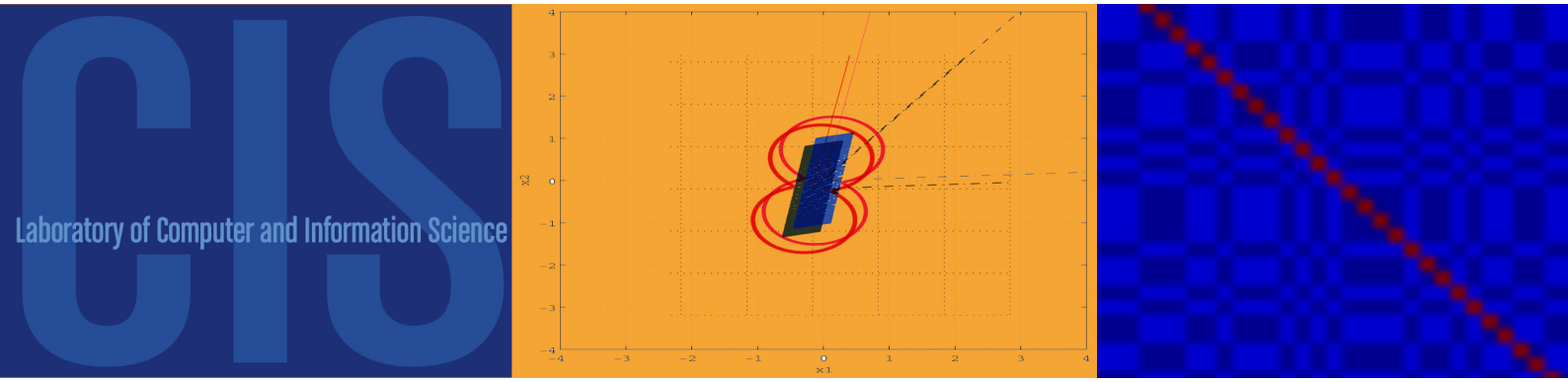
Helsinki University of Technology
Dissertations in Computer and Information Science
Espoo 2006 Report D14

BLIND SOURCE SEPARATION FOR INTERFERENCE CANCELLATION IN CDMA SYSTEMS

Karthikesh Raju



TEKNILLINEN KORKEAKOULU
TEKNISKA HÖGSKOLAN
HELSINKI UNIVERSITY OF TECHNOLOGY
TECHNISCHE UNIVERSITÄT HELSINKI
UNIVERSITE DE TECHNOLOGIE D'HELSINKI

Helsinki University of Technology
Dissertations in Computer and Information Science
Espoo 2006 Report D14

BLIND SOURCE SEPARATION FOR INTERFERENCE CANCELLATION IN CDMA SYSTEMS

Karthikesh Raju

Dissertation for the degree of Doctor of Science in Technology to be presented with due permission of the Department of Computer Science and Engineering for public examination and debate in Auditorium T2 at Helsinki University of Technology (Espoo, Finland) on the 11th of August, 2006, at 12 o'clock noon.

Helsinki University of Technology
Department of Computer Science and Engineering
Laboratory of Computer and Information Science
P.O.Box 5400
FIN-02015 HUT
FINLAND

Distribution:
Helsinki University of Technology
Laboratory of Computer and Information Science
P.O.Box 5400
FIN-02015 HUT
FINLAND
Tel. +358-9-451 3272
Fax +358-9-451 3277
http://www.cis.hut.fi/

Available in pdf format at http://lib.hut.fi/Diss/2006/isbn9512282615/

ISBN 951-22-8260-7 (printed version)
ISBN 951-22-8261-5 (electronic version)
ISSN 1459-7020

Otamedia Oy
Espoo 2006

Raju, K. (2006): **Blind Source Separation for Interference Cancellation in CDMA Systems**. Doctoral thesis, Helsinki University of Technology, Dissertations in Computer and Information Science, Report D14, Espoo, Finland.

Keywords: blind source separation, independent component analysis, denoising, denoising source separation, interference cancellation, code division multiple access, direct sequence code division multiple access, jammer mitigation, switching, priors, conjugate priors.

Abstract

Communication is the science of "reliable" transfer of information between two parties, in the sense that the information reaches the intended party with as few errors as possible. Modern wireless systems have many interfering sources that hinder reliable communication. The performance of receivers severely deteriorates in the presence of unknown or unaccounted interference. The goal of a receiver is then to combat these sources of interference in a robust manner while trying to optimize the trade-off between gain and computational complexity.

Conventional methods mitigate these sources of interference by taking into account all available information and at times seeking additional information *e.g.*, channel characteristics, direction of arrival, *etc.* This usually costs bandwidth. This thesis examines the issue of developing mitigating algorithms that utilize as little as possible or no prior information about the nature of the interference. These methods are either *semi-blind*, in the former case, or *blind* in the latter case.

Blind source separation (BSS) involves solving a source separation problem with very little prior information. A popular framework for solving the BSS problem is *independent component analysis* (ICA). This thesis combines techniques of ICA with conventional signal detection to cancel out unaccounted sources of interference. Combining an ICA element to standard techniques enables a robust and computationally efficient structure. This thesis proposes switching techniques based on BSS/ICA effectively to combat interference. Additionally, a structure based on a generalized framework termed as *denoising source separation* (DSS) is presented. In cases where more information is known about the nature of interference, it is natural to incorporate this knowledge in the separation process, so finally this thesis looks at the issue of using some prior knowledge in these techniques. In the simple case, the advantage of using priors should at least lead to faster algorithms.

Preface

The work related to this thesis has been carried out at the Neural Networks Research Centre (NNRC), hosted by the Laboratory of Computer and Information Science, at Helsinki University of Technology (HUT). I am deeply grateful to Professor Erkki Oja, for first accepting me in 1999 to work for my Master's thesis and later for extending a warm welcome again in 2001 for starting work on my doctoral studies. It has been a true privilege in having him as my supervisor, and I would like to thank him for that. In addition to being my supervisor, he has been a fatherly and trustworthy figure, and has always helped me with all the bureaucratic issues.

Professor Tapani Ristaniemi, Tampere University of Technology, my adviser for this thesis has been prolific in guiding me throughout. His enormous experience, energy and sense of humor are some things that I would wish to strive for in life. His support, both as an adviser and a good friend has been instrumental in completing my thesis and making my stay here a very rewarding one.

My second adviser Professor Juha Karhunen of NNRC, has always helped me to understand and polish my ideas. With a passion for perfection, he has always fine tuned all our papers, and has always improved the quality of my work. This manuscript has had a definite and significant face lift after his stringent quality control. It is a privilege to have good advisers, I have had the rare luxury of three, and I am truly obliged to all of them.

Most of this work had been carried out under the auspices of the Telectronics II programme of the Academy of Finland. I am grateful to the Academy for the funding received. In addition, the personal grants received from the Nokia Foundation and the Emil Aaltosen Foundation are truly acknowledged.

Many thanks to Professor Olli Simula, and to our secretaries Leila Koivisto and Tarja Pihamaa for ensuring that my stay here has been smooth. Leila's numerous letters and documents have been of great help in me obtaining the various visas. She has always been cheerful and I am deeply indebted to her.

Special thanks to all my co-authors: Professors Erkki Oja, Tapani Ristaniemi, Juha Karhunen, Dr. Aapo Hyvärinen, Dr. Jaakko Särelä, M.Sc. Toni Huovinen, and M.Sc. Phani Sudheer.

I have had the chance of sharing several anecdotes and experiences with many people of the laboratory and several friends. I wish to thank everyone in general and Dr. Ricardo Vigário, Dr. Ella Bingham, Mr. Alexander Ilin, Mr. Jan-Hendrik Schleimer, Mr. Matti Aksela, Mr. Ganesh Sivaraman, Mr. Subramanian Ramjee and Dr. Christopher Latham in particular.

This work has been pre-examined by Professor Lars K. Rasmussen and Professor Asoke Nandi. I wish to thank you both for your patience in reading the manuscript and for your insightful comments.

I would also like to thank my parents, my uncles, and my first boss and good friend Dr. Seshu Bhagavathula. I am sure they have all played a great role in motivating me to take up this big task and seeing to it that I don't flinch and turn back.

Finally, I am deeply grateful to Gayathri for her love and understanding during the writing of the manuscript and for her skill in managing things effectively and efficiently, so that I could devote my energy to writing the manuscript.

Otaniemi, August 2006 Karthikesh Raju

Contents

Mathematical notation

lower- or upper-case		scalar, constant or scalar function
bold face lower-case		column or row vector, vector-valued function
bold face upper-case		matrix, matrix-valued function
α	scalar	scalar, or the prior weight
γ_b	scalar	average Signal-to-Noise Ratio per bit
Γ_{ICA}	scalar	Signal-to-Interference Ratio after processing by independent component analysis algorithms
$\mathbf{\Lambda}_s$	$K \times K$	diagonal matrix containing the eigenvalues
ν	scalar	Gaussian variable
$\boldsymbol{\nu}_i$	$1 \times T$	ith additive noise row vector
$\boldsymbol{\nu}$	$M \times T$	additive noise matrix
Ω	function	log-likelihood function
ρ	scalar	jammer duty factor
θ_j	scalar	arrival angle of the jammer
θ_q	scalar	arrival angle of the interference term
θ_r	scalar	arrival angle of the information term
$\mathbf{\Theta}$	$N_a \times L$	array steering matrix
$\mathbf{\Upsilon}$	$K(2M+1)\times K(2M+1)$	diagonal matrix containing the amplitudes
$\mathbf{A}$	$K \times N$	mixing matrix
$\mathbf{A}$	$K \times K$	diagonal matrix containing the amplitudes of all users
$\mathbf{a}$	$K \times 1$	a column mixing vector
$\mathbf{a}^{'}$	$1 \times (K-1)$	a row vector containing symbols of all users other than the desired user
$\mathbf{a}_j$	$K \times 1$	ith column mixing vector
$\overline{\mathbf{a}}_m$	$1 \times 2KL$	row vector containing all early and late parts of the symbols and their path gains
a_{ij}	scalar	mixing coefficient of the jth source in ith observation
$\mathbf{B}$	$N \times K$	demixing matrix
$\mathbf{b}_k$	$1 \times M$	kth user's symbols, row vector
$\tilde{\mathbf{b}}_k$	$1 \times M$	soft-decision estimates of kth user's symbols, row vector

$\hat{\mathbf{b}}_k$	$1 \times M$	hard-decision estimates of kth user's symbols, row vector
b_{km}	scalar	kth user's mth bit
C	scalar	code length
$\mathbf{D}$	$T \times T$	linear denoising matrix applied to the source estimate
$\mathbf{D}^*$	$T \times T$	linear denoising matrix applied to the entire data
d_{kl}	scalar	delay of the kth user along lth path
$\mathbf{E}$	$L \times K$	matrix of the eigenvectors
$\mathbf{e}$	$K \times 1$	a column eigenvector
$\mathbf{e}_l$	$K \times 1$	lth column eigenvector
$\mathbf{f}(.)$	$1 \times T$	denoising function, row vector
F	function	Fourier transform
F^{-1}	function	Inverse Fourier transform
f_c	scalar	carrier frequency
f_j	scalar	jammer frequency
f_ξ	scalar	frequency of the out-of-cell interference component
$\mathbf{G}$	$C \times 2KL$	code matrix containing all the path gains and path delays
$\mathbf{G}$	$N_a \times (L+1)$	a matrix containing the response of a uniform linear array
$\mathbf{G}'$	$C \times 2(K-1)$	matrix containing all the path gains and delays excluding the desired user
$\mathbf{G}$	$N \times T$	gain matrix
$g(.)$	scalar	objective function
H_o	$-$	null hypothesis that the jammer is absent
H_1	$-$	alternative hypothesis that the jammer is present
$\mathcal{I}$	scalar	global rejection index
i, j, l	scalar	general purpose indices, usually i refers to observations, j to sources and l to sphered data
$L(.)$	function	likelihood function
L	scalar	number of retained principal components of $\mathbf{y}_l$
L	scalar	number of significant paths in the channel
$\mathcal{L}$	$K \times K$	lower Cholesky factor of the matrix $\mathbf{R}$
l	scalar	lth path
M	scalar	number of observations $\mathbf{x}_i$

M	scalar	block size
$\mathbf{n}$	$1 \times CM$	row vector containing the noise components
N	scalar	number of sources $\mathbf{s}_j$
N_a	scalar	number of elements in the antenna array
N_t	scalar	length of the training sequence
$O(t)$	$1 \times CM$	out-of-cell interference component
$P(t)$	function	pulse shaping function
P_b	scalar	average bit error probability
$\overline{P}_b$	scalar	average bit error probability after some processing
$\overline{P}_b^{\dagger}$	scalar	"threshold" average bit error probability for determining the presence of a jammer
$Q(.)$	function	Q-function used for probability of error calculations
$q(.)$	function	posterior density function
$\mathbf{R}$	$K \times K$	cross correlation matrix of the spreading codes
$r(t)$	scalar	received signal along with interference before down conversion
$\mathbf{S}$	$N \times T$	matrix of N sources with T samples
$\mathbf{S}$	$K \times C$	matrix containing the spreading codes of all users
$\mathbf{S}^{\dagger}$	$N \times T$	matrix of N sources that have been normalized and centered
$\mathbf{s}$	$1 \times T$	a row vector consisting of a source
$\mathbf{s}_j$	$1 \times T$	a row vector consisting of the jth source
$\mathbf{s}_k$	$1 \times C$	a row vector containing the spreading code of user k
$\overline{\mathbf{s}}_{kl}$	$1 \times 2C$	a row vector containing the early parts of kth user's code along path l
$\underline{\mathbf{s}}_{kl}$	$1 \times 2C$	a row vector containing the late parts of kth user's code along path l
$s_j(t)$	scalar	value of the jth source at (time) index t.
$s_k(t)$	scalar	value of the kth user's spreading code at (time) index t.
$\mathbf{s}(t)$	$N \times 1$	a column vector containing the values of all sources at time instance t
T	scalar	number of samples in sources $\mathbf{s}_i$, and observations $\mathbf{x}_i$
T	scalar	symbol duration
T_c	scalar	duration of the chip sequence
$\mathbf{U}_s$	$C \times K$	matrix containing the eigenvectors

$\mathbf{V}$	$L \times K$	sphering matrix
$\mathbf{v}$	$K \times 1$	a column sphering vector
$\mathbf{v}_l$	$K \times 1$	lth column sphering vector
v_{li}	scalar	sphering coefficient of the ith observation in the lth principal component
$\mathbf{W}$	$N \times L$	demixing (separating) matrix (from the sphered data)
$\mathbf{w}$	$N \times 1$	a column demixing vector
$\mathbf{w}_j$	$N \times 1$	jth column demixing vector
w_{ji}	scalar	demixing coefficient of the ith observation in the jth source
$\mathbf{X}$	$K \times T$	matrix of M observations with T samples after whitening
$\mathbf{X}_{orig}$	$K \times T$	matrix of M observations with T samples before whitening
$\mathbf{x}$	$1 \times T$	a row vector consisting of an observation
$\mathbf{x}_i$	$1 \times T$	a row vector consisting of the ith observation sequence over time
$x_i(t)$	scalar	value of the ith observation at (time) index t.
$\mathbf{V}$	$T \times K$	whitening matrix
$\mathbf{x}(t)$	$K \times 1$	a column vector containing the values of all observations at time instance t
$\mathbf{x}(t)$	$CM \times 1$	vector containing the spread data
$\mathbf{Y}$	$L \times T$	matrix of sphered components
$\mathbf{y}$	$1 \times T$	a row vector consisting of a sphered component $\mathbf{y}_1$
$\mathbf{y}$	$C \times 1$	vector containing the received signals after down conversion
$\mathbf{y}_l$	$1 \times T$	a row vector consisting of time observations/samples of the lth sphered component
$y(t)$	scalar	received signal before down conversion
$\mathbf{Z}$	$K \times T$	denoised data or the whitened data

Abbreviations

AIC	Akaike information criterion
AWGN	additive white Gaussian noise
BER	bit-error-rate
BLER	block-error-rate
BSS	blind source separation
CDMA	code division multiple access
CLT	central limit theorem
CM	constant modulus
CMA	constant modulus algorithm
CMOE	constrained minimum energy output
CSI	channel state information
DD	decorrelating detector
DOA	direction-of-arrival
DS-CDMA	direct sequence code division multiple access
DSS	denoising source separation
DR	despread-respread
EM	expectation-maximization
EVD	eigenvalue decomposition
FCC	forward error correction
FDMA	frequency division multiple access
FOBI	fourth-order blind identification
GSM	groupe special mobile
HOS	higher-order statistics
ICA	independent component analysis
IDMA	interleave-division multiple access
ISI	intersymbol interference
JADE	joint approximative diagonalization of eigenmatrices
LMMSE	linear minimum mean-square error
LMS	least mean-square
LTI	linear time-invariant
MAI	multiple access interference
MAP	maximum a posteriori
ML	maximum likelihood
MC-CDMA	multi carrier code division multiple access
MDL	minimum description length
MF	matched filter
ML	maximum-likelihood
MLSD	maximum-likelihood sequence detector

MMSE	minimum mean-square error
MRC	maximum ratio combining
MSE	mean-square error
MUD	multiuser detection
NBI	narrow band interference
NMT	nordic mobile telephone
NW	noise whitening
PCA	principal component analysis
pdf	probability density function
PIC	parallel interference cancellation
PN	pseudo-noise
QPSK	quadrature phase shift keying
RLS	recursive least-squares
SDD	subspace decorrelating detector
SDMA	space-division multiple access
SIC	serial/successive interference cancellation
SIR	signal-to-interference ratio
SINR	signal-to-interference-plus-noise
SMMSE	subspace minimum mean-square error
SJR	signal-to-jammer ratio
SNR	signal-to-noise ratio
SOBI	second-order blind identification
STFT	short-term Fourier transform
TDMA	time division multiple access
TDSEP	temporal decorrelation source separation
ULA	uniform linear array
UMTS	universal mobile telecommunication system
WBI	wide band interference

Chapter 1

Introduction

Man's cohesive existence depends on the ability to communicate audio-visually. This basic urge has led to the development of several methods for expression. Signals, signs, symbols, icons and gestures initially satisfied this need, but given that man is inventive, these instruments for communication were limited. This led to the invention of the hieroglyphics, the alphabet and the language. The industrial revolution kick-started the process of mass dissemination of knowledge mainly due to the invention of the printing press, a significant milestone that slowly started a process of accumulation that culminated in the discovery of radio waves in the 19th century. This process is still not over, and the world has moved from the simple diplex to radio broadcasts to television to the (modern) wireless communications. The internet and mobile personal communications are mere milestones in this evolution of communication. Indeed this urge to communicate, has brought humans a long way from smoke and light signals to modern day fourth generation (4G) systems, but the key aspect of all this is error-less communication that is dependent on effectively reproducing what was to be communicated at the other end of the communication chain.

Spread spectrum technology forms the back bone of the third generation systems and possible future ones. Spread spectrum involves the use of a very large bandwidth for transmitting the information as compared to real bandwidth of the information. This is achieved by the use of unique codes for "spreading" and "despreading" the information. When these codes satisfy certain criteria (See Chapter 2 for a more precise description), several users can simultaneously communicate leading to the access methodology called Code Division Multiple

Access (CDMA).

Naturally, spreading and despreading information requires frequency resources. These resources are not only limited, but are also heavily controlled by regulatory authorities. Hence, bandwidth efficiency forms a very important factor for future wireless networks. Bandwidth efficiency can be achieved at different levels of design. Good modulation schemes, and network planning and design can help in increasing the bandwidth efficiency.

One of the major limiting factors to bandwidth efficiency is *interference*. Interference can originate from the other users in the same system, coexisting other systems (overlayed in frequency), non-linearities in the transmitter and receiver, man made interference, external interference, *etc.* Interference must be estimated and appropriately compensated for, else it will severely degrade the performance of the system. Hence, interference *cancellation, suppression* and *mitigation* form a core part in receiver design.

Fig. 1.1 shows a block diagram of a digital communication system from the data modulator to the data demodulator (encryption and channel coding have been left out as this is not the focus of this thesis). The content of this thesis is confined to the receiver design that helps in mitigating interference while helping to conserve bandwidth. Receivers form the lower block of the diagram indicated by the dark blocks in the Fig. 1.1. Receivers typically mitigate interference by utilizing the available knowledge about the interference structure. The more refined this knowledge is, the better the performance of the receiver. Typical assumptions on the nature of interference are that its components are additive white Gaussian noise (AWGN), meaning that interference is neither time structured nor frequency structured. This is a worst case assumption from both the detection and information theoretic viewpoint. Another assumption is that the interference is impulsive. Such non-Gaussian interference can be multi-access interference (MAI) and inter-symbol interference (ISI). Receivers for mitigating these classes of interference usually require channel state information (CSI). Non-Gaussian assumptions also arise when transmission is assumed to be uncoordinated. In this case receivers exploit the fact that transmissions are discontinuous in time. Finally, interference is assumed to be either wide-band (WBI) or narrow-band (NBI) based on signal correlation properties. In this case receivers use the interference correlation in the time or frequency or spatial domain.

Another class of receivers has evolved recently. It tends to make minimal assumptions on the interference structure or has a fairly limited *a priori* system

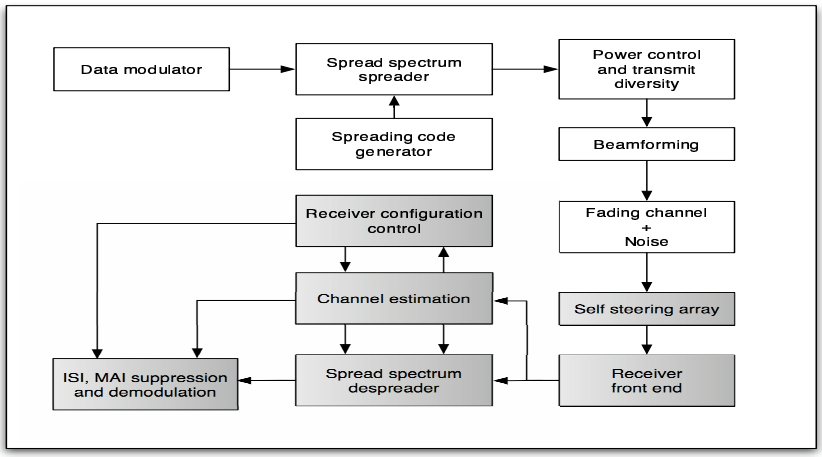


Figure 1.1: *Generic block diagram of a spread spectrum communication system. Adapted from [49].*

information. These receivers are termed as *blind* receivers. This notion of "blind" has long existed in the signal processing community. A typical example of it are blind source separation (BSS) techniques or the blind deconvolution/equalization techniques. BSS is commonly used to recover sources that have been *mixed* in some manner without explicit knowledge of either the mixing process or the source signals. The notion of blind receivers has gained importance in the communications community for the following reasons:

1. very high demand exists for high bandwidth applications such as video *etc.* ;
2. wireless communications is growing rapidly;
3. spectrum is limited.

The motivation for using blind receivers in communications stems from the fact that they do not require the knowledge of the CSI usually needed for conventional detectors. This information is usually obtained from training sequences. The use of these sequences reduces the data rate. For practical channels (*e.g.*,. time-varying), this has to be performed periodically, which has a detrimental effect on throughput. For example, in the groupe special mobile (GSM) system around 20% of the symbols are training symbols. Hence, by using blind methods, this

overhead can be eliminated — data symbols can be sent during this period instead of training sequences. Even though this is in principle appealing, the practical use of blind methods is restricted due to several reasons. Most blind methods make some assumptions on the structure of the data which are not always true. Performance of these methods is mostly linked to these assumptions. They always have some ambiguities that make the final detection complicated. However, the use of some *a priori* information in blind methods has led to practical algorithms called "semi-blind" methods. These methods are quite similar to the blind methods, but usually incorporate some known information for overcoming the problems faced by blind methods. One way of incorporating some information is to use shorter training sequences (essentially to overcome the ambiguities). This allows for increase in the effective data rate. This makes semi-blind techniques popular and practical for future receivers.

1.1 Scope and contributions of this thesis

This thesis was born out of the need to apply the principles of blind source separation (BSS) to interference cancellation. The growth in computational power has led to a surge in popularity of blind methods in various fields of science and engineering. This has helped to convert several principles into elegant algorithms. Blind methods are successfully used in the fields of biomedical engineering [194, 195] for identification of artifacts, and in non-invasive fetal ECG extraction [209], in astrophysics, to analyze multispectral images [46] and in speech processing [184], just to mention a few applications. More examples of BSS applications can be found in [48, 75].

In the field of communication engineering, the popularity of blind methods stems from the fact that they enable processing of received signals with minimal assumptions of the system. With increases in computational power this can lead to enhancements in the quality of communications. Here, blind methods are mainly used in desired user separation (summarized in Chapters 2 and 3 of [202]), or channel identification [47] and in deconvolution [56, 60]. The article [109] provides a good summary on the existing blind methods. Most methods here use the linear *minimum mean-square error* (MMSE) detector as a starting point and formulate adaptations of this detector [64]. Another approach has been to use neural networks for estimating the interference. These methods mainly use training data [2, 121] or adaptation of the learning rate by using recurrent neural networks [179]. Subspace based approaches [109] try to estimate the interference

subspace and try to find a linear decorrelator in the direction orthogonal to the interference subspace. Other blind approaches include adaptations of the *constant modulus algorithm* (CMA) to cancel interferences [50, 185]. Even though several blind approaches exist most of these approaches are based on second-order statistics. These methods use conventional methods as their starting points and hence do not fully harness the power of higher-order statistics. Neural approaches are essentially non-linear but are not entirely blind as they use training data. Sec. 3.5 of Chapter 3 provides a summary of these methods.

Independent component analysis (ICA) [20, 26, 27, 75, 77, 90] is a popular technique for solving the BSS problem. ICA relies on higher-order statistics (typically the fourth-order statistic - *kurtosis*) to solve the BSS problem. Even though these methods are very popular in other domains, they are yet to attract attention in the field of communication engineering. A reason may be that some of these methods have an artificial neural network [58] background. Some of the first works to apply ICA in symbol demodulation are [157, 158]. Papers [29, 30] address the issue of delay estimation with ICA. Jammer mitigation with a BSS principle is first discussed in [12], where an ICA method called JADE [20] (briefly explained in Sec. 4.1.7 of Chapter 4) and second-order methods are used to mitigate a temporally correlated jammer. The paper [11] also addresses the issue of blind beamforming to jammer mitigation. The papers [212, 213] investigate anti-jamming for GPS receivers based on subspace projections and analysis of signal distributions.

However, the principle of applying ICA towards jammer mitigation has not been very clearly addressed before. The foundations of BSS and ICA provide a strong motivation for applying them in the area of jammer mitigation and interference cancellation. The focus of this area is clearly defined in Fig. 1.2 as an intersection of the fields of Communications, Signal Processing (Blind Signal Processing) and Interference Cancellation. Hence, the following are defined as the scope of this thesis:

1. To develop methods that incorporate the principles of blind source separation for interference cancellation. ICA algorithms are a natural choice as they provide a framework for these methods and utilize higher-order statistics instead of conventional second-order statistics. It is also assumed that "independence", being a stronger assumption than "decorrelation", should lead to better algorithms;
2. To effectively utilize available information in the system. This leads to the development of semi-blind algorithms. Semi-blind because these algorithms bring in all the benefits of BSS while at the same time retaining some of the benefits of conventional algorithms. Semi-blind algorithms utilize shorter

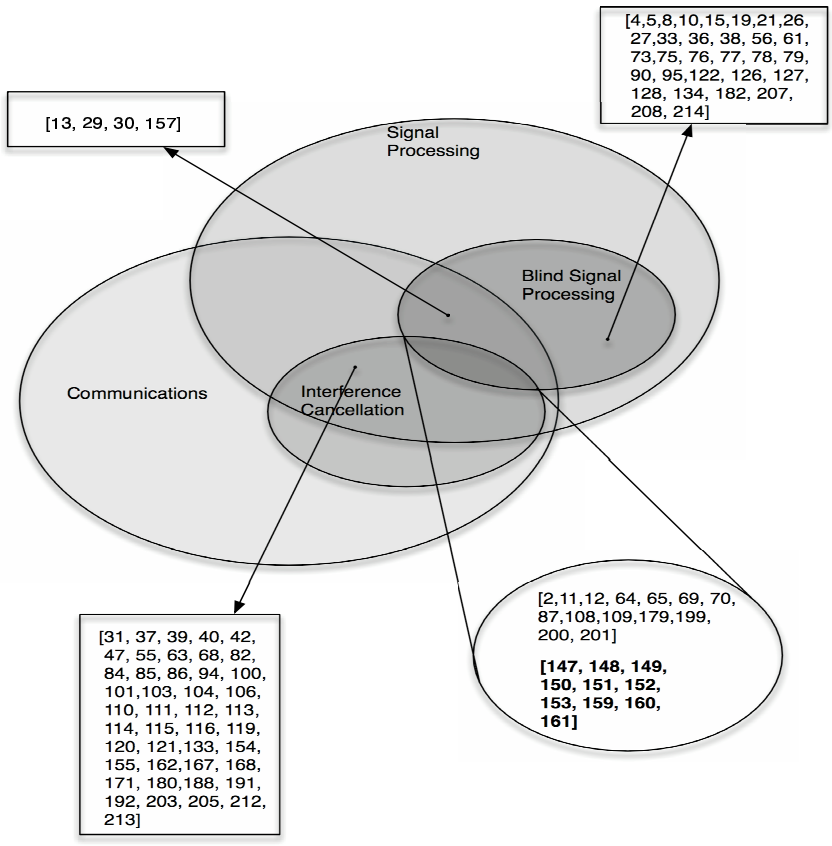


Figure 1.2: *Scope of the thesis. The references in this thesis have been classified based on their area. Though several signal processing methods are well established in communications, blind methods have not been very popular. This thesis focuses on blind methods in interference cancellation indicated by the darkest intersection. Citations in bold correspond to the articles that form the basis of this thesis.*

training sequences than conventional methods, which can then be used to transmit the user's data;

3. To combine traditional and blind methods in an efficient manner by which they both complement each other. This again leads to semi-blind structures;

4. To develop blind structures that are simple and efficient and are based on the principles of BSS and additionally denoising. Denoising source separation [166] is a framework for constructing source separation algorithms around a denoising function;
5. To utilize "prior" information efficiently. It is noted in an earlier work [78] that the use of priors leads to better results with less data. Additionally, priors also increase the computational speed of the estimation methods. Though this research was carried out in the field of image analysis, they ideas are equally applicable here.

Based on the above scope the following novel methods were developed and analyzed as a part of this thesis:

1. **The plain ICA-RAKE receiver structure.**

   The plain ICA-RAKE is a combination of the RAKE receiver and an ICA block. The ICA block separates the sources, while the RAKE receiver identifies, despreads and demodulates the desired source.

2. **ICA-RAKE Pre-Switch for cancellation of jammers and out-of-cell interference.**

   The ICA-RAKE Pre-Switch structure builds upon the plain ICA-RAKE structure. It is based on a "simple" distance measure, and improves the performance of the receivers when the interference level is low *i.e.*, at high signal-to-interference (SIR) levels.

3. **ICA-RAKE Post-Switch for cancellation of jammers and out-of-cell interference.**

   ICA-RAKE Post-Switch is a correlation-based receiver that improves the performance of the ICA-RAKE receiver structure at low SIR levels. It is better suited to handle multipath interference, but is computationally more complex than ICA-RAKE Pre-Switch.

4. **Denoising Source Separation (DSS) technique for interference cancellation.**

   DSS based receiver - IC-DSS - is a blind technique that uses only the spreading code for reconstructing the "possible" transmitted sequence. It does not use any other conventional techniques for identification other than the spreading code. It does not require the use of training data, and hence is blind.

5. **Effective use of prior information in these receivers.**

   Prior information should be used as and when available. It helps in reducing the data requirements, improving the quality of the results and finally in increasing the computational speed of these algorithms. This is examined with the help of illustrative examples from image analysis, and a possible means to use them in interference cancellation is provided.

The above methods 1–3 have been built on the basis of the independent component analysis (ICA) algorithm - FastICA [73, 75, 76]. Another ICA method considered was JADE [21]. Additionally, temporally jammed signals were recovered by a BSS technique based on the TDSEP method [214]. The first three methods are semi-blind as they use minimal pilot sequences for identification of the signals. These methods address issues 1–3 of the above scope.

Issue 4 of the above scope, the denoising source separation based structure, was developed additionally as a blind technique. This method has the knowledge of only the spreading code of the desired user.

Additionally, prior information on the nature of the interference can be effectively used in order to speed up the algorithms and to improve them. This information has also been incorporated into the development of these methods as laid out in the issue 5 of the scope.

These methods generally provide around 5 to 8 dB gain over conventional techniques depending on the nature of the interference. Additionally, they handle multipath interference well, and typically provide better block error performance.

This work is presented in the form of a monograph for the sake of clarity and to make it easier to read. However, the following publications [78, 147, 148, 149, 150, 151, 152, 153, 159, 160, 161] form the basis of the research that is collected in this thesis. These publications are also listed below.

1.2 List of related publications

Publication 1. T. Ristaniemi, K. Raju, J. Karhunen, Jammer Mitigation in DS-CDMA Array System Using Independent Component Analysis, *Proc. IEEE Int. Conference on Communications (ICC 2002)*, New York City, NY, USA, April 28 - May 2, 2002, pp. 232–236.

Publication 2. K. Raju, T. Ristaniemi, J. Karhunen, E. Oja, Suppression of Bit-Pulsed Jamming Using Independent Component Analysis, *In Proc. 2002 IEEE Int. Symp. on Circuits and Systems (ISCAS 2002)*, Phoenix, Arizona, USA, May 26-29, 2002, pp. I-189/I192.

