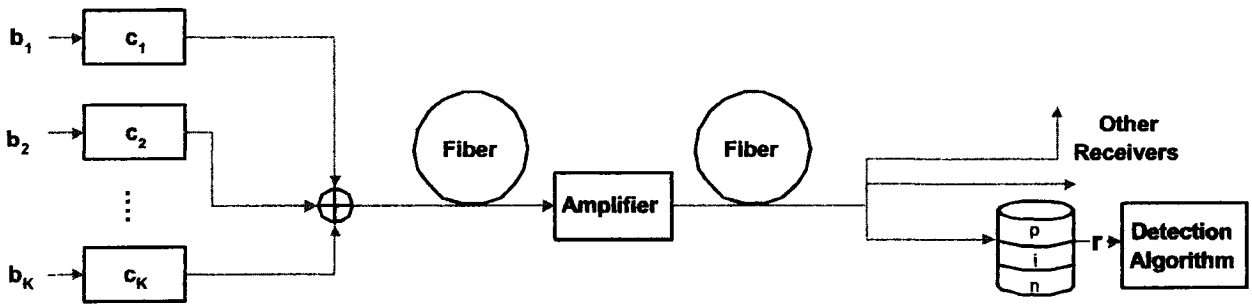


Figure B.1: Multiuser O-CDMA system model

- Cross-correlation property:

$$\sum_{l=0}^{N} c_i(l)c_j(l+k) \leq \lambda_b \tag{B.1}$$

for any integer $k \in [0, N]$, and $i \neq j$.

- Auto-correlation property:

$$\sum_{l=1}^{N} c_i(l)c_i(l+k) \leq \lambda_a \tag{B.2}$$

for any integer $k \in [1, N]$.

Hence λ_a is the maximum auto-correlation of a code sequence with its shifted versions, and λ_b is the maximum cross-correlation of a code sequence with any shifted version of any other code in the code set. For the most widely used case of $\lambda_a = \lambda_b = 1$, the number of codes in the set is upper bounded by $\lfloor (N-1)(w-1) \rfloor / w \rfloor$ [113]. In order to illustrate the OOC structures more clearly, we proceed with an example.

Example I: Consider an OOC denoted by $\Phi(13, 3, 1, 1)$, there are 2 distinct code words (using the notation of [87]):

$$\Phi = \{1101000000000, 10100001000000\}.$$

It can easily be verified that they both satisfy the auto-correlation and cross correlation properties stated in (B.1) and (B.2). The codes can be equivalently represented by their nonzero positions as $\{0, 1, 3\}$ and $\{0, 2, 7\}$.

The symbol period is denoted by T_s, the chip period is T_c with the relation $N = T_s/T_c$. The spreading code for user k is denoted by $\{c_k(l)\}_{l=0}^{N-1}$.

The baseband representation of the received O-CDMA signal after coherent reception is given by

$$s(t) = \sum_{n=-\infty}^{\infty} \sum_{k=1}^{K} A_k \bar{c}_k(t - nT_s) b_k(n) + d(t), \tag{B.3}$$

where $d(t)$ is the dark current. The transmitted amplitude for user k is A_k. The symbol sequence of the kth user, $b_k(n)$ is independent and identically distributed and modulated by on-off keying (OOK). In the equation above, the effective spreading waveform of the kth user is represented by $\bar{c}_k(t)$. This is formed by convolution of the spreading waveform with the fiber dispersion impulse response as $\bar{c}_k(t) = c_k(t) \star h(t)$. The spreading waveform can be given as

$$c_k = \sum_{l=1}^{N} c_i(l) \psi(t - lT_c), \tag{B.4}$$

where $\psi(t)$ is the chip pulse shape of duration T_c, where $\|\psi(t)\| = 1$. For simplicity, we consider rectangular pulse shapes. The transmitter of user k is a coherent laser with amplitude A_k. The output of the coherent laser can be modeled as a Poisson process (even a non-coherent source can be approximated as Poisson [115]). As the optical detector, we consider the p-i-n diode structure. The dark current of a p-i-n diode is a sum of many independent random electron generations, and since it is discrete, it can also be modeled as a Poisson process.

In order to apply the OMP algorithm to O-CDMA detection, we introduce the following notation for the discrete time model for single bit duration

$$\mathbf{r} = \sum_{k=1}^{K} A_k b_k \mathbf{c}_k + \mathbf{d}, \tag{B.5}$$

where b_k is the transmitted bit of the k^{th} user, $\mathbf{c}_k$ is a column vector of length N formed by OOC chip sequence of user k, and $\mathbf{d}$ represents the noise from the p-i-n diode. If we define the dictionary matrix $\mathcal{D} = [\mathbf{c}_1, \mathbf{c}_2, \ldots, \mathbf{c}_K]$, the received vector $\mathbf{r}$ is a linear combination of the columns of $\mathcal{D}$ with coefficients $A_k b_k$, $k = 1...K$. These coefficients are actually

the transmitted bits multiplied by the channel gains. For the case of OOK, it is straight forward to recover b_k from $A_k b_k$ [83]. Once b_k are recovered, the detection for all K users is accomplished. Although there may be more efficient implementation methods, for the sake of simplicity, OMP based multi-user detector is modeled as in Fig. B.2. The values at the chip positions are used in the algorithm after they are detected by a p-i-n diode. There are

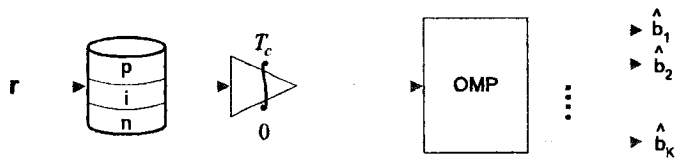


Figure B.2: Block diagram of OMP multi-user detector.

two main differences when compared to the dictionary solution for wireless CDMA systems that were discussed in the previous chapter. First, the columns of the dictionary now consists of O-CDMA codes. Second, the noise and the power variations in the channel are Poisson distributed rather than Gaussian. Hence for each case, optimal thresholds are deduced for detection. However the details of the threshold are beyond the scope of this work, and we assume that perfect threshold values are decided for each algorithm.

In this work, we assume that prior to detection, time delays of individual users are estimated. If the delays are unknown to the detector during the detection process, every cyclic shift of each user's code should be included in the dictionary which would make the complexity formidable (a dictionary size of $511 \times 85 = 43435$ columns). However, that is not the most practical way of doing it. At the receiver prior to detection, a time-delay estimation matrix can be formed (size of $N = 511$ columns) for each user and acquire the delays for each user. Then these delay values can be used to form the dictionary matrix (size of $K = 85$ columns). This method has two advantages: the maximum dictionary size is 511 and the

delay estimation is not necessary at each bit duration, since we are not considering a mobile system.

On the other hand, the other detection methods reported in the literature, do not address the issue of time-delay estimation either. Besides, any other delay estimation method that can be used in the other detectors is perfectly well applicable to the proposed approach. The effect of the delay estimation methods to the system performance are beyond the scope of this chapter.

B.2 Conventional Optical-CDMA Receiver Structures

In this section we briefly summarize some of the existing O-CDMA receiver structures. These are the decorrelator receiver, the optical hard limiter, the chip level detector and the interference canceler. The details of these algorithms can be found in [8, 115]. In our discussion, we assume that the desired user is user 1.

B.2.1 Decorrelator Receiver

The decorrelator receiver acts as a matched filter by comparing the input signal with the transmitted signature sequence. The incoming signal is detected at the nonzero positions of the desired user's code by photo-detection and then integrated. Next the output is compared with a threshold in order to estimate the transmitted bit.

There are two types of decorrelator receivers: passive and active. The block diagrams of these decorrelators are given in Fig. B.3 and Fig. B.4 respectively. Passive decorrelators employ w delay lines and combine the output of these delay lines after photo-detection and integration. Although this is a simple and effective method, it requires extremely high

speed electronic circuitry. Hence, active decorrelators are proposed where the decorrelation is achieved totally in optical domain by means of complex optical elements.

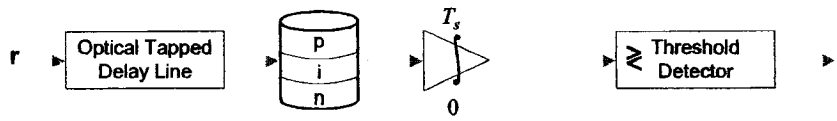


Figure B.3: Block diagram of passive decorrelator.

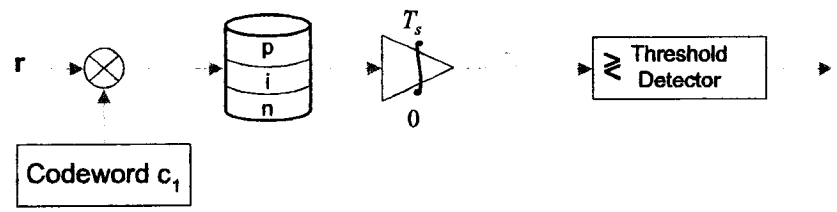


Figure B.4: Block diagram of active decorrelator.

B.2.2 Optical Hard-Limiter

The optical hard-limiter receiver limits the energy in each chip duration to a single transmitted bit energy, hence canceling some of the interference. The input output characteristic of the hard limiter that is employed in the simulations is given in Fig B.5.

The receiver structure is basically the afore-mentioned decorrelator structure preceded by the hard-limiter as shown in Fig. B.6.