Publication 3. K. Raju, T. Ristaniemi, ICA-RAKE-Switch for Jammer Cancellation in DS-CDMA Array Systems, *In Proc. 2002 IEEE Int. Symp. on Spread Spectrum Techniques and Applications (ISSSTA 2002)*, Prague, Czech Republic, September 2-5, 2002, pp. 638–642.

Publication 4. T. Ristaniemi, K. Raju, J. Karhunen, E. Oja, Jammer Cancellation in DS-CDMA Array Systems: Pre and Post Switching of ICA and RAKE, *In Proc. 2002 IEEE Int. Symposium on Neural Networks for Signal Processing (NNSP 2002)*, Martigny, Switzerland, September 4-6, 2002, pp. 495–504.

Publication 5. K. Raju, T. Ristaniemi, J. Karhunen, E. Oja, Jammer Cancellation in DS-CDMA Array Systems using Independent Component Analysis, *IEEE Transactions on Wireless Communications*, vol. 5, no. 1, January 2006, pp. 77–82 .

Publication 6. T. Ristaniemi, K. Raju, J. Karhunen, E. Oja, ICA-Assisted Inter-Cell-Interference Cancellation in CDMA Array Systems, *In Proc. Fourth International Symposium on Independent Component Analysis and Blind Signal Separation (ICA 2003)*, Nara, Japan, April 1-4, 2003, pp. 739–744.

Publication 7. K. Raju, T. Ristaniemi, Exploiting Independence to Cancel Interference due to Adjacent Cells in a DS-CDMA System, *In Proc. 14th IEEE Int. Symposium of Personal Indoor Mobile Radio Communications (PIMRC 2003)*, Beijing, China, Sept 7-10, 2003, pp. 2130 – 2134.

Publication 8. K. Raju, T. Ristaniemi, J. Karhunen, Semi-Blind Interference Suppression on Coherent Multipath Environments, *In Proc. First IEEE Int. Symposium of Control, Communications and Signal Processing (ISCCSP 2004)*, Hammamet, Tunisia, March 21-24, 2004, pp. 283–286.

Publication 9. K. Raju, J. Särelä, A Denoising Source Separation based approach to Interference Cancellation for DS-CDMA Array Systems, *In Proc. 38th Asilomar Conference on Signals, Systems and Computers*, Pacific Grove, USA, Nov 07-10, 2004, pp. 1111–1114.

Publication 10. K. Raju, B. Phani Sudheer, Blind Source Separation for Interference Cancellation - A Comparison of Several Spatial and Temporal Statistics Based Techniques, *In Proc. 3rd Workshop on the Internet, Telecommunications and Signal Processing*, Adelaide, Australia, Dec 20-22, 2004.

Publication 11. K. Raju, T. Huovinen, T. Ristaniemi, Blind Interference Cancellation Schemes for DS-CDMA Systems, *In Proc. IEEE International Symposium on Antennas and Propagation and USNC/URSI National Radio Science Meeting*, Washington, USA, July 3-8, 2005, pp.

Publication 12. A. Hyvärinen, K. Raju, Imposing Sparsity on the mixing matrix in independent component analysis, *Neurocomputing*, 49:151–162, 2002 (Special Issue on ICA and BSS).

1.3 Summary of the related publications

Publication 1 is the first paper to address the issue of jammer mitigation in the presence of signals with low temporal correlation. Continuous wave jamming is considered both at the carrier frequency and with a frequency offset. This paper studies the behaviour of the scheme when the signals are not correlated in time. The selection of the user is based on training sequences. Finally conventional detection is performed for the selected source. This paper introduces the ICA-RAKE structure.

In **Publication 2**, the case of a wide-band jammer is examined. It is assumed that both the jammer and the information signal are temporally uncorrelated at the chip level. The jammer is a bit-pulsed jammer. The ICA-RAKE structure provides some improvements especially at low Signal-to-Jammer (SJR) ratios. The results are compared with conventional matched filtering, and with maximum ratio combining (MRC) RAKE receiver. In the ICA-RAKE case, the dimensionality of the data from multiple sensors is reduced.

Pre-Switch is introduced in **Publication 3**. Plain ICA-RAKE structures perform poorly when the contribution of the jammer is low, *i.e.*, the results saturate around 0 dB SJR. This is because the contribution of the jammer is low enough for plain ICA to ignore it. In Pre-switch, an estimation of the SJR is performed and on the basis of the SJR, ICA is used to separate the interfering source. This results in improvements at high SJR's. The results now are almost identical to

MF or MRC when the SJR is high, but substantially better when there is significant jamming. The complexity of the receiver is also substantially reduced due to less processing at high SJRs.

A soft decision based receiver called Post-Switch is presented in **Publication 4**. In Post-Switch, two branches of the data are processed independently by plain ICA and MF. The soft decision outputs along with the training symbols, are used to decide the decoding branch. Post-Switch provides better results even at high SJR but at a higher complexity. As a comparison the results of Pre-Switch are also used here.

Publication 5 is a combination of the above papers in a comprehensive and thorough manner. The experiments and the results in the above paper have been revised to include some more settings.

Interference from adjacent cells is addressed in **Publication 6**. Additionally, the issue of loads is also examined here. The result of this study is that ICA viewed external interference (irrespective of the number of sources) as a composite interference. Along with this interference, the effect of increasing the load in the given cell-of-interest is also examined.

In **Publication 7**, some soft lower-bounds (experimental) for adjacent cell interference cancellation are determined. The results are examined in view of these bounds. Moreover, the effect of increasing the load of the interfering sources is also studied. The result was that ICA always viewed this external interference as a composite interference. This also makes it possible to improve estimates of users in the other cell. All the results in this paper uses the Post-Switch ICA-RAKE detector structure.

Multipath propagation and the issue of coherent jammers are addressed in **Publication 8**. ICA-RAKE Post-Switch is used to examine the effects of multipath propagation. Both the interference and the information signal consist of multipaths. In previous simulations, the use of small block sizes degraded the results due to the accuracy of the covariance matrix used in the eigenvalue decomposition during PCA whitening. By examining the slopes of the bit-error-rates it is possible to examine the right block size given a channel. Most experiments in this paper are performed with the "adequate" block size.

The DSS-based interference cancellation scheme forms the theme of **Publication 9**. The basic DSS algorithm [166] is adapted and applied with a non-linearity

to suppress the bit-pulsed jammer. The results show improvements over conventional detection. This is a purely blind approach to interference cancellation in the sense that the DSS-detector uses only the spreading codes of the desired user, and does not require training sequences for estimation of the channel or identification of the user.

All the above ICA-RAKE detection methods use FastICA [75] for separating the sources. In the case of multipath propagation, it is meaningful to examine other temporal decorrelating schemes too. **Publication 10** studies some temporal and spatial statistics based source separation methods. FastICA and JADE [21] used spatial statistics while TDSEP [214] used temporal statistics. The degradation of performance with multipath was not significant with TDSEP. Additionally a DSS-based algorithm was also used in the comparisons.

Publication 11 is a joint paper summarizing work on both uplink and downlink interference cancellation in DS-CDMA systems. This paper examines the improved non-linear RAKE-ICA receiver [70, 71] for the uplink in addition to the ICA-RAKE structures for downlink interference cancellation. This paper is an effort to present both uplink and downlink methods under a unified framework.

The motivation of using prior information stems from the fact that it prevents over-learning in cases where data are not sufficient. ICA already uses a prior on the statistical properties of the source. The theory behind prior information on the mixing matrix is examined in **Publication 12**. It is shown that a class of priors known as "Bayesian priors" can be used to improve the maximum likelihood estimates of ICA. Illustrative examples from the estimation of basis functions of images prove the viability of this concept. In CDMA however, priors are useful in the estimation of users symbols in cases of over-saturated systems.

1.4 Author's contributions to the research

Working on a thesis usually is a two prong methodology. The first is the identification and carving of a small area of research within a larger framework, and the second task involves developing methods to address issues in the chosen area. The fact that very little work was done previously on utilizing ICA in communications helped in identifying this area. Prof. Tapani Ristaniemi's work on symbol demodulation with ICA [157] along with Profs. Belouchrani and Amin's paper [12] formed the starting point of this work. Most of the code (developed

in Matlab and Python) in the papers has been developed by the author. All the experiments have been planned and carried out by the author. In Publication 1, Publication 4, Publication 6, the author mainly conducted the experiments while the manuscript was prepared by Prof. Ristaniemi. All the other manuscripts have been mostly written by the author. This naturally involved discussions with other authors. Pre- and Post-Switch was originally the author's idea, which was further refined during discussions. The use of ICA to cancel adjacent cells emerged from a discussion with Dr. Jorma Lilleberg from Nokia Mobile Phones, during a conference. The planning and experimentation was performed by the author. The DSS based receivers were a joint work with Drs . Jaakko Särelä and Harri Valpola. In addition to applying the algorithm and developing it for this application, the author's role here has been in discussions, planning, conducting and writing most of the manuscript. The planning of the experiments, and guidance for Publication 10 was the author's while Mr. Phani Sudheer performed the experiments. The analysis of the results and the manuscript were written by the author. Most of this work was performed as a collaborative work with Profs. Tapani Ristaniemi, Juha Karhunen and Erkki Oja.

1.5 Structure of this thesis

This thesis is presented as a monograph for clarity and to make it easier to read. Chapter 2 introduces the field of multiple access communications and in particular code division multiple access (CDMA). A model for the received signal is developed in this chapter. This model is used in all the simulations in the later chapters. Interference structures are described and modeled. Particularly, the two kinds of interference used in this thesis *i.e.*, *out-of-cell interference* and *jammers*, are described in detail.

A concise review of the existing methods for mitigating these sources of interference forms Chapter 3. Conventional techniques including the maximal ratio combining (MRC) RAKE receiver are described. A summary of optimal and suboptimal multiuser detection techniques is provided. A general synopsis of blind methods is provided and a classification of blind methods is made. Existing subspace based versions of the linear methods are reviewed. Serial and parallel interference cancellation techniques are presented. Finally, a review of jammer mitigation techniques is provided and summarized.

Chapter 4 starts the novel contributions of the thesis by presenting the re-

ceivers based on ICA. A brief review of the concepts of ICA is provided in this chapter. The methods FastICA and JADE are described and pointers are provided to additional techniques. Another BSS technique called TDSEP is also summarized. The next part of this chapter involves description and development of the ICA-RAKE, ICA-RAKE Pre-Switch and ICA-RAKE Post-Switch algorithms. Each algorithm is defined and described. A short pseudo code is provided for each algorithm, and finally several appropriate scenarios applying the algorithm are presented. Each scenario looks at the effect of a particular parameter on the algorithm *e.g.*, the effect of varying signal-to-jammer ratio (SJR) while keeping other parameters constant, or the effect of a particular interference on the performance of the algorithm *e.g.*, the effect of narrow-band jamming. A short analysis of the complexity of the algorithms is made. Finally, theoretical lower-bounds are investigated in a "semi-analytical" manner. The algorithms and the results of this chapter are in part included in **Publications 1-8, 10**.

A denoising source separation (DSS) based approach is described in Chapter 5. A short summary of the generic DSS algorithm precedes the development of the DSS algorithm for interference cancellation. Again, the algorithm is evaluated for different scenarios. **Publication 9** includes some of the results in this chapter.

Chapter 6 examines these methods in the context of existing methods. ICA can often be compared to adaptive beamforming techniques, while the denoising source separation based receiver is similar to the despread-respread [162] concept. Additionally, the notion of blindness is examined in relation to both the signal processing and communication technology fields.

Open questions, future trends and directions are examined in Chapter 7, which also concludes this thesis.

An approach for using prior information in source separation and interference cancellation is presented in the Appendix A. A general framework and motivation for the use of prior information is provided following the class of priors described. These priors, called *conjugate priors* require minimal changes on the algorithm. The ICA-RAKE Post-Switch is adapted to incorporate the priors, and the algorithm is evaluated with the help of a scenario. **Publication 12** describes this theory in relation to image analysis.

Finally, in Appendix B the experimental setup, the data structure, and the simulation methodology are described briefly.

Chapter 2

A Mathematical Model for CDMA Mobile Communications

Allocation of resources, enabling several transmitters to share a common channel is the essence of *multiaccess communication*. This idea dates back to the late 19th century when Thomas A. Edison invented the *diplex*, which enabled the simultaneous transmission of two telegraphic messages in the same direction through the same wire [192]. Modern day examples of several transmitters using a common channel include mobile telephones, pocket-radio networks, interactive cable televisions, *etc.*

Traditionally, each source — referred to as *user* — is provided with a certain resource in a multiaccess system. This resource can either be frequency or time slots, or both. The main question is how this common resource is allocated among several users keeping in mind that each user in the system must be able to communicate despite the fact that other users occupy the same resources, perhaps simultaneously. As the users in the system increase, the demand to allocate optimally these resources as efficiently as possible also increases.

The first multiple access technique, *frequency-division multiple access* (FDMA), refers to a principle where several users coexist by having different carrier frequencies, so that the resulting spectra do not overlap and the users are separated orthogonally. An example of FDMA-based system is the Nordic mobile telephone (NMT) system [102] in Scandinavia during the early '80s.

Time-division multiple access (TDMA) is FDMA's equivalent in the time domain. Here, orthogonality between users is maintained by allocating each user a separate time slot. Hence, one user can transmit and receive data during his predetermined time interval. The popular mobile communication technique GSM (groupe special mobile), is based on the principle of dividing a given frequency slot among several users by means of TDMA. A comprehensive review on the GSM system can be found in [174]. While TDMA requires the users to maintain time synchronism, FDMA allows for completely uncoordinated transmissions in the time domain.

A rather disjoint method of multiple access is the *random multiaccess* method where the user transmits the message as if he or she were the sole user in the system. If indeed nobody else is transmitting, the message is received. However, since the users are uncoordinated, if another user is transmitting simultaneously, the messages will interfere — called *collisions*. If collisions occur, the users wait for a random period and then retransmit their data. This random communication system is essentially not orthogonal, but provides a level of orthogonality due to its inherent randomness. A modification of the above random multiaccess scheme is the ethernet scheme used in local area networks [178].

The fundamental weakness in the above methods is the assumption that all users transmit continuously for the same length of time. This might be nearly true for packet data as in random multiaccess, but is not true in the case of voice transmission. Time slots are wasted when users' activity ceases prematurely and frequency reuse is non-optimal. These factors lead to wastage of resources and reduce channel capacity [198].

As seen in the above multiaccess techniques, the ability to separate signals orthogonally is of paramount importance. While FDMA and TDMA use unique frequency and distinct time slots, a completely different approach is to use orthogonal codes. The multiaccess technique using codes is called *code-division multiple access* (CDMA). Each user now has a unique code. Heuristically speaking, every user applies a unique code — called *spreading* — to the information signal before transmitting it through the common medium. Since all users can transmit at all times, the signals interfere with each other. Again by applying this unique code, the transmitted signal of each user is recovered from the mixture by the receiver. Examples of systems using CDMA are the American IS-95 system [181] and the just evolved European third generation (3G) universal mobile telecommunication system (UMTS) [131]. Fig. 2.1 shows the various multiple access schemes.

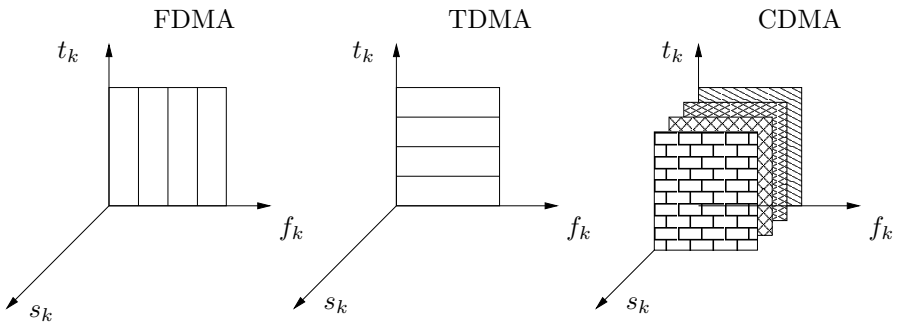


Figure 2.1: *Multiple access schemes: FDMA, TDMA, CDMA. While FDMA uses distinct frequencies, and TDMA uses separate time slots, CDMA uses orthogonal codes to separate the users.* f_k *is the frequency along the* X*-axis,* t_k *is the time along the* Y*-axis and* s_k *is the code along the* Z*-axis.*

CDMA, in a very simple form consists of using a sequence of ± 1 as the code to spread the information signal. This code is then called the *spreading code* or *chip sequence*. In this case the access method is termed as *direct sequence-CDMA* (DS-CDMA) [146]. *Frequency hopping* (FH) is another technique, where each user's code is a pattern of carrier frequencies that are used, hopping from one to another in predetermined order and time [146]. DS-CDMA is used in the 802.11b *Wifi* standard [17], while frequency hopping is used in the *bluetooth* system [1]. Enhanced CDMA systems include the combination of multi-carrier systems with CDMA and is termed *MC-CDMA*. A full classification of the various CDMA systems is presented in Fig. 2.2.

Another multiple access scheme particularly relevant in multiple antenna communications is the *space-division multiple access* (SDMA) scheme [163, 186]. SDMA is based on the fact that users are seldom at the same spatial location. The idea of SDMA is to exploit all the information collected in the spatial domain in addition to the temporal dimension in order to achieve improvements in wireless transmission [163]. Hence, SDMA can be seen to use direction (angle) as another dimension in the signal space, which can be channelized and assigned to different users [51]. To this effect SDMA makes use of directional antenna as shown in Fig. 2.3. Orthogonal channels can be assigned only if the angular separation between users exceeds the angular resolution of the directional antenna. When using an antenna array, precise angular resolution requires a large array. This is impractical for the base station or access point and is certainly un-

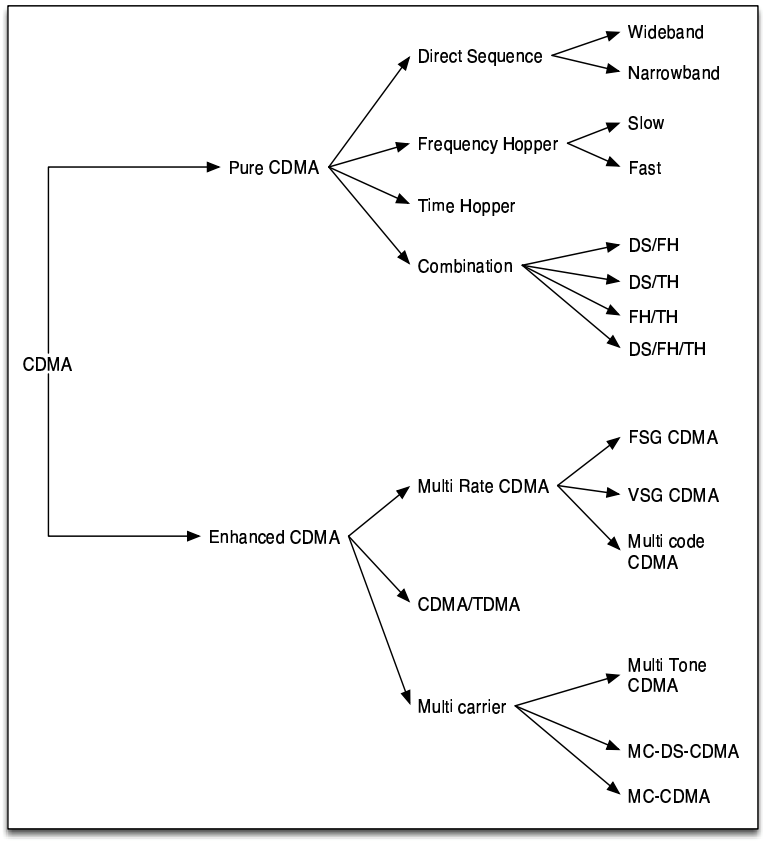


Figure 2.2: *Classification of various CDMA systems. Direct Sequence and Multicarrier CDMA are very popular in the area of mobile wireless communications. Adapted from [43, 192, 198].*

feasible in small user terminals [51]. In practice, SDMA is usually implemented using sectorized antenna arrays, where the 360° angular range is divided into N sectors. There is high directional gain in each sector and little interference between sectors.

Finally, IDMA [142, 143, 144] or *interleave-division multiple access* is a multi-

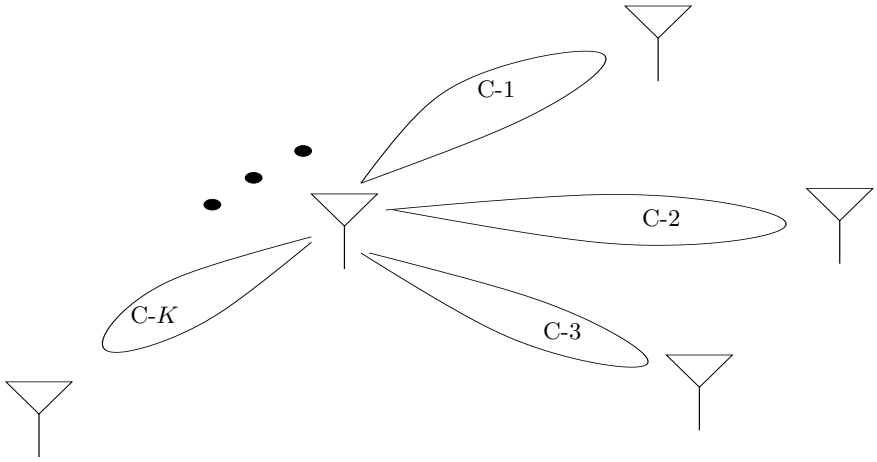


Figure 2.3: *Space-division multiple access. SDMA uses direction to separate users. This is achieved by the use of multiple antenna systems. By using antenna arrays each user is assigned a separate channel (C-1, C-2, . . . , C-K). SDMA is usually implemented by dividing the* 360° *angular range into N sectors, termed as a sectorized antenna array. Adapted from [51].*

ple access scheme that uses interleaving as a technique to distinguish the users. Here, the users are separated by different chip-level interleaving methods instead of the different signatures as in a conventional CDMA scheme. IDMA is a wideband scheme, and inherits many of the advantages from CDMA, in particular, diversity against fading. Random interleaving, combined with forward error correction (FEC) combats the fading effect [142]. IDMA is particularly effective in narrowband applications with a small number of users [18, 142].

2.1 Direct sequence spread spectrum

2.1.1 Simplified Model

Assuming a system with only one user, before transmission each symbol $b[i]$ of the binary data (narrow band information sequence) is 'directly' multiplied with a sequence s – termed as *spreading sequence* – which is independent of the binary

data to produce a baseband signal as

$$x[i] = s.b[i], \tag{2.1}$$

The information sequence of the user $b[i]$ has a symbol duration T and a symbol rate $R = 1/T$, while the chip duration of the spreading sequence is T_c with a chip rate $R_c = 1/T_c$. This effect of multiplication of the information sequence $b[i]$ with the spreading sequence s is to spread the baseband bandwidth R of $b[i]$ to a baseband bandwidth of R_c.

The chip sequence is also termed the *pseudo noise* (PN) sequence. The ratio between the durations $C = T/T_c$ (T is the symbol duration and T_c is the chip duration) is the processing gain. From simple algebra, a C-dimensional vector space has at most C-orthogonal vectors, and hence for orthogonal synchronous DS-CDMA systems the processing gain gives a theoretical upper limit for the number of users that can be supported [192]. The basic operation of spreading in the time domain is illustrated in Fig. 2.4.

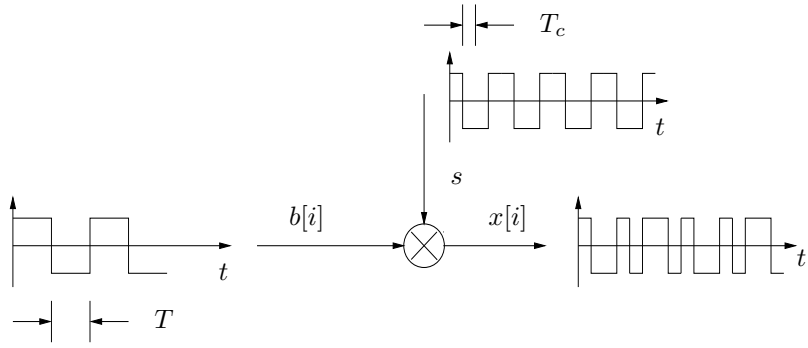


Figure 2.4: *Illustration of spreading for a single user. T_c is the chip duration and T is the symbol duration. The spread signal is x. Here the original bit sequence b is (1,-1,1,-1), spreading code s is $(1, -1, -1, 1)$, and the spread signal is $(1, -1, -1, 1, -1, 1, 1, -1, 1, -1, -1, 1, -1, 1, 1, -1)$.*

The PN sequences are usually generated by very simple shift registers as in [192]. These sequences exhibit noise like random properties and should satisfy three important properties — balance property, run-length property and shift-and-add property [198]. Certain PN sequences called *m-sequences* have the following properties:

- the period of the sequence is $C = 2^n - 1$.
- autocorrelation of the sequence is,

$$E(s[i]s[i+j]) = \begin{cases} 1 & , \quad \text{when} \quad j = 0, C, 2C, \ldots \\ -1/C & , \quad \text{otherwise.} \end{cases} \tag{2.2}$$

Hence, m-sequences have an impulse-like autocorrelation. This implies that a m-sequence is nearly-orthogonal to its time shifted replica of equal length. Another widely used PN sequences are Gold codes. Gold codes have a three-valued autocorrelation and cross-correlation function with values $\{-1, -\zeta(n), \zeta(n) - 2\}$ [140], where,

$$\zeta(n) = \begin{cases} 2^{(n+1)/2} + 1 & , \quad \text{for odd } n, \\ 2^{(n+2)/2} + 1 & , \quad \text{for even } n. \end{cases} \tag{2.3}$$

Kasmami codes [140] are another class of PN sequences. Codes other than PN sequences can also be used for spreading. One such example are orthogonal codes, that have zero cross-correlation when the offset between codes is zero. However, they suffer from poor cross-correlation properties with different offsets. Some examples of these codes are Walsh codes, and orthogonal Gold codes.

In the case where the data would only be subjected to noise, the received data is of the form,

$$\begin{aligned} y[i] &= x[i] + n[i] \\ &= sb[i] + n[i], \end{aligned} \tag{2.4}$$

where,

- $x[i]$ is the *transmitted* or *desired* data,
- $n[i]$ is the *interfering* noise.

2.1.2 DS-CDMA signal model

The previous section dealt with a single user case with direct sequence modulation. In this section a concrete discrete-time model of the DS-CDMA system is

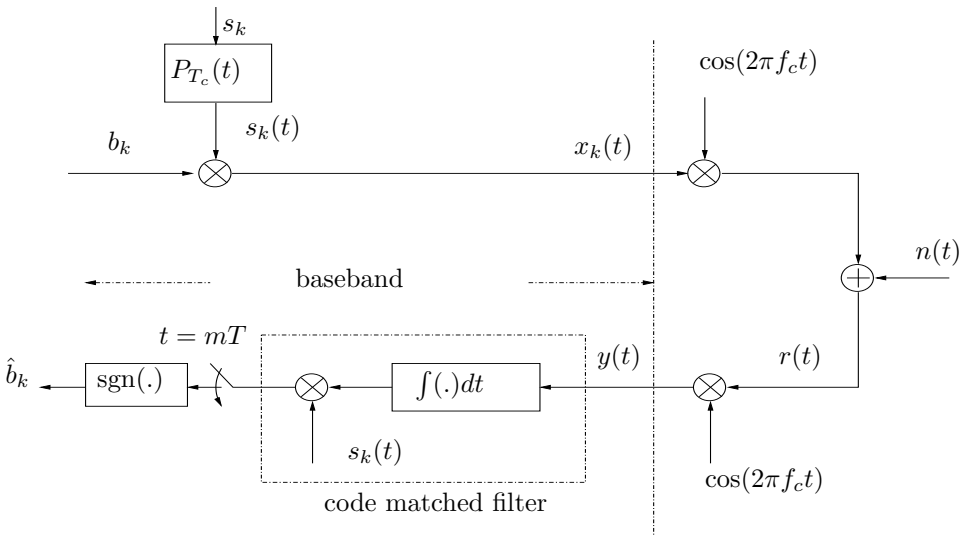


Figure 2.5: *Notations used in the thesis. b_k is the kth user's information bit. $P_{T_c}(t)$ is the rectangular pulse (See* (2.8)*) of duration T_c and s_k is the kth user's spreading code. $x_k(t)$ is the spread signal. $r(t)$ is the received signal and $y(t)$ is the received signal at the baseband.*

constructed for the purposes of digital signal processing. This model is a baseband model with fading/non-fading single or multipath channel, and AWGN. (For simplicity, the notations of this thesis are laid out in Fig. 2.5.) This channel is modeled as in Fig. 2.6. Assuming that the system now has K users, the signal, now sent by user k, is [192],

$$x_k(t) = \sum_{m=0}^{M-1} b_{km} s_k(t - mT), \tag{2.5}$$

which contains the information of M symbols b_{km}. $s_k(.)$ is the kth user's binary chip sequence *i.e.,* the spreading code, supported by $[0, T)$, where T is the symbol duration. The signal passes through a multipath channel (see Fig. 2.6) which is assumed to be fixed during one symbol duration. This implies that $a_{kl}(t) = a_{klm}, t \in [mT, (m+1)T]$. The variable a_{klm} is called the attenuation factor of the lth path, which is a complex number and may vary from symbol to symbol.

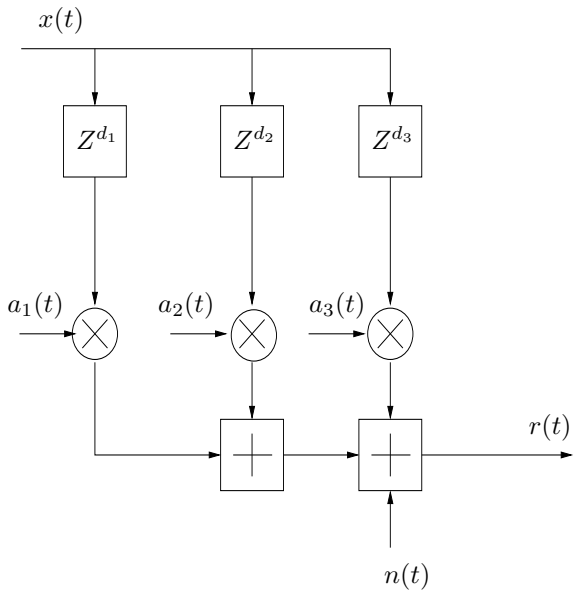


Figure 2.6: *Channel model for a $L = 3$ path channel. $x(t)$ is the data to be sent along the channel. Z^{d_i} is the delay along path i, and is a delay by d_i chips. $a_i(t)$ is the attenuation for path i at time instant t.*

Hence, the received signal has the form,

$$y(t) = \sum_{m=0}^{M-1} \sum_{k=1}^{K} b_{km} \sum_{l=1}^{L} a_{klm} s_k(t - mT - \zeta_{kl} T_c) + n(t), \tag{2.6}$$

where L is the number of resolvable paths, T_c is the chip duration, and $\zeta_{kl} T_c$ is the delay of the lth path of user k, where $\zeta_{kl} = d_{kl} + \delta_{kl}$, with d_{kl} integer and $\delta_{kl} \in [0, 1)$. The delay of each path is assumed to remain constant during the interval of M symbols. $n(t)$ denotes the noise, and the chip sequence length (*i.e.*, processing gain) is $C = \frac{T}{T_c}$. For simplicity, from now it is assumed that $T_c = 1$.

Since the chip sequence $s_k(.)$ is now continuous by definition, it includes not only the binary chips $s_k[i]$, but also a chip waveform $p(t)$. More precisely,

$$s_k(t) = \sum_{i=0}^{C-1} s_k[i] P(t - iT_c), \tag{2.7}$$

where $P(t)$ is supported by $[0, T_c]$ only. This thesis assumes rectangular waveforms for each user, and hence,

$$P(t) = \begin{cases} 1 & , \quad \text{when} \quad 0 \leq t \leq T_c, \\ 0 & , \quad \text{otherwise.} \end{cases} \tag{2.8}$$

In distortionless transmission, the transmitted signal is only subject to scaling and delaying [23]. In this case, the received signal would be of the form,

$$y_{\text{dist}}(t) = \sum_{m=0}^{M-1} \sum_{k=1}^{K} b_{km} a_k s_k(t - mT - \zeta_k). \tag{2.9}$$

Hence, in addition to noise, the channel distortions in (2.6) are modeled by multiple resolvable paths L and possibly time varying path strengths a_{klm}.

Continuous-to-discrete time conversion of the above model can be realized by a chip-matched filter, which is a simple integrate-and-dump device. Using chip-rate sampling, which means integrating over a chip-duration T_c,

$$y[i] = \int_{iT_c - T_c + \tau}^{iT_c + \tau} P(t - (i-1)T_c - \tau) y(t) dt, \tag{2.10}$$

where the timing offset $0 \leq \tau \leq T_c$ is due to the fact that the receiver might not have accurate information about the chip edges. Sampling the same yields C-vectors,

$$\mathbf{y}_m^{[sc]} \stackrel{\text{def}}{=} [y[mC], \quad y[mC+1], \quad \ldots, \quad y[(m+1)C-1]]^T, \tag{2.11}$$

where $[.]^T$ denotes the transpose operation and $[sc]$ denotes that this is a "short-code" model.