For example, if $w = 5$ chip positions are received as $(5, 0, 3, 0, 0)$, a decorrelator would result in a bit estimation of '1' by summing all the positions, since $5 + 3 > w/2$. However

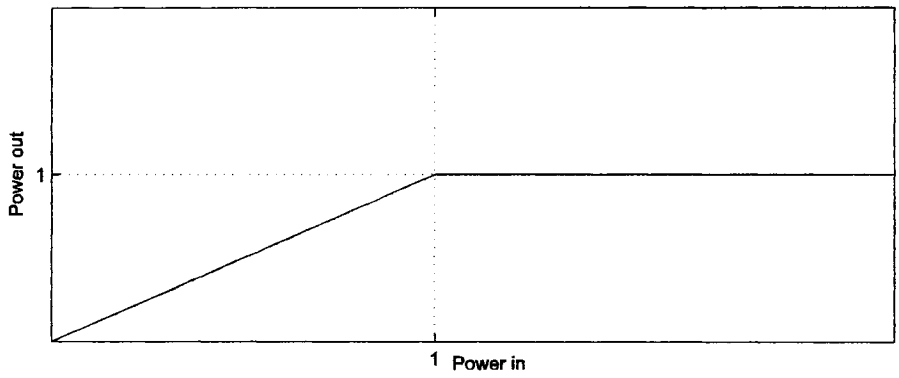


Figure B.5: The input-output characteristics of the optical hard limiter.

Figure B.6: Block diagram of optical hard limiter receiver.

a hard limiter limits the maximum value at each chip duration to '1'. Hence, after the hard-limiter, received signal becomes $(1, 0, 1, 0, 0)$ and a bit estimation of '0' occurs since $2 < w/2$.

B.2.3 Chip-Level Receiver

The chip-level receiver investigates each chip position of the code in the optical domain. If all w chip positions are detected as '1' then a '1' is decided, otherwise a '0' is decided.

The block diagram of the receiver is given in Fig. B.7. After each chip duration is detected by photo-detectors, the signal integrated over a chip duration for all nonzero code locations, then compared to a threshold and either a '1' or '0' is decided for every chip duration. If at

Figure B.7: Block diagram of chip-level receiver.

least one of them is '0', then a '0' is decided. For example, detection of $(5, 8, 3, 1, 0)$ would result in a bit estimation of '0'.

Although it has very good performance especially in long codes, chip-level receiver requires very fast detection circuitry, increasing the system cost. It also has poor performance under low signal to noise ratio since the single chip detections become less reliable.

B.2.4 Interference Canceler

The interference canceler repeatedly applies decorrelator, every time taking the last iteration's decisions into consideration. After an estimate of each user's transmitted bit is acquired with a decorrelator receiver, these estimates are used to generate an estimate of interference on the desired user. This interference estimate is employed in the decorrelator receiver of the desired user. This is accomplished in an iterative way, where at each step, the interference estimates become more reliable. The interference estimate from the previous steps is also employed in the estimation of interference procedure. The level of the interference canceler is given according to the number of iterations. The block diagram of a 2-level interference canceler is given in Fig. B.8. The details about the interference canceler can be found in [8].

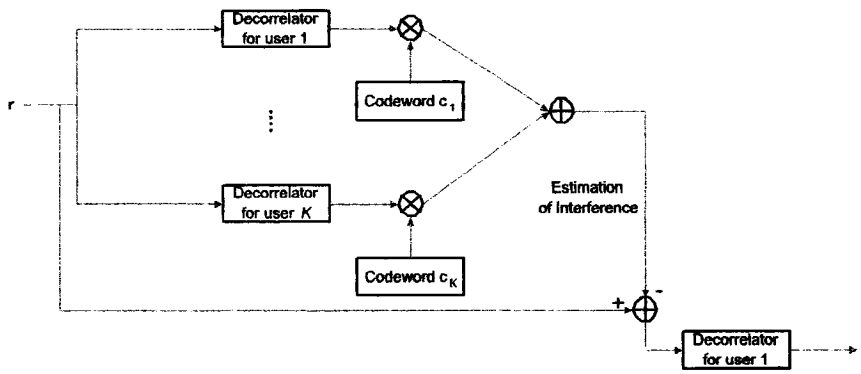


Figure B.8: Block diagram of 2-level interference canceler.

B.3 Simulation Results

In simulations, other than the proposed OMP method, four different receiver structures that are summarized in Section B.2 are considered. These are the decorrelator receiver, the optical hard-limiter, the chip level detector, and the interference canceler.

In the simulations, we build our OOC as discussed in [18], using the parameter set $\Phi(511, 3, 1, 1)$. There are 85 distinct codes available for this set. The codes c_i $(i = 1...85)$ can be represented by their non-zero positions. These positions can be found by the method described in [87] as $\{0, i, x_i + i + 86\}$ where x_i is defined as:

$$x_i = \begin{cases} 43 - j & i = 2j, 1 \leq j \leq 42 \\ 127 - j & i = 2j + 1, 1 \leq j \leq 19 \\ 128 - j & i = 2j + 1, 20 \leq j \leq 41 \\ 43 & i = 85 \\ 86 & i = 43 \\ 148 & i = 1 \end{cases}.$$

Table B.1: O-CDMA System Parameters

O-CDMA Parameters	
Chip duration (T_c)	0.5 ns
Symbol duration(T_s)	0.255 μs
Chip rate (R_c)	2 GHz
Bit rate (R_b)	4 Mbps
Processing gain	511

As a result the 85 codewords are constructed as:

$$\{0,1,235\},\{0,2,130\},\{0,3,215\},\{0,4,131\}...$$

The main imperfections considered are the dark current, the output distribution of the p-i-n diode and the multiple user interference. Since both dark current and output of the laser are assumed to have a Poisson distribution, the probability of a photon count F can be identified solely by the mean photon count μ as:

$$P(F) = e^{-\mu}\frac{\mu^F}{F!}. \tag{B.6}$$

We consider two distinct system models. First we investigate the case without multimode fiber dispersion. Next we employ multimode fibers.

B.3.1 Performances of Receivers without Fiber Dispersion

Fiber dispersion is assumed to be negligible for the moment. At the receiver, we assume for an interval of T_c, the average effect of dark current is on the order of 5 photons. The main system parameters are given in Table B.1.

Initially, we investigate the error performances of the mentioned algorithms, as the number of users increase. As depicted in Fig. B.9, the bit error rates (BER) for hard limiter and decorrelator become unacceptable for highly loaded systems. The OMP based detector's error performance is logarithmically affected by the increase while the others are affected almost linearly. Hence for high number of active users, OMP based detection outperforms the conventional detection methods.

Next, we investigate the effect of the average number of photons received for 32 active users. As can be observed in Fig. B.10, especially for low number of photons, which we can interpret as low signal to noise ratio (SNR), OMP based detector performs better than the conventional methods. This is due to the fact that the OMP algorithm is not very sensitive to the SNR as stated previously in [49, 51].

Finally, in Fig. B.11, we investigate the effect of unequal received power from different users. In this simulation, we also consider 32 simultaneously active users in the system. It is well known that wireless CDMA systems are critically affected by unequal received powers. This phenomena is known as the near-far effect. A similar case of unequal power levels could arise also in optical systems. To the best of authors' knowledge all of the proposed detector architectures investigate the case when all users transmit with equal powers. However, this is not necessarily the case. In our investigation, we keep the power of the desired user the same, while multiplying the powers of the interferers with the coefficient indicated in the x-axis of Fig. B.11. As can be seen in the figure, although the OMP based detector cannot perform as well as optical chip level detector, it out-performs the rest up to a near-far ratio of 6. Also, as in the case of wireless systems, decorrelator without interference canceler performs poorly as the near-far ratio increases.

The hard-limiter detectors discard the extra interference power as a direct result of their input output characteristics shown in Fig. B.5. Hence they are unaffected from the near-far ratio effect. The main problem about the optical hard-limiters is that, once the received energy

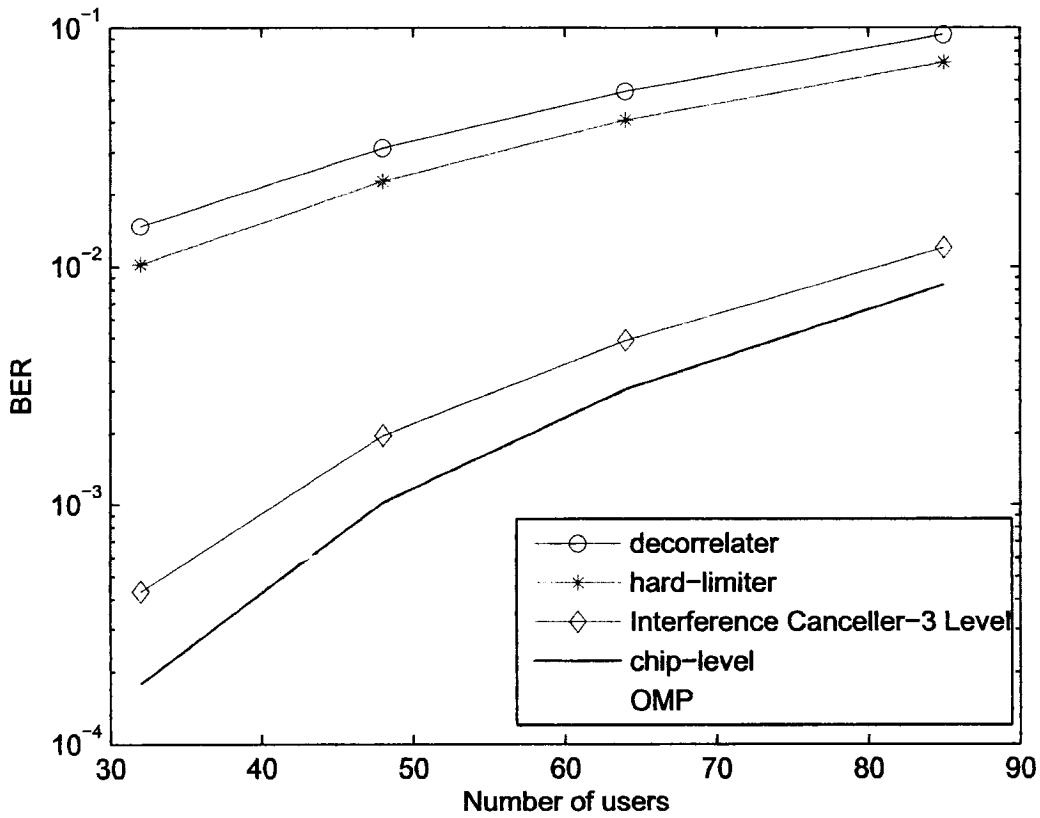


Figure B.9: Dependence of BER to interference for different detection methods.

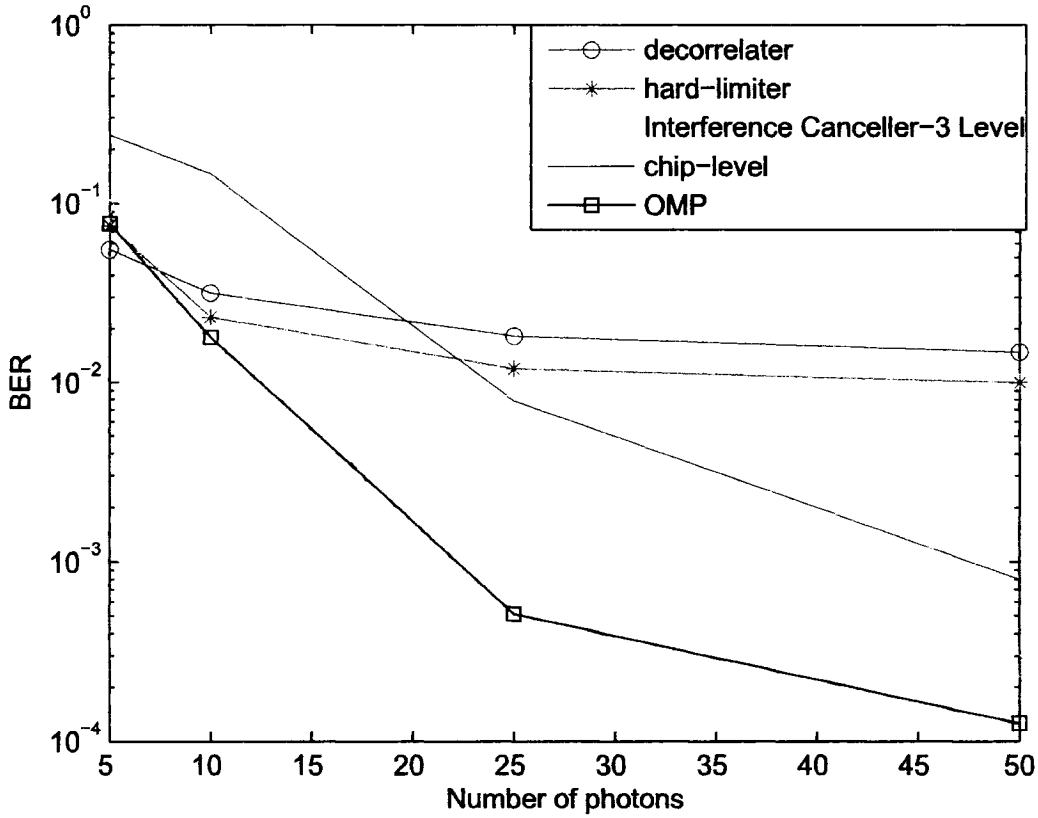


Figure B.10: The effect of number of photons received for different detection methods on BER.

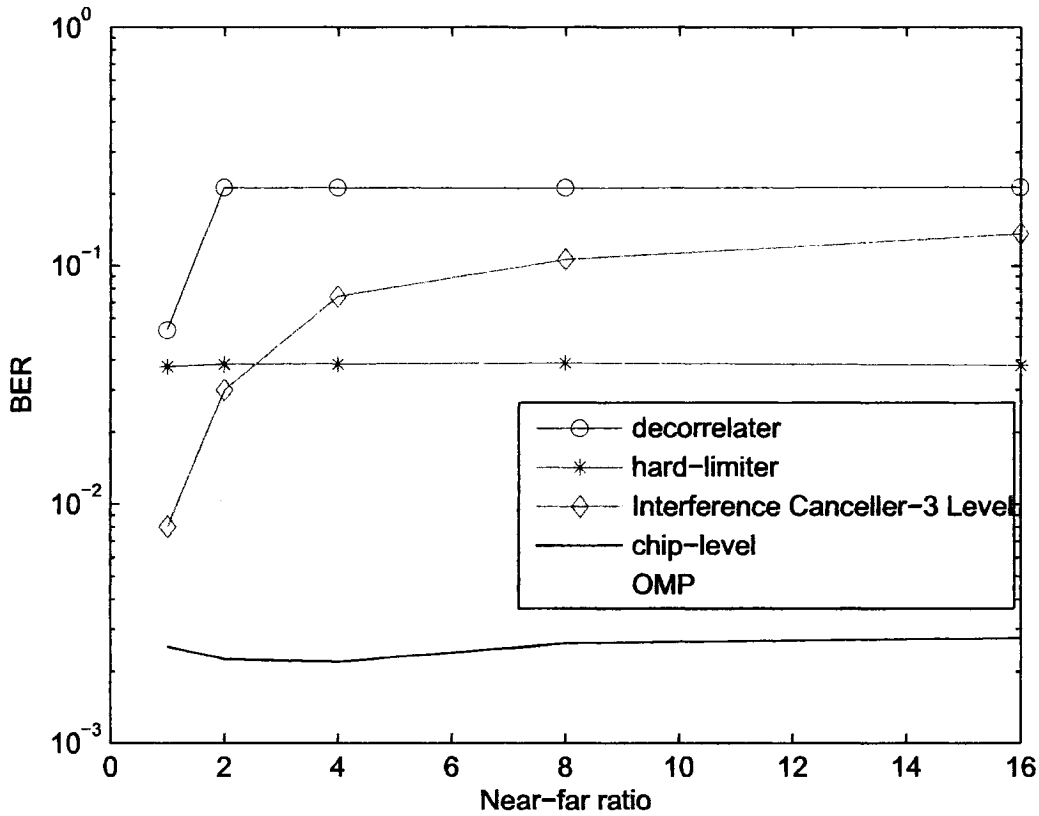


Figure B.11: The near-far effect for different detection methods on BER.

for single chip duration is limited, the possibility of interference cancelation by estimating other user's signals is eliminated. This causes performance loss when the number of active users in the system increases.

Although chip-level detectors perform well, they need to work in an optical chip duration, which is practical (but not cost-efficient) for only all-optical systems. Furthermore, it is well known that chip-level detectors are highly affected by the dark current and the increase in the number of users in the channel.

B.3.2 The Effect of Fiber Dispersion on Receiver Performances

Here we investigate the effect of fiber dispersion on the performances of O-CDMA receiver structures. Recently the effect of dispersion to the O-CDMA systems has been investigated for single mode fibers [46, 80].

Although single mode fiber comes with high performance and much less pulse spread, it is both expensive and fragile. Multimode fibers are cheap, easy to handle and widely deployed in the field. In this dissertation, we consider the short distance multimode fiber communications.

There are mainly two types of multimode fibers: step index and graded-core index [43]. The graded-core index is the second generation optical fibers, developed for canceling most of the pulse spread by gradually changing the core index of the fiber according to a profile, the most popular one being α-profile [38]. An α-profile has the index of refraction profile defined by:

$$n(r) = \begin{cases} n_0[1 - 2\Delta(r/a)^\alpha]^{1/2} & r < a \\ n_0[1 - 2\Delta]^{1/2} & r > a \end{cases}, \tag{B.7}$$

where a is the core radius of the fiber, n_0 is the maximum refraction index in the core region, n_c is the refraction index in the cladding region, r is the distance from the origin of the fiber

Table B.2: Optical Fiber Parameters

Parameters for dispersion	
Maximum Refraction Index	1.5
Laser Frequency	1320 nm
Δ	0.02
α	2.05
Maximum error in the profile	0.003
Fiber length	500 m

and Δ can be given by:

$$\Delta = \frac{n_0^2 - n_c^2}{2n_c^2}. \tag{B.8}$$

Hence the value of α effects the decreasing index profile inside the core region of the fiber. In [38] more information about α-profile multimode fibers is given and the impulse response is formulated.

In our simulations, the fiber impulse response given in [38] is adopted as:

$$h(t) = \begin{cases} \frac{\alpha+2}{\alpha} \left| \frac{\alpha+2}{\Delta(\alpha-2)} \right|^{(2/\alpha)+1} |t|^{2/\alpha} & \text{except for } \alpha \approx 2 \\ \frac{2}{\Delta^2} & \text{for } \alpha \approx 2 \end{cases} \tag{B.9}$$

Time t changes from 0 to T;

$$T = \begin{cases} \frac{\alpha-2}{\alpha+2}\Delta & \text{except for } \alpha \approx 2 \\ \frac{\Delta^2}{2} & \text{for } \alpha \approx 2 \end{cases} \tag{B.10}$$

where Δ is a function of refractive indexes of the core and the cladding regions of the fiber.

By employing this model and the system parameters given in Table B.2, we compare the performances of O-CDMA detectors under fiber dispersion. In our case, we consider the short

distance, high data rate and low cost optical fibers. We investigate the 32 user case in Fig. B.12 and 16 user case in Fig. B.13. As it is seen in Fig. B.12 and Fig. B.13, OMP based detection outperforms all of the other algorithms under dispersion as well.

There is a performance loss in both interference canceler and OMP methods under dispersion. However in this case no dispersion compensation was considered. The results can easily be extended to dispersion compensation case. A simple method for compensation can be adding a few of the cyclic-shifted versions of the chip sequences to the dictionary $\mathcal{D}$.