The sampled chip sequence of user k is denoted as

$$\mathbf{s}_k \stackrel{\text{def}}{=} [s_k[1], \quad s_k[2], \quad \ldots, \quad s_k[C]]^T. \tag{2.12}$$

Thus each vector sample $\mathbf{y}_m$ contains all the information from time duration CT_c, *i.e.*, one symbol duration. This is termed as having a *processing window* of one symbol duration. It is possible that this sampling is *asynchronous* with respect to the symbol, and the vector might contain samples of two successive symbols. As it is the spread data which is integrated, it means that the first element(s) in $\mathbf{y}_m$ corresponds to the *late* part of the chip sequence $\mathbf{s}_k$.

$$\underline{\mathbf{s}}_{kl} \quad = \quad \underline{\mathbf{s}}_k(d_{kl}) \quad \stackrel{\text{def}}{=} \quad [\quad s_k[C-d_{kl}+1], \quad \ldots, \quad s_k[C], \quad 0, \quad \ldots \quad 0 \quad]^T, \tag{2.13}$$

while the rest correspond to the *early* part

$$\overline{\mathbf{s}}_{kl} \quad = \quad \overline{\mathbf{s}}_k(d_{kl}) \quad \stackrel{\text{def}}{=} \quad [\ \ 0, \quad \ldots, \quad 0, \quad s_k[1], \quad \ldots, \quad s_k[C-d_{kl}] \quad]^T. \quad (2.14)$$

In addition, the sampling is *chip asynchronous*, which is due to τ in (2.10). Chip-synchronous sampling (with respect to the user k's path l) would mean integration with $\tau = \delta_{kl}$, $\delta_{kl} \in [0, 1)$. This would mean examining the effect of $y[i]$ having $\tau \in [\delta_{kl}, \delta_{kl}+T_c]$. But, as sampling is just integration, the chip sequences that are related to those two successive symbols (with respect to each path) are neither exactly (2.13) nor (2.14), but their convex combinations [109, 13]:

$$\underline{\mathbf{c}}_{kl} \stackrel{\text{def}}{=} (1-\delta_{kl})\underline{\mathbf{s}}_k(d_{kl}) + \delta_{kl}\underline{\mathbf{s}}_k(d_{kl}+1), \quad (2.15)$$

$$\overline{\mathbf{c}}_{kl} \stackrel{\text{def}}{=} (1-\delta_{kl})\overline{\mathbf{s}}_k(d_{kl}) + \delta_{kl}\overline{\mathbf{s}}_k(d_{kl}+1). \quad (2.16)$$

Hence, the vector sample has the known form [13, 75],

$$\mathbf{y}_m^{[sc]} = \sum_{k=1}^{K}\sum_{l=1}^{L}\left(a_{kl,m-1}b_{k,m-1}\underline{\mathbf{c}}_{kl} + a_{klm}b_{km}\overline{\mathbf{c}}_{kl}\right) + \mathbf{n}_m, \quad (2.17)$$

where $\mathbf{n}_m$ denotes noise. The model in (2.17) is the *subspace* model based on short codes. This model was introduced by [13], and is used widely in subspace-based detection techniques [65, 109, 201]. This model is used in this thesis as the methods proposed are subspace based.

A more compact representation of the data in (2.17), can be obtained by defining the code matrix $\mathbf{G}$ with dimensions $C \times 2KL$ as,

$$\mathbf{G} = [\underline{\mathbf{c}}_{11}, \quad \overline{\mathbf{c}}_{11}, \quad \ldots, \quad \underline{\mathbf{c}}_{1L}, \quad \overline{\mathbf{c}}_{1L}, \quad \ldots \quad , \quad \underline{\mathbf{c}}_{KL}, \quad \overline{\mathbf{c}}_{KL}], \quad (2.18)$$

and a 2×1 vector $\mathbf{z}_{klm}$ where,

$$\mathbf{z}_{klm} = [a_{kl,m-1}b_{k,m-1}, a_{klm}b_{km}]^T. \quad (2.19)$$

Stacking all the vectors $\mathbf{z}_{klm}$ into a $2KL$-vector $\mathbf{a}_m$,

$$\mathbf{a}_m = [\mathbf{z}_{11m}^T, \ldots, \mathbf{z}_{1Lm}^T, \ldots, \mathbf{z}_{KLm}^T]^T, \quad (2.20)$$

(2.17) can be rewritten in a matrix form as:

$$\mathbf{y}_m^{[sc]} = \mathbf{G}\mathbf{a}_m + \mathbf{n}_m. \quad (2.21)$$

In this representation, $\mathbf{G}$ depends on the spreading codes, and on paths via their gains and delays, while $\mathbf{a}_m$ depends on the symbols, and on paths via their numbers and strengths.

If the channel remains fixed during the period of observation, *i.e.*, $a_{klm} = a_{kl}, \forall m = 1, \ldots, M$, then this model can be simplified further. Now, each path is a scaled and delayed copy of the first path. Formally, this can be represented as

$$\mathbf{y}_m^{[sc]} = \sum_{k=1}^{K} [b_{k,m-1} \sum_{l=1}^{L} a_{kl} \underline{\mathbf{c}}_{kl} + b_{km} \sum_{l=1}^{L} a_{kl} \overline{\mathbf{c}}_{kl}] + \mathbf{n}_m. \tag{2.22}$$

The data model in (2.21) still holds, but the dimensions are reduced. The matrix $\mathbf{G}$ is of size $C \times 2K$, $\mathbf{G} \stackrel{\text{def}}{=} [\mathbf{g}_1, \ldots, \mathbf{g}_{2K}]$ where,

$$\mathbf{g}_{2k-1} = \sum_{l=1}^{L} a_{kl} \underline{\mathbf{c}}_{kl}, \tag{2.23}$$

$$\mathbf{g}_{2k} = \sum_{l=1}^{L} a_{kl} \overline{\mathbf{c}}_{kl}, \tag{2.24}$$

$$\tag{2.25}$$

and the vector $\mathbf{a}_m$ contains only binary symbols,

$$\mathbf{a}_m = [b_{1,m-1}, \quad b_{1,m}, \quad \ldots, \quad b_{K,m-1}, \quad b_{K,m}]^T. \tag{2.26}$$

This data model holds well for several blind source separation tasks. This thesis uses this data model as a starting point and uses an array structure to deal with external interferences as explained in Sections 2.2.2 and 2.2.3.

The second order statistics of the data can be obtained from the autocorrelation matrix as,

$$\mathbf{R} = E\{\mathbf{y}_m^{[sc]} \mathbf{y}_m^{[sc]H}\} = \mathbf{G}\mathbf{A}\mathbf{G}^T + \sigma^2 \mathbf{I}, \tag{2.27}$$

where $\mathbf{A} = E\{\mathbf{a}_m \mathbf{a}_m^H\}$, σ^2 is the noise variance, $\mathbf{I}$ is a $2KL \times 2KL$ identity matrix and H denotes the conjugate transpose.

2.2 Interference sources in CDMA

As with any (digital) communication system, interference is common. It can range from plain noise to smart jammers. Interference can be categorized into the following types:

- interchip interference component, that depends on $b_k[m+n], n \neq 0$. This is usually termed *intersymbol interference* (ISI) for non-spread digital modulation;
- the component due to background noise $n[m]$;
- the other-user interference components in the cell-of-interest, which depend on $b_j[m], j \neq k$ (for user k). This is the *multiuser interference*;
- interference due to users in the neighboring cells - *inter-cell interference*;
- external interference due to coexisting systems in the same band or due to adversaries that intentionally disrupt transmission, usually termed as *jammers*.

This section examines some of these types of interference, and casts them into a model that can be used later in the proposed suppression or mitigation methods.

2.2.1 In-cell interference

Due to the nature of CDMA, all users coexist at all times using both the resources of time and frequency. The other users naturally tend to interfere with the desired user and this interference can be expressed as:

$$I(t) = \sum_{m=0}^{M-1} \sum_{k=1;k\neq k_d}^{K} b_{km} \sum_{l=1}^{L} a_{klm} s_k(t - mT - d_{kl}T_c), \tag{2.28}$$

or in discrete representation as,

$$\mathbf{i}_m = \mathbf{G}^{'} \mathbf{a}_m^{'}. \tag{2.29}$$

Here, k_d is the desired user, $\mathbf{G}^{'}$ is a $C \times 2(K-1)$ code matrix containing all codes except that of user k_d, and $\mathbf{a}_m^{'}$ contains all the elements of $\mathbf{a}_m$ except $a_{k_d m} b_{k_d m}$.

2.2.2 Inter-cell interference

Inter-cell interference refers to interference originating from the adjacent or neighboring cells. This interference is still a part of the network, but it can cause

significant loss in performance. A typical macro-cell geometry with a physical layout of the cell-of-interest and the first two layers of interfering cells is shown in Fig. 2.7. B_0 is the cell-of-interest. The cell B_1 and its neighbors in the radius of d_{in} form the first layer of interfering cells. The outer layer of interfering cells are at the distance of d_{out} from the cell-of-interest.

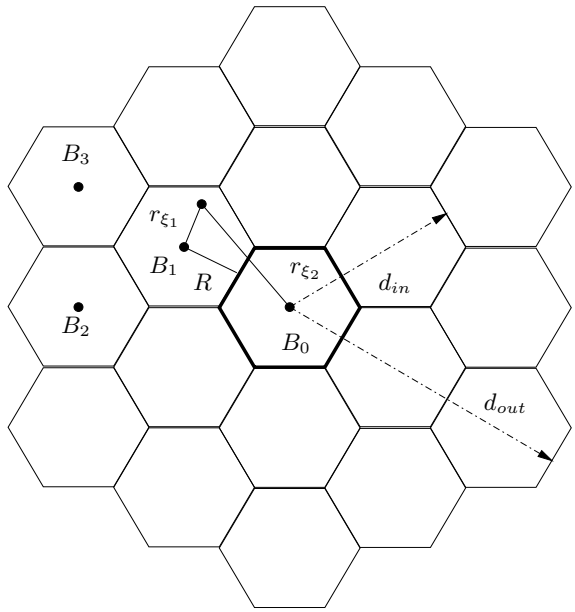


Figure 2.7: *Macro-cell geometry with a physical layout of the cell-of-interest and the first two layers of interfering cells. The inner layer of interfering cells is within a radius* $d_{in} = 2000$ *m while the outer layer is within a distance of* $d_{out} = 3300$ *m. The cell with dark boundaries is the cell-of-interest.*

This interference originating from the users of the neighboring cell depends on the difference due to shadowing and distance loss, between the received energy at the base station-of-interest and the active users' own base station. In a simplified model, each cell can be assumed to have a hexagonal shape, and the out-of-cell interference can be assumed to be limited to the first two layers of interference cells, as in Fig. 2.7. Under these assumptions, the out-of-cell interference can be

expressed as [156, 161] [1],

$$O(t) = \sum_{\xi,m} a_\xi b_{\xi,m} (\frac{r_{\xi_1}}{r_{\xi_2}})^{\gamma/2} 10^{\frac{\chi_{\xi_1} - \chi_{\xi_2}}{20}} s_\xi(t - mT - d_\xi). \tag{2.30}$$

Here, the users are labeled with the symbol ξ, regardless of the cell where the user is. The complex path gain a_ξ includes the target signal strength. r_{ξ_1} and r_{ξ_2} represent the distances of a user ξ to its own base station and to the base station of the cell-of-interest. These distances are assumed to be constant over the observation interval. The parameter γ indicates the power loss. This is assumed to be of fourth order ($\gamma = 4$). The shadowing terms for the ξth user are modeled as log-normally distributed random variables and are constant over the observation interval. Therefore, the variables χ_{ξ_1} and χ_{ξ_2} are Gaussian random variables. These variables are assumed to be independent and to have equal mean $\bar{\xi}$ and standard deviation σ_ξ. Since every user is not shadowed by the same object, due to geographical separation, the shadowing process can be assumed to be independent for each user.

Now the received signal (2.6) contains this interference in addition to information bearing term and can be expressed as:

$$y(t) + O(t), \tag{2.31}$$

where $y(t)$ is the information bearing term and $O(t)$ is the interference term.

Assuming a uniform linear array (ULA) [85] at the receiver, the received signal at the nth sensor ($n = 1, \ldots N_a$), converted to the baseband can be written as,

$$y_n(t) = y(t) e^{j(n-1)\theta_r} + \sum_\xi O_\xi(t) e^{j2\pi(f_\xi - f_c)t} e^{j(n-1)\theta_\xi}, \tag{2.32}$$

where, without loss of generality, the carriers can be assumed to have the same initial phase and arrive from a direction θ_r. In practice, the interfering signals $O_\xi(t)$ arrive from somewhat different directions θ_ξ, which may be closer or farther away from each other depending on the situation. The interference is often fairly weak and the sum constitutes a kind of additive noise term which disturbs reception of the information signal $r(t)$. To simplify the situation, it can be said that all these interfering signals come from the same *average* direction θ_q. Then the interference term in (2.32) can be rewritten as,

$$e^{j(n-1)\theta_q} \sum_\xi O_\xi(t) e^{j2\pi(f_\xi - f_c)t}. \tag{2.33}$$

[1] An exact derivation of the model can be found in [156].

As in Sec. 2.1.2, the short-code model can be derived. At the receiver, a uniform linear array is used, and hence the received data after the array can be represented as,

$$\mathbf{y}_i = \mathbf{\Theta}\mathbf{y}_i^{[sc]} + \mathbf{n}_i, \qquad i = 1, \ldots, N_a, \tag{2.34}$$

where $\mathbf{\Theta}$ is the $N_a \times L$ *array steering matrix* defined as [85],

$$\mathbf{\Theta} = \begin{bmatrix} 1 & \ldots & 1 \\ e^{j\theta_{b1}} & \ldots & e^{j\theta_{bL}} \\ \vdots & \vdots & \vdots \\ e^{j(N_a-1)\theta_{b1}} & \ldots & e^{j(N_a-1)\theta_{bL}} \end{bmatrix}. \tag{2.35}$$

2.2.3 External interference

CDMA, or rather older versions of spread spectrum systems were primarily designed for military communications to overcome a jamming situation, *i.e.*, when an adversary intends to disrupt the communication. In order to disrupt the communication, the adversary needs to do two things: (a) detect that a communication is taking place, and (b) to transmit a jamming signal which is designed to confuse the receiver. Spread spectrum communications are therefore designed to make these tasks as difficult as possible. For an interesting review of the military history of spread spectrum systems refer to [170] or chapter 2 of [173].

In modern CDMA systems, there is no longer a real adversary that wishes to disrupt communications. Jamming here takes place more due to the coexistence of multiple systems in the same band. This jamming is mostly unintentional by nature. Such jammers can either be continuous wave (CW, narrow-band) or pulsed (wide-band) [140]. The jamming signal can be expressed as

$$j(t) = \delta_p(t)\sqrt{J}\exp(i2\pi f_j t + \phi), \tag{2.36}$$

where $i = \sqrt{-1}$, and $\delta_p(t) = 1$ with a probability p during a symbol[2]. It defines the nature of jammer. The jammer is a continuous wave when $p = 1$ and pulsed at the symbol level otherwise[3]. The power, frequency, and the phase of the jammer signal $j(t)$ are denoted by J, f_j, and ϕ respectively. The phase is assumed to be uniformly distributed over the interval $[0, 2\pi)$. Examples of continuous and pulsed jammers are illustrated in Fig. 2.8. The received signal (2.6) is now jammed and

[2]Here, i is used as the imaginary constant as opposed to j used widely in the thesis. This is to differentiate it with the jammer j.

[3]Another example of wide-band jamming is chip-pulsed jamming.

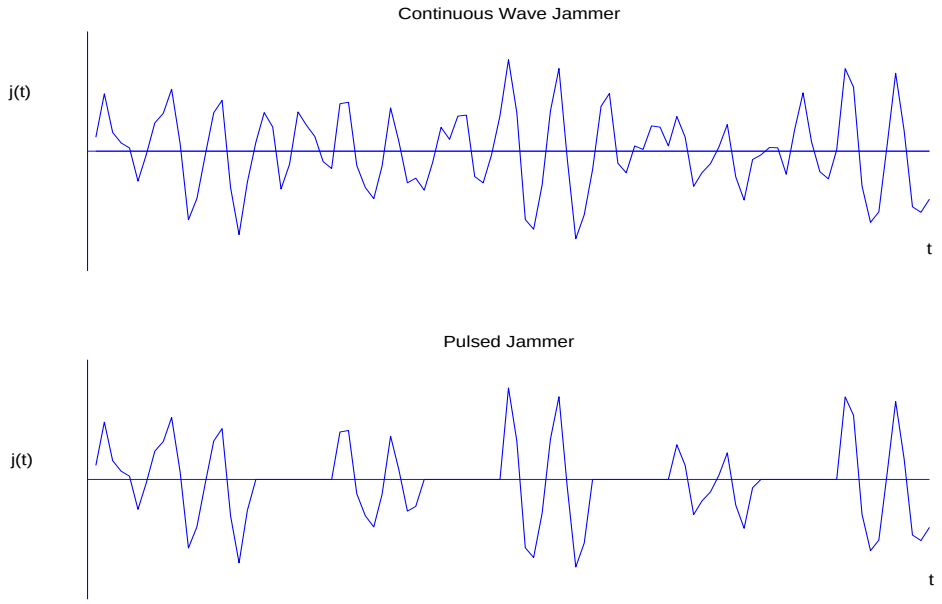


Figure 2.8: *Examples of jammers. The upper sub-figure is a continuous wave (CW) jammer. The jammer at the lower sub-figure is pulsed.*

is of the form:

$$y(t) + j(t), \tag{2.37}$$

where $y(t)$ is the information bearing term and $j(t)$ is the jammer.

Assuming an antenna array at the receiver, and following the approach of Sec. 2.2.2, the model can be expressed as

$$\mathbf{y}_i = \mathbf{\Theta}\mathbf{y}_i^{[sc]} + \mathbf{n}_i, \qquad i = 1, \ldots, N_a. \tag{2.38}$$

where $\mathbf{\Theta}$ is the array steering matrix as defined in (2.35) and $\mathbf{y}_i^{[sc]}$ is the received data according to the short-code model, and N_a is the number of sensors in the antenna array.

2.3 Relation between the models

The relation between the subspace-based short-code model in (2.21) and the array models of (2.34) and (2.38) is shown in Fig. 2.9. Every block of data in the short-code model obtained from the channel along every path is now combined by the array steering matrix $\mathbf{\Theta}$. Assuming that the antenna array has N_a sensors, the

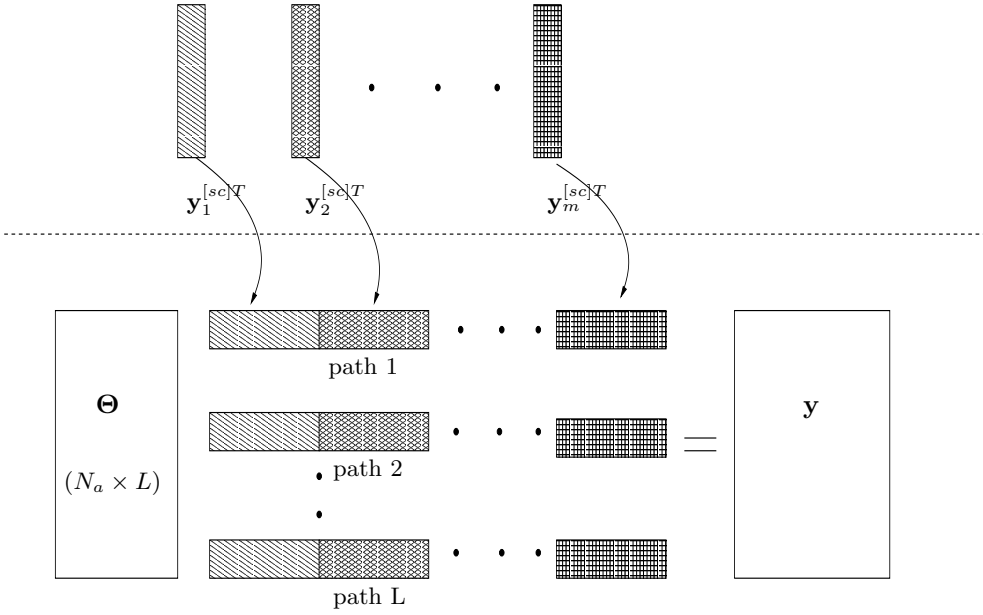


Figure 2.9: *Relation between the short code model of* (2.21) *and the array models of* (2.34) *and* (2.38). *The short-code vectors (above)* $\mathbf{y}_1^{[sc]}, \ldots, \mathbf{y}_m^{[sc]}$ *are of dimensions* $C \times 1$. *In the array model (below), N short-code vectors are stacked together to form a* $1 \times NC$ *received vector (N is the number of blocks considered). Assuming an antenna array with* N_a *sensors, the final received vector* $\mathbf{y}$ *has a dimension of* $N_a \times NC$.

received vector at every sensor is a combination of the array steering matrix with all the data blocks $\mathbf{y}_1^{[sc]}, \ldots, \mathbf{y}_m^{[sc]}$ arranged next to the other. Supposing there are N blocks, the dimension of the received vector $\mathbf{y}$ at each antenna array is now $1 \times NC$, where C is the length of the code, while in the short-code model,

each block of the received vector was of dimension C.

Hence, both the models are quite similar. In the short-code model, the (mixing) structure is contained in the code matrix $\mathbf{G}$, while in the array matrix, the (mixing) structure is contained in the array steering matrix $\mathbf{\Theta}$. Rather than processing each block separately, now every vector along each path is first processed and then the blocks are separated and processed.

Chapter 3

Review of Detection and Interference Rejection Methods

Shannon [172] defined precisely communication as a methodology of *reproducing at one point either exactly or approximately a message selected at another point*, and termed this as the fundamental problem in communication. Today after nearly 60 years, the problem is as fundamental as it was then. Detectors and interference cancelers are those blocks in the chain that either aim to reproduce or approximate the message selected at one point. Here, in the context of modern communications, detection techniques, in addition to demodulating the multiuser data stream, must reject the interference from various sources. These receiver structures can either make a joint detection of the symbols of different users — *centralized*, or can demodulate a signal of one desired user only — *decentralized*. The former class of detectors are sometimes termed joint detectors, while the latter are also called single user detectors. Algorithmically, these detectors can additionally be linear, or can be based on estimation of non-linear functions. Considering the practical importance of the detectors, they can also be classified on the basis of their complexity. Most of these receivers employ some form of information (*training sequences, channel estimation*) for reliable estimation of the transmitted symbols. Another class of detection techniques using as little information as possible has emerged in recent times. These structures are called *blind* receivers. Hybrid structures that utilize some information while still following the framework of blind detectors, have been popular presently. These techniques are collectively called *semi-blind*. Fig. 3.1 provides a classification and summary of various detection techniques.

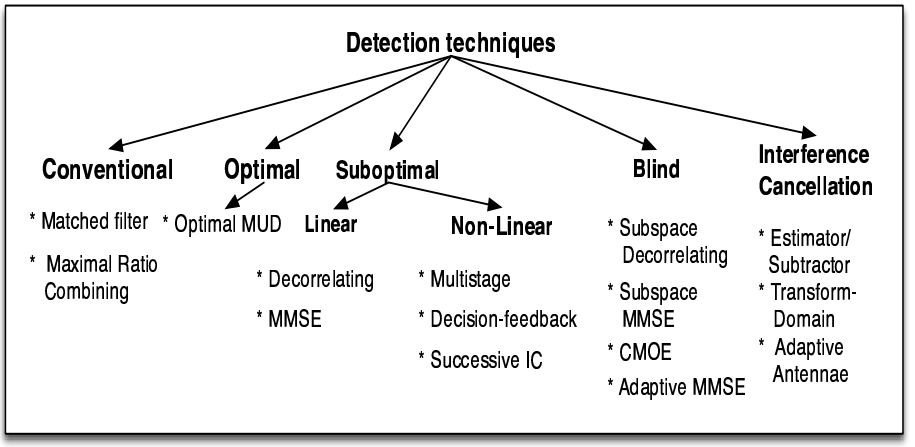


Figure 3.1: *Classification of detectors. This is a general classification. There exist several adaptations and combinations of the above detectors. Partially adapted from [41, 88, 109, 192].*

3.1 Conventional detection

Conventional detection is based on the matched filter detector [192, 198]. This detector simply correlates the received signal with the desired user's spreading code, and samples the output at the bit rate. This is a straightforward extension of the single user design. This detector does not take into account any other users in the system or channel dynamics and is therefore not robust to asynchronous, fading channels or pseudo noise (PN) sequences with substantial cross-correlations. For a synchronous case with a single path channel ($L = 1$), the data (2.21) is of the form[1],

$$\mathbf{y}_m = a_{km} b_{km} \mathbf{s}_k + \mathbf{G}^{'} \mathbf{a}_m^{'} + \mathbf{n}_m, \tag{3.1}$$

[1]Since this chapter is an overview of several detection and interference rejection methods, the models used by all the methods are not necessarily the same. Many of these methods use more specific models. The subspace model developed in Chapter 2 is applicable to sections 3.1, 3.5 and to the following chapters.

where the $C \times 2(K-1)$ matrix $\mathbf{G}^{'}$ contains all the columns of $\mathbf{G}$ except $\mathbf{s}_k$, and the $2(K-1)$ vector $\mathbf{a}_m^{'}$ contains all the elements of $\mathbf{a}_m$ except $a_{km}b_{km}$. Now conventional detection or matched filter involves correlating with the spreading code as,

$$\begin{aligned}\hat{b}_{km} &= \text{sign}(\mathbf{s}_k^T \mathbf{y}_m) \\ &= \text{sign}(\mathbf{s}_k^T a_{km} b_{km} \mathbf{s}_k + \mathbf{s}_k^T \mathbf{G}^{'} \mathbf{a}_m^{'} + \mathbf{s}_k^T \mathbf{n}_m) \\ &= \text{sign}(b_{km} + \boldsymbol{\rho}_k^T \mathbf{a}_m^{'} + \tilde{n}_{km}) \\ &= \text{sign}(b_{km} + \text{MAI}_{km} + \tilde{n}_{km}).\end{aligned} \tag{3.2}$$

The vector $\boldsymbol{\rho}_k^T$ represents the *multiple access interference* (MAI) for the kth user and is represented as,

$$\rho_{i,k} = \mathbf{s}_k^T \mathbf{G}_i^{'}. \tag{3.3}$$

For slowly fading channels, when the impulse response can be estimated accurately and is assumed to be known, these receivers (now the filter is matched to the convolution of the signature waveform of the user s_{km} and the channel impulse response a_{km}) perform well [192]. For multipath channels such a matched filter is called the coherent RAKE receiver [145, 146, 198]. The output of the coherent RAKE receiver for user k is obtained by *maximal ratio combining* (MRC) the matched filter outputs for different propagation paths (Fig. 3.2), *ie.* by,

$$\hat{b}_{km} = \text{sign}(\sum_{l=1}^{L} \hat{a}_{klm}^{*} (\overline{\mathbf{c}}_k^T(\hat{\zeta}_{kl}) \mathbf{y}_m + \underline{\mathbf{c}}_k^T(\hat{\zeta}_{kl}) \mathbf{y}_{m+1})), \tag{3.4}$$

where $\hat{a}_{klm}^{*}$ is the estimate for the path gain during the mth symbol, and $\hat{\zeta}$ is the estimated path delay.

If the delay spread is significantly smaller than the symbol interval, the intersymbol interference can be assumed to be negligible and a "hard decision" on the RAKE output $y_k^{[MRC]}$ yields (near) optimal solution [88]. If the channel introduces ISI, this receiver is no longer optimal, and a receiver that minimizes the error probability is significantly more complicated to implement.

3.2 Optimal multiuser detection

Multiuser detection (MUD) is the method of demodulating a particular user by

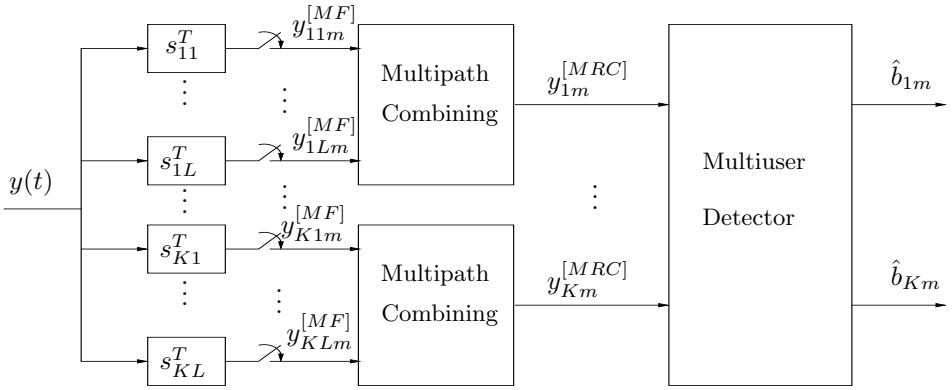


Figure 3.2: *Multiuser receiver structure. The outputs of L paths are combined coherently at this receiver before multiuser signal processing. The superscripts* $[MF]$ *and* $[MRC]$ *refer to outputs after matched filtering and maximal ratio combining respectively. The coherent RAKE receiver is similar in structure. The signal processing can also be performed before multipath combining, in which case the "multipath combining" and "multiuser" blocks in the above figure are interchanged.*

attempting to exploit the structure of the multiple access interference [109, 192, 215]. The optimal detection strategies are referred to individual optimum and jointly optimum respectively [192]. The former criterion results in the minimum bit-error-rate. Jointly optimum decisions are obtained by a *maximum-likelihood sequence detector* (MLSD) that selects the most likely sequence of transmitted bits given the observations. Other conceivable intermediate optimality criteria exist such as demodulating only one user *i.e.,* the receiver selects the most likely sequence of bits for that user, rather than making individually optimal decisions.

Assume that $\mathbf{\Upsilon}_M$ is the $K(2M+1) \times K(2M+1)$ diagonal matrix whose $k+iK$ element is the amplitude of user k's ith symbol. Let $\mathbf{R}$ be the $K(2M+1) \times K(2M+1)$ matrix with coefficients

$$r_{kl}[i] = \int_{-\infty}^{\infty} s_k(t - iT - \zeta_k) s_l(t - iT - \zeta_l), \tag{3.5}$$

with $s_k(), s_l()$ denoting the k and lth user's spreading codes. The matrix $\mathbf{R}$ can

be written as a function of $K \times K$ crosscorrelation matrices,

$$\mathbf{R} = \begin{bmatrix} \mathbf{R}[0] & \mathbf{R}^T[1] & \mathbf{0} & \dots & \mathbf{0} & \mathbf{0} \\ \mathbf{R}[1] & \mathbf{R}[0] & \mathbf{R}^T[1] & \dots & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{R}[1] & \mathbf{R}[0] & \dots & \mathbf{0} & \mathbf{0} \\ \dots & \dots & \dots & \dots & \dots & \dots \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \dots & \mathbf{R}[1] & \mathbf{R}[0] \end{bmatrix}. \tag{3.6}$$

Furthermore, define

$$\mathbf{H} = \mathbf{\Upsilon}_M \mathbf{R} \mathbf{\Upsilon}_M. \tag{3.7}$$

Now, the objective is to compute the $\mathbf{b}$ that maximizes the likelihood function $\exp\left(\frac{-1}{2\sigma^2} \int [y(t) - y^{[MF]}(t)]^2 dt\right)$ [192] (For a detailed derivation, refer to page 167 of [192]),

$$\Omega(\mathbf{b}) = 2\mathbf{b}^T \mathbf{\Upsilon}^T \mathbf{y}^{[MF]} - \mathbf{b}^T \mathbf{H} \mathbf{b}. \tag{3.8}$$

As the observations in the function to be maximized by the jointly optimum decisions are from the matched filter outputs, $\mathbf{y}^{[MF]}$ forms a sufficient statistic for $\mathbf{b}$ [192].

The maximization of $\Omega(\mathbf{b})$ using combinatorial optimization techniques is usually not preferred as the complexity is exponential in the product $K(2M + 1)$ [192]. This complexity is prohibitive in practice due to typically large values of M. By exploiting the structure of the matrix $\mathbf{H}$, the solution can be cast into a dynamic programming problem [192]. This can then be solved by the use of the Viterbi algorithm [44, 197].

The applicability of the optimal MUD is rather limited due to two limitations:

- it is computationally complex;
- it is not applicable in downlink environments, due to the fact that the detector should know all the codes of the interfering users, which is not the case in the downlink receiver.

3.3 Linear suboptimal detection

Linear multiuser detectors are equalizer type detectors that process the matched filter output vector $\mathbf{y}^{[MF]} = [\mathbf{s}_1^T\mathbf{y}_m, \ldots, \mathbf{s}_K^T\mathbf{y}_m]^T$ (or the maximal ratio combined vector $\mathbf{y}^{[MRC]}$) by a linear operator. In other words, the output of the linear multiuser detector is [41],

$$\tilde{\mathbf{y}}_m = \mathbf{W}\mathbf{y}_m^{[MF]}. \tag{3.9}$$

Different choices of the matrix $\mathbf{W}$ yield different multiuser receivers. The identity matrix $\mathbf{W} = \mathbf{I}_{MKL}$, is equivalent to a conventional single user receiver.

3.3.1 Decorrelating detector

The *decorrelating detector* (DD), which completely removes the MAI, is obtained when the choice of the matrix $\mathbf{W}$ is the inverse of the cross-correlation of the codes $\mathbf{R}$ (the cases for the invertibility of $\mathbf{R}$ are discussed in [106]) as [192],

$$\mathbf{W}_{DD} = \mathbf{R}^{-1}. \tag{3.10}$$

The decorrelating detector has several desirable features. It does not require the knowledge of the users' powers, and its performance is independent of the power of the interfering users. The only requirement is the knowledge of the timing information. In addition, when the users' energies are not known, and the objective is to optimize performance for the worst case MAI scenario, the decorrelating detector is the optimal approach [106]. However, this receiver involves matrix inversion and its performance degrades as the cross-correlations between users increase [41]. The decorrelating detector is also closely related to the zero-forcing equalizer [146, 192].

A partial decorrelator, which also makes the additive channel noise component white, is the *noise-whitening* (NW) decorrelator. It is obtained when,

$$\mathbf{W}_{NW} = \mathcal{L}, \tag{3.11}$$

where $\mathcal{L}$ is the lower Cholesky factor of $\mathbf{R}$ such that $\mathbf{R} = \mathcal{L}^{\mathbf{T}}\mathcal{L}$ [40].