B.4 2-D Optical Codes

In this section, we extend the BS based detection scheme to two dimensional (2-D) O-CDMA systems presented in [94], namely temporal/spatial single pulse per row (T/S-SPR) codes. When 2-D codes are used instead of 1-D, the cross-correlation is reduced. Also autocorrelation sidelobe problem of the 1-D does not exist in 2-D codes [100]. In addition to those, the cardinality of the code family is higher in 2-D. The 2-D systems also prove to be more secure than 1-D systems.

B.4.1 2-D System Model

An O-CDMA system with K users is considered as shown in Fig. B.14. In the figure, each user transmits the information sequence $b_i \in \{0,1\}$, for $i = 1,2,\ldots K$. Optical orthogonal codes (OOC) of alternating (0,1) sequences are employed. Each user spreads its code in both space and time. Sticking to the notation of [94], T/S-SPR codes are defined as $\Phi(N,R,p,w,\lambda_a,\lambda_b)$, where N is the code length, w is the code weight (i.e. the number of '1's in a block of N), and they satisfy the correlation properties for any code sequence $\mathbf{c}_i, \mathbf{c}_j \in \Phi$ as defined in [94]. If we define the autocorrelation and cross correlation of 2-D

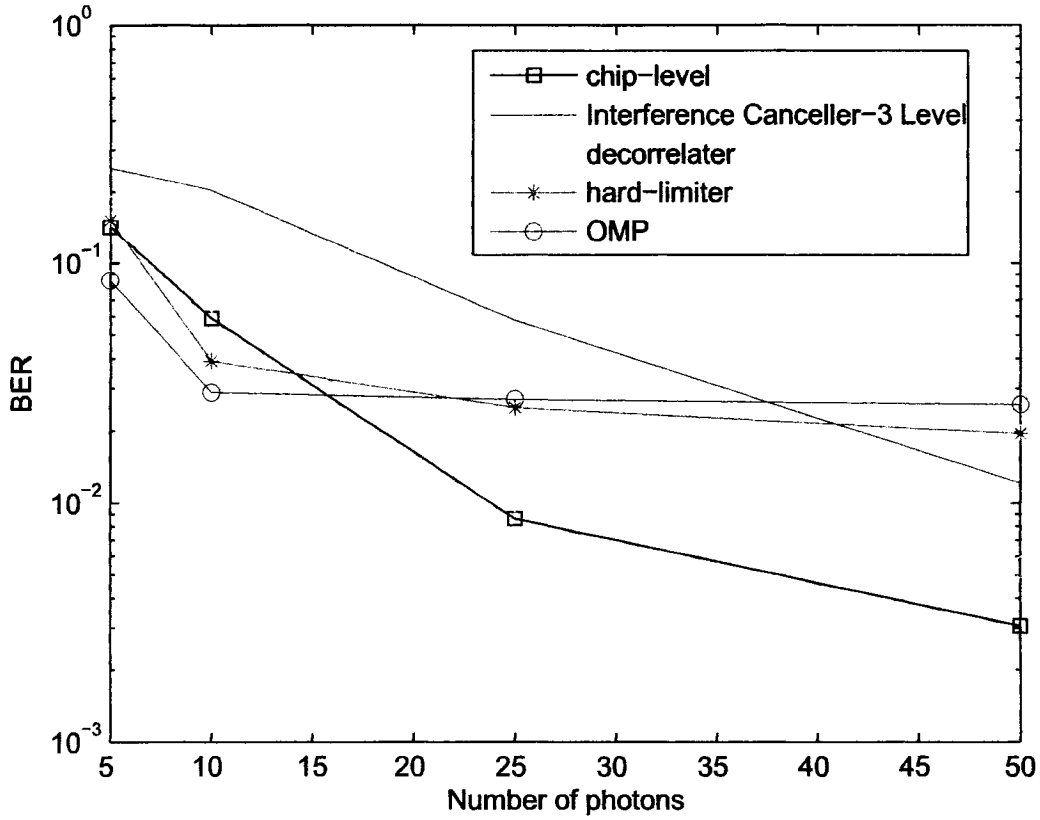


Figure B.12: The effect of number of photons received for different detection methods on BER under the effect of fiber dispersion (32 users).

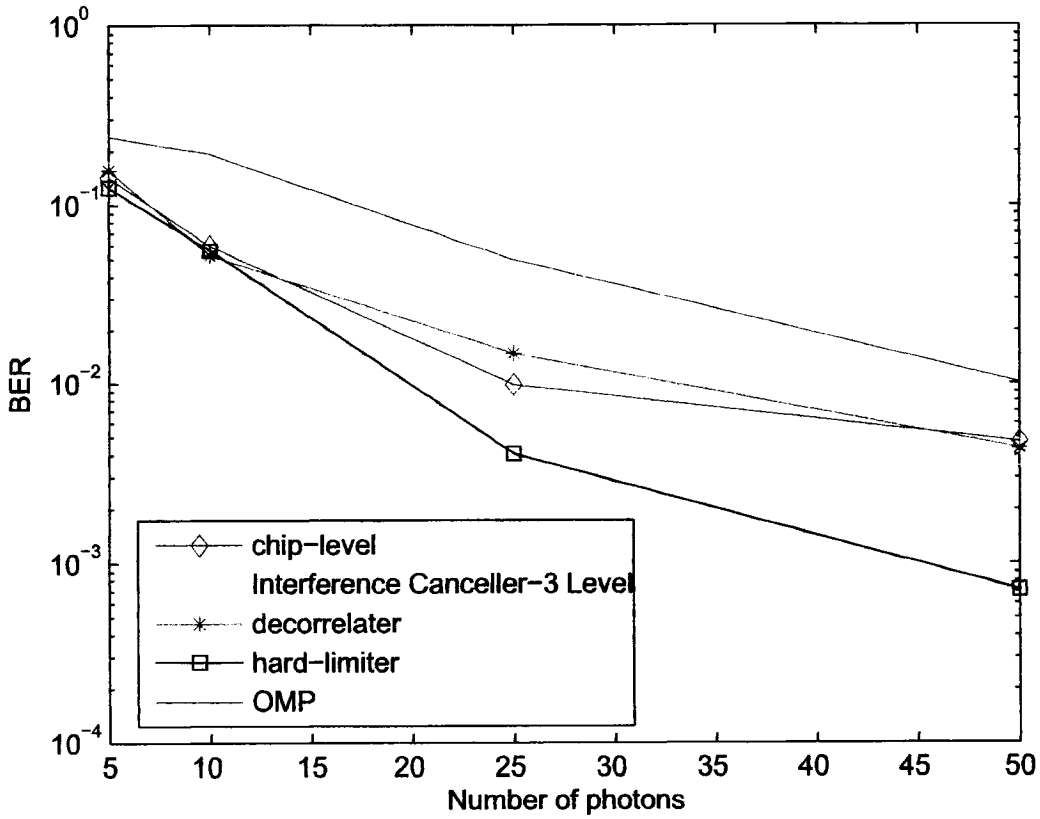


Figure B.13: The effect of number of photons received for different detection methods on BER under the effect of fiber dispersion (16 users).

codes $x(t)$ and $y(t)$ as

$$\begin{aligned} Z_{x,x}(l) &= \sum_{m=0}^{R-1}(\sum_{n=0}^{N} x_{m,n}x_{(m,n+l)_{modN}}) \\ Z_{x,y}(l) &= \sum_{m=0}^{R-1}(\sum_{n=0}^{N} x_{m,n}y_{(m,n+l)_{modN}}), \end{aligned} \tag{B.11}$$

then we can write the necessary conditions as

$$\begin{aligned} Z_{x,x}(l) &= W \quad l = 0 \\ Z_{x,x}(l) &\leq \lambda_a \quad 1 \leq l \leq N-1 \\ Z_{x,y}(l) &\leq \lambda_b \quad 0 \leq l \leq N-1. \end{aligned} \tag{B.12}$$

The number of spatial components is denoted by R, and p is the number of nonzero elements in R.

The baseband representation of the received O-CDMA signal after coherent reception is given by

$$s(t) = \sum_{n=-\infty}^{\infty}\sum_{k=1}^{K} A_k\bar{c}_k(t-nT_s)b_k(n) + d(t), \tag{B.13}$$

where $d(t)$ is the dark current. The transmitted amplitude for user k is A_k. The symbol sequence of the kth user, $b_k(n)$ is independent and identically distributed and modulated by on-off keying (OOK). In the equation above, the effective spreading waveform of the kth user is represented by $\bar{c}_k(t)$. The transmitter of user k is a coherent laser with amplitude A_k. The output of the coherent laser can be modeled as a Poisson process. As the optical detector, we consider the p-i-n diode structure. The dark current of a p-i-n diode is modeled as a Poisson process as well.

In order to apply the OMP algorithm to 2D O-CDMA detection, we use the same solution methodology employed for 1D case. However the 2D codes are converted to 1D by appending their columns to a single colum and stacking them as column vectors of the dictionary matrix.

B.4.2 Simulation Results

We build our simulations as discussed in [94], using the parameter set $\Phi(49, 7, 1, 7, 1, 1)$. The users are assumed to be asynchronous. The main imperfections considered are the dark current, the output distribution of the p-i-n diode and the multiple user interference. Since both dark current and output of the laser is assumed to have a Poisson distribution. Hence the probability of a photon count F can be identified solely by the mean photon count μ as:

$$P(F) = e^{-\mu}\frac{\mu^F}{F!}. \tag{B.14}$$

In Fig. B.15, performance with respect to received number of photons is presented. The OMP method is shown to be much more efficient than the hard-limiter that was employed in [94] at low received power levels, which are critical in terms of bit error rate . There are 7 simultaneous users in the system. As shown in Fig. B.15, the performance of OMP based detector is much better than decorrelator and limiter receiver structures and comparable to the performance of chip-level detector.

Next we investigate the effect of number of users. Here we assume, on the average 25 photons are received.