3.3.2 MMSE multiuser detector

Another approach from estimation theory is to assume that the information symbols b_{km} are independent and uniformly distributed. When the channel is known, $\mathbf{W}$ can be chosen in a way to minimize the mean square errors at the detector outputs (the so called *linear minimum mean-square error* (LMMSE) [92] detectors) *i.e.*,

$$E\{(b_k - \mathbf{w}_k^T \mathbf{y})^2\}. \tag{3.12}$$

The $\mathbf{W}_{opt}$ that solves this cost function is of the form [192],

$$\mathbf{W}_{LMMSE} = [\mathbf{R} + \sigma^2 \boldsymbol{\Upsilon}^{-2}]^{-1}. \tag{3.13}$$

The LMMSE receiver performs well at low *signal-to-noise* (SNR) ratios and maximizes the *signal-to-interference-plus-noise* (SINR) ratio [111]. It is attractive due to its applicability to decentralized implementation. Adaptive versions of the LMMSE algorithm exists, and can be used for slowly fading channels. For an exhaustive treatment of MMSE detectors please refer to [192].

3.4 Non-linear suboptimal detection

Detectors that utilize feedback to reduce MAI in the received signal widely belong to a class of non-linear suboptimal detectors. These algorithms can be divided into three broad classes[2]:

- multistage detectors [189, 190, 205];
- decision-feedback detectors [39, 40];
- successive interference cancelers [138, 189].

The first two classes of algorithms utilize previously made decisions on the other users to cancel the interference present in the signal of the desired user. These algorithms are termed *decision-directed* algorithms. They require the estimation

[2]These categories are not actually disjoint and some realizations of suboptimal receivers may use a combination of these classes. The classification here is to facilitate the presentation of these detectors.

of channel parameters and coherent detection. The third class of detectors use the soft decision outputs rather than hard decision outputs to remove the MAI components and are described in Sec. 3.6. These algorithms lead themselves to non-coherent implementation. Several representatives of these three classes of non-linear detectors are described in Chapter 7 of [192].

3.5 Blind detection

Blind detectors or blind adaptive receivers are receivers that do not require knowledge of the signal waveforms and propagation channels of the interfering sources. They usually require a minimal amount of information about the desired signal. Formally, blind schemes can be classified based on the knowledge the receiver assumes as [109][3]:

(C1) the receiver knows the timing (or more generally, the channel) and the spreading waveform of the desired user;
(C2) the receiver knows only the spreading waveform of the desired user;
(C3) the receiver does not know any information about the desired user, other than the fact that the desired user is digitally modulated at a given symbol rate.

An example of the receiver in category (C1) is the adaptation of the linear MMSE receiver with a *constrained minimum energy output* (CMOE) criterion [64, 65]. Some of the receivers in this category are similar to minimum variance beamforming for adaptive antennas [85]. The scheme involves determining the direction of arrival of the desired signal, and then the interference can be nulled by ensuring that the output variance is minimized. However, this requires the knowledge of the spreading codes and the propagation channel of the desired user, and is analogous to knowing the desired user's array response in beamforming. Another device in this category is the ICA-RAKE receiver [159]. It uses the principles of category (C3) while still employing some timing information.

Receivers in category (C2) can be formulated by adapting the linear MMSE receiver. There are also several training-based adaptive schemes based on criteria other than the linear MMSE. It is also possible to obtain recursive decorrelating detectors [24]. Neural network approaches using training-data exist [2, 121, 179], though they outperform linear approaches only when the training data is available

[3]See Sec. 6.1 of Chapter 6 for discussions about the idea of "blind" algorithms.

for all users. The RAKE-ICA receiver [158] is an example of systems in this category that employ some ideas from category (C3).

Blind equalization for a single receiver [87, 183], and blind source separation [20] for a multiuser system form the general area into which receivers from category (C3) fit in. The DSS receiver [152] is based on the principles of blind source separation (Sec. 4.1 of Chapter 4 contains an overview on blind source separation.) These methods fit into the CDMA system quite well, and if additional information is available, then they can be exploited to simplify the implementation and to improve the results. Some applications of these receivers are mitigation of the effects of non-cooperative applications. This is one of the applications targeted in this thesis.

3.5.1 Subspace based detectors

In (3.1), matrix $\mathbf{G}^{'}$ contributes to the interference, and hence the columns of $\mathbf{G}^{'}$ span a subspace termed as the *interference* space, denoted by $\mathcal{G}_{\mathcal{I}}$. Let ${\mathcal{G}_{\mathcal{I}}}^{\perp}\mathbf{s}_k$ denote the projection of $\mathbf{s}_k$ orthogonal to the interference subspace, as illustrated in Fig. 3.3.

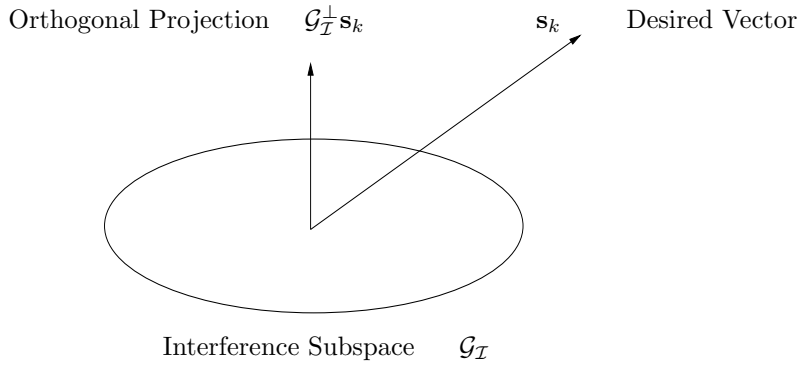


Figure 3.3: *Illustration of the interference subspace $\mathcal{G}_{\mathcal{I}}$ and the orthogonal projection ${\mathcal{G}_{\mathcal{I}}}^{\perp}\mathbf{s}_k$ of the code. If this projection is non-zero, then it is possible to choose a linear correlator that preserves the energy of the desired vector, while ensuring that the interference contribution is zero.*

If the projection $\mathcal{G}_{\mathcal{I}}{}^{\perp}\mathbf{s}_k$ is nonzero, then a linear correlator chosen along this direction, forces the contribution of the interference to be zero, while preserving some energy of the desired vector (due to its nonzero projection in the desired direction). This is the subspace version of the decorrelating detector described in Sec. 3.3.1. Formally, the subspace decorrelating detector is a solution to,

$$\arg\min_{\mathbf{w}} E\{|\mathbf{w}^H\mathbf{G}\mathbf{a}_m|^2\}, \quad \text{subject to} \quad \mathbf{w}^H\mathbf{s}_k = 1. \tag{3.14}$$

Such a subspace decorrelating detector has the form [201],

$$\mathbf{w}_{SDD} = \frac{1}{\mathbf{s}_k^T\mathbf{U}_s(\mathbf{\Lambda}_s - \sigma^2\mathbf{I})^{-1}\mathbf{U}_s^H\mathbf{s}_k}\mathbf{U}_s(\mathbf{\Lambda}_s - \sigma^2\mathbf{I})^{-1}\mathbf{U}_s^H\mathbf{s}_k, \tag{3.15}$$

where σ^2 is the estimated noise variance. The dimensions of signal subspace parameters (eigenvectors and eigenvalues) $\mathbf{U}_s$ and $\mathbf{\Lambda}_s$ are $C \times K$ and $K \times K$ respectively. The mth symbol can then be estimated as,

$$\hat{b}_{km} = \text{sign}(\hat{a}_{km}^*\mathbf{w}_{SDD}^H\mathbf{y}_m). \tag{3.16}$$

Although the decorrelating detector totally eliminates MAI, there is some loss of signal energy relative to the noise energy at the output of the detector *i.e.*, immediately preceding the sign decision. This is because the correlator is along the direction of the orthogonal projection $\mathcal{G}_{\mathcal{I}}{}^{\perp}\mathbf{s}_k$. This implies that for Gaussian noise, the exponent of decay of error probability with E_b/N_o is reduced by the factor by which the signal energy is reduced [109, 157].

To overcome this effect, the detector must estimate a symbol from the noisy observations which gives the best fit to the actual symbol. This leads to minimizing the mean-square error (MSE) of the objective function of the form,

$$\arg\min_{\mathbf{w}} E\{|b_{km} - \mathbf{w}^H\mathbf{y}_m|^2\}, \quad \text{subject to} \quad \mathbf{w}^H\mathbf{s}_k = 1. \tag{3.17}$$

So, the detector is of the form [201]:

$$\mathbf{w}_{SMMSE} = \frac{1}{\mathbf{s}_k^T\mathbf{U}_s\mathbf{\Lambda}_s^{-1}\mathbf{U}_s^H\mathbf{s}_k}\mathbf{U}_s\mathbf{\Lambda}_s^{-1}\mathbf{U}_s^H\mathbf{s}_k. \tag{3.18}$$

Both the detectors in (3.15) and (3.18) become identical when the background noise tends to zero. They can be viewed as matched filters operating in the signal subspace. The symbol estimate for both the detectors are:

$$\hat{b}_{km} = \text{sign}(\alpha(\mathbf{T}\mathbf{y}_m)^H(\mathbf{T}\mathbf{s}_k)), \tag{3.19}$$

where α is a scalar, and in the MMSE case, $\mathbf{T} = \mathbf{\Lambda}_s^{-1/2}\mathbf{U}_s^H$. Hence, both the data and the code are first projected to the signal subspace which is followed by matched filtering.

In the case of multipaths, the processing window must be enlarged. With a window size of $D+1$ symbols, the subspace DD and MMSE detectors are of the form [200]:

$$\begin{aligned}\tilde{\mathbf{w}}_{SDD} &= \sum_{l=1}^{L} \frac{1}{\tilde{\mathbf{c}}_{kl}^T \tilde{\mathbf{U}}_s (\tilde{\mathbf{\Lambda}}_s - \sigma^2\mathbf{I})^{-1} \tilde{\mathbf{U}}_s^H \tilde{\mathbf{c}}_{kl}} \tilde{\mathbf{U}}_s \\ &\qquad (\tilde{\mathbf{\Lambda}}_s - \sigma^2\mathbf{I})^{-1} \tilde{\mathbf{U}}_s^H \tilde{\mathbf{c}}_{kl} \qquad (3.20)\\ \tilde{\mathbf{w}}_{SMMSE} &= \sum_{l=1}^{L} \frac{1}{\tilde{\mathbf{c}}_{kl}^T \tilde{\mathbf{U}}_s \tilde{\mathbf{\Lambda}}_s^{-1} \tilde{\mathbf{U}}_s^H \tilde{\mathbf{c}}_{kl}} \tilde{\mathbf{U}}_s \tilde{\mathbf{\Lambda}}_s^{-1} \tilde{\mathbf{U}}_s^H \tilde{\mathbf{c}}_{kl}. \qquad (3.21)\end{aligned}$$

Now, the dimensions of $\tilde{\mathbf{U}}_s$ and $\tilde{\mathbf{\Lambda}}_s$ are $(D+2)C \times (D+2)K$ and $(D+2)K \times (D+2)K$, respectively.

These methods assume that the desired user's timing is known, which is not the case in practice. However, an accurate estimate of the timing is required to avoid performance loss [137]. The timing can be performed by N_t training symbols [109],

$$\sum_{l=1}^{L} \tilde{\mathbf{c}}_{kl} = \frac{1}{N_t} \sum_{m=1}^{N_t} b_{km} \tilde{\mathbf{y}}_m. \qquad (3.22)$$

Other methods for joint timing and symbol demodulation are based on the *multiple signal classification* (MUSIC) as in [13, 176] or as in [108] where the CMOE criterion is used. Multipath scenarios have been considered in [200].

3.5.2 Other blind approaches

The MMSE criterion is a starting point for several blind approaches. Adaptive MMSE [111] and improved MMSE [25] use training symbols and coarse timing estimation [111]. Paper [110] describes training symbols for the initial timing acquisition, and hence avoids coarse timing information completely for estimating the desired user's symbols. Bounds for the average near-far resistance have been derived in [112] with random (short) codes.

When the channel is fading quickly compared with the symbol rate, *least mean-squares* (LMS) and *recursive least squares* (RLS) based adaptive algorithms do not work [109, 146]. Paper [7] considers the behavior of the MMSE receiver in frequency flat-fading channels. In [99] linear MMSE receivers for frequency selective channels were considered. Here, adequate training for the LMS was provided by the conventional RAKE receiver.

The receivers of category (C3) are the most challenging approach due to the fact that neither training symbols nor the code information is used. The only assumptions are on the symbol rate and the model order. By using the *Akaike information criteria* (AIC) or the *minimum description length* (MDL) [28, 107] a coarse estimate for the model order can be made. The constant modulus (CM) criterion [87] can then be used for estimating the desired user's symbols. The constant modulus criterion tries to find a filter that yields a symbol whose modulus is as close as possible to some constant value. Papers [50, 185] were the first to apply this principle to ISI suppression. The algorithms were based on a linearly constrained and multistage *constant modulus algorithm* (CMA). CMA algorithms are particularly interesting due their close relationship to MMSE [33, 34, 35]. The local minima of the CMA contrast function correspond to the MMSE solutions. Since there is another minimum — which is the global minimum — this usually means that CMA results in better performance. This local minimum is due to the fact that MMSE heavily relies on the accuracy of the timing or channel parameter estimation. This relation between CM algorithms and Wiener (or MMSE) receivers has been shown in [211]. CM algorithms can be regarded as another method to solve the blind source separation problems for which they can act as objective functions [20]. Sec. 4.1 of Chapter 4 deals with blind source separation. Some blind information-theoretic algorithms are discussed in [199]. Several blind methods for signal reception are summarized in [202].

3.6 Interference cancellation

Interference Cancellation[4] is quite similar to multiuser detection[5]. These receivers generally estimate the multiple-access or the multipath induced interference and then subtract the interference estimate from the MF (or MRC) output. This can be viewed as an approximation of the MLSD receiver with the assumption that the data, the amplitude, and the delays of the interfering users are known [189]. Several principles for estimating the interference lead to several IC techniques. The interference can be canceled on a user by user basis leading to *serial* or *successive* interference cancellation (SIC). The other approach is to estimate the interference and cancel it simultaneously from all the users, leading to *parallel* interference cancellation (PIC).

In terms of multiple access interference, the received signal can be described as,

$$\mathbf{y}_k = \text{desired signal} + \text{MAI due to stronger users}(1, \ldots, k-1) + \text{noise}. \quad (3.23)$$

This actually means that the strongest user often interferes with the desired user to the maximum extent. The strongest user is also least affected by MAI. These two points form the basis of *successive interference cancellation* (SIC) [138]. Hence, SIC cancels signals in the order of their strength *i.e.*, estimates of the interference from user 1 (the strongest user) are subtracted from the MF outputs of the other users. The second strongest user is estimated next and subtracted. This requires that the users must be ordered before cancellation of the MAI. This is a problem in fast fading channels. The SIC has the inherent problem in asynchronous CDMA systems, that the processing window of user 1 must ideally be K symbols so that the MAI caused by the users $1, \ldots K-1$ can be canceled from the matched filter output of user K [124]. Another problem is that the SIC might not yield good results in heavily loaded systems where the performance of conventional receiver is poor. This is often the reason why the initial estimate for the SIC is obtained via a conventional receiver. If the MAI

[4]Cancellation, rejection and suppression are terms used interchangeably in literature quite often. From a purist point of view, these are different even though they all perform essentially similar tasks. Cancellation involves estimating and subtracting, while rejection and suppression does not necessarily mean subtraction. Rejection and suppression are in effect filtering techniques, that block the interfering source.

[5]The similarity between interference cancellation and multiuser detection stems from the fact that they both cancel interferences. Multiuser detection deals with canceling other user's interference in the system. Interference cancellation can be thought of canceling all other interferences *e.g.*, jammers *etc.*

estimate of the signal of user 1 is poor in the cancellations, the estimation errors propagate to all users. However, the SIC has good performance in systems where the users have distinctly different strengths [63]. This cannot be the case though in systems with a very large number of users. The SIC has also been applied to multicarrier CDMA systems [42].

Another method is to subtract simultaneously the signals originating from all the interfering users from the target user's signals. This is the *parallel interference cancellation* (PIC) [189]. The PIC techniques in general are multistage, where channel and data estimates of the $(n-1)$th stage aid the estimation process in the nth stage. The PIC receiver performs quite well when all the users are of equal strength. Several adaptations of the PIC method exist for both slowly fading channels [125, 165, 190] and relatively fast fading channels [62, 68, 101, 100].

Since both SIC and PIC must operate fast enough to keep up with the bit rate and not introduce intolerable delay, the number of cancellation stages must be limited. This is consistent with the objective of controlling complexity by choosing an appropriate performance/complexity trade-off. Some practical methods for implementing them have been proposed in [89]. Matrix-algebraic approaches to SIC's and linear PIC's have been proposed in [154, 177], where it has been shown that linear SIC schemes correspond to matrix filtering on the received chip-matched filtered signal. Combinations of SIC-PIC receivers also exist [133, 169].

Groupwise interference cancelers separate the users into groups, detect the symbols of the users within some group, and form an estimate of all the MAI caused by the users within that group based on the symbol decisions. The MAI estimate is then subtracted from the other users' MF outputs. The groupwise interference cancellation can be performed either in serial [188] or in parallel [3, 54].

3.6.1 Narrow-band and wide-band interference cancellation

While the above class of interference cancelers concentrated on MAI, CDMA is considered to be robust to other sources of interference such as jammers (Sec. 2.2.3). This is due to the use of spreading codes (also known as *bandwidth expansion*) in CDMA systems. In general, this is acceptable, however active suppression of external sources of interference can lead to improvement in

performance gains [119]. This section provides a general overview of *narrow-band interference* (NBI) suppression techniques based on signal processing paradigms.

Estimator-Subtractor Methods

The estimator-subtractor methods perform time-domain notch filtering. The estimator-subtractor implementation is shown in Fig. 3.4. These methods essentially form a replica of the NBI which is then subtracted from the received signal to enhance the wide-band components. Examples of such systems are described in [69, 83, 84, 94, 103, 104, 113, 114, 115, 167, 168, 180, 203]. Approaches to form a replica tend to exploit the predictability of the NBI and the spread spectrum signal. Since the spread data has a nearly flat spectrum, accurate prediction from past values are not possible. However, the narrow-band interfering signal can be predicted on the time scale of the PN signal. Hence, predictions of the received signal based on the previous values are in effect estimates of the interfering signal, provided the prediction horizon is beyond the correlation time of the PN sequence. By subtracting predicted values of the received signal, obtained this

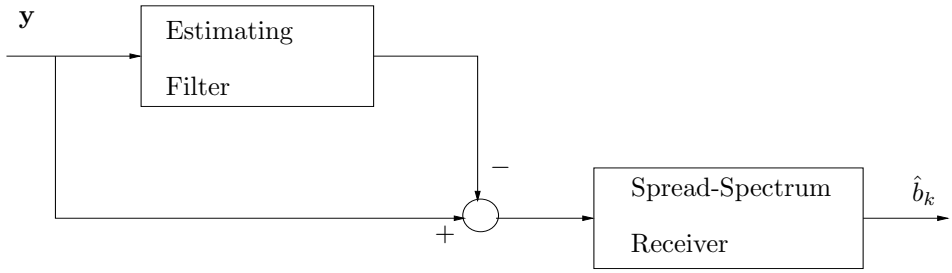


Figure 3.4: *Estimator-Subtractor model for interference cancellation. These methods form a replica of the NBI and subtract it from the received signal.*

way from the actual received signal and using the resultant prediction error as the input to the PN correlator, the effect of the interfering signal can be reduced significantly. This procedure is generally a *whitening* operation on the received signal.

Alternative methods for implementing the estimator-subtractor scheme involve

obtaining an estimate of the NBI by interpolation [103, 104, 114]. These methods provide a better processing gain and have desirable phase profiles. The disadvantage is the delay of order of a few correlation times of the PN sequence, which is generally less than the duration of a symbol, however.

These estimator-subtractor approaches form a notch filter. These formalisms employ linear transversal prediction or interpolation filters to create a replica of the interference. They form a linear prediction of the received signal based on a fixed number of previous samples, or a linear interpolation based on a fixed number of previous and future samples. This estimate is then subtracted from the received signal — which is timed approximately — to obtain the error signal. This error signal can be used as an input to the PN correlator. The filter taps are updated using an adaptive algorithm, such as the least mean-square algorithm. Such processing invariably introduces some distortion into the spread spectrum signal, which can be overcome if PN sequences of sufficient length are used during the spreading and despreading process. In general, this length must be greater than the length of the filter [113]. This requirement is usually met in practical systems.

Transform-domain methods

Another approach to NBI cancellation is to transform the domain in which interference cancellation is performed. These transform-domain approaches use the Fourier transform of the received signal, remove the NBI in the frequency domain and then transform the signal back to the time domain by the corresponding inverse transform for despreading [31, 94, 116, 120, 141] as shown in Fig. 3.5. Filtering out the NBI can be achieved by using an adaptive mask that removes those Fourier components with energy levels above a certain threshold [119]. Alternatively, a whitening mask can be used by first applying a non-parametric spectrum estimator to the received signal, from which the mask can be derived [94]. This very much depends on the overall system bandwidth and the processing speed requirements. Instead of transforming the signal to the frequency domain, it is possible to obtain a suitable time-frequency representation of the signal. Some distributions, such as the Wigner-Ville distribution, can be used to represent the signal in the frequency domain. The interference can then be suppressed by using a suitable mask. The paper [37] uses *short-term Fourier transform* (STFT) [118, 175]. The paper [96] uses the Wigner-Ville distribution to obtain time-frequency representations of the received signal. Filter-banks can be used to reduce the signal distortion and can be implemented in situations

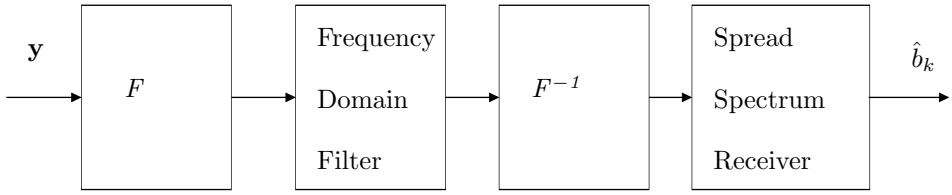


Figure 3.5: *Transform-domain methods. Interference cancellation is performed by transforming the signal to the frequency domain and filtering out the NBI via a notch filter.*

that require low power consumption [86]. Wavelet transforms [175] are another approach to obtain time-frequency representations of the received signal. Papers [55, 213] consider estimating the instantaneous frequency of the NBI to construct a signal subspace to enable the removal of the jammer.

Adaptive antennae

The third approach is to use spatial filtering techniques to eliminate the interference. These techniques unlike the estimator-subtractor and the transform-domain techniques can handle both narrow and wide-band interference. Antenna arrays eliminate the interference by either null steering or beamforming [139, 171]. The antennae in the array are weighted so that any particular signal can be nulled, thus null steering constantly computes the weights in order to maximize the received signal energy level. This method actually steers the antenna away from the interference source. Beamforming is another approach that tries to adjust the antenna in order to maximize the SNR. The antenna array steers in the direction of the desired signal. However, it is possible to end up in situations where the interference is in the same location as the signal source. This implies that the interference is coherent. In such cases these methods fail to eliminate the interference [155], and additionally they require an estimate of the *direction-of-arrival* (DOA) of the desired signal.

Table 3.1 is a summary of the various interference cancellation techniques discussed above. They are classified according to cost of processing, size, power

requirements, flexibility to adapt to changing environments, resolution, and complexity. Techniques based on the estimator-subtractor methods (adaptive filtering) and transform-domain methods generally have low cost and low complexity. They are adaptable to environment changes, and can offer varying degrees on resolution. Adaptive antennae techniques have generally high cost, are of large size and usually have high power requirements, and are of high complexity.

Type	Cost	Size	Power	Flexibility	Complexity
Estimator-Subtractor	low	small	low		
Transform-Domain					
STFT	low	small	low	*env*	low
Filter Banks	low	small	low	*env*	low
Wavelet Transform	low	small	low	*env, res*	low
Subspace Processing	low	small	low		
Adaptive Antennae					
Null steering	high	large	high		high
Beam forming	high	large	high		high

Table 3.1: Summary of various interference cancellation schemes; *env* refers to a changing environment and *res* indicates multiresolution characteristics.

Other Approaches

There are several other approaches to canceling NBI and WBI. Iterative methods (similar to the iterative SIC/PIC) to combat them have been proposed in [193]. Approaches based on using bilinear signal representations are examined in [212] (considered for GPS spread spectrum signals). Blind techniques for cancellation of narrow-band-interference based on second-order statistics have also been proposed in [11, 12]. Comparison of BSS based interference rejection techniques and the SIC and PIC techniques is made in [70]. In [71] a group-wise SIC based on BSS is presented. Paper [98] provides a general summary of both narrow-band and wide-band interference rejection techniques.

Chapter 4

ICA Assisted Interference Cancellation

Blind adaptive receivers generally do not require knowledge of the waveforms and the channel characteristics of the interfering sources. These detectors operate with minimal information about both the desired signal and the interfering source. A classification of blind schemes was presented in Sec. 3.5. Blind signal separation is an area of signal processing where mixtures of several sources are separated without the knowledge of the mixing process. A familiar example of this is the so-called "cocktail party problem" [75], where several different, simultaneous conversations are separated by a listener. A popular framework for solving such problems is Independent Component Analysis (ICA). Several schemes exist for interference suppression that are based upon a similar principle. In this chapter, receivers that combine the above principles with conventional detectors are presented. These techniques fall under category (C1) of the classification in Sec. 3.5, and are hence *semi-blind*.

4.1 A brief overview of ICA

This section presents the generic mixture model used in blind source separation and independent component analysis. Note that the notations in this section are different from the remainder of the thesis and this is due to the fact the these notations are widely used in the BSS/ICA community. A connection between the generic model described here and the variables in Chapter 2 is established in Sec. 4.2.

4.1.1 Introduction to ICA

Consider the following linear model ,

$$\mathbf{X}_{orig} = \mathbf{AS} + \boldsymbol{\nu}, \tag{4.1}$$

where

$$\mathbf{X}_{orig} = \begin{bmatrix} \mathbf{x}_{orig,1} \\ \mathbf{x}_{orig,2} \\ \vdots \\ \mathbf{x}_{orig,K} \end{bmatrix}, \quad \mathbf{S} = \begin{bmatrix} \mathbf{s}_1 \\ \mathbf{s}_2 \\ \vdots \\ \mathbf{s}_N \end{bmatrix}, \quad \boldsymbol{\nu} = \begin{bmatrix} \boldsymbol{\nu}_1 \\ \boldsymbol{\nu}_2 \\ \vdots \\ \boldsymbol{\nu}_K \end{bmatrix}.$$

This model consists of N sources of T samples, *i.e.*, $\mathbf{s}_i = [s_i(1) \dots s_i(t) \dots s_i(T)]$. The symbol t represents time, but could represent some other variable, *e.g.*, space. The observations $\mathbf{X}_{orig}$ consists of K mixtures of the sources, where, $\mathbf{x}_{orig,i} = [x_i(1) \dots x_i(t) \dots x_i(T)]$. Usually it is assumed that there are at least as many observations as sources *i.e.*, $K \geq N$. The sources and the observations are related by a $K \times N$ matrix $\mathbf{A} = [\mathbf{a}_1 \ \mathbf{a}_2 \dots \ \mathbf{a}_N]$ consisting of the vectors $\mathbf{a}_i = [a_{1i} \ a_{2i} \dots \ a_{Ki}]^T$. This linear mapping is called the *mixing matrix*. The model assumes some noise $\boldsymbol{\nu}$ considered to be Gaussian. Fig. 4.1 shows the schematics of the above problem. Fig. 4.2 shows two original signals and their mixtures and the scatter plot of these mixtures is shown in Fig. 4.3.

Recovering the unknown parts $\mathbf{A}$ and $\mathbf{S}$ of (4.1) is a problem under the domain of *linear source separation*. Solution to the linear source separation problem is not possible, if there is no information on some of the variables $\mathbf{A}$ or $\mathbf{S}$, in addition to the observed (*known*) data $\mathbf{X}_{orig}$. If the mixing $\mathbf{A}$ is known and the noise

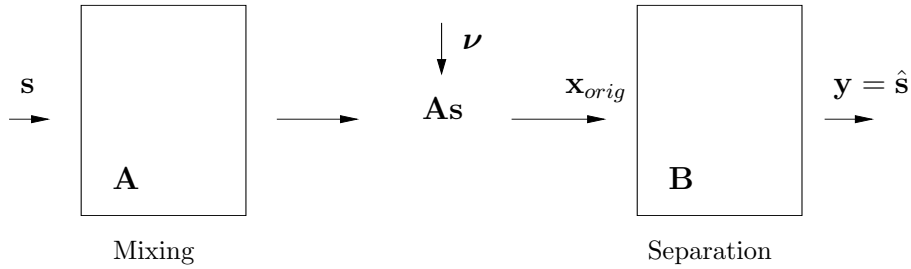


Figure 4.1: *Schematic illustration of mixing and separation.* $\mathbf{A}$ *is called the mixing matrix and its estimated inverse* $\mathbf{B}$ *is the unmixing matrix.*

is negligible, the sources can be estimated by finding the (*unmixing*) matrix $\mathbf{B}$, the (pseudo) inverse of the mixing matrix $\mathbf{A}$, for which $\mathbf{BX}_{orig} = \mathbf{BAS} = \mathbf{S}$ (Fig. 4.1). If there are as many observations as sources, then $\mathbf{A}$ is square and has full-rank, hence $\mathbf{B} = \mathbf{A}^{-1}$. The full-rank assumption is the necessary and sufficient condition for the existence of the pseudo-inverse of $\mathbf{A}$. When there are more observations than sources, there exist several matrices $\mathbf{B}$ that satisfy the condition $\mathbf{BA} = \mathbf{I}$. In this case, the choice of $\mathbf{B}$ depends on the components of $\mathbf{S}$ that we are interested in. For cases where there are fewer observations than sources, a solution does not exist unless further assumptions are made. Now the rank of $\mathbf{A}$ is less than the number of sources. There are some redundancies in the mixing matrix, and hence further information is required.

On the other hand, if no non-trivial prior information about the mixing matrix $\mathbf{A}$ is known or assumed, this problem of estimating the matrices $\mathbf{A}$ and $\mathbf{S}$ is referred to as *blind source separation* (BSS). The model defined in (4.1), with negligible noise $\boldsymbol{\nu}$, is then separable under the following fundamental restrictions [27]:

(R1) the components of $\mathbf{S}$ are statistically independent;
(R2) at most one component of $\mathbf{S}$ is Gaussian distributed;
(R3) the mixing matrix $\mathbf{A}$ is of full rank.

One very popular technique for solving the BSS problem is *independent component analysis* (ICA) [20, 27, 75, 77, 90]. In BSS, the main focus is to determine the underlying independent sources. The best known applications of ICA are in

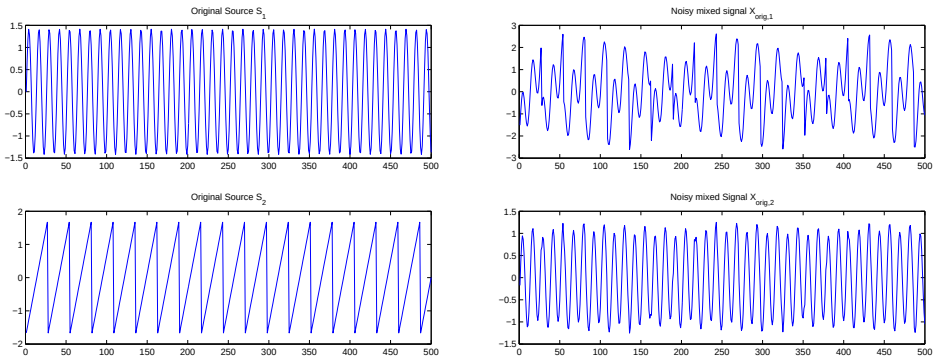


Figure 4.2: *Effect of mixing. The original sources S_1 and S_2 are shown in the left plot, and the mixed signals $X_{orig,1}$ and $X_{orig,2}$ are shown in the right plot. The mixtures also contain additive white Gaussian noise. The signal-to-noise ratio is about* $10dB$.

the field of signal and image processing *e.g.*, biomedical engineering [195, 209], speech processing [184], multispectral image processing [46] *etc.*

Hence, ICA can be defined as the computation method for separating a multivariate signal into its subcomponents assuming that all of these subcomponents are mutually independent. Alternatively, ICA can also be defined as a linear transformation on a multivariate signal $\mathbf{X}_{orig}$: $\mathbf{S} = \mathbf{B}\mathbf{X}_{orig}$, so that the components are as independent as possible, in the sense of maximizing some function $F(\mathbf{s}_1, \ldots, \mathbf{s}_N)$, that measures independence.

This section first reviews the concepts of independence and various measures to quantify independence. After this a review of several methods to perform ICA are presented in subsection 4.1.7. Then some preprocessing techniques usually used in ICA are reviewed briefly in subsection 4.1.6. Subsection 4.1.8 ends this section with a brief discussion of the ambiguities of BSS.

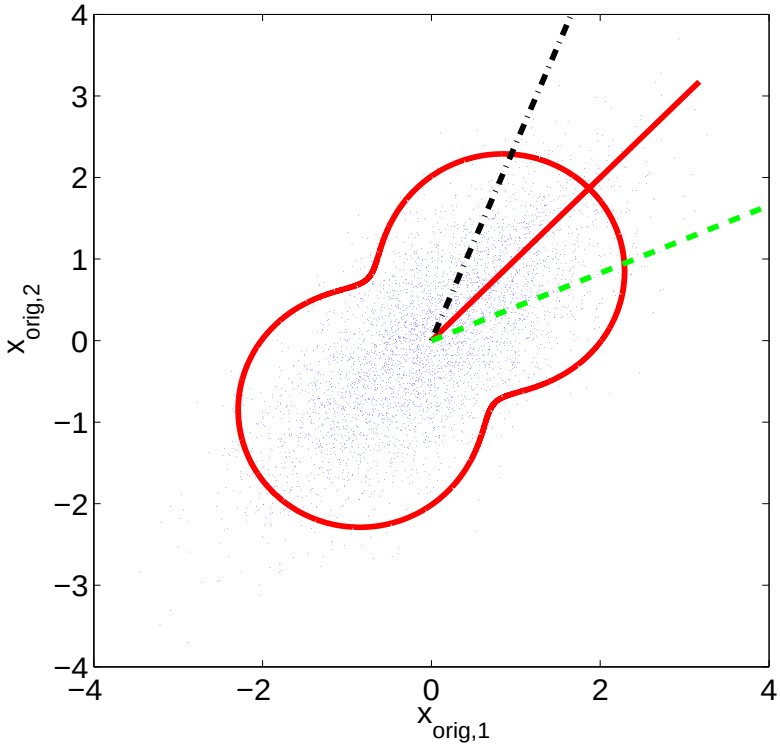


Figure 4.3: *The scatter plot of the mixtures shown in Fig. 4.2. The curves denote the standard deviation of the projection of the data in different directions. The dashed lines depict the directions of the columns of mixing matrix and the solid line is the largest eigenvector.*

4.1.2 Statistical independence

Two random variables s_1 and s_2 are said to be independent if their joint probability density is a product of their respective marginal densities [136]. Intuitively, this appears to be correct: having observed one random variable, all the statistics of the other, independent random variables remain unchanged. Formally, a random vector $\mathbf{s} = [s_1, \cdots, s_N]^T$, with a multivariate density $p(\mathbf{s})$ has statistically

independent components if the density can be factorized as

$$p(\mathbf{s}) = \prod_{i=1}^{N} p_i(s_i). \tag{4.2}$$

It is assumed that a density function for each random variable exists. This definition (4.2) leads to the notion of independence based on conditional densities. Hence, for two random variables s_1 and s_2,

$$p(s_1|s_2) = p(s_1, s_2)/p(s_2) = p(s_1)p(s_2)/p(s_2) = p(s_1). \tag{4.3}$$

In other words, the density of s_1 is unaffected by observing s_2, when the two variables are independent.

4.1.3 Measuring independence

Since the definition of independence involves density functions of random variables, measuring independence is a non-trivial task. It requires estimating density functions of the variables. A measure of independence can be obtained from the Kullback-Leibler divergence [97],

$$K(p\|q) = \int p(\mathbf{s}) \log \frac{p(\mathbf{s})}{q(\mathbf{s})} d\mathbf{s}, \tag{4.4}$$

where p and q are n-dimensional probability density functions. This divergence is always non-negative, and is zero if and only if the two distributions are equal [28]. This divergence — sometimes called the Kullback-Leibler distance — is not a proper distance measure, however, as it is not symmetric.

If the random variables s_i are independent, their joint probability density can be factorized, according to (4.2). Thus one might measure the independence of s_i as the Kullback-Leibler divergence between the real density $p = p(\mathbf{s})$ and the factorized density $q = p_1(s_1)p_2(s_2)\cdots p_n(s_n)$, where $p_i(.)$ are the marginal densities of s_i. This form of the Kullback-Leibler divergence is called the *mutual information* and is defined as [28],

$$I(s_1, s_2, \cdots, s_n) = K(p(\mathbf{s})\|p_1(s_1)\cdots p_n(s_n)). \tag{4.5}$$

The distance in (4.5) is zero, if and only if the variables are independent, and is non-negative otherwise. Therefore mutual information can be used as a measure of independence.

The conventional way of defining mutual information though is by differential entropy [28, 107]:

$$I(s_1, \cdots, s_n) = \sum H(s_i) - H(\mathbf{s}), \tag{4.6}$$

where $H(s_i) = -\int_{-\infty}^{\infty} p(s_i) \log p(s_i) ds_i$ is the differential entropy of the random variable s_i. [1]. Mutual information of this form was already proposed in [172], where it was called the rate of information transmission.

The idea of minimal entropy can lead to another important measure of independence. A fundamental property of a Gaussian distribution is that it has the largest entropy among all distributions with a given covariance matrix. This is called the *maximal property* of the Gaussian distribution. This measure shows that entropy can be used to measure the non-Gaussianity of a given random variable. Such a measure can be zero for a Gaussian variable, and non-negative otherwise. This measure is termed as *negentropy* and is defined as,

$$N(\mathbf{s}) = H(\mathbf{s}_{gauss}) - H(\mathbf{s}), \tag{4.7}$$

where $\mathbf{s}_{gauss}$ is a Gaussian random variable having the same covariance matrix Σ as $\mathbf{s}$ [75].

Unfortunately, all the three measures discussed above require the estimation of the density function, and this has severely restricted the use of these measures. Some authors have used approximation of mutual information based on polynomial density expansions [5, 27], which leads to the use of higher-order cumulants.

4.1.4 Cumulants and independence

Cumulants [93] provide a practical way to describe distributions using simple scalar functions. Consider a zero mean random variable s. The characteristic function of s is defined as $\hat{h}(t) = E\{\exp(its)\}$. Expanding the logarithm of the characteristic function as a Taylor series gives,

$$\log \hat{h}(t) = \kappa_1(it) + \frac{\kappa_2(it)^2}{2} + ... + \frac{\kappa_r(it)^r}{r!} + ..., \tag{4.8}$$

where the κ_r are some constants. These constants are called the cumulants (of the distribution) of s. In particular, the first few (four) cumulants have simple

[1] (4.6) can also be derived starting from (4.5); for more information refer to [28], page 19.

expressions,

$$\kappa_1 = E\{s\} \tag{4.9}$$
$$\kappa_2 = E\{s^2\} \tag{4.10}$$
$$\kappa_3 = E\{s^3\}. \tag{4.11}$$

(4.9) is the *mean*, zero in this case, (4.10) is the *variance*, (4.11) is the *skewness* — that defines the how symmetric the distribution is .

Of special interest is the fourth order cumulant, called *kurtosis*, expressed as [93, 135],

$$\text{kurt}(s) = E\{s^4\} - 3(E\{s^2\})^2. \tag{4.12}$$

Kurtosis can be considered a measure of non-Gaussianity of s. Moreover, negentropy can be approximated using skewness and kurtosis as [27, 75],

$$N(s) \approx \frac{1}{12}E\{s^3\}^2 + \frac{1}{48}\text{kurt}(s)^2, \tag{4.13}$$

assuming s is zero mean and has unit variance. Many distributions are sparse and non-symmetric and hence using either of these measures is roughly equivalent to using negentropy as a measure of non-Gaussianity. Kurtosis is a natural choice here, but care has to be taken as it is a non-robust measure. It is extremely sensitive to outliers (distributions with long tails).

For a Gaussian random variable, kurtosis is zero; it is typically positive for distributions with heavy tails and a peak at zero, and negative for flatter densities with lighter tails. Distributions of positive (negative) kurtosis are thus called super-Gaussian (sub-Gaussian). Thus, non-Gaussianity can be measured either by the absolute value of kurtosis or by the square of kurtosis.

4.1.5 Central limit theorem

The central limit theorem states that a sum of n independent, identically distributed non-Gaussian random variables tends asymptotically to the normal distribution, as n tends to infinity [117, 135]. In other words, this sum is more Gaussian than the individual random variables. Therefore, by inverse intuition, it can be said that the less Gaussian the linear combination of mixtures, the more independent the result of this transformation. Hence, the linear combination $y = \sum_i w_i x_i$ of the observed variables x_i will be maximally non-Gaussian if it equals to one of

the independent components s_j [75]. Thus, the task now is to find a transformation $\mathbf{w}_j$ such that the distribution of $y_j = \mathbf{w}_j^T \mathbf{x}$ is as non-Gaussian as possible. Hence, it is plausible to look at maximization of a measure of non-Gaussianity as an objective for designing an ICA algorithm.

4.1.6 Data pre-processing for ICA

The most common preprocessing technique for multidimensional data is to reduce the dimensionality of the data, before performing ICA. There are obvious benefits in dimensionality reduction. Performing ICA on high dimensional data may lead to poor results due to the fact that such data usually contain very few latent components. The data might be corrupted by noise, and several original dimensions might contain only noise. Thus, finding the subspace where the data exist, termed as the principal subspace, helps in reducing noise. Additionally, when the number of parameters is too large when compared with the number of available data points, the estimation of the parameters becomes difficult, and may lead to over-learning. In ICA, this leads to estimates that have a single spike or bump and are practically zero everywhere else [79].

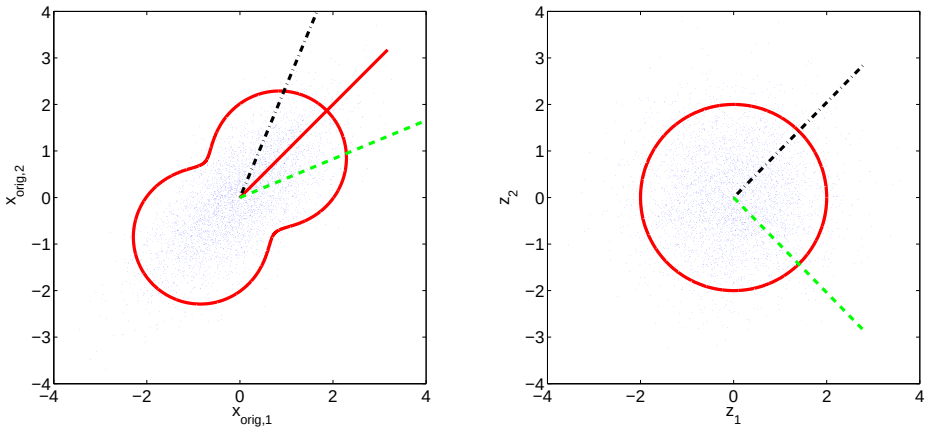
In addition to dimensionality reduction, the observed signals are centered and decorrelated. Decorrelation removes the second-order dependencies between the observed signals. A popular technique to achieve this is the *principal component analysis* (PCA) [67, 75]. Other names for PCA include Hotelling transformation [67] or Karhunen-Loéve transformation [105].

PCA is a transform where the observed vectors $\mathbf{X}_{orig}$ are first centered by removing their mean. This in practice is the average value of the sample. The vector is then linearly transformed into another vector $\mathbf{X}$ with m elements (possibly $m < n$), so that the elements are uncorrelated with each other. This linear transformation is found by computing the eigenvalue decomposition of the correlation matrix $\mathbf{C} = E\{\mathbf{X}_{orig}\mathbf{X}_{orig}^T\}$. This can also be computed using the empirical estimates as $\mathbf{X}_{orig}\mathbf{X}_{orig}^T/T$. Their eigenvectors form a new coordinate system into which the data are projected.

Combined with normalization of variances to unity, this process is called *whitening* or *sphering* and is performed by PCA as,

$$\mathbf{X} = \mathbf{\Lambda}^{-1/2}\mathbf{E}^T\mathbf{X}_{orig} = \mathbf{V}\mathbf{X}_{orig}, \tag{4.14}$$

where $\mathbf{X}$ is the whitened data matrix, $\mathbf{\Lambda}$ is the diagonal matrix corresponding to the eigenvalues of the correlation matrix, and $\mathbf{E}$ contains the corresponding eigenvectors of the correlation matrix as its columns. $\mathbf{V}$ is called the whitening matrix. Now, ICA estimation is performed on the whitened data $\mathbf{X}$ instead of the original data $\mathbf{X}_{orig}$. For whitened data it is sufficient to find an orthogonal separation matrix if the independent components are assumed white. The effect of whitening is illustrated in Fig. 4.4.



Scatter plot of the mixtures $\mathbf{X}_{orig}$ *Whitened data* $\mathbf{Z}$

Figure 4.4: *Effect of whitening. The curved boundaries represent the standard deviation of the projection of the data in different directions. The figure on the right shows the effect of whitening on the standard deviation. The dashed lines depict the mixing matrix while the solid line denotes the direction of the largest eigenvector. Adapted from [166].*

The whitening matrix is not unique; another choice would be $\mathbf{V} = \mathbf{E}\mathbf{\Lambda}^{-1/2}\mathbf{E}^T = \mathbf{C}^{-1/2}$. This will be used in the performance analysis in Section 4.10.

Dimensionality reduction is performed by PCA simply by projecting the K dimensional data to a lower dimensional space spanned by the m ($m < K$) dominant eigenvectors (*i.e.*, the eigenvectors corresponding to the largest eigenvalues) of the correlation matrix. The eigenvector matrix $\mathbf{E}$ and the diagonal matrix of eigenvalues $\mathbf{\Lambda}$ are of dimension $K \times m$ and $m \times m$ respectively. Identifying the lower dimensional subspace correctly is a non-trivial task in practice. For noiseless data, this involves finding a subspace corresponding to the non-zero eigenvalues. In reality, the data are corrupted by the noise, and they are not con-

tained exactly within the subspace. Now, the eigenvectors corresponding to the largest eigenvalues should describe the data well; however, in general "weak" independent components may be lost in the dimension reduction process. Trial and error are often needed to determine the number of eigenvectors, and subsequently the number of independent components.

PCA is by no means the only method to perform dimensionality reduction. Other methods include Local PCA [36, 45] and random projection [16]; however, PCA is the most popular. This is mainly due to the fact that PCA finds the orthogonal directions in which the data have maximal variance, and PCA is also an optimal method of dimensionality reduction in the mean-square error sense.

The separation matrix $\mathbf{B}$ in Fig. 4.1 can now be viewed as a two step process *i.e.*, whitening and rotation as,

$$\mathbf{B} = \mathbf{W}^T\mathbf{V}. \tag{4.15}$$

The whitening matrix $\mathbf{V} = \mathbf{\Lambda}^{-1/2}\mathbf{E}^T$ can be found by PCA, while the rotation matrix $\mathbf{W}$ is estimated by one of the ICA methods described next in Sec. 4.1.7. This process is shown in Fig. 4.5.

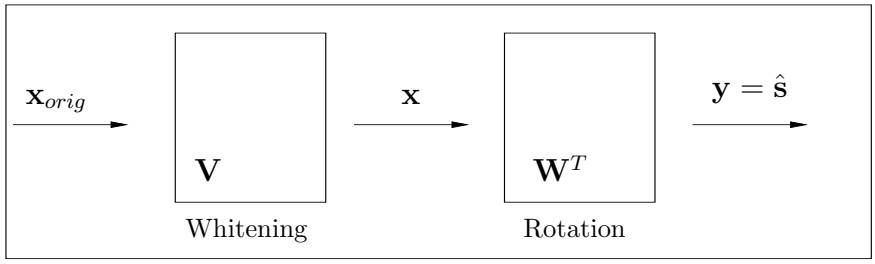


Figure 4.5: *Schematic of separation: whitening and rotation. The unmixing matrix* $\mathbf{B}$ *in 4.1 can be regarded as a concatenation of the whitening matrix* $\mathbf{V}$ *and the (orthogonal) rotation matrix* $\mathbf{W}$

4.1.7 Several routes to independence

From the above discussion non-Gaussianity can be a viable property to perform ICA. FastICA [73, 75, 76] is one of the popular algorithms using non-Gaussianity to perform separation. While kurtosis is a measure for non-Gaussianity, it is possible to perform ICA via *higher-order statistics* (HOS) methods [126] and higher-order cumulant tensors. The JADE method [21] uses the principle of diagonalization of cumulant tensors to separate the sources. A brief discussion of other algorithms that are based on implicit *probability density function* (pdf) estimation is also made in the following subsection. Finally, this definition of ICA as in (4.1) assumes that the mixtures are random variables; however there are cases when mixtures of signals containing more structure than simple random variables — e.g., *time signals* — are observed. TDSEP [214] is an algorithm that separates such sources. This section reviews this algorithm, too.

ICA by maximization of non-Gaussianity

FastICA[2] [73, 75, 76] is an iterative fixed point algorithm, derived from a general objective or a *contrast* function. Assume $\mathbf{x}$ is the whitened data vector (see Section 4.1.6) and $\mathbf{w}^T$ is one of the rows of the rotation/separating matrix $\mathbf{W}$. Consider the function,

$$J_G(\mathbf{w}) = E\{G(\mathbf{w}^T\mathbf{x})\}. \tag{4.16}$$

The function G is chosen to be a smooth scalar valued function that measures the deviation of the source estimate $\mathbf{y} = \mathbf{W}^T\mathbf{x}$ from the Gaussian distribution. Different G leads to different approaches. Its derivative g, known as the *activation* function, typically appears in the algorithm as a non-linear function.

Maximizing the contrast function (4.16) using an appropriate Newton approximation leads to the following update for $\mathbf{w}$ [75]:

$$\mathbf{w} \leftarrow E\{\mathbf{x}g(\mathbf{w}^T\mathbf{x})\} - E\{g'(\mathbf{w}^T\mathbf{x})\}\mathbf{w}, \tag{4.17}$$

where $\mathbf{w}^T$ is one of the rows of $\mathbf{W}$, the orthogonal separating matrix. In an on-line stochastic algorithm the expectations can be omitted, while in practice they are replaced by their empirical estimates. The nonlinear function g is chosen so that it is the derivative of the non-quadratic contrast function G that either measures non-Gaussianity via the kurtosis measure or (an approximation of the)

[2] `http://www.cis.hut.fi/projects/ica/fastica`

negentropy measure. Choosing $G(y) = y^4$, one obtains a kurtosis based algorithm as described in [76]. More robust choices of G are non-polynomial functions such as $\log\cosh(y)$ or $\exp(-y^2)$ [73, 75].

Before evaluating the algorithm (4.17), the data is pre-processed as described in Sec. 4.1.6. The unit norm vector $\mathbf{w}$ is chosen randomly at first. $\mathbf{w}$ is normalized again to have unit norm after each iteration step of (4.17). The iteration is continued until the direction of $\mathbf{w}$ converges.

The above estimation algorithm is for real valued data. In the case of complex signals (like the QPSK modulated signals), the complex version of the FastICA [15] can be used. Here the $\mathbf{w}$, the row of the unmixing matrix $\mathbf{W}$ is estimated as,

$$\mathbf{w} \leftarrow E\{\mathbf{x}[\mathbf{w}^H\mathbf{x}]^* g(|\mathbf{w}^H\mathbf{x}|^2)\} - E\{g(|\mathbf{w}^H\mathbf{x}|^2) + |\mathbf{w}^H\mathbf{x}|^2 g'(|\mathbf{w}^H\mathbf{x}|^2)\}\mathbf{w}, \quad (4.18)$$

where the asterisk denotes complex conjugation and $\mathbf{w}^H$ is the vector $\mathbf{w}$ transposed and complex conjugated. Choosing a kurtosis based function, the above algorithm for complex cases can be written as,

$$\mathbf{w} \leftarrow E\{\mathbf{x}[\mathbf{w}^H\mathbf{x}]^*|\mathbf{w}^H\mathbf{x}|^2\} - 2\mathbf{w}. \quad (4.19)$$

The independent components (ICs) $\mathbf{s}_j$ can be estimated one by one — *deflationary* approach — or can be estimated simultaneously — *symmetric* approach. In the former case, one must ensure that the rows $\mathbf{w}_j$ of the separating matrix $\mathbf{W}$ are orthogonal. This can be done after every iteration step (4.17) or (4.19) by subtracting from the current estimate $\mathbf{w}_p$ the projections of all previously estimated $p-1$ vectors before normalization [75]:

$$\mathbf{w}_p \leftarrow \mathbf{w}_p - \sum_{j=1}^{p-1} (\mathbf{w}_p^H \mathbf{w}_j)\mathbf{w}_j. \quad (4.20)$$

In the latter case, the iteration step (4.17) or (4.19) is computed for all $\mathbf{w}_p$, and after that the matrix $\mathbf{W}$ is orthogonalized as

$$\mathbf{W} \leftarrow (\mathbf{W}\mathbf{W}^H)^{-1/2}\mathbf{W}. \quad (4.21)$$

Papers [73, 130] discuss the convergence properties of the FastICA algorithm. The asymptotic convergence of the algorithm is at least quadratic, and usually cubic when the ICA model (4.1) holds. This rate is much faster than that of

gradient-based optimization algorithms. With a kurtosis based contrast function, FastICA can be shown to converge globally to the independent components [76].

FastICA can also be derived from the maximum likelihood principle, which is more sound compared to the heuristic derivation used in this section. The exact details can be found in [75] or very briefly in Sec. A.2.1. Other ICA algorithms based on the maximum likelihood principle are the Bell-Sejnowski algorithm [8], and the natural gradient algorithm [4].

ICA by tensorial methods

Another approach for the estimation of independent components consists of using higher-order *cumulant tensors*. Tensors are generalizations of matrices, or linear operators. Cumulant tensors are then generalizations of the covariance matrix $\mathbf{X}_{orig}\mathbf{X}_{orig}^T/T$. The covariance matrix is the second order cumulant tensor, and the fourth order tensor is defined by fourth-order cumulants as $\mathrm{cum}(x_i, x_j, x_k, x_l)$. Sec. 4.1.6 used *eigenvalue decomposition* (EVD) to whiten the data. By whitening, the data is transformed so that its second-order correlations are zero. This principle can be generalized so that the off-diagonal elements of the fourth-order cumulant tensor can be minimized. This kind of (approximate) higher-order decorrelation results in a class of methods for ICA estimation.

Joint approximate diagonalization of eigenmatrices (JADE) [21] refers to the principle of computing several cumulant tensors $\mathbf{F}(\mathbf{M}_i)$, where $\mathbf{F}$ represents the cumulant tensor and $\mathbf{M}_i$ represents the eigenmatrices. These tensors are diagonalized jointly as well as possible. Consider a matrix $\mathbf{W}$, that diagonalizes $\mathbf{F}(\mathbf{M})$ for any $\mathbf{M}$. In other words, $\mathbf{WF}(\mathbf{M})\mathbf{W}^T$ is diagonal. This is due to the fact that the matrix $\mathbf{F}$ is a linear combination of the terms $\mathbf{w}_i\mathbf{w}_i^T$, assuming that the model in (4.1) holds.

A possible objective measure for the diagonality of $\mathbf{Q} = \mathbf{WF}(\mathbf{M}_i)\mathbf{W}^T$ can be the sum of squares of the off-diagonal elements: $\sum_{k\neq l} q_{kl}^2$. Equivalently, since $\mathbf{W}$ is orthogonal, and it does not change the total sum of squares of a matrix, minimization of the sum of squares of the off-diagonal elements is equivalent to the maximization of the squares of the diagonal elements. Thus, the following objective function can act as a good measure of the joint diagonalization process:

$$J_{JADE}(\mathbf{W}) = \sum_i \|\mathrm{diag}(\mathbf{WF}(\mathbf{M}_i)\mathbf{W}^T)\|^2, \tag{4.22}$$

which is the sum of squares of all the diagonal elements of all the diagonalized cumulant tensors.

Choosing $\mathbf{M}_i$ as the eigenmatrices of the cumulant tensor is natural consequence of the fact that the n eigenmatrices span the same subspace as the cumulant tensor, and hence they contain all the relevant information on the cumulants. Moreover, with this choice, the objective function in (4.22) can be restated as [75]

$$J_{JADE}(\mathbf{W}) = \sum_{ijkl \neq iikl} \mathrm{cum}(y_i, y_j, y_k, y_l)^2, \tag{4.23}$$

where $\mathbf{y}$ is the estimate of the independent sources obtained by the unmixing matrix $\mathbf{W}$ as $\mathbf{y} = \mathbf{W}\mathbf{x}$. Eq. (4.23) means that minimizing J_{JADE} also minimizes the sum of squared cross-cumulants of y_i.

Since JADE uses explicit tensor EVD it is restricted to fairly small dimensions for computational reasons. Its statistical properties are inferior to methods using likelihood or non-polynomial cumulants [75]. However, with low dimensional data, JADE is a competitive alternative to FastICA.

A similar approach that uses the EVD is the fourth-order blind identification (FOBI) method [19] which is simpler, and deals with the EVD of the weighted correlation matrix. It is of reasonable complexity, and is probably the most efficient of all the ICA methods. However, it fails to separate the sources when they have identical kurtosis. Other approaches include maximization of squared cumulants [61], and fourth-order cumulant based methods as described in [127, 128].

ICA by other estimation techniques

As mentioned earlier, the above algorithms are by no means the only ones to perform ICA. Based on the approach one chooses, there exist several algorithms for separation of independent sources. The Jutten-Hérault algorithm [90] uses the fact that correlations between non-linear functions of independent variables are zero. Explicit fourth-order cumulant based methods for ICA are described in [128, 127]. The Bell-Sejnowski algorithm [8] is based on maximization of entropy between the outputs of a nonlinear network. Several gradient algorithms for maximum likelihood estimation exist. The natural gradient algorithm by Amari, Cichocki and Yang [5] is based on Amari's natural gradient principle. The

extended Bell-Sejnowski algorithm is based on maximum likelihood estimation with an adaptive non-linearity [48]. Source separation algorithms for arbitrary source PDFs have been considered in [207, 208, 210]. For a more detailed list of algorithms, refer to [26, 75, 126].

Exploiting time structure

The model in (4.1) assumes linear mixing of independent random variables. There, the index t is just a sampling index. In fact, the basic model does not change if the samples are re-ordered. However, in many applications, this is not the case. What are mixed are not random variables but time signals, or time series. The fact that time signals have much more structure than simple random variables, can be used to improve the estimation of the model. This information can also make the estimation of the model possible, even in cases where basic ICA fails — *e.g.*, if the independent components are Gaussian but correlated over time.

In order to account for the time structure, the model in (4.1) can be restated as,

$$\mathbf{x}_{orig}(t) = \mathbf{A}\mathbf{s}(t) + \boldsymbol{\nu}(t), \tag{4.24}$$

where $\mathbf{x}_{orig}(t)$ and $\mathbf{s}(t), t = 1, \ldots, T$ are column vectors of $\mathbf{X}_{orig}$ and $\mathbf{S}$, respectively and $\boldsymbol{\nu}(t)$ is the noise. It should be noted here that $\mathbf{x}_{orig}(t)$ denotes the tth column, while in (4.1) $\mathbf{x}_{orig,i}$ represents the ith row.

Now the independent components are time signals, $s_i(t), t = 1, \cdots, T$, with t as the time index, and $\mathbf{A}$ is the mixing matrix. The independent components are not restricted to be non-Gaussian. Here, the assumptions for separation are that the independent components have different autocovariances, or the variances are non-stationary [75].

The simplest form of time structure is given by linear autocovariances, given by: $\text{cov}(x_i(t)x_i(t-\tau))$, the covariances between the values of signal at different time points. τ is some constant time lag, $\tau = 1, 2, \cdots,$. If the data has time dependencies, these are non-zero.

In addition to the autocovariances of one signal, we also need the covariances between two signals: $\text{cov}(x_i(t)x_j(t-\tau))$ where $i \neq j$. All these statistics for a

given time lag can be grouped together in the time-delayed covariance matrix,

$$\mathbf{C}_{\tau}^{\mathbf{x}} = E\{\mathbf{x}_{orig}(t)\mathbf{x}_{orig}(t-\tau)^T\}. \tag{4.25}$$

The key idea here is to use this information instead of higher-order statistics [122]. Now, the matrix $\mathbf{B}$ must ensure that both the instantaneous and lagged covariances of $\mathbf{y}(t) = \mathbf{B}\mathbf{x}_{orig}(t)$ go to zero,

$$E\{y_i(t)y_j(t-\tau)^T\} = 0, \quad \forall i, j, \tau, \quad i \neq j. \tag{4.26}$$

The motivation for this is that for independent components $s_i(t)$, the delayed covariances are all zero due to independence. These delayed cross-covariances provide extra information to estimate the model. Note that (4.26) means that $E\{\mathbf{y}(t)\mathbf{y}(t-\tau)^T\}$ is diagonal for all T.

Hence, separation algorithms use several time delays, and simultaneously diagonalize all the corresponding delayed covariance matrices. Exact diagonalization is not possible, since the eigenvectors of different covariance matrices are unlikely to be identical, except in the theoretical case where the data are exactly generated by the model in (4.24). The objective function [214] is now to minimize the sum of the off-diagonal elements of several delayed covariances of $\mathbf{y} = \mathbf{W}\mathbf{x}_{orig}$, as

$$J_1 = \sum_{\tau \in S} \text{off}(\mathbf{W}\mathbf{C}_{\tau}^{\mathbf{x}}\mathbf{W}^T), \tag{4.27}$$

where $\mathbf{C}_{\tau}^{\mathbf{x}}$ is the delayed covariance matrix and S is the set of chosen time delays τ. Minimizing J_1 under the constraint that $\mathbf{W}$ is orthogonal gives the estimation method. The minimization can be performed using existing EVD methods to diagonalize simultaneously these matrices approximately. The *second-order blind identification* (SOBI) [10] and *temporal decorrelation source separation* (TDSEP) [214] algorithms are based on the above principle.

4.1.8 Ambiguities of BSS

The previous section described the principles of a technique for blind source separation - ICA. While separation is possible, examining the model that (4.1) represents, it is possible to see that the following ambiguities exist [75]:

1. *The energies of the independent components cannot be determined,*

   Since both $\mathbf{A}$ and $\mathbf{S}$ are unknown, the effect of multiplication of one of the

source estimates with a scalar constant α is canceled by dividing its corresponding column in the mixing matrix by α. This indeterminacy can be solved by ensuring that the random variables have unit variance *i.e.*, $E\{s_i^2\} = 1$. This still leaves the *ambiguity of sign*. While this is insignificant in certain applications, care has to be taken in applications where sign plays a crucial role.

2. *The order of the independent components cannot be determined.*

   The noiseless version of the model in (4.1) can be written as,

$$\mathbf{X}_{orig} = \sum_{i=1}^{N} \mathbf{a}_i \mathbf{s}_i = \mathbf{AS}. \tag{4.28}$$

   Now, both $\mathbf{A}$ and $\mathbf{S}$ being unknown, the order of the terms can be changed freely in (4.28) and any of the independent components can be called the first one. This implies that the correspondence between a physical signal and the estimated independent component is not one-to-one. This indeterminacy is particularly severe in many applications where *identification* of the estimated components is of very high importance. Formally, this means that the following relation between the mixing matrix $\mathbf{A}$ and the separation matrix $\mathbf{B}$ holds:

$$\mathbf{AB} = \mathbf{P}, \tag{4.29}$$

   where $\mathbf{P}$ is a permutation matrix.

4.2 ICA model and CDMA

The linear mixing model in (4.1) or (4.24) consists of a set of sources $\mathbf{S}$, mixed by a mixing matrix $\mathbf{A}$ to generate the observed (mixed) signals $\mathbf{X}_{orig}$. When the sources are independent, this corresponds to the ICA model, which can then be solved blindly by any BSS technique generally, and more specifically ICA.

In the models considering interference as defined in (2.34), and (2.38), the received data is mixed by the matrix $\mathbf{\Theta}$, the array steering matrix. This mixing due to the array is linear. Therefore, comparing the model in (4.24) to (2.34), (2.38) the following relations can be observed:

- $\mathbf{\Theta}$ consisting of the array steering matrix is the mixing matrix and corresponds to $\mathbf{A}$ in (4.24);
- $\mathbf{y}_i^{[sc]}$ containing the short-code data at each element of the antenna array corresponds to the sources $\mathbf{s}(t)$ in (4.24);
- $\mathbf{y}_i$ the received data and corresponds to $\mathbf{x}(t)$ in (4.24).

On the other hand, the model in (2.21) consists of a linear transformation of both the independent variables (the transmitted symbols) and their delayed version, and also resembles the generic ICA model. However, this mixing model is convolutional. Chapter 23 of [75] deals with convolutional separation when the channel is assumed to stay constant during a block of M symbols *i.e.*, the fading coefficients a_l depend only on the path l but not on the symbol index m. This convolutional mixing due to the channel is still present in the models defined in (2.34), and (2.38), but the interest is in the mixing produced by the antenna array. All the algorithms in this thesis assume the array model as defined in (2.34) or (2.38).

4.3 Simple interference cancellation scheme: plain ICA-RAKE

ICA can be used to separate all the sources of interference. Assuming that the interference is independent of the user-of-interest, ICA can act as a basic algorithm that separates the sources. Conventional detection can then be used to process the user-of-interest. The detection in this case can be depicted by the following diagram. The key issue in the use of simple ICA as in Fig. 4.6 is identifying the separated users. As seen in Section 4.1.8, it is not possible to identify the user directly due to the permutation ambiguity. Hence, it is required to use some other information to identify the user. Pilot sequences or training sequences (Section 3.5) can achieve this. A simple method to match and identify the user can be based on the distance measure defined as,

$$\text{dist}_H[b_k, \hat{b}_k^t] = b_k \circledcirc b_k^t = \sum_{m=1}^{N_t} |b_{km}^t - \hat{b}_{km}^t|, \tag{4.30}$$

where b_{km}^t is the training sequence of the kth source, and $\hat{b}_{km}^t$ is the extracted binary training sequence after separation by ICA. N_t is the length of the training

Figure 4.6: *Interference cancellation with ICA. The sources are separated by ICA. Identification of the sources (signals and the jammers) requires additional information. After identification, conventional detection sufficiently recovers the users' signals.*

sequence. This distance, termed as *Hamming distance* [28], gives a measure of the number of bits (points) at which b_{km}^t and $\hat{b}_{km}^t$ differ.

Now, the user-of-interest is that user possessing the minimum Hamming distance with respect to the training sequence and can be identified as,

$$I_k = \arg\max_l \{N_t - \text{dist}_H[b_k, \hat{b}_k^l]\}, \quad l = 1, \ldots, N_s, \tag{4.31}$$

where N_t is the length of the training sequence, and N_s is the number of sources estimated by the ICA algorithm, and I_k corresponds to the selected source index.

Finally, conventional detection is performed for the data of the selected source I_k. This leads to a simple *ICA-RAKE* formulation.

In **Algorithm 1**, the section of the code marked FastICA can be replaced with either JADE or TDSEP or any other ICA algorithm, as long as the data are pre-processed accordingly.