As depicted in Fig. B.16, as the number of users increase, OMP outperforms the other detection techniques.

Up to this point, we assumed, the received power from all users are equal. In Fig. B.17 we consider the case where the desired user has less power than the others. The difference is referred to as near-far ratio.

In Fig. B.17, it is shown that OMP based detector is more resistant to near-far ratio problem, which is a very frequent problem in CDMA systems.

Although in some cases, OMP based detector performs worse than the chip-level detector,

the main advantage of OMP is its complexity. Although it has a similar complexity to the other detectors, OMP based detector recovers the transmitted data from all users, where as the other detectors only detect a single user. Hence OMP detector has the critical advantage of multi-user detection.

B.5 Conclusions

In this appendix, we proposed a novel O-CDMA detector employing OMP algorithm. The proposed detector performs well under both low SNR and high user load conditions. It is very flexible in terms of extending the algorithm to more general cases such as asynchronous O-CDMA or dispersion cancelation by simply modifying the dictionary matrix stored in the receiver.

Furthermore, the proposed detector is more robust to users with unequal received power than the other methods in the literature. This is extremely important for the cases where near-far effect is an issue, and power control is hard to accomplish.

It also has a complexity in the order of a two-level interference canceler. The comparison of complexities of the basis selection methods to some other algorithms in the literature were given in [51].

The results show that the proposed method is a promising candidate with its high performance under the extreme conditions such as low SNR, unequally distributed user powers and high number of simultaneous users. The proposed method's performance can be improved further by employing more complex basis selection algorithms with the complexity-performance trade-off.

Next, we proposed a novel O-CDMA detector employing OMP algorithm and 2-D codes. The proposed detector performs well under low SNR a conditions. It is very flexible in terms of

extending the algorithm to more general cases such as fiber-dispersion cancelation by simply modifying the dictionary matrix stored in the receiver. The main advantage of the proposed detector lies in the fact that it is a multi-user detector with a performance that is close to single user systems. Simulation results indicate that the proposed scheme is very robust to near-far ratio and performs well with high number of simultaneous users.

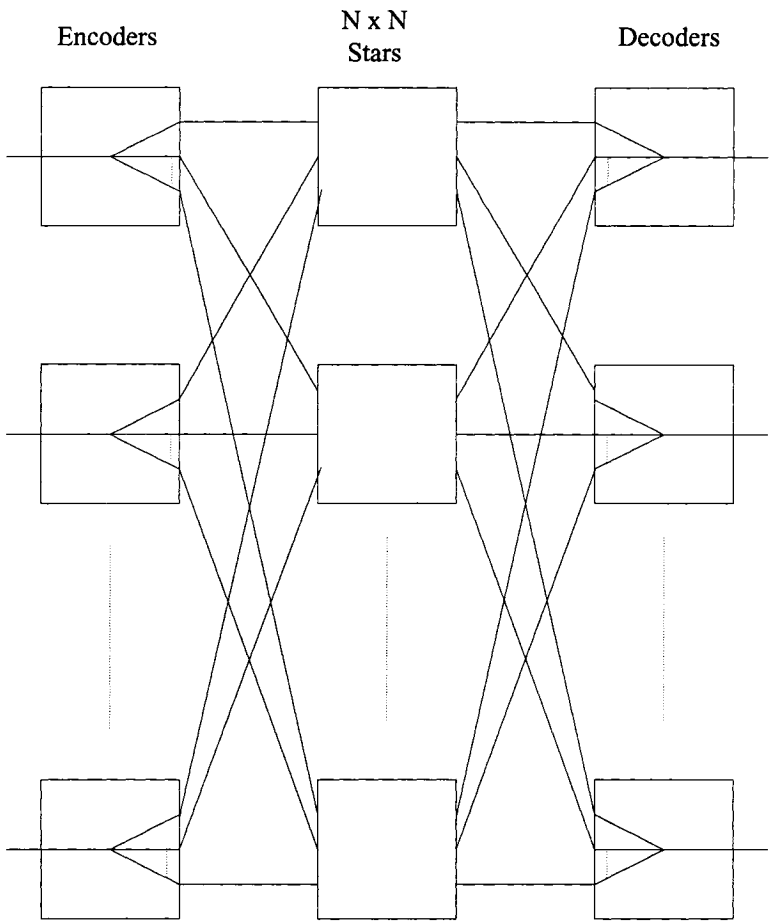


Figure B.14: Multiuser O-CDMA system model

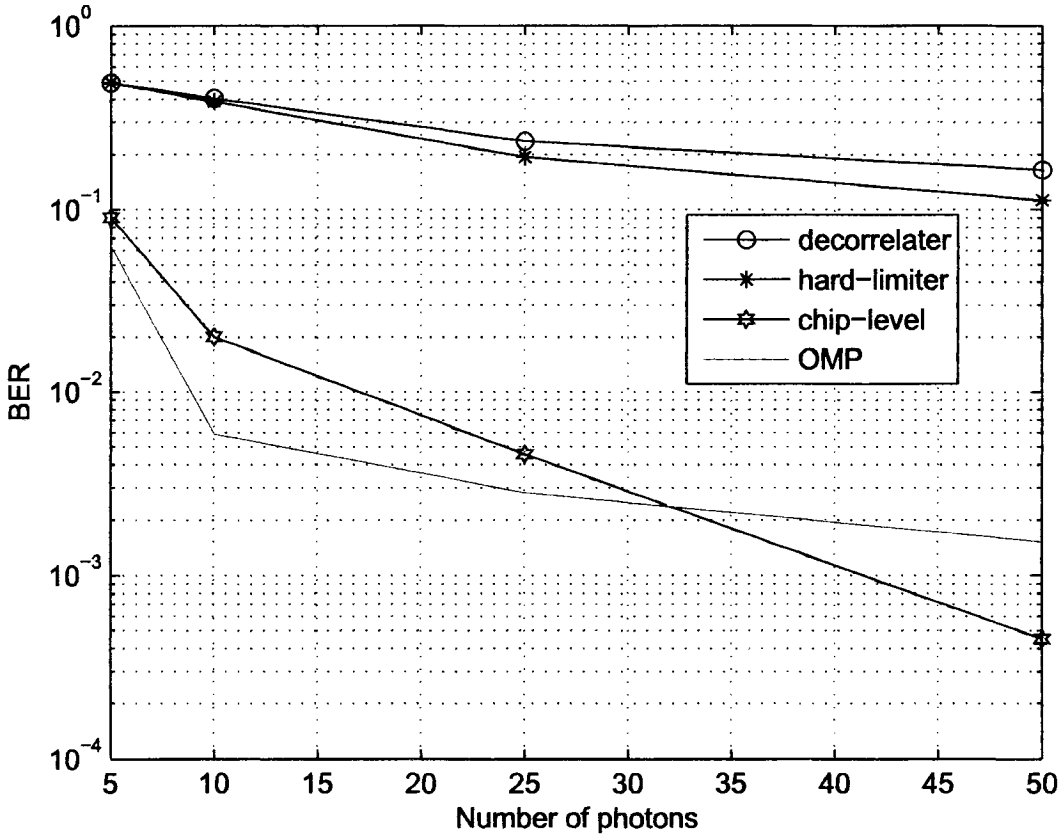


Figure B.15: Bit error rate performances of detectors with varying number of photons.

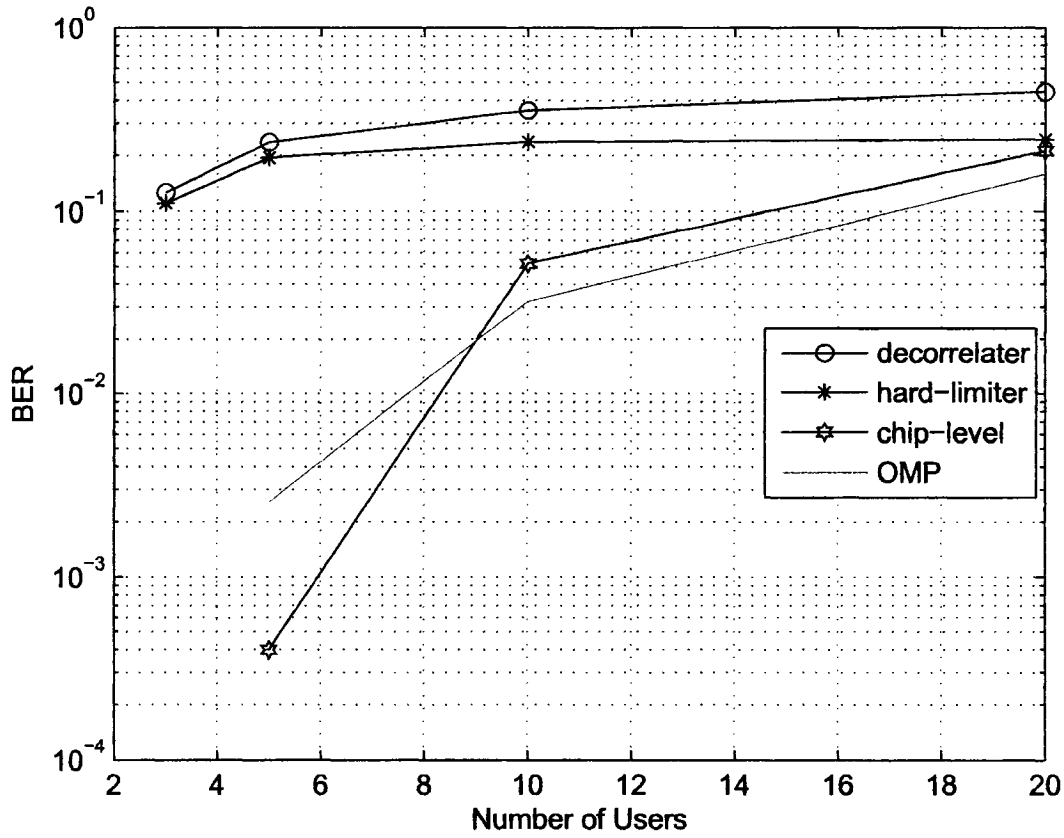


Figure B.16: Bit error rate performances of detectors with varying number of users.