4.3.1 Choice of ICA Algorithm in plain ICA-RAKE

As illustrated in Section 4.1.7, there are several algorithms that maximize a particular criterion to achieve source separation. Hence, it is possible to choose any of the ICA algorithms for interference cancellation as illustrated in Fig. 4.7.

Algorithm 1 ICA-RAKE Receiver

$w \Leftarrow \texttt{rand}(N_a, L)$
$\hat{w}_{old} \Leftarrow \texttt{rand}(N_a, L)$
$\epsilon \Leftarrow 10^{-6}$
$\Delta = \texttt{norm}(w - w_{old}) + \epsilon$
while $\Delta \geq \epsilon$ **do**
  $w_{old} \Leftarrow w$ {# — FastICA Begins —— #}
  $z \Leftarrow y' * w$
  $w^+ \Leftarrow \frac{y}{M} * (\texttt{abs}(z)^2 * z) - 2w$
  $w = w^+ / \texttt{norm}(w^+)$
  $\Delta = \texttt{norm}(w - w_{old})$ {# — FastICA Ends — #}
end while
$\tilde{y} \Leftarrow w' * y$
for $i = 1$ to N_s **do**
  $\tilde{b}_i \Leftarrow \texttt{mrc}(\tilde{y}_i)$
  $d_i \Leftarrow N_t - \texttt{dist}_H[b_k^t, \tilde{b}_i^t]$
end for
$\tilde{k} \Leftarrow \texttt{argmax}(d)$
$\hat{b}_k \Leftarrow \texttt{sign}(\tilde{b}_{\tilde{k}})$

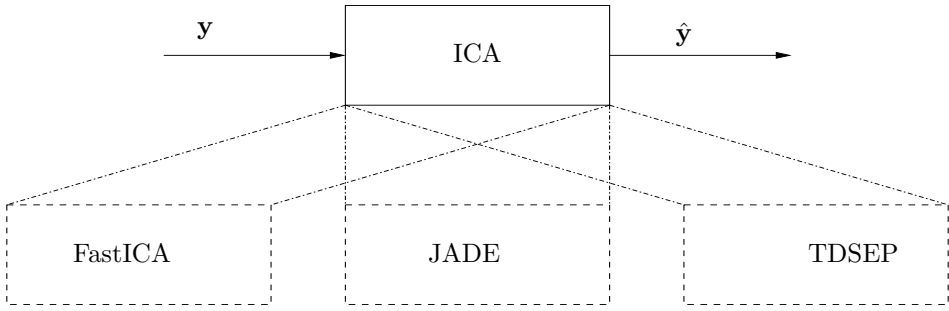


Figure 4.7: *Choice of various ICA algorithms. The choices include kurtosis or* tanh *based FastICA, the tensorial method JADE, and the temporal decorrelation source separation technique TDSEP.*

However, certain algorithms are better suited for certain tasks. It is therefore not realistic to choose an algorithm and perform cancellation when the algorithm

is ill-suited for the task. The most suitable algorithm can thus be determined by performing trial numerical simulations[3].

This thesis presents examples of interference cancellation with the aid of *scenarios*. In the topic of interference cancellation, the parameter space is large and hence careful selections are made. Each scenario consists of altering a set of important parameters *e.g.*, the number of users K, the block size M, the *signal-to-interference* (SIR) ratio, *signal-to-jammer* (SJR) ratio, *signal-to-noise* (SNR) ratio *etc.,* with a specific aim of evaluating the effect of that parameter on the algorithm(s) in question. These scenarios are specifically intended for situations that are not favorable for communications. These situations arise in hostile environments, especially in sensor networks when sensors are deployed in enemy territory. Heavy jamming here, prevents any form of communications. Moreover, in civilian applications, there are several instances where non-cooperative behavior among devices can be observed. The chosen scenarios reflect this. Scenarios 1,2,3,4,9 are typical situations in military applications, while Scenarios 5,6,7,9 are situations that concentrate on interferences due to non-cooperative behavior between devices operating in an adjacent cell. Scenarios 8, 10 are used to fine tune the above parameters.

Scenario 1 consists of evaluating the different algorithms against an external interferer.

All the results are evaluated by using the measures *bit-error-rate* (BER) and the coded *block-error-rate* (BLER). The BER calculates the proportion of erroneous bits in a block. With the coded block-error-rate, a block is assumed to reach the target BER in blocks if 95% of the bits within a block are correct[4]. By this definition it is possible to assess how the bit errors are differentially distributed over the blocks. The use of BLER helps in distinguishing methods that might have similar BER, but have erroneous bits distributed over a much smaller set of blocks.

Scenario 1: External interference cancellation. An external interference is

[3]The simulation set-up is described in Appendix B.

[4]Typically, a communication system has a RF front-end, IF stage, and then down conversion and signal processing implemented in a baseband application specific IC or field programmable gate array. Several stages of RF and IF processing help in boosting the received signal, before any baseband signal processing is performed. All communication systems also have some form of channel specific coding along with several stages of error correction. In order to model the effects RF processing, coding and error correction, the definition of 95% is chosen in this thesis. As stated in [151] the results are quantitatively similar for thresholds of 98% and 99%.

considered as a test case here. The interference is a bit pulsed jammer. The down-link system consists of $K = 8$ users, spread with Gold codes of length $C = 31$. The channel is an AWGN channel with $L = 1$ path. Heavy jamming is usually observed in military applications.

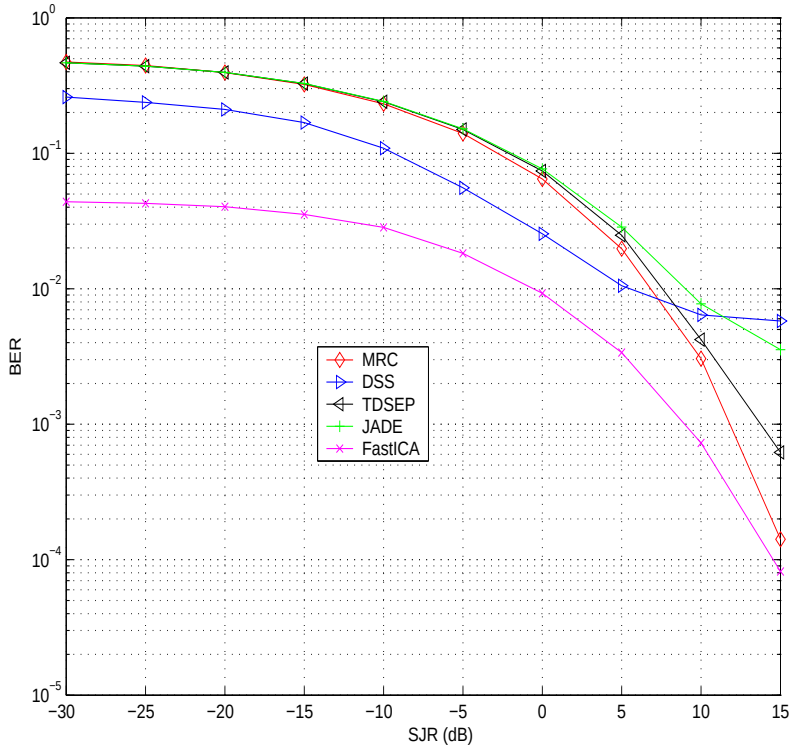


Figure 4.8: *Performance of the different ICA techniques. For the scenario considered in the simulation, FastICA based methods provide the best BER performance. JADE and TSDEP have a BER performance equivalent to that of MRC. These results have been previously published in [153].*

Fig. 4.8 illustrates the performance differences between various ICA algorithms. It is seen that (a variant of) FastICA provides the best results compared with JADE and TDSEP. JADE based on cumulants requires a larger number of samples ($M > 5000$) to provide reasonable results. TDSEP performs well when

the number of paths is high[5]. From the performance results above, FastICA is the preferred choice of ICA algorithm for the methods developed later in this chapter.

4.4 Examples of interference cancellation with ICA-RAKE receiver

The performance of the basic ICA-RAKE receiver (Algorithm 1 in Section 4.3) is evaluated by simulating a DS-CDMA down-link system. Several simulation parameters are varied which lead to the several scenarios outlined below. The interference to be canceled includes an external jammer. The model is then similar to the one outlined in Sec. 2.2.3. The performance of the ICA-RAKE receiver is compared with the standard MRC described in Sec. 3.1. Other obvious candidates are chip-rate equalization [66] and the MMSE type algorithms described in Chapter 3. Since, post-processing in the proposed algorithms are based on the popular MRC receiver, the algorithms are compared with these methods. BER and BLER are the performance measures used in the analysis.

Scenario 2: Suppression of a continuous jammer. The jammer can be either locked or unlocked to the carrier frequency of the source. In the former case, $f_j = f_c$, *i.e.*, the jammer is assumed to operate exactly at the carrier frequency. In the latter case, the frequency offset is set to $f_j - f_c = \frac{1}{200T_s}$, where T_s is the duration of a symbol. This now results in a total phase shift of 2π during a block of $M = 200$ symbols. The SNR is fixed to 10 dB, while the Signal-to-Jammer (SJR) ratio is varying. The system has $K = 8$ users, spread with codes of length $C = 31$.

Fig. 4.9 shows the results of suppression achieved with the ICA-RAKE receiver. The BER behavior of ICA-RAKE is quite equal in this case to MRC with $N_a = 2$ antenna elements up to -10 dB SJR after which it saturates. This saturation is expected due to the fixed level of thermal noise. However, Fig. 4.10 shows that ICA-RAKE is able to enhance the BLER behavior. Compared to MRC with $N_a = 3$ antenna elements, approximately a 3 dB gain is achieved when taking either 10^{-1} or 10^{-2} as the target BLER.

[5]This scenario shall be revisited later in the thesis, when dealing with coherent jammers in multipath cases.

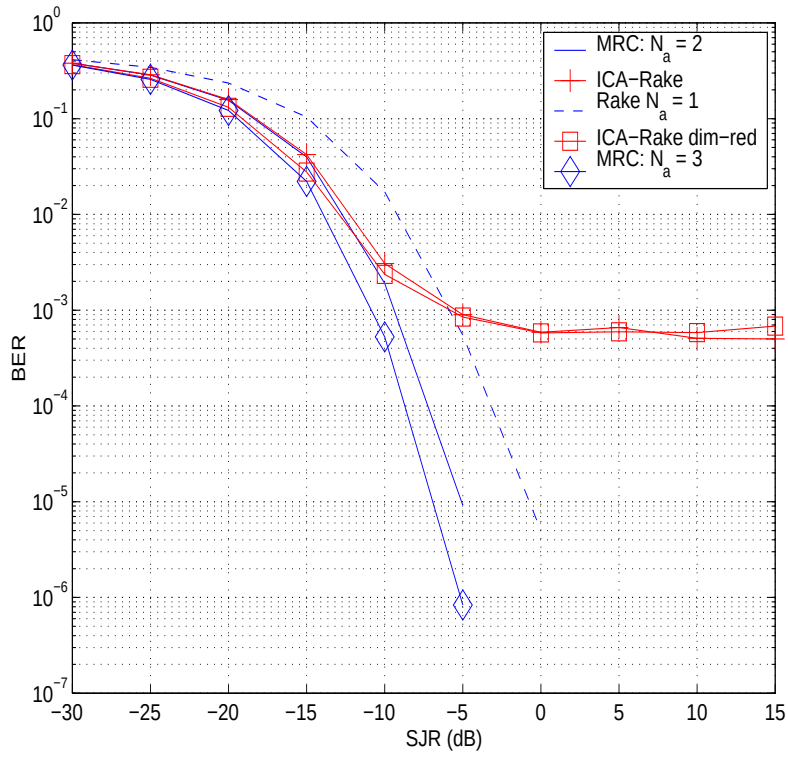


Figure 4.9: *Suppression of a continuous wave jammer locked exactly to the carrier frequency. The graph plots the obtained BER for different values of SJR. ICA-RAKE cancels the interference up to about −10 dB after which, it saturates due to the fixed level of thermal noise. These results have been previously published in [151].*

Clearly, it can be seen from the BER curves in Fig. 4.11 that ICA-RAKE provides an advantage of $0--5$ dB over the other systems, depending on whether 2 or 3 antenna elements are in use. For example, with $N_a = 3$ antenna elements, the performance is better by 5 dB when a target BER of 10^{-1} is chosen. In Fig. 4.12, the BLER performance is similar, its gains being over 2 dB at the BLER levels of 10^{-1} or 10^{-2}.

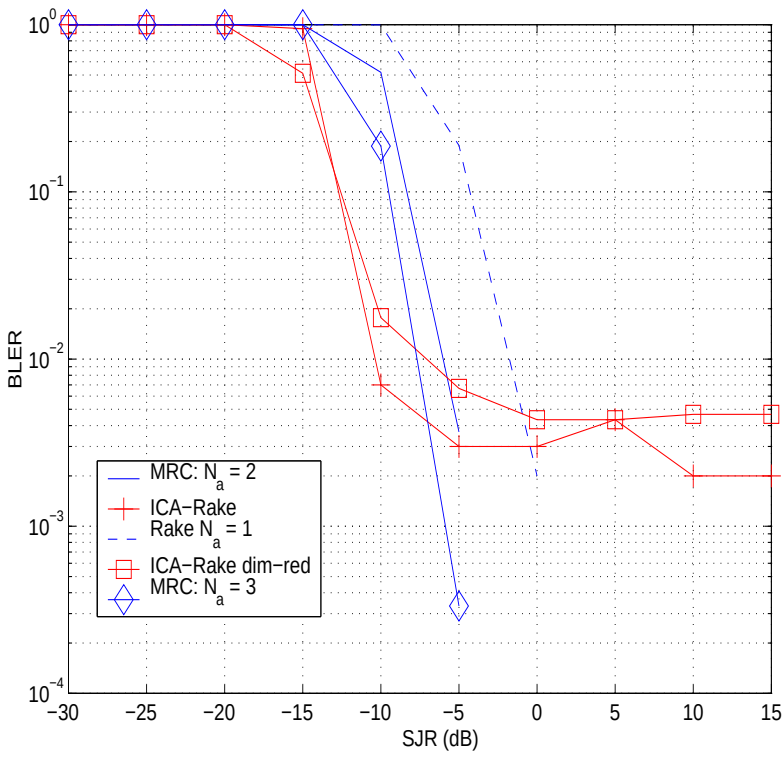


Figure 4.10: *Suppression of a continuous wave jammer locked exactly to the carrier frequency. The graph plots the obtained BLER for different values of SJR. ICA-RAKE cancels the interference up to about −10 dB after which, it saturates due to the fixed level of thermal noise. These results have been previously published in [151].*

Scenario 3: Suppression of a bit-pulsed jammer. The jammer is now bit-pulsed. This makes it uncorrelated with the sources at the chip level. Bit-pulsed jammers can either be locked ($f_j = f_c$), or unlocked ($f_j \neq f_c$). In the unlocked case, the frequency offset is set to $f_j - f_c = \frac{1}{200T_s}$, where T_s is the duration of a symbol. This now results in a total phase shift of 2π during a block of $M = 200$ symbols. The SNR is fixed to 10 dB, while the Signal-to-Jammer (SJR) ratio is varying. The system has $K = 8$ users, spread with codes of length $C = 31$.

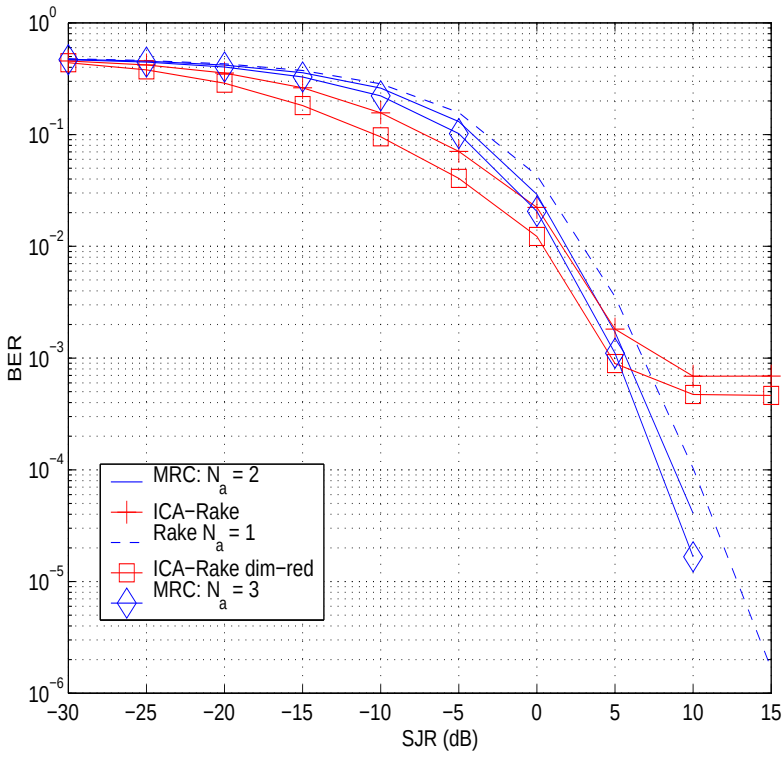


Figure 4.11: *Bit-error-rate (BER) in the case of suppression of a continuous wave jammer with a frequency offset set to* $f_j - f_c = \frac{1}{200T_s}$. *The symbol duration is* T_s. *The means that there is a total phase shift of* 2π *during a block of* $M = 200$ *symbols. Saturation effects after SJR* $= 5$ *dB can be seen here, too. These results have been previously published in [151].*

Fig. 4.13 shows the results. The case where this frequency offset exists is chosen for simplicity. Now, ICA-RAKE combination performs comparably with MRC having similar dimensions in the regions of high jamming (SJR = 0 dB). When the number of antenna elements is increased, the ICA-RAKE combination now outperforms MRC by at-least 2 dB in the SJR region between -10 to -5 dB. The BLER figure (Fig. 4.14), indicates a similar trend. This combination is better by about 6 dB when 10^{-2} is used as a target BLER and as much as 3 dB with a target of 10^{-1}.

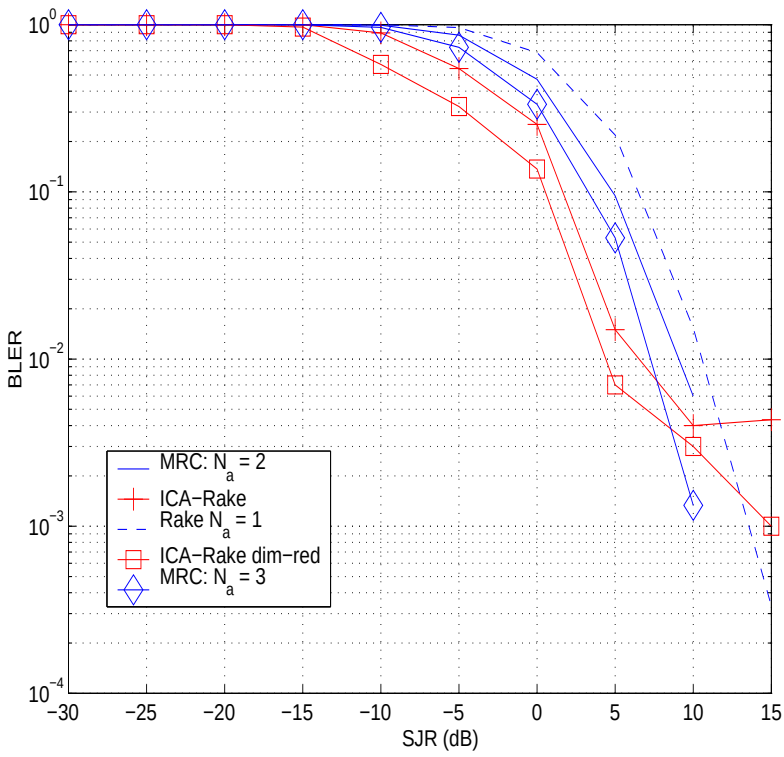


Figure 4.12: *Block-error-rate (BLER) in the case of suppression of a continuous wave jammer with a frequency offset set to* $f_j - f_c = \frac{1}{200T_s}$. *The symbol duration is* T_s. *This means that there is a total phase shift of* 2π *during a block of* $M = 200$ *symbols. Saturation effects after SJR* $= 5$ *dB can be seen here, too. These results have been previously published in [151].*

4.5 Semi-blind schemes for interference cancellation

The plain ICA-RAKE receiver [159, 150] is only the first step in the evolution of ICA based solutions. The ICA-RAKE receiver is semi-blind for the following reasons:

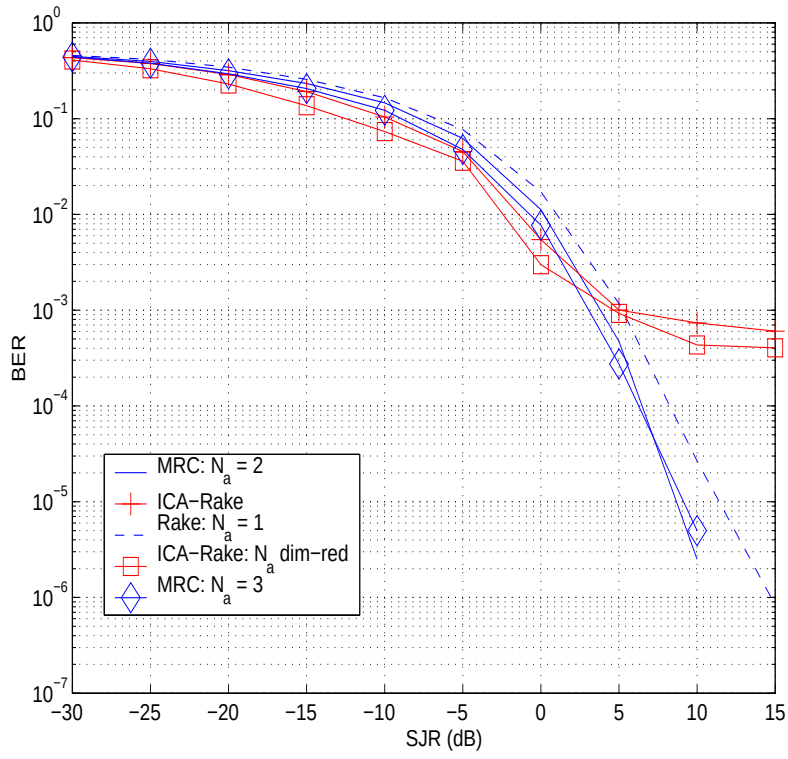


Figure 4.13: *Achieved BER for the suppression of a bit-pulsed jammer with a frequency offset set to* $f_j - f_c = \frac{1}{200T_s}$. *The symbol duration is* T_s. *These results have been previously published in [150].*

- the separation of the user-of-interest and the interfering source is blind;
- identification of the separated source requires the use of training sequences;
- since it uses the conventional MRC for post-processing, timing information is still required, and is obtained by the use of the above training sequence.

Hence, the use of training sequences solves two main problems. As with conventional detection it helps in extracting the timing information, and additionally, it helps in identifying the user-of-interest. Hence, the ICA-RAKE receiver falls

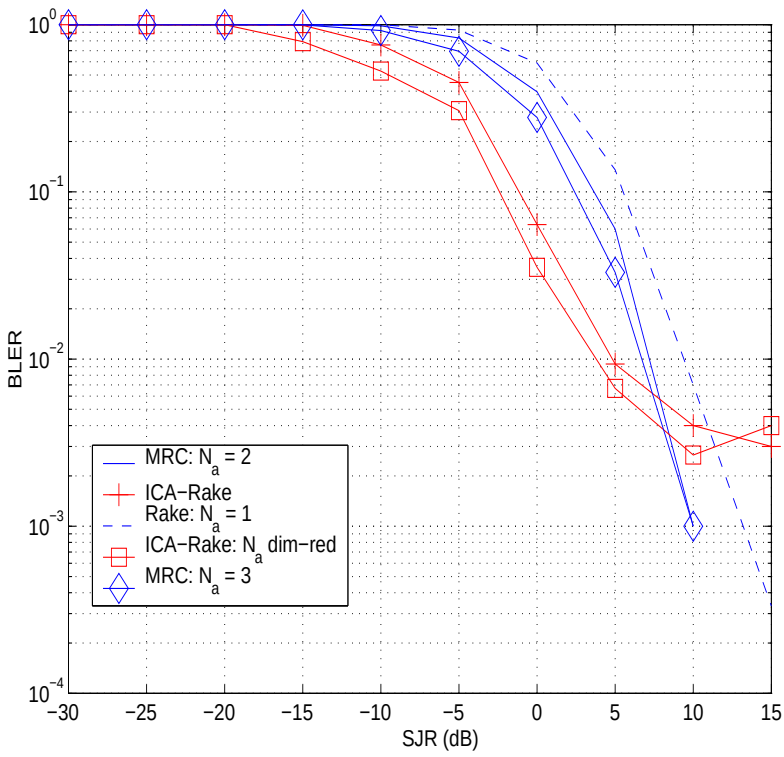


Figure 4.14: *Suppression of a bit-pulsed jammer with a frequency offset set to* $f_j - f_c = \frac{1}{200T_s}$. *The symbol duration is* T_s. *The BLER performance of ICA-RAKE is superior to that of MRC. These results have been previously published in [150].*

into category (C1) (Sec. 3.5).

From the results in Figs. 4.9, 4.11, and 4.13, it is evident that pre-processing the received signal with ICA does not provide additional suppression capability if the powers of the interfering sources and the users' signal are almost equal. The performance saturates in all the above cases when the SJR is near 0 dB, and does not improve further. This saturation is due to the fixed level of the thermal noise. This can further be observed in the following scenario.

Scenario 4: Cancellation with varying noise. Consider a situation similar to **Scenario 3**. The contribution of the jammer is now fixed *i.e.*, SJR $= -5$ dB, and the thermal noise varies. All other parameters are similar to that of **Scenario 3**.

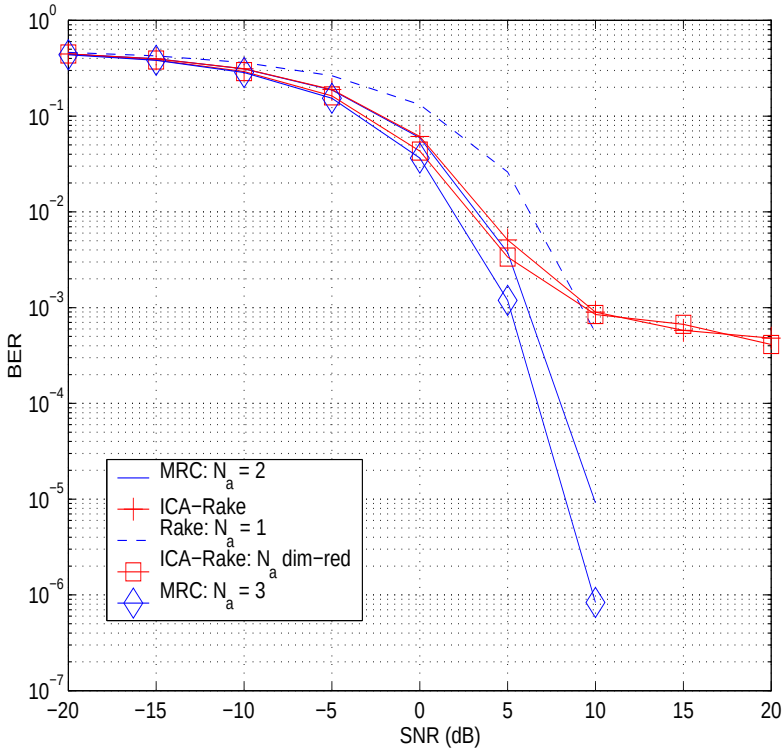


Figure 4.15: *BER in the case of suppression of a bit-pulsed jammer with a frequency offset set to* $f_j - f_c = \frac{1}{200T_s}$. *The symbol duration is* T_s. *The contribution of the jammer was now fixed to SJR* $= -5$ *dB, with the thermal noise varying. The saturation effect is visible after SNR* $= 5$ *dB. These results have been previously published in [150].*

The result is shown in Fig. 4.15. The BER performance of ICA-RAKE is similar to that of MRC with multiple antenna elements. Moreover, the BLER performance, in Fig. 4.16, is better by around 2 dB at a target of 10^{-2}. However, ICA-RAKE receivers saturate beyond 5 dB SNR. Conventional detectors, on

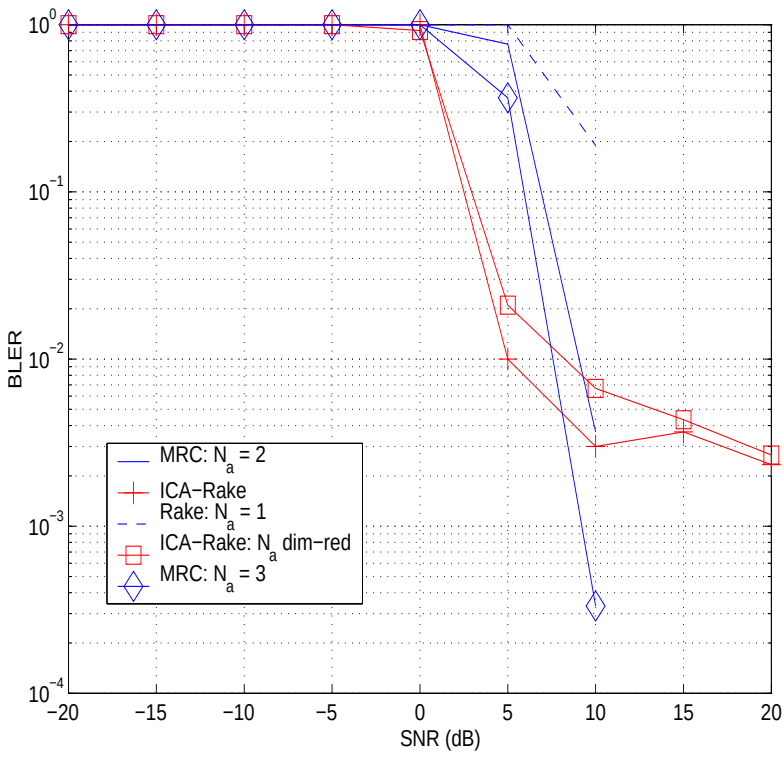


Figure 4.16: *Achieved BLER for the suppression of a bit-pulsed jammer with a frequency offset set to* $f_j - f_c = \frac{1}{200T_s}$. *The symbol duration is* T_s. *The contribution of the jammer was now fixed to SJR* $= -5$ *dB, with the thermal noise varying. The saturation effect is visible after SNR* $= 5$ *dB. These results have been previously published in [150].*

the other hand, do not saturate. They offer improvements when the SNR or the SJR is high *e.g.*, $>$ 10 dB. Hence, for low interference it is natural to use conventional detection, while ICA based structures are suitable for situations where the interference is high. This leads to the switching structures described in the following section.

4.6 The pre-switching detector: ICA-RAKE Pre-Switch

The *ICA-RAKE Pre-switch* is a structure that uses the training sequences effectively to overcome saturation effects. These sequences can be used to determine the extent of the interference. Based on this estimate the choice between simple detection and pre-processing is made. Thus, the Pre-switch estimates the training sequence based on the received signal. By comparing the estimated training symbols with the "true" training symbols, an estimate on the extent of the interference is obtained,

$$\Delta^t = \text{sign}(N_t - \max_{l=1,\ldots,N_a}(\text{d}ist_H[b_k^t, \hat{b}_k^l]) - \delta_{sw} N_t), \tag{4.32}$$

where N_t is the length of the training sequence, b_k^t is the training sequence of user k and $\hat{b}_k^l$ is the estimated sequence. δ_{sw} is the switching constant, usually a fraction of the number of training sequences N_t. N_a corresponds to the number of elements in the antenna array, which in-turn is the number of available mixtures. The illustration in Fig. 4.17 demonstrates this switching mechanism. The

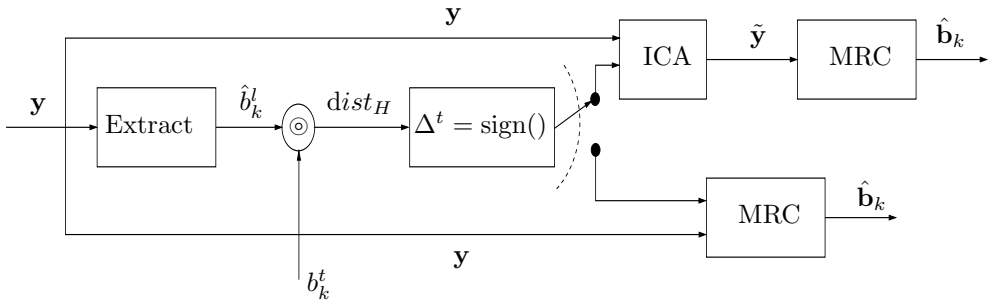


Figure 4.17: *Illustration of ICA-RAKE Pre-switch. The training sequences are extracted from the received data* $\mathbf{y}$*, which is then compared to the "true" training sequence. This comparison yields a raw estimate on the amount of interference, which is then used to decide the right detection technique.*

switch between conventional detection and ICA-RAKE detection is based on the estimate of the interference, for each block of the data. The motivation behind this is that when the signal is jammed, estimates of the training sequence are

poor, and hence further processing and suppression is required. If the estimates are good — usually a case when the SIR is quite high — Δ^t is negative, indicating that filtering and processing with ICA will not yield any better results. This further avoids the computational burden of filtering every block of received data by ICA and intuitively selects only blocks of data that are severely jammed.

An outline for the ICA-RAKE Pre-switch is shown in **Algorithm 2**. The training sequences are first extracted and an estimate on the training-bit-error-rate is made. Based on the estimate a choice is made between using ICA for separating the interference and simple detection by MRC. The `doICA` portion of the algorithm is based on the kurtosis based FastICA approach as in the basic ICA-RAKE detector. By switching between the two states, it is possible to offset the saturation effects seen in the plain ICA-RAKE detector. The performance of

Algorithm 2 ICA-RAKE Pre-Switch

for $i = 1$ to N_a **do**
  $\hat{b}_i \Leftarrow \texttt{mrc}(y_{i,1:N_t})$
  $d_i \Leftarrow \texttt{dist}_H[b_k^t, b_i]$ {# — Extract Training Symbols — #}
end for
$\tilde{k} \Leftarrow \texttt{argmax}(d)$
$\Delta^t \Leftarrow \texttt{sign}(N_t - d_{\tilde{k}} - \delta_{sw} N_t)$ {# — Calculate the distance — #}
if $\Delta^t = 1$ **then**
  $\tilde{b}_k \Leftarrow \texttt{mrc}(y)$ {# — MRC Detection — #}
  $\hat{b}_k \Leftarrow \texttt{sign}(\tilde{b}_k)$
else
  $w = \texttt{doICA}(y)$ {# — ICA-RAKE Detection — #}
  $\tilde{y} \Leftarrow w' * y$
  for $i = 1$ to N_s **do**
    $\tilde{b}_i \Leftarrow \texttt{mrc}(\tilde{y}_i)$
    $d_i \Leftarrow N_t - \texttt{dist}_H[b_k^p, \tilde{b}_i^t]$
  end for
  $\tilde{k} \Leftarrow \texttt{argmax}(d)$
  $\hat{b}_k \Leftarrow \texttt{sign}(\tilde{b}_{\tilde{k}})$
end if

the ICA-RAKE Pre-switch is shown in Fig. 4.18. The setup is as in **Scenario 2**. It can now be seen that saturation effects are not present here.

In portions where the SJR ratio is high, the ICA-RAKE branch is not activated, as MRC is able to cope with the interference. The performance of the detector in regions of high interference is quite similar to that of the ICA-RAKE detector.

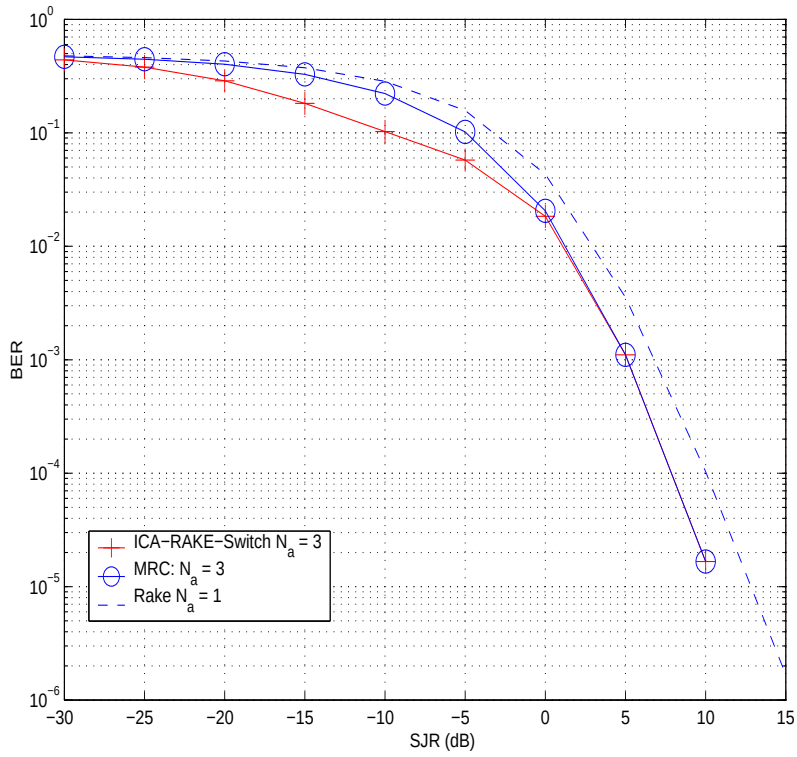


Figure 4.18: *BER in the case of suppression of a continuous wave jammer. The effect of using the Pre-Switch results in removing the saturation effects of the plain ICA-RAKE receiver. When the SJR is high, the ICA portion of the receiver is no longer triggered. This also leads improvement in the speed of the detector. These results have been previously published in [147].*

The BLER curves in Fig. 4.19 also exhibit a similar behavior. Methods for improving the performance in regions of high jamming are presented in later sections.

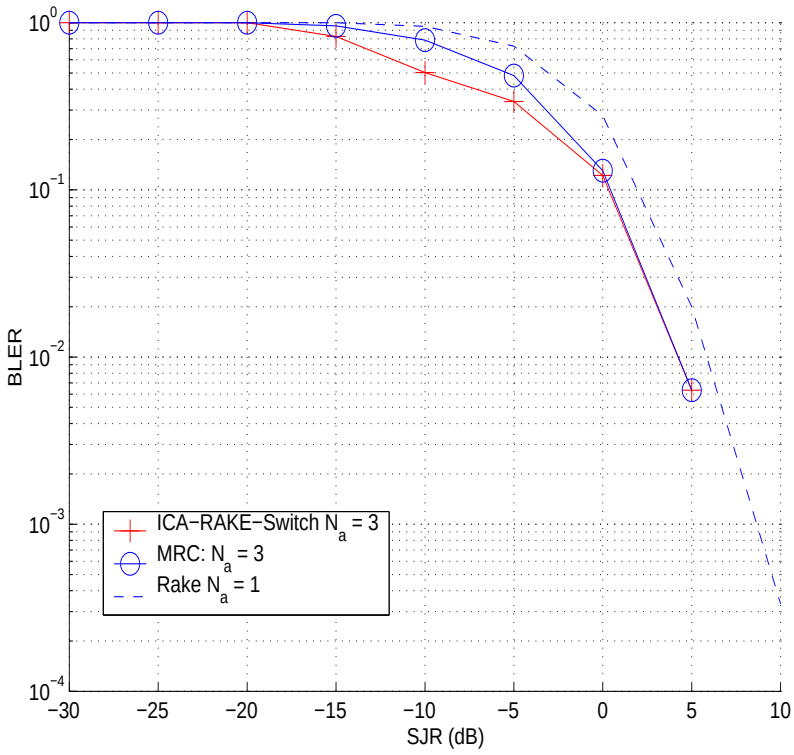


Figure 4.19: *BLER in the case of suppression of a continuous wave jammer. The effect of using the Pre-Switch results in removing the saturation effects of the plain ICA-RAKE receiver. When the SJR is high, the ICA portion of the receiver is no longer triggered. This also leads improvement in the speed of the detector. These results have been previously published in [147].*

4.7 The semi-blind correlating detector: ICA-RAKE Post-Switch

The ICA-RAKE Pre-switch detector is a first-order detector, in the sense that it uses simple distance measures to estimate the interference strength and switch between ICA-RAKE and MRC. A better method is to use correlations as a mea-

sure of similarity between the estimated training sequences and the true training sequence. This leads to the *ICA-RAKE Post-switch* detector configuration, shown in Fig. 4.20. The ICA-RAKE Post-switch operates by first separating the

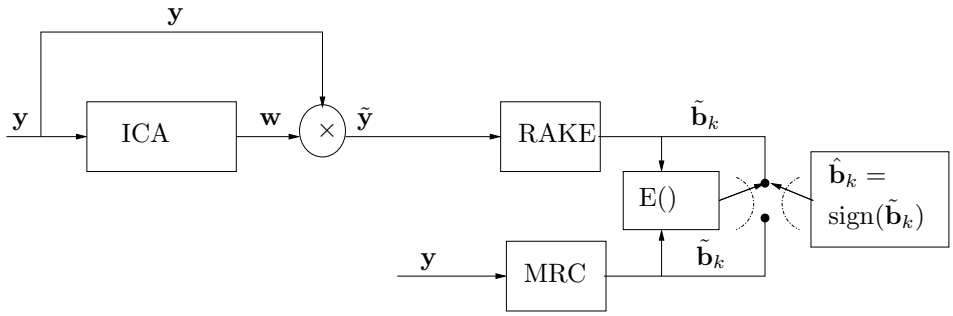


Figure 4.20: *Schematic of the ICA-RAKE Post-switch. The similarity measure is now based on correlations between the estimated and "true" training sequence. Both the ICA portion and the MRC portion of the receivers are active, and the final hard decision is based on the method that provides the best similarity to the "true" training sequence.*

sources. All the estimated sources are de-spread and the soft-decision outputs of the user-of-interested are obtained. Simultaneously, soft-decision outputs without the ICA separation are obtained separately by a maximum ratio combiner. The training sequences $\hat{b}^t_{k[1]}$ and $\hat{b}^t_{k[2]}$ are extracted from these streams. The subscript $[1], [2]$ refers to the branch from which they originate. These sequences are correlated with the true training sequence b^t_k. The branch that is chosen is based on the sequence that provides a better match with the training sequence b^t_k as,

$$i = \arg\max_{b=1,2} \left(\left[\frac{b^t_k \cdot (b^t_{k[1]})^T}{N_t} \right], \left[\frac{b^t_k \cdot (b^t_{k[2]})^T}{N_t} \right] \right). \tag{4.33}$$

A schematic of the ICA-RAKE Post-switch is in **Algorithm 3**. `doICA, mrc` are the functions as described in **Algorithms 1,2**. The ICA-RAKE Post-switch works on tentative estimates for the mth soft symbol of the desired user k from both the branches and decodes the branch that provides higher correlation. This is equivalent to estimating the interference level in the ICA-RAKE Pre-switch.

Though the ICA-RAKE Pre- and Post-switch look similar in concept - they estimate the interference level and switch between separating the interference out

Algorithm 3 ICA-RAKE Post-Switch

$w \Leftarrow \texttt{doICA}(y)$
$\tilde{y} \Leftarrow y^{'} * w$
$\tilde{b}_{k[1]} \Leftarrow \texttt{mrc}(\tilde{y})$
$\tilde{b}_{k[2]} \Leftarrow \texttt{mrc}(y)$
$\hat{b}^t_{k[1]} \Leftarrow \texttt{sign}(\tilde{b}_{k[1],1:N_t})$
$\hat{b}^t_{k[2]} \Leftarrow \texttt{sign}(\tilde{b}_{k[2],1:N_t})$
$\Delta_{[1]} \Leftarrow \texttt{corr}(b^t_k, \hat{b}^t_{k[1]})$
$\Delta_{[2]} \Leftarrow \texttt{corr}(b^t_k, \hat{b}^t_{k[2]})$
$i \Leftarrow \texttt{argmax}(\Delta)$
$\hat{b}_k = \texttt{sign}(\tilde{b}_{k[i]})$

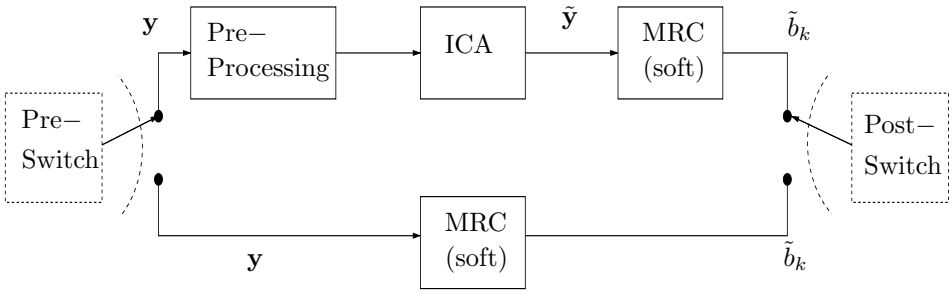


Figure 4.21: *Illustration of the ICA-RAKE Pre- and Post-Switch. The switching that takes place at the front end of the structure is termed as Pre-Switch. It is the simpler of the two structures, while the switch based on post-processing is called the Post-Switch. Post-Switch requires soft estimates from both the branches.*

and processing the signal without separation - Pre-Switch is based on accurately estimating the training sequence from the interfered sources. In Post-Switch on the other hand, the training sequences are more or less free of all interference as they are already processed and then estimated. Both these switches can be illustrated with the following schematic diagram (Fig. 4.21).

The Pre-Switch is faster than the Post-Switch as it bypasses the separation-identification phase of the Post-Switch when the interference is low. On the

other hand, since the matching is based on a first-order distance measure, the performance is limited to the performances of the individual detectors *i.e.*, ICA-RAKE and MRC. Computational considerations of ICA-RAKE, ICA-RAKE Pre- and Post-Switch are postponed to Section 4.9.

4.8 Examples of interference cancellation with the Pre- and Post-Switch

Pre- and Post-switch based examples of interference cancellation are presented in this section[6]. These scenarios include cancellation of inter-cell interference and cancellation of the external jammer.

In canceling the interference due to adjacent cells, the model used is as described in Sec. 2.2.2. The number of interfering layers is either one or two (which is equivalent to having either 6 or 12 interfering cells). The case of external interference involves canceling a jammer as explained in the previous scenarios.

Scenario 5: Effect of inner and outer layer interference. The cell-of-interest is surrounded by two layers of interfering cells at a distance of $d_{in} = 2000$ m. Each cell has a radius of $d_c = 666$ m. All cells have a fixed number of users (*i.e.*, are loaded similarly), and the spreading factor is $C = 31$. The length of the data block for each user is $M = 200$ symbols and is modulated by QPSK modulation. An antenna array with $N_a = 3$ elements is assumed in the receiver.

Cancellation of the inner and outer layer interference is effective with both ICA-RAKE Pre- and Post-Switch (Fig. 4.22 and Fig. 4.23). The ICA-RAKE Post-Switch offers better performance than the Pre-Switch from -15 dB onwards, before which both methods perform equivalently. From about -20 dB SIR to around 0 dB SIR, the gain is around 3 dB for the interference structure consisting of the inner layer of cells.

Fig. 4.23 is the achieved block-error rate for an interference structure involving both the inner and the outer layers. The results are similar. ICA considers this interference as a single block interference, which helps in mitigating it. While all

[6]ICA-RAKE Pre- and Post-Switch and the abbreviations Pre- and Post-Switch are equivalent.

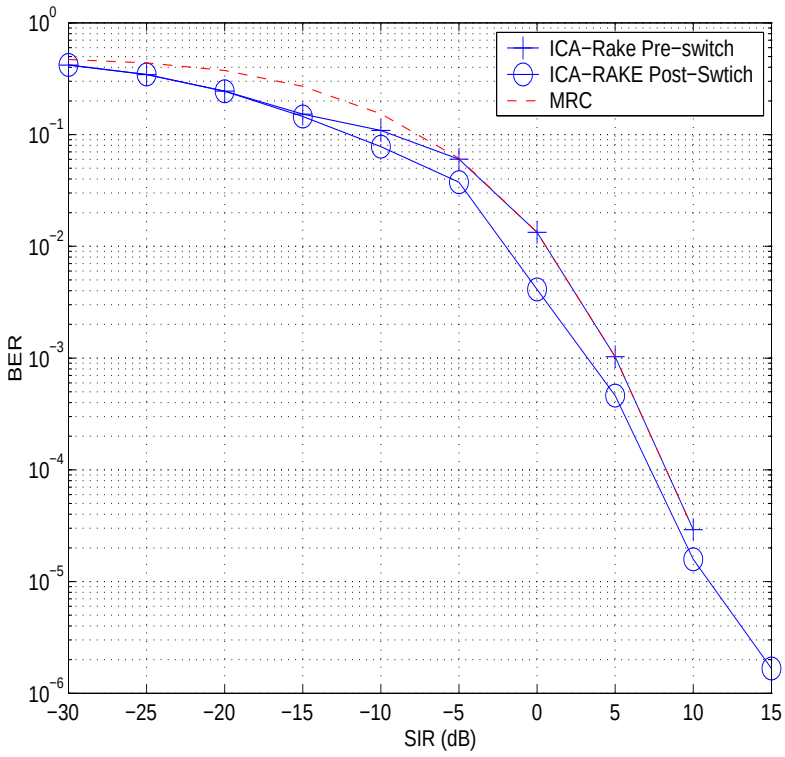


Figure 4.22: *Cancellation of interference due to the inner layer. The BER performance of the Pre-Switch is equivalent to that of MRC in this case. On the other hand, the performance of the Post-Switch is better. These results have been previously published in [161].*

the three methods compared perform equally, Post-Switch is marginally better as illustrated in **Scenario 6**.

Scenario 6: Load increases in the adjacent cells. Assuming that the cell-of-interest has a fixed number of users, the load of the adjacent cells is increased from 0.25 load to full load. This causes variation in the nature of interference now at the cell of interest. All other parameters are similar to **Scenario 5**.

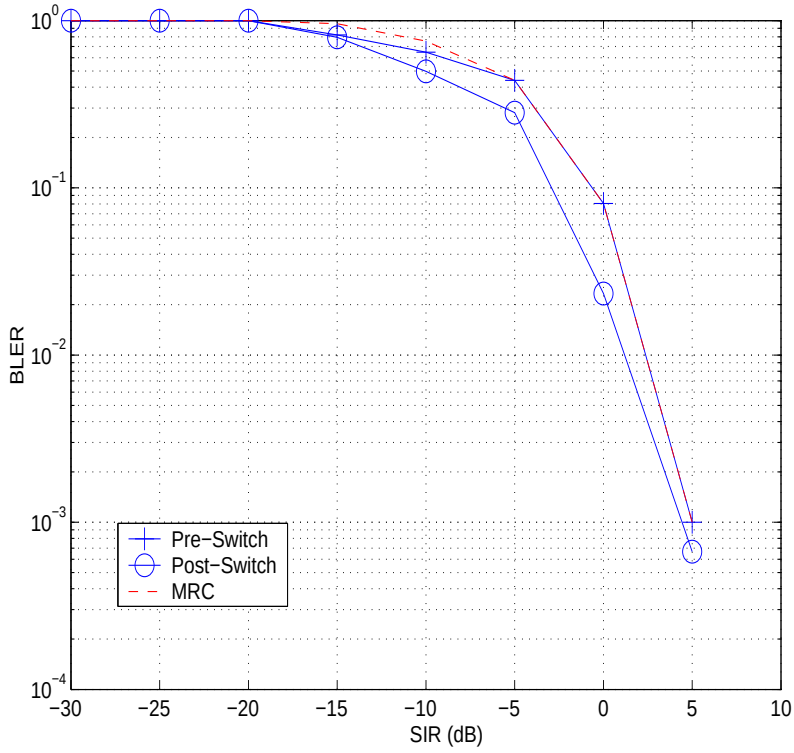


Figure 4.23: *The BLER performance for cancellation of interference due to the both the layers. In this case, all the three methods provide the same level of performance, with the Post-Switch marginally better. These results have been previously published in [161].*

In Fig. 4.24, the effect of the increase in load of the adjacent cells on the performance of the algorithm is seen. The ICA-RAKE Post-Switch is only considered in this experiment. As expected, the performance of all receivers is good if the interference is minimum. As the load increases from 0.25 to 1.0 *i.e.*, the number of interfering users in the cells increases from 8 to 32, the performance drops. In spite of this loss, the gain induced by the ICA-RAKE Post-Switch can be seen clearly. For example at around 5 dB SIR, the performance of the fully loaded ICA-RAKE Post-switch is equivalent to that of MRC with half load. This results in improved bandwidth utilization. In **Scenario 7**, the effect of increases of load

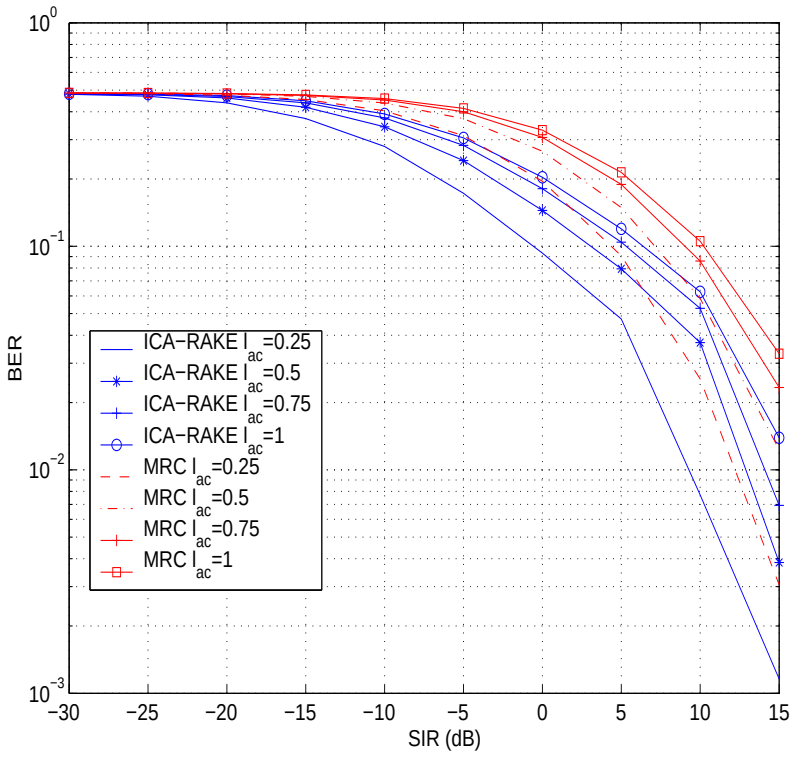


Figure 4.24: *Load increases in the adjacent cells. These increases cause a degradation in the performance of both the detectors. Post-Switch still outperforms MRC with a considerably less interference load. These results have been previously published in [148].*

in the home cell can be seen which helps in understanding the ICA induced gain on conventional detection.

Scenario 7: Load fluctuation in home cell with constant interfering cells. With the simulation parameters similar to the ones in **Scenario 5**, the number of interfering cells and the load in those cells were fixed to half load. The load in the cell-of-interest is now fluctuated. These effects are examined here. Additionally, the loss due to such interference is quantified with the help of soft lower- bounds. These bounds are the simulated

lower-bounds obtained with a single source in the home cell and a single interfering cell with a single user.

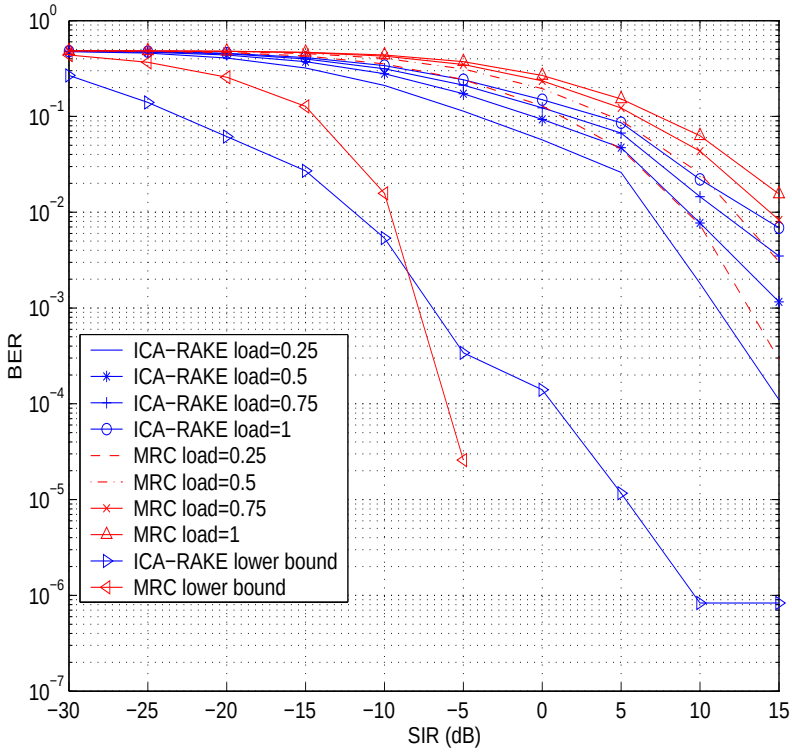


Figure 4.25: *Effect of increasing load in the home cell. Post-Switch at full load is able to provide better performance than MRC with considerably less load. The curves on the left are experimentally determined "soft" lower bounds. These results have been previously published in [148].*

With constant external interference, load fluctuations in the home cell are examined in Figs. 4.25 and 4.26. In the BER figure (Fig. 4.25), there is a gain of around 3 dB for every successive load increase with respect to MRC. The performance of the ICA-RAKE Post-Switch with full load is better (or roughly) equivalent to the half load performance of MRC. Hence, at around 5 dB SIR, ICA-RAKE Post-Switch based receiver can handle and support twice the number of

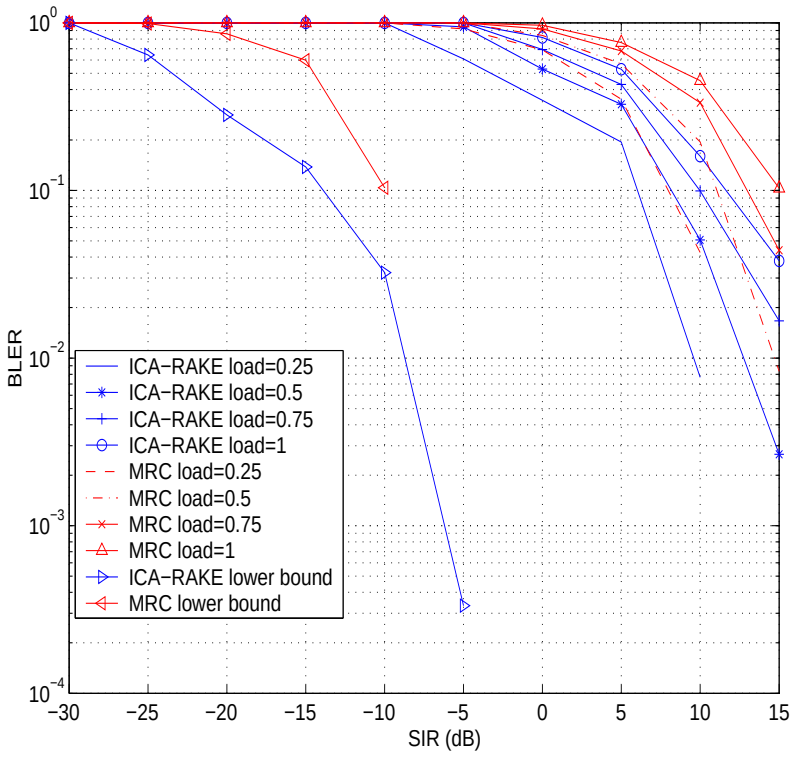


Figure 4.26: *Effect of increasing load in the home cell. Post-Switch at full load is able to provide better performance than MRC with considerably less load. The curves on the left are experimentally determined "soft" lower bounds. These results have been previously published in [148].*

users as compared to MRC. The BLER performance in Fig. 4.26 is quite similar. The leftmost curves in both these figures are the "soft-lower" bounds. Up to around −8 dB SIR, ICA-RAKE Post-Switch offers good performance compared to MRC for the lower bound. At SIR= −20 dB, the gain is as high as 7 dB. After this, the contribution of the interfering signal falls to such a level that it is no longer practical to use ICA for canceling the interference. Moreover, the fixed level of noise causes the ICA-RAKE Post-Switch performance to saturate. Similar conclusions can be drawn from the BLER curves as the slopes of the ICA-RAKE curves are much lower than the MRC equivalent.

While ICA-RAKE algorithms provide good performance in the ranges between -20 to 10 dB SIR/SJR, they still are poor at the extremes. Both Pre- and Post-Switch can handle multiple sources of interference, but their performance is not very good when the SIR $= -30$ dB or at very high interference ratios. This performance can be attributed to the whitening process (Sec. 4.1.6). The whitened data is a projection of the original data onto a space spanned by the eigenvectors of the covariance matrix as,

$$\mathbf{y} = \mathbf{\Lambda}^{-1/2}\mathbf{E}^T\mathbf{y}_{orig}, \tag{4.34}$$

with ,

$$[\mathbf{E}, \mathbf{\Lambda}] = \text{eig}(E\{\mathbf{y}_{orig}\mathbf{y}_{orig}^T\}). \tag{4.35}$$

The accuracy of this projection is hence dependent on the quality of the covariance matrix $E\{\mathbf{y}_{orig}\mathbf{y}_{orig}^T\}$.

One way to improve the accuracy of the covariance matrix is to use a larger block size. Another possibility is to use ICA algorithms which do not require pre-whitening. Such algorithms, for example the natural gradient algorithm [5], are suitable for a small number of sources. For larger numbers of sources this algorithm, too, requires pre-whitening to perform approximately.

Scenario 8: Increase in block sizes. The setup here is similar to **Scenario 3**. The interference is a bit-pulsed jammer. The block size is now $M = 5000$ QPSK modulated symbols. All the other system parameters remain the same. The noise in the system is also considerably reduced here. The detector settings are based on FastICA, JADE, TDSEP and MRC. JADE, and TDSEP are re-examined here again due to the increase in block sizes. Increased block size could help in the performance of all the data driven algorithms.

The effect of increasing the block size results in better performance of the Pre- and Post-Switch detectors as seen in Figs. 4.27 and 4.28. While JADE based and TDSEP based receiver performances are equivalent to the MRC, Pre-Switch provides improved performance up to about -5 dB SJR, after which the MRC portion of the receiver takes over. Post-Switch on the other hand has a significantly improved performance.

Now, post-switch provides a favorable improvement over all the other compared schemes. The BLER curves again demonstrate the superiority of the Pre-

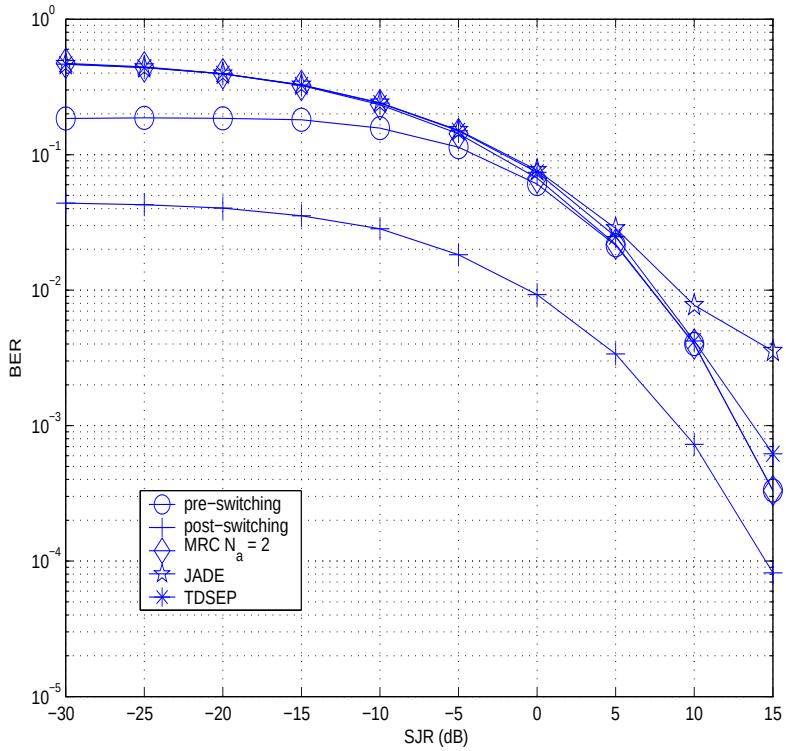


Figure 4.27: *BER improved interference cancellation with extended block sizes. Both the Pre- and Post-Switch benefit from the extended block size. However, this size is not yet sufficient for a tensorial method. In the absence of significant temporal structure, TDSEP has a poor performance. Block size increases do not affect MRC, which is not dependent on it. These results have been previously published in [151].*

and Post-Switch detectors. With a target of 10^{-2} BLER, the improvement of the Pre-Switch over MRC is nearly 5 dB while Post-Switch provides a gain of 10 dB.

All the above scenarios considered a single path channel. Multipath on the other hand introduces temporal correlations for a single user, which can be utilized effectively by some algorithms. Generally, multipath interference causes a degradation in performance when they are not utilized effectively.

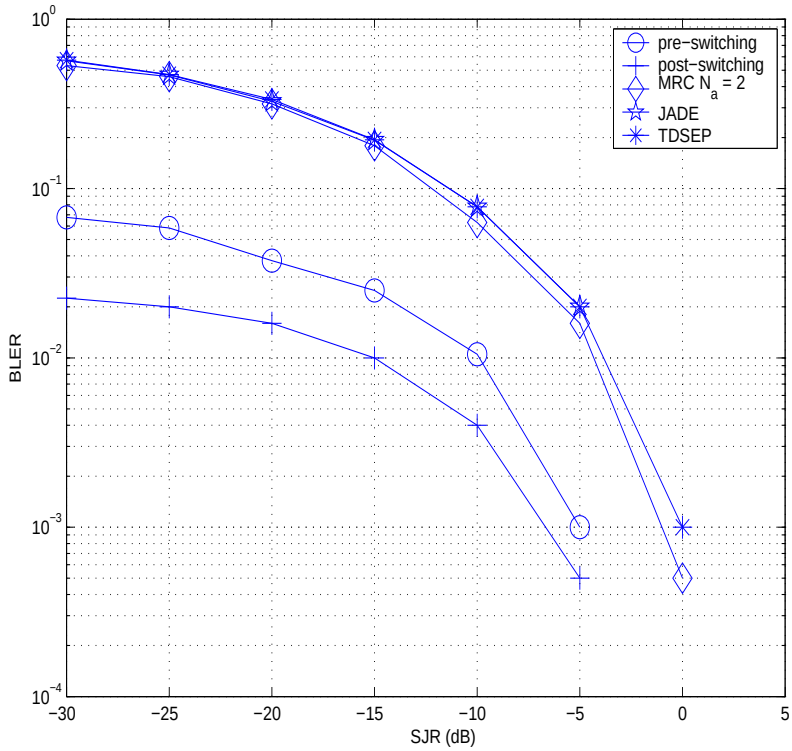


Figure 4.28: *BLER improved interference cancellation with extended block sizes. Both the Pre- and Post-Switch benefit from the extended block size. However, this size is not yet sufficient for a tensorial method, such as JADE. In the absence of significant temporal structure, TDSEP has a poor performance. Block size increases do not affect MRC, which is not dependent on it. These results have been previously published in [151].*

Scenario 9: Multipath interference cancellation. **Scenario 3** is the setup used in this scenario. The interference structure is the same. The system now has a fixed multipath channel [91, 82, 81] consisting of up to $L = 5$ paths. The interference is also multipath. This leads to a coherent jammer [53].