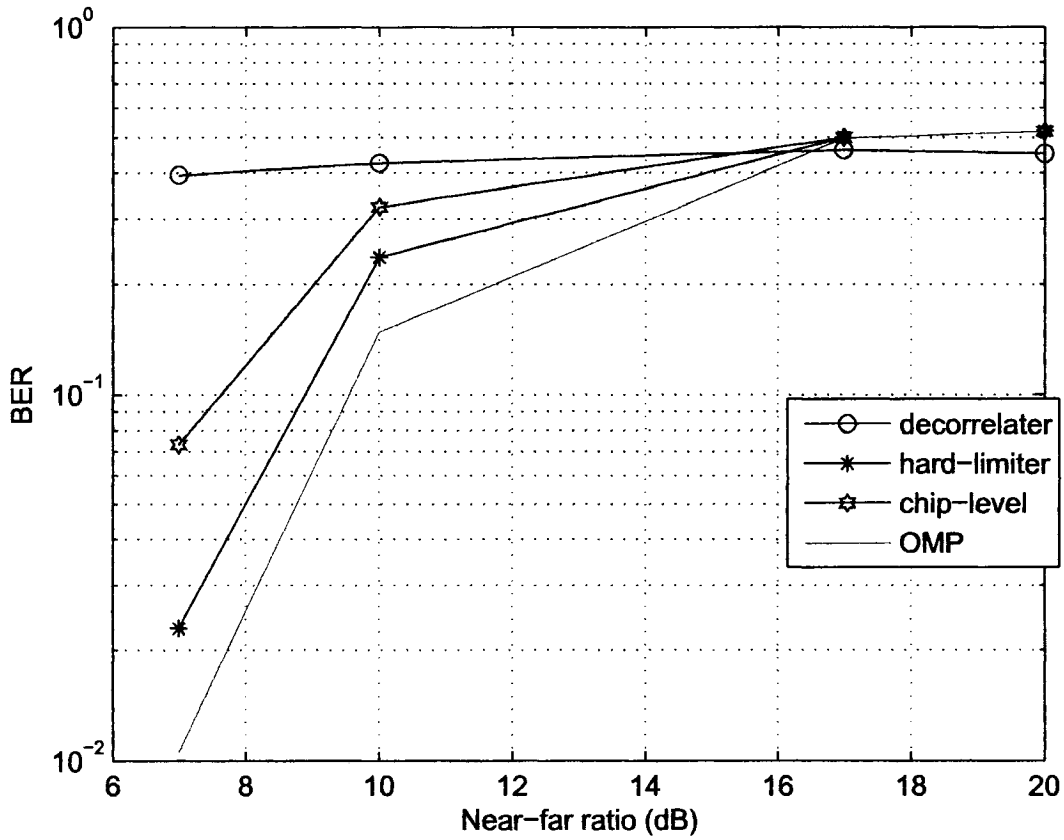


Figure B.17: Bit error rate performances of detectors with varying levels of near-far effect.

Appendix C

List of Publications

PEER REVIEWED JOURNAL PUBLICATIONS

1. Güneş Karabulut, Tolga Kurt and Abbas Yongaçoğlu, "Estimation of Direction-of-Arrival by Matching Pursuit (EDAMP)", EURASIP Journal on Wireless Communications and Networking - Special Issue on Advanced Signal Processing Algorithms for Wireless Communications, Vol.2 , April 2005, pp. 197 - 205.

2. Güneş Karabulut, Tolga Kurt and Abbas Yongaçoğlu, "Optical CDMA Detection by Sequential Basis Selection Algorithms", Journal of Lightwave Technology, Vol.23, No.11, November 2005 pp. 3708 - 3715.

3. Güneş Karabulut and Abbas Yongaçoğlu, "A New Perspective for Successive Interference Cancellation in CDMA Systems: Greedy Multi-user Detection", submitted to IEEE Transactions on Signal Processing.

4. Güneş Karabulut, Lucia Moura, Daniel Panario and Abbas Yongaçoğlu, "Flexible Tree-Search for Matching Pursuit Algorithms", accepted by IEE Proceedings, Vision, Image and Signal Processing.

5. Güneş Karabulut, Tolga Kurt and Abbas Yongaçoğlu, "Angle of Arrival Detection by Basis Selection Algorithms", submitted to IEEE Transactions on Wireless Communications.

6. Güneş Karabulut, Tolga Kurt and Abbas Yongaçoğlu, "A New Perspective for Successive Interference Cancellation in CDMA Systems: Greedy Multi-user Detection", submitted to IEEE Transactions on Signal Processing.

INTERNATIONAL CONFERENCE PUBLICATIONS

1. Güneş Karabulut, Tolga Kurt and Abbas Yongaçoğlu, "Blind Channel Estimation and Multi-user Detection for Wireless CDMA Using Matching Pursuit", International Conference on Communications (ICC), Istanbul, Turkey, June 2006.

2. Güneş Karabulut, Tolga Kurt and Abbas Yongaçoğlu, "Applications of Matching Pursuit Algorithms to Telecommunication Problems", IEEE-SIU2006, Antalya, Turkey, April, 2006.

3. Güneş Karabulut and Abbas Yongaçoğlu, "Adaptive Resolution Angle of Arrival Estimation", European Conference on Signal Processing (EUSIPCO), Antalya, Turkey, September 2005.

4. Güneş Karabulut and Abbas Yongaçoğlu, "Estimation of Fast Fading Channels with Orthogonal Matching Pursuit Algorithm", International Conference on Sampling Theory and Applications (SAMPTA), Samsun, Turkey, July 2005.

5. Güneş Karabulut, Tolga Kurt and Abbas Yongaçoğlu, "2-D Optical CDMA Detection by Orthogonal Matching Pursuit Algorithm", Canadian Workshop on Information Theory (CWIT), Montreal, Canada, June 2005.

6. Güneş Karabulut, Tolga Kurt and Abbas Yongaçoğlu, "High Resolution Estimation of Directions of Arrival'", IEEE Vehicular Technology Conference (VTC), Stockholm, Sweden, May 2005.

7. Güneş Karabulut and Abbas Yongaçoğlu, "Estimation of Fast Fading Channels with Orthogonal Matching Pursuit Algorithm", IEEE Sarnoff Symposium, Princeton, USA, April 2005.

8. Tolga Kurt, Güneş Karabulut and Abbas Yongaçoğlu, "Optical CDMA Detection by Orthogonal Matching Pursuit", Int. Conf. on Acoustics, Speech, and Signal Processing (ICASSP) , Philadelphia, USA, March 2005.

9. Güneş Karabulut, Lucia Moura, Daniel Panario and Abbas Yongaçoğlu, "Flexible Tree-Search based Orthogonal Matching Pursuit", Int. Conf. on Acoustics, Speech, and Signal Processing (ICASSP), Philedelphia, USA, March 2005.

10. Güneş Karabulut, Tolga Kurt and Abbas Yongaçoğlu, "Angle of Arrival Detection by Matching Pursuit Algorithm", IEEE 60th Vehicular Technology Conference (VTC), Los Angeles, USA, September 2004.

11. Güneş Karabulut, Tolga Kurt and Abbas Yongaçoğlu, "Sparse Channel Estimation by Matching Pursuit Algorithm", IEEE 60th Vehicular Technology Conference (VTC), Los Angeles, USA, September 2004.

12. Güneş Karabulut and Abbas Yongaçoğlu,"Multiple Description Coding Using Orthogonal Matching Pursuit", Third Annual Mediterranean Ad-Hoc Networking Workshop (MED-HOC), Bodrum, Turkey, June 2004.

Bibliography

[1] S. Adireddy, Lang Tong, and H. Viswanathan. Optimal placement of training for frequency-selective block-fading channels. *IEEE Trans. on Info. Theory*, 48(8):2338–2353, Aug. 2002.

[2] J. Adler, B.D. Rao, and K. Kreutz-Delgado. Comparison of basis selection methods. In *Asilomar Conference on Signals, Systems and Computers*, volume 1, pages 252–257, 1996.

[3] E. Aktas and U. Mitra. Blind channel estimation for multi-user CDMA systems. *IEEE International Conference on Communications*, pages 1064–1068, June 1998.

[4] O.K. Al-Shaykh, E. Miloslavsky, T. Nomura, R. Neff, and A. Zakhor. Video compression using matching pursuits. *IEEE Trans. Circuit Sys. Video Tech.*, 9(1):123–143, February 1999.

[5] M. Andrle, L. Rebollo-Neira, and E. Sagianos. Backward-optimized orthogonal matching pursuit approach. *IEEE Signal Proc. Letters*, 11(9):705–708, Sept. 2004.

[6] D.K. Borah. Estimation of fading channels with a parallel matching pursuit structure. In *IEEE Vehicular Technology Conference, Fall*, 2004.

[7] D.K. Borah and B.D. Hart. Frequency-selective fading channel estimation with a polynomial time-varying channel model. *IEEE Transactions on Communications*, 47(6):862–873, June 1999.

[8] M. Brandt-Pearce and B. Aazhang. Multiuser detection for optical code division multiple access systems. *IEEE Transactions on Communications*, 42(2/3/4):1801–1810, Feb./Mar./Apr. 1994.

[9] R.M. Buehrer, N.S. Correal-Mendoza, and B.D. Woerner. A simulation comparison of multiuser receivers for cellular CDMA. *IEEE Trans. on Vehicular Tech.*, 49(4):1065–1085, July 2000.

[10] M. Buhren, M. Pesavento, and J.E. Bohme. A new approach to array interpolation by generation of artificial shift invariances: interpolated ESPRIT. In *IEEE International Conference on Acoustics, Speech, and Signal Processing*, volume 5, pages 205–208, 2003.

[11] J.K. Cavers. An analysis of pilot symbol assisted modulation for Rayleigh fading channels. *IEEE Trans. on Vehicular Tech.*, 40(4):686–693, Nov. 1991.

[12] P. Charge, Y. Wang, and J. Saillard. An extended cyclic MUSIC algorithm. *IEEE Trans. on Signal Proc.*, 51(7):1695–1701, July 2003.

[13] S. Chen and J. Wigger. Fast orthogonal least squares algorithm for efficient subset model selection. *IEEE Transactions. on Signal Proc.*, July 1995.

[14] S.S. Chen and D.L. Donoho. Application of basis pursuit in spectrum estimation. In *IEEE International Conference on Acoustics, Speech, and Signal Processing*, volume 3, pages 1865–1868, 1998.

[15] S.S. Chen, D.L. Donoho, and M.A. Saunders. Atomic decomposition by basis pursuit. *SIAM Review*, 43(1):129–159, 2001.

[16] V. Cherkassky and F. Mulier. *Learning from Data: Concepts, Theory, and Methods.* Wiley-Interscience, 1998.

[17] O. Christensen. *An Introduction to Frames and Riesz Bases.* Birkhauser, 2003.

[18] F. R. K. Chung, J. A. Salehi, and V. K. Wei. Optical orthogonal codes: design, analysis and applications. *IEEE Transactions on Information Theory*, 35(3):595–604, May 1989.

[19] R.R. Coifman and M.V. Wickerhauser. Entropy-based algorithms for best basis selection. *IEEE Trans. on Info. Theory*, 38(2):713–718, March 1992.

[20] N.S. Correal, S.F. Swanchara, and B.D. Woerner. Implementation issues for multiuser DS-CDMA receivers. *International Journal of Wireless Information Networks*, 5(3):257–279, July 1998.

[21] S.F. Cotter. *Subset Selection Algorithms with Applications.* PhD thesis, University of California, San Diego, 2001.

[22] S.F. Cotter and B.D. Rao. The adaptive matching pursuit algorithm for estimation and equalization of sparse time-varying channels. In *Asilomar Conference on Signals, Systems and Computers*, volume 2, pages 1772–1776, 2000.

[23] S.F. Cotter and B.D. Rao. Matching pursuit based decision-feedback equalizers. In *IEEE International Conference on Acoustics, Speech, and Signal Processing*, volume 5, pages 2713–2716, 2000.

[24] S.F. Cotter and B.D. Rao. Application of tree-based searches to matching pursuit. In *IEEE International Conference on Acoustics, Speech, and Signal Processing*, volume 6, pages 3933–3936, 2001.

[25] S.F. Cotter and B.D. Rao. Sparse channel estimation via matching pursuit with application to equalization. *IEEE Trans. on Comm.*, 50(3):374–377, March 2002.

[26] S.N. Crozier, D.D. Falconer, and S.A Mahmoud. Least sum of squared errors (LSSE) channel estimation. *IEE Proceedings Radar and Signal Processing*, 138(4):371–378, Aug 1991.

[27] Z. Cvetkovic and M. Vetterli. Overcomplete expansions and robustness. In *IEEE-SP International Symposium on Time-Frequency and Time-Scale Analysis*, pages 325–328, 1996.

[28] I. Daubechies. Time-frequency localization operators: a geometric phase space approach. *IEEE Trans. on Info. Theory*, 34(4):605–612, July 1988.

[29] G. Davis and S. Mallat. Wavelet vector quantization with matching pursuit. In *IEEE-IMS Workshop on Information Theory and Statistics*, volume 3, page 55, 1994.

[30] G. Davis, S. Mallat, and Z. Zhang. Adaptive time-frequency decompositions. *Opt. Eng.*, 33(7):2183–2191, July 1994.

[31] D.L. Donoho and M. Elad. Optimally sparse representation in general (nonorthogonal) dictionaries via l^1 minimization. *PNAS*, 100(3):2197–2202, March 2000.

[32] D.L. Donoho and X. Huo. Uncertainty principles and ideal atomic decomposition. *IEEE Trans. on Info. Theory*, 47(7):2845–2862, Nov. 2001.

[33] J.H. Friedman and W. Stuetzle. Projection pursuit regression. *J. of Amer. Stat. Assoc.*, 76:817–823, 1981.

[34] A. El Gamal and T.M Cover. Multiple user information theory. *Proceedings of IEEE*, 68:1466–1483, Dec. 1980.

[35] A. Gersho and R.M. Gray. *Vector Quantization and Signal Compression*. Kluwer Academic Publishers, 1992.

[36] G.B. Giannakis and C. Tepedelenlioglu. Basis expansion models and diversity techniques for blind identification and equalization of time-varying channels. *Proceedings of IEEE*, 86:1969–1986, Oct. 1998.

[37] A.A. Giordano and F.M. Hsu. *Least Square Estimation with Applications to Digital Signal Processing*. John Wiley & Sons, 1985.

[38] D. Gloge and E. A. J. Marcatili. Multimode theory of graded-core fibers. *Bell Systems Technical Journal*, pages 1563–1578, Nov. 1973.

[39] L.C. Godara. Application of antenna arrays to mobile communications, part II: Beamforming and direction of arrival considerations. *Proceedings of IEEE*, 85:1195–1245, August 1997.

[40] I.F. Gorodnitsky and B.D. Rao. Sparse signal reconstruction from limited data using focuss: a re-weighted minimum norm algorithm. *IEEE Transactions. on Signal Proc.*, 45(3):600–616, Mar. 1997.

[41] V.K. Goyal, M. Vetterli, and N.T. Thao. Quantized overcomplete expansions in R^N: analysis, synthesis, and algorithms. *IEEE Trans. on Info. Theory*, 44(1):16–31, Jan. 1998.

[42] R. Gribonval and E. Bacry. Harmonic decomposition of audio signals with matching pursuit. *IEEE trans. on Signal Processing*, 51(1):101–111, Jan. 2003.

[43] J. Hecht. *Understanding Fiber Optics*. Prentice Hall, 2002.

[44] J.M. Holtzman. DS/CDMA successive interference cancellation. In *IEEE International Symposium on Spread Spectrum Techniques and Applications*, volume 1, pages 69–78, 1994.

[45] J.M. Holtzman. Successive interference cancellation for direct sequence code division multiple access. In *IEEE Military Communications Conference*, volume 3, pages 997–1001, 1994.

[46] Y. Hur, M. Maeng, E. Gebara, K. Lim, and J. Laskar. A novel optical CDMA system using PAM-4 signaling with improved chromatic dispersion tolerance and spectral efficiency. In *33rd European Microwave Conference*, volume 2, pages 603–606, 2003.

[47] E. Inaty, P. Fortier, and L. A. Rusch. SIR performance evaluation of a multi-rate OFFH-CDMA system. *IEEE Communication Letters*, 5(5):1485–1487, May. 2001.

[48] E. Inaty, H. M. H. Shalaby, P. Fortier, and L. A. Rusch. Multirate optical fast frequency hopping CDMA system using power control. *Journal of Lightwave Technology*, 20(2):166–177, Feb. 2002.

[49] G. Karabulut, T. Kurt, and A. Yongaçoglu. Angle of arrival detection by matching pursuit algorithm. In *Vehicular Technology Conference, Fall, Los Angeles, USA*, volume 1, pages 324–328, September 2004.

[50] G. Karabulut, T. Kurt, and A. Yongaçoglu. 2-D optical CDMA detection by orthogonal matching pursuit algorithms. In *Canadian Workshop on Information Theory, Montreal, Canada*, June 2005.

[51] G. Karabulut, T. Kurt, and A. Yongaçoglu. Estimation of direction-of-arrival by matching pursuit (EDAMP). *Eurasip Journal on Wireless Communications and Networking, Special Issue on Advanced Signal Processing Algorithms for Wireless Communications*, 2:197–205, April 2005.

[52] G. Karabulut, T. Kurt, and A. Yongaçoglu. High resolution estimations of directions of arrival. In *IEEE Vehicular Technology Conference, Stockholm, Sweden*, May 2005.

[53] G. Karabulut, T. Kurt, and A. Yongaçoglu. Blind channel estimation and multi-user detection for wireless CDMA using matching pursuit. In *International Conference on Communications, Istanbul, Turkey*, June 2006.

[54] G. Karabulut, L. Moura, D. Panario, and A. Yongaçoglu. Flexible tree-search based orthogonal matching pursuit algorithm. *IEE Proceedings, Vision, Image and Signal Processing.*, Accepted for Publication.

[55] G. Karabulut, L. Moura, D. Panario, and A. Yongaçoglu. Flexible tree-search based orthogonal matching pursuit algorithm. In *IEEE International Conference on Acoustics, Speech, and Signal Processing, Philadelphia, USA*, volume 4, pages 673–676, March 2005.

[56] G. Karabulut and A. Yongaçoglu. Sparse channel estimation using orthogonal matching pursuit algorithm, los angeles, usa. In *Vehicular Technology Conference, Fall*, September 2004.

[57] G. Karabulut and A. Yongaçoglu. Estimation of time-varying channels with orthogonal matching pursuit algorithm. In *IEEE Sarnoff Symposium, Princeton, USA*, April 2005.

[58] G. Karabulut and A. Yongaçoglu. Training sequence design for channel estimation with orthogonal matching pursuit. In *International Conference on Sampling Theory and Applications (SAMPTA), Samsun, Turkey*, July 2005.

[59] G. Karabulut, A. Yongaçoglu, and T. Kurt. Adaptive resolution angle of arrival estimation. In *European Conference on Signal Processing, Antalya, Turkey*, pages 197–205, September 2005.

[60] G. Z. Karabulut, T. Kurt, and A. Yongaçoglu. Optical CDMA detection by matching pursuit algorithms. *Journal of Lightwave Technology*, 23(11):3708–3715, November 2005.

[61] A. Keshavarzian and J. A. Salehi. Optical orthogonal code acquisition in fiber-optic CDMA systems via the simple serial-search method. *IEEE Transactions on Communications*, 50(3):473–483, Mar. 2002.

[62] B. K. Kim, S. Park, Y. Yeon, and B. W. Kim. Radio-over-fiber system using fiber-grating-based optical CDMA with modified PN codes. *IEEE Photonics Technology Letters*, 15(10):1485–1487, Oct. 2003.

[63] R. Kohno and M. Hatori. Cancellation techniques of co-channel interference in asynchronous spread spectrum multiple access systems. *Electronics and Communications in Japan*, 66-A(5), 1983.

[64] A.M. Kondoz. *Digital Speech*. Wiley, 1996.

[65] D. Kreher and D. Stinson. *Combinatorial Algorithms: Generation, Enumeration and Search*. CRC Press, 1998.

[66] K. Kreutz-Delgado and B.D. Rao. Measures and algorithms for best basis selection. In *IEEE International Conference on Acoustics, Speech, and Signal Processing*, volume 3, pages 1881–1884, 1998.

[67] T. Kurt, G. Z. Karabulut, and A. Yongaçoglu. Optical CDMA detection by orthogonal matching pursuit. *IEEE International Conference on Acoustics, Speech, and Signal Processing, Philadelphia, USA*, March 2005.

[68] W. C. Y. Lee. *Mobile Communications Engineering*. Mc-Graw Hill, 1998.

[69] Jingdong Lin, J.G. Proakis, and H. Lev-Ari. Optimal tracking of time-varying channels: a frequency domain approach for known and new algorithms. *IEEE Journal on Selected Areas in Comm.*, 13(1):141– 154, Jan. 1995.

[70] Yang Liu and D.K. Borah. Estimation of fading channels with large possible delay spreads. *Electronic Letters*, 39(1):130–131, Jan. 2003.

[71] Yang Liu and D.K. Borah. Estimation of time-varying frequency-selective channels using a matching pursuit technique. In *IEEE Wireless Communications and Networking*, volume 2, pages 941–946, 2003.

[72] S. Mallat. *A Wavelet Tour of Signal Processing*. Academic Press, 1998.

[73] S.G. Mallat and Z. Zhang. Matching pursuits with time-frequency dictionaries. *IEEE Trans. on Signal Proc.*, 41(12):3397–3415, Dec. 1993.

[74] C.D. Meyer. *Matrix Analysis and Applied Linear Algebra*. SIAM, 2000.

[75] A. S. Motahari and M. Nasiri-Kenari. Multiuser detection for optical CDMA networks based on expectation-maximization algorithm. *IEEE Transactions on Communications*, 52(4):652–660, Apr. 2004.

[76] B.K. Natarajan. Sparse approximate solutions to linear systems. *SIAM Journal on Computing*, 24(2):227–234, Apr. 1995.

[77] R. Neff, T. Nomura, and A. Zakhor. Decoder complexity and performance comparison of matching pursuit and DCT-based MPEG-4 video codecs. In *IEEE International Conference on Image Processing*, volume 1, pages 783–787, 1998.

[78] R. Neff and A. Zakhor. Very low bit-rate video coding based on matching putsuits. In *Data Compression Conference*, pages 411–420, 1995.

[79] R. Neff and A. Zakhor. Dictionary approximation for matching pursuit video coding. In *IEEE International Conference on Image Processing*, volume 2, pages 828–831, 2000.

[80] S.X. Ng, F. Guo, J. Wang, L.-L. Yang, and L. Hanzo. Joint source-coding, channel-coding and modulation schemes for AWGN and Rayleigh fading channels. *Electronic Letters*, 39(17):1259–1261, August 2003.

[81] O. Ozdemir and M. Torlak. Blind channel estimation without eigen decomposition. *IEEE International Conference on Communications*, pages 2621–2625, June 2004.

[82] Y.C. Pati, R. Rezaiifar, and P.S. Krishnaprasad. Orthogonal matching pursuit: recursive function approximation with applications to wavelet decomposition. In *Asilomar Conference on Signals, Systems and Computers*, volume 1, pages 40–44, 1993.

[83] J. Proakis. *Digital Communications*. McGraw-Hill, 2000.

[84] T.S. Rappaport. *Wireless Communications Principles and Practice*. Prentice Hall, 2002.

[85] L. Rebollo-Neira and D. Lowe. Backward-optimized orthogonal matching pursuit approach. *IEEE Signal Proc. Letters*, 9(4):137–140, April 2002.

[86] R. Roy and T. Kailath. ESPRIT-estimation of signal parameters via rotational invariance techniques. *IEEE Trans. on Acoust. Speech, Signal Processing*, ASSP-37:984–995, 1989.

[87] J. A. Salehi. Code division multiple-access techniques in optical fiber networks. I. fundamental principles. *IEEE Transactions on Communications*, 37(8):824–833, Aug. 1989.

[88] J. A. Salehi and C. A. Brackett. Code division multiple-access techniques in optical fiber networks. II. systems performance analysis. *IEEE Transactions on Communications*, 37(8):834–842, Aug. 1989.

[89] H. Sawada, R. Mukai, and S. Makino. Direction of arrival estimation for multiple source signals using independent component analysis. In *International Symposium on Signal Processing and Its Applications*, volume 2, pages 411–414, 2003.

[90] K. Sayood. *Introduction to Data Compression.* Morgan Kaufmann Publishers, 2000.

[91] R. O. Schmidt. Multiple emitter location and signal parameter estimation. *IEEE Trans. on Antennas and Propagation*, 34:276–280, April 1986.

[92] W.F. Schreiber. Advanced television systems for terrestrial broadcasting: Some problems and some proposed solutions. *IEEE Proc.*, 83:958–981, June 1995.

[93] D.S. Shen and S. Roy. An adaptive multiuser receiver for CDMA systems. *IEEE Journal on Selected Areas in Comm.*, 12(5):808–816, June 1994.

[94] E. S. Shivaleela, K. N. Sivarajan, and A. Selvarajan. Design of new family of two-dimensional codes for fiber optic CDMA networks. *Journal of Lightwave Technology*, pages 501–508, 1998.

[95] P.K. Shukla and L.F. Turner. Channel-estimation-based adaptive DFE for fading multipath radio channels. *IEE Proceedings Vision, Image and Signal Processing*, 138(6):525–543, Dec. 1991.

[96] S.S. Skiena. *The Algorithm Design Manual.* Springer-Verlag New York, 1997.

[97] K. Skretting and J. Husoy. Partial search vector selection for sparse signal representation. In *Nordic Conf. in Signal Processing*, 2003.

[98] R. Steele. *Mobile Radio Communications.* IEEE Press, 1992.

[99] P. Stoica and K.C. Sharman. Maximum likelihood methods for direction-of-arrival estimation. *IEEE Trans. on Acoust. Speech, Signal Processing*, 38(7):1132–1143, 1990.

[100] L. Tancevski and L. A. Rusch. Impact of the beat noise on the performance of 2-D optical CDMA systems. *IEEE Communication Letters*, 4(8):264–266, Aug. 2000.

[101] L. Tong and S. Perreau. Multichannel blind identification: From subspace to maximum likelihood methods. *IEEE Proc.*, 86:1951–1968, Oct. 1998.

[102] M. Torlak and G. Xu. Blind multi-user channel estimation in asynchronous CDMA systems. *IEEE Trans. on Signal Processing*, 45(1):137–147, Jan 1997.

[103] H. L. Van Trees. *Optimum Array Processing*. Wiley, 2002.

[104] J.A. Tropp. Greed is good: Algorithmic results for sparse approximation. *IEEE Trans. on Info. Theory*, 50(10):2231–2242, Oct. 2004.

[105] J.A. Tropp. *Topics in Sparse Approximation*. PhD thesis, University of Texas at Austin, 2004.

[106] M.K. Tsatsanis and G.B. Giannakis. Equalization of rapidly fading channels: self-recovering methods. *IEEE Transactions on Communications*, 45(5):619–630, May 1996.

[107] M.K. Tsatsanis and G.B. Giannakis. Modeling and equalization of rapidly fading channels. *Int. J. Adaptive Control Signal Processing*, 109:159–176, Mar. 1996.

[108] M. J. Rendas V. A. N. Barroso and J. P. Gomes. Impact of array processing techniques on the design of mobile communication systems. In *IEEE 7th Med. Electrotechnical*, pages 1291–1294, 1994.

[109] S. Verdu. Multiple-access channelswith point process observation: optimum demodulation. *IEEE Transactions on Information Theory*, 32:642–651, Sep 1986.

[110] S. Verdu. *Multiuser Detection*. Cambridge Univ. Press, 1998.

[111] M. Vetterli and T. Kalker. Matching pursuit for compression and application to motion compensated video coding. In *IEEE International Conference on Image Processing*, pages 725–729, 1994.

[112] M. Visintin. Karhunen-Loève expansion of a fast Rayleigh fading process. *Electronic Letters*, 32(18):1712–1713, Aug. 1996.

[113] Chi-Shun Weng and Jingshown Wu. Optical orthogonal codes with large crosscorrelation and their performance bound for asynchronous optical CDMA systems. *Journal of Lightwave Technology*, 21(3):735–742, March 2003.

[114] J. Xin and A. Sano. Computationally efficient subspace-based method for direction-of-arrival estimation without Eigendecomposition. *IEEE Trans. on Signal Proc.*, 52(4):876–893, Apr. 2004.

[115] S. Zahedi and J. A. Salehi. Analytical comparison of various fiber-optic CDMA receiver structures. *Journal of Lightwave Technology*, 18(12):1718–1727, Dec. 2000.