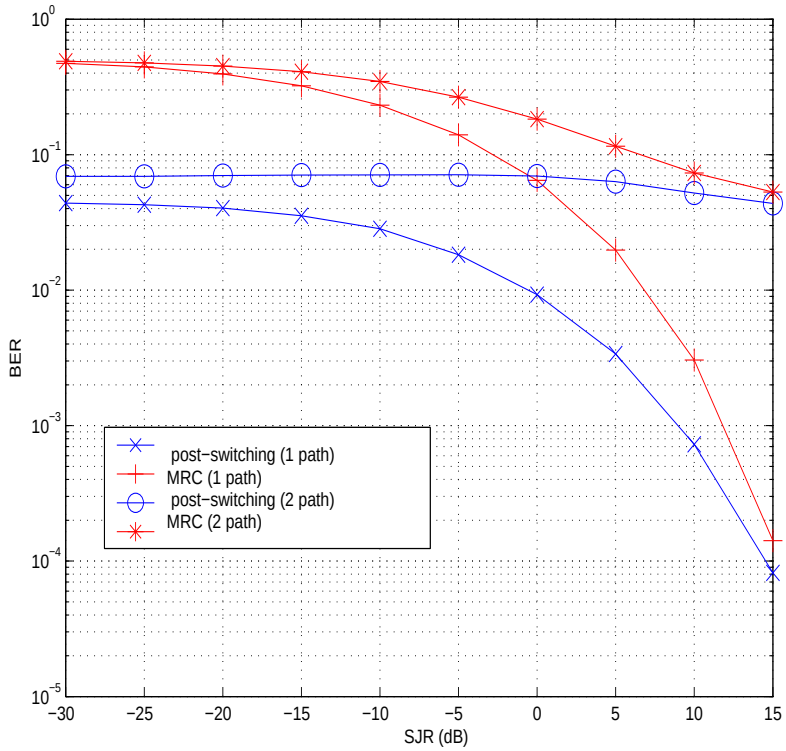


Figure 4.29: *Cancellation in the presence of fixed multipaths [91, 82, 81]. The BER of Post-Switch is better than that of MRC even though it exhibits saturation effects in the case of $L = 2$ paths. The performance of the Post-Switch is effective even in the presence of multipaths. These results have been previously published in [151].*

In Fig. 4.29 an $L = 2$ path channel is considered. The BER performance of the Post-Switch now saturates when the number of paths in the system increases. Overall, the performance is still better than MRC even under a single path channel. Up to 0 dB SJR, the performance of the Post-Switch with a two path channel is significantly better than that of MRC with a single path. The BLER curves in Fig. 4.30 reinforce the above facts.

Another way to visualize the results is to view the distribution of the correct bits in one block. With a block size of $M = 5000$ QPSK symbols, there are 10000

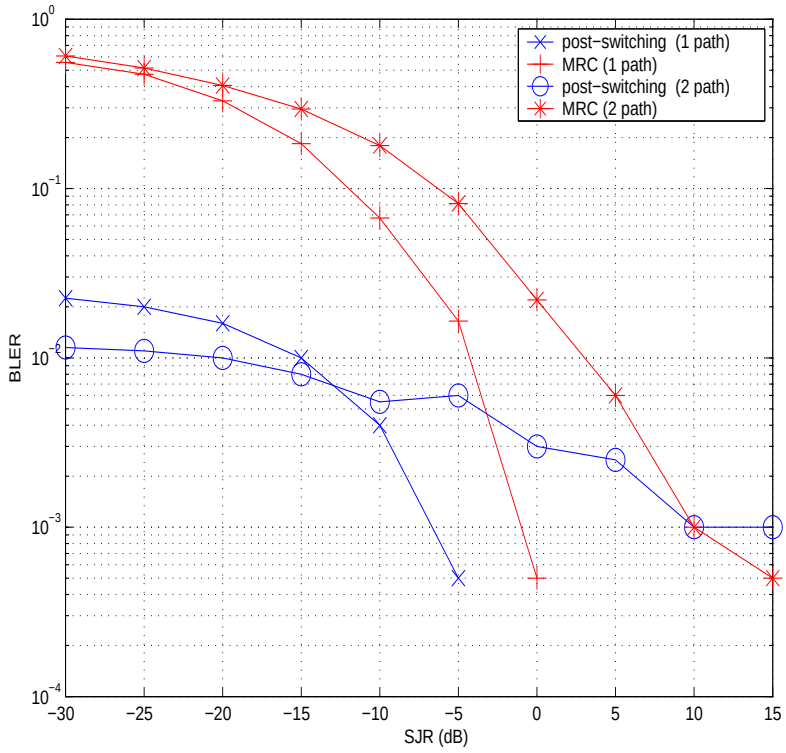


Figure 4.30: *Cancellation in the presence of fixed multipaths [91, 82, 81]. The performance of the Post-Switch, measured by the BLER, is effective even in the presence of multipaths. These results have been previously published in [151].*

bits. Each scenario is simulated with 2000 trials. The distribution for the correct bits for **Scenario 9** is shown in Fig. 4.31 for a single SJR value (SJR $= -30$ dB). The left column of Fig. 4.31 is for an $L = 1$ path system and the right column is for an $L = 5$ path system. For a single path case, Post-Switch has nearly all blocks (> 1900) having almost all bits correct. This is not the case with MRC. Around 40% (≈ 800) blocks have completely erroneous bits. 25% of the blocks have around 4000 correct bits, while only a very small portion of the blocks have $> 60\%$ correct bits. When the number of paths increases, the distribution tends to move to the left *i.e.*, the number of blocks with a large number of erroneous bits increases. This trend is clearly visible in the right column of Fig. 4.31. With

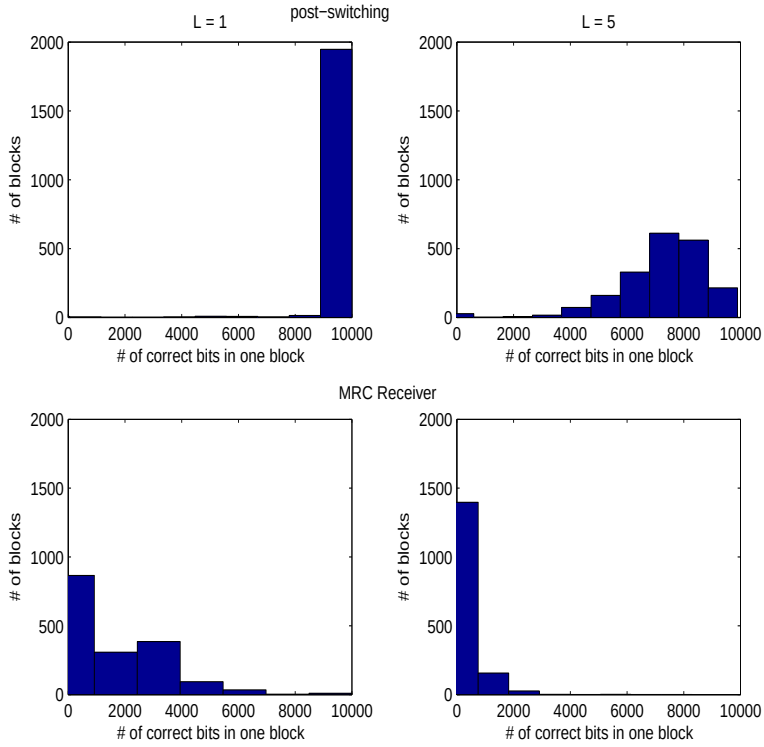


Figure 4.31: *Distribution of the correct bits. The upper two figures show results for Post-Switch, the two lower ones for MRC. The left subfigures show the case of one path ($L = 1$) and the right subfigures, the case of $L = 5$ paths. With a block size of $M = 5000$ QPSK symbols, there are* 10000 *bits. As the distribution shifts to the left, there are more blocks with a large portion of erroneous bits. This effect can be seen in the MRC case when the number of paths is $L = 5$. These results have been previously published in [151].*

Post-Switch, the mean, which in this case is the number of correct bits, is around 8000. With MRC, the distribution is towards the other end. More than 70% of the blocks have a large number of erroneous bits. These distribution graphs demonstrate the vulnerability of these algorithms to multipath propagation.

The sample size (or in this case the block size) affects the performance of

subspace based algorithms. When this is low the parameters of the subspace are inaccurately estimated and this leads to poor performance. Increasing the sample size hence improves the estimates, but beyond a certain value these increases do not really help. Determination of the optimal sample size is still an unresolved topic. Several principles like the Akaike information criteria (AIC) or the minimum description length [107, 28] can be used to obtain a coarse estimate of the order. Experimentally, the "optimal" sample size can be determined by calculating the slope of the BLER curves.

Scenario 10: Block size determination. The interference level is now fixed to SJR= -30 dB. The block size M varies. All the other parameters other are fixed as in **Scenario 9**.

Fig. 4.32 shows the BLER as a function of varying block sizes. In the case of MRC, the changes in block sizes have no impact on its performance. Post-Switch on the other hand is very much dependent of the block size. There is a drastic improvement in performance as the block size increases from $M = 200$ to $M = 500$, and performance continues to improve until about $M = 1000$. Examining the slopes of the BLER curves in Fig. 4.33, the correct block size for a single path channel corresponds to value of $M = 1000$. In the $L = 2$ path case, the block size still corresponds to $M = 1000$ for Post-Switch while it is has no bearing at all on MRC, as expected.

In Fig. 4.34, the distribution of the correct bits is depicted for Post-Switch and JADE. Distributions for MRC and TDSEP are shown in Fig. 4.35. JADE exhibits poor performance for all paths. This is most likely due to the insufficient sample size for the estimation of the tensor matrices. Post-Switch is capable of providing good performance for $L = 2$ paths, and has at least 50% correct bits in all blocks. TDSEP in Fig. 4.35, has an interesting behavior. As the number of paths increases, the graph tends to move towards the right, contrary to the others. The performance actually improves. This is because the increase in paths tend to produce temporally correlated mixtures and TDSEP effectively uses these temporal correlations to separate the data. This is seen in more detail in Fig. 4.36.

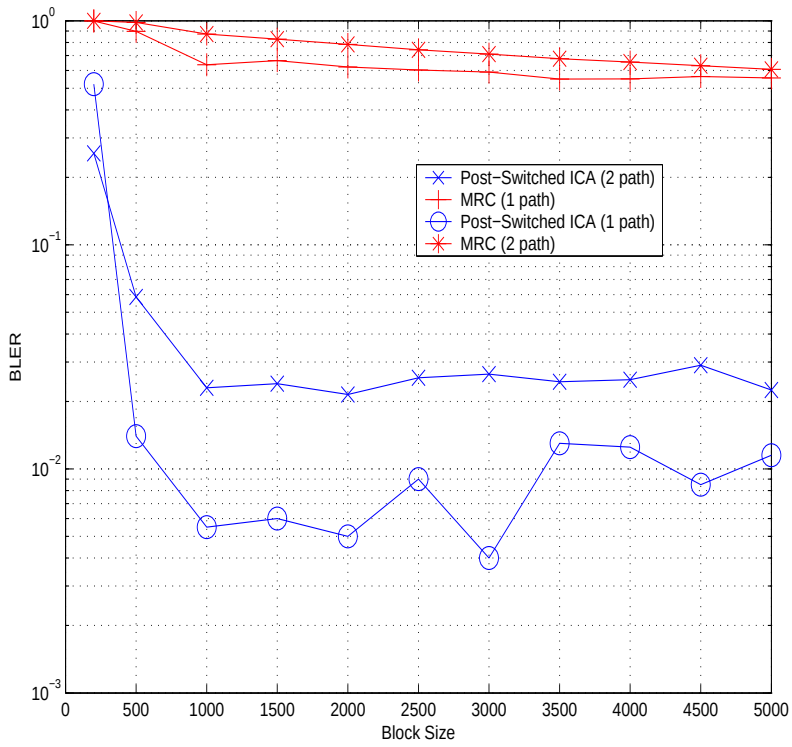


Figure 4.32: *Block error rates as a function of block size. For a single path, Post-Switch tends to saturate after a block size of $M = 1000$. While there is still some fluctuation in the performance for $L = 2$ paths, from the slopes (right), $M = 1000$ is a sufficient sample size for these experiments. These results have been previously published in [149].*

4.9 Computational considerations

The computational complexity of plain ICA-RAKE, Pre-Switch and Post-Switch schemes is expressed as the number of *multiply and add*-operations needed. Assume that the number of users is K, the length of the data vector is C, the block size is M, and L is the number of correlated paths on the channel.

The proposed methods are subspace methods. They require that the signal

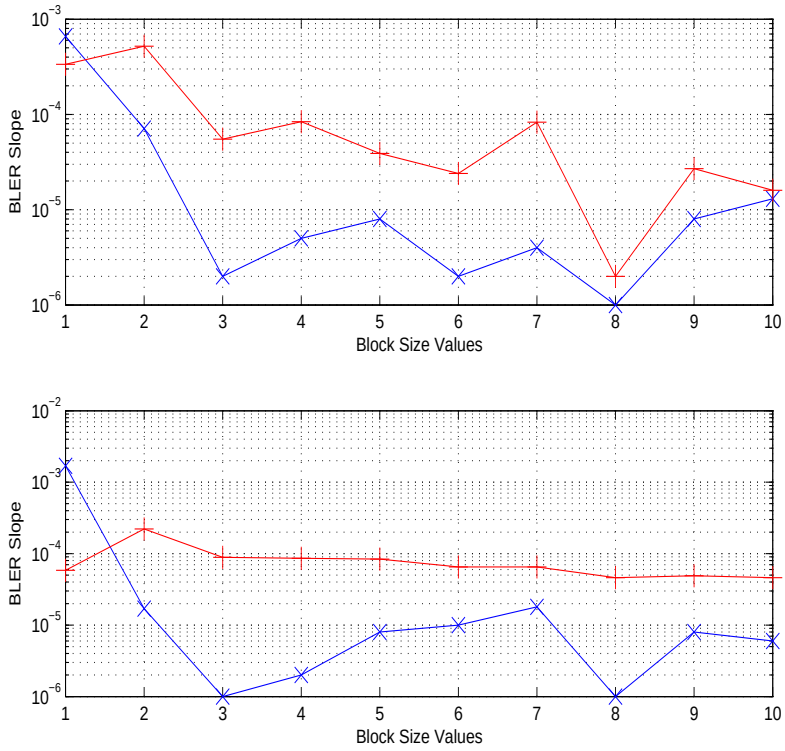


Figure 4.33: *Block error rates as a function of block size. The x-axis indices correspond to the block sizes starting from* 500 *to* 5000 *at steps of* 500. *The* + *represents MRC, while the x represents Post-Switch. The above figure is for* $L = 1$ *paths, which the figure below represents the* $L = 2$ *path case. For a single path, Post-Switch tends to saturate after a block size of* $M = 1000$. *While there is still some fluctuation in the performance for* $L = 2$ *paths.* $M = 1000$ *is a sufficient sample size for these experiments. These results have been previously published in [149].*

subspace is estimated *i.e.*, the vectors that span the signal subspace must be found. The classical approach for calculating the signal subspace is either eigenvalue decomposition (EVD) or the singular value decomposition (SVD) of the data autocorrelation matrix. The computational complexity of estimating eigenvectors of a $C \times C$ dimensional matrix is $\mathcal{O}(C^3)$ [52]. This complexity can be reduced to $\mathcal{O}(C^2K)$ when only K eigenvectors are computed. Several further

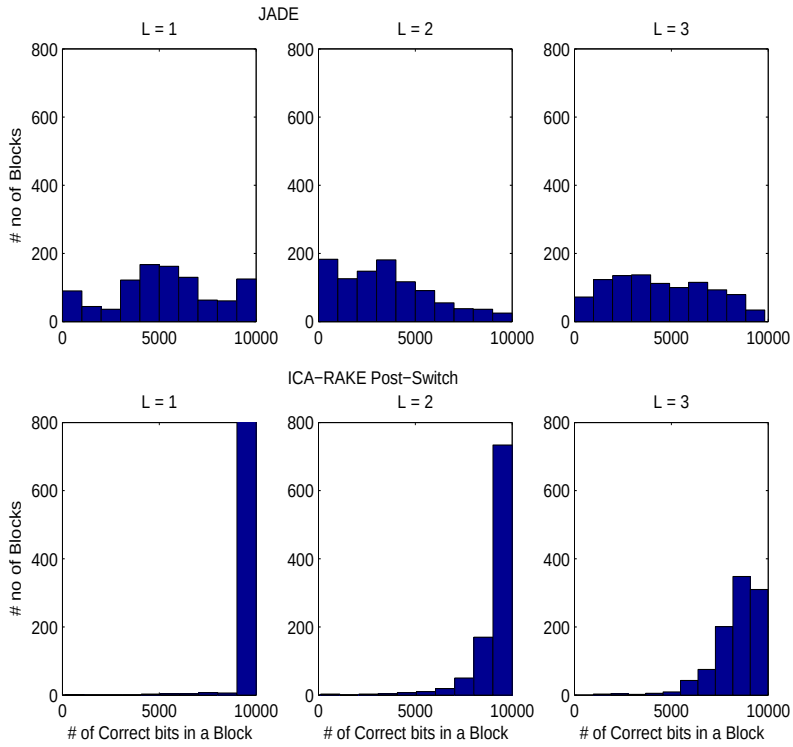


Figure 4.34: *Distribution of correct blocks for JADE and ICA-RAKE Post-Switch. The interference level is now set to SJR = −10 dB. The performances tend to fall as the number of paths increases. Post-Switch has a good performance even as the number of paths increases. These results have been previously published in [153].*

algorithms have emerged [206], that can reduce this complexity to $\mathcal{O}(CK)$ per iteration, often at the expense of accuracy of these estimates.

In ICA, the data is pre-processed by whitening as in (4.14). The computational complexity in this case is $\mathcal{O}(CK) + \mathcal{O}(K^2) = \mathcal{O}(CK)$, since $K < C$ [158]. Additional operations follow from the iterations of the FastICA algorithm's update rule defined as,

$$\mathbf{w}(t) = \frac{1}{M}\mathbf{y}\beta(t-1)^H - \gamma\mathbf{w}(t-1), \tag{4.36}$$

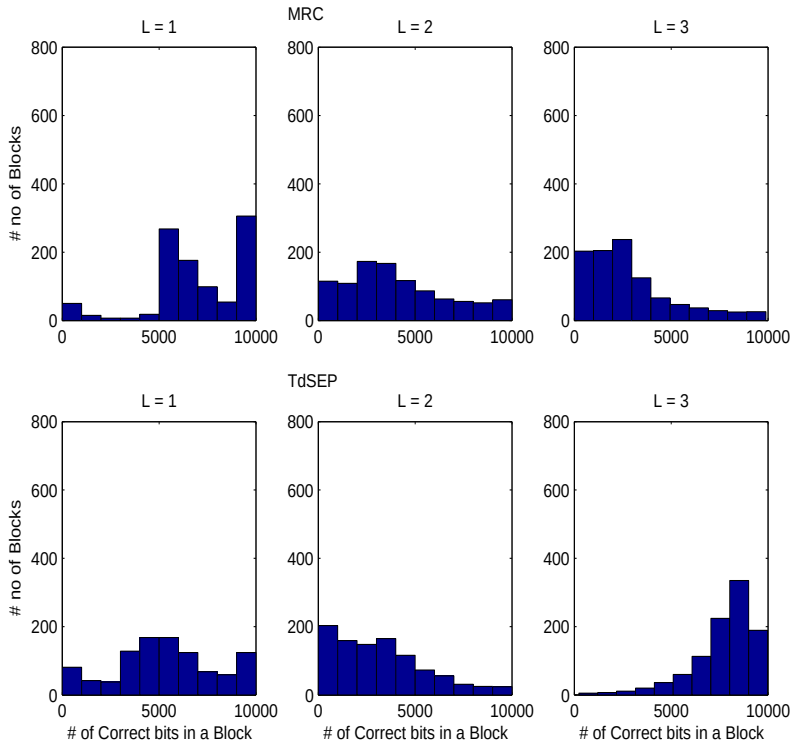


Figure 4.35: *Distribution of the correct blocks for MRC and TDSEP. The interference level is now set to SJR = −10 dB. The performance of MRC tends to fall as the number of paths increases, but for TDSEP it increases. These results have been previously published in [153].*

where M is the size of the one block of data *i.e.*, the number of samples, $\mathbf{y} \stackrel{\text{def}}{=} [\mathbf{y}_1 \dots \mathbf{y}_m]$ and $\beta(t) \stackrel{\text{def}}{=} |\mathbf{w}(t)^H \mathbf{y}|^2 \times (\mathbf{w}(t)^H \mathbf{y})$. This multiplication is defined elementwise. Thus, $\mathbf{w}^H \mathbf{y}$ needing KM multiplications and summations, is of the order $\mathcal{O}(KM)$, and $\beta(t)$ is of the order $\mathcal{O}(KM) + \mathcal{O}(M)$ where the last term is due to element-wise multiplications. Hence, the final complexity of one FastICA iteration is in the order of $\mathcal{O}(KM)$. With whitening, this complexity is of the order $\mathcal{O}(CK) + \mathcal{O}(KM)$.

Method	Complexity
Plain ICA-RAKE	
Whitening	$\mathcal{O}(CK)$
Fast-ICA	$\mathcal{O}(KM)$
Matched filter	$\mathcal{O}(CL)$
Overall	$\mathcal{O}(CK + KM + CL)$
Pre-Switch ICA-RAKE	
Whitening	$\mathcal{O}(CK)$
Fast-ICA	$\mathcal{O}(KM)$
Switching	$\mathcal{O}(N_t)$
Matched filter	$\mathcal{O}(CL)$
Overall	$\approx \mathcal{O}(CK + KM + CL)$
Post-Switch ICA-RAKE	
Whitening	$\mathcal{O}(CK)$
Fast-ICA	$\mathcal{O}(KM)$
Switching	$\mathcal{O}(N_t)$
Matched filter	$\mathcal{O}(CL)$
Overall	$\approx \mathcal{O}(CK + KM + N_t + CL)$

Table 4.1: Computational complexity of ICA methods. Here, C is the code length, K is the number of users, the block size is M, L is the number of correlated paths, and N_t is the length of the training sequence.

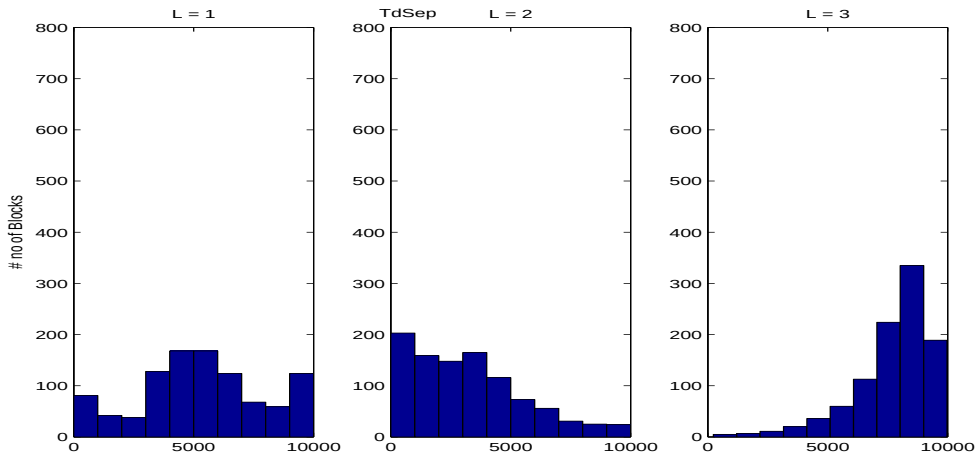


Figure 4.36: *Distribution of correct blocks with a TDSEP based detector. The presence of multipath increases the temporal correlation. This helps in the TDSEP detector's performance. These results have been previously published in [153].*

All the structures rely on the matched filter configuration, which requires C computations for one path, and hence has a complexity of $\mathcal{O}(CL)$ for the L considered paths. Pre-Switch and Post-Switch require the calculation of distance measures for switching. The complexities for these are of the order $\mathcal{O}(N_t)$ additions for the Pre-switch that uses N_t training sequences, and a first order distance measure as defined in (4.30). This is negligible since N_t is small, and additions do not significantly increase the complexity. Hence the Pre-Switch has a complexity in the order of $\mathcal{O}(CK) + \mathcal{O}(KM) + \mathcal{O}(CL)$. In the case of the Post-Switch, $\mathcal{O}(N_t)$ multiplications and additions are required for calculating the second-order distance. Hence, the order is $\mathcal{O}(CK) + \mathcal{O}(KM) + \mathcal{O}(CL) + \mathcal{O}(N_t)$. Many modern microprocessor designs execute a combined (floating point) multiply-add instruction in one clock cycle. If this can be exploited, then the complexities of both the methods can be similar. The complexities for the algorithms are summarized in the Table 4.1.

4.10 Analysis of the ICA-RAKE schemes

In this section, the plain ICA-RAKE receiver, the Pre-Switch ICA-RAKE and the Post-Switch ICA-RAKE are analyzed. Bounds for the performance of the above receivers are derived in a semi-analytical manner. These bounds will be derived in two steps:

1. Lower bound for a QPSK modulated system in the presence of a jammer [146],
2. ICA induced gain due to the separation of the jammer.

A combination of the above two will lead to a semi-analytic bound for the plain ICA-RAKE receiver. Additionally, by analyzing these theoretical results, a criterion for switching will be derived.

4.10.1 Performance of QPSK modulated system in the presence of a jammer

The average bit error probability for binary orthogonal signals (QPSK modulation) is given by [140]

$$P_b = Q\left(\sqrt{\frac{E_b}{N_o}}\right), \tag{4.37}$$

where E_b is the energy per bit and N_o is the one-sided noise power spectral density. $\frac{E_b}{N_o}$ is the SNR per bit, now represented by γ_b. The $Q()$ function is defined as [146]:

$$Q(x) = \frac{1}{\sqrt{2\pi}} \int_x^\infty e^{-t^2/2} dt, \quad x \geq 0. \tag{4.38}$$

The $Q()$ function is generally defined by using the complementary error function as:

$$Q(x) = \frac{1}{2}\text{erfc}\frac{x}{\sqrt{2}} \tag{4.39}$$

where $\text{erfc}(x) = 1 - \text{erf}(x)$ is the complementary error function, and it describes the area under the tail of the Gaussian pdf for $x > \mu$, the mean of the pdf. The error function is defined as $\text{erf}(x) = \frac{2}{\sqrt{\pi}} \int_0^x e^{-t^2} dt$.

When transmitting, the jammer increases the noise power spectral density of the noise from N_o to $N_o + N_J/\rho$. The jammer transmits using the duty factor ρ,

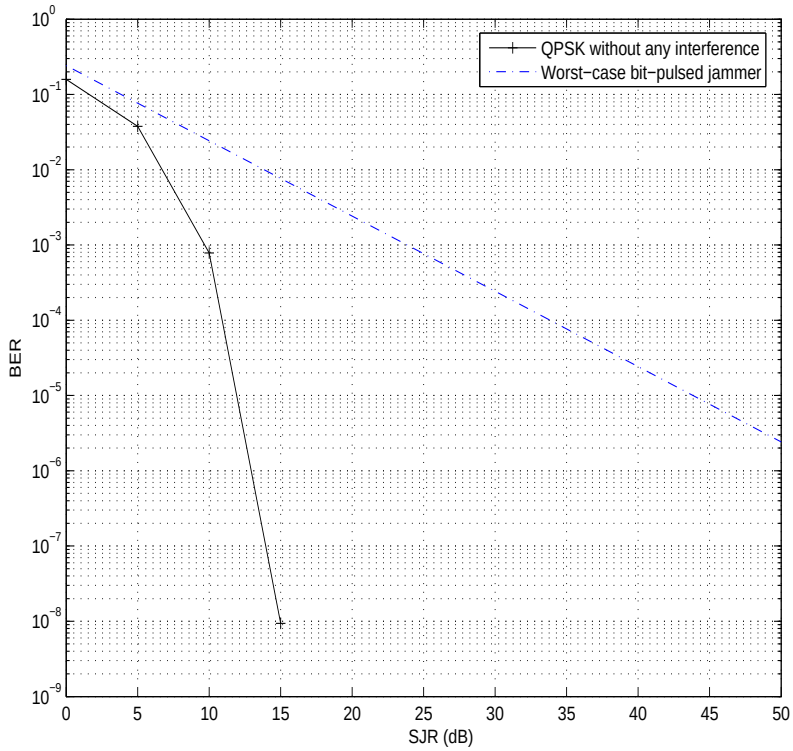


Figure 4.37: *Probability of error for orthogonal binary signals (QPSK). While the lower-bound has an exponential relation with the signal-to-noise ratio, the upper-bound due to a worst case jammer ha s a linear relation with respect to the signal-to-jammer ratio.*

so that the average bit error probability is

$$\overline{P_b} = (1-\rho)Q\left(\sqrt{\gamma_b}\right) + \rho Q\left(\sqrt{\frac{E_b}{N_o + N_J/\rho}}\right). \tag{4.40}$$

The jammer chooses ρ to maximize $\overline{P_b}$. A system that operates in a jamming environment uses the maximum possible transmitter power. Hence the first term

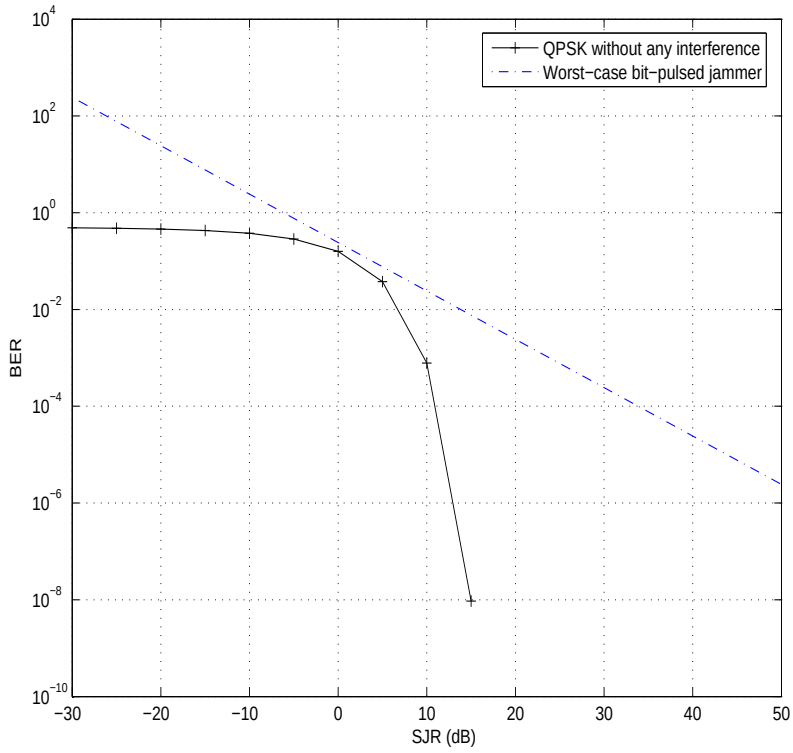


Figure 4.38: *Probability of error for orthogonal binary signals (QPSK). While the lower-bound has an exponential relation with the signal-to-noise ratio, the upper-bound due to a worst case jammer has a linear relation with respect to the signal-to-jammer ratio. This figure uses the SJR range used in the other experiments.*

of (4.40) vanishes [140] and it can be approximated by

$$\overline{P_b} \cong \rho Q\left(\sqrt{\frac{E_b\rho}{N_J}}\right). \tag{4.41}$$

The Q-function can be bounded by an exponential yielding [140]

$$\overline{P_b} \leq \frac{\rho}{\sqrt{4\pi E_b\rho/N_J}} e^{-E_b\rho/N_J}. \tag{4.42}$$

Maximizing (4.42) over ρ leads to $\rho = N_J/2E_b$. Thus the worst case performance is bound by

$$\overline{P}_{b,max} = \frac{1}{\sqrt{2\pi e}} \frac{1}{2E_b/N_J}, \tag{4.43}$$

where E_b/N_J is the signal-to-jammer (SJR) ratio, denoted now by β_b. This upper bound is valid when $\beta_b \geq 0.5$ [140]. While (4.37) represents an exponential dependence of P_b with the SNR, this relation is now inverse-linear in the upper-bound given in (4.43). The loss in performance is plotted in Fig. 4.37, where it can now be observed that an optimized bit pulsed jammer causes a degradation of approximately 10 dB at a BER level of 10^{-2}. This loss is as high as 31.5 dB at a BER of 10^{-5}. Fig. 4.38 depicts this loss for a SJR range from -30 dB to 50 dB.

4.10.2 Performance gain due to ICA

ICA involves the separation of source signals from their mixtures. All other signals in the mixture other than the signal-of-interest, are sources of interference on that specific source. Thus, ICA can be viewed as a procedure that enhances this signal-to-interference ratio. Hence, the improvement in SIR depends on how well the sources are separated. If the source signals are estimated fairly accurately, the SIR improvement is high, for other cases this improvement can be marginal. If the model is as described in (4.1) and the estimated separation matrix is $\mathbf{B}$, then the product of the matrices $\mathbf{AB}$ gives a good idea of the quality of the separation matrix. In the ideal case this is a permutation matrix $\mathbf{P}$ or when the sources are ordered then the identity matrix $\mathbf{I}$.

Quantitatively, the demixer's performance can be measured by the global rejection index [5, 58] defined by:

$$\mathcal{I} = \sum_{i=1}^{m} \left(\sum_{j=1}^{m} \frac{|p_{ij}|}{max_k |p_{ik}|} - 1 \right) + \sum_{j=1}^{m} \left(\sum_{i=1}^{m} \frac{|p_{ij}|}{\max_k |p_{ki}|} - 1 \right), \tag{4.44}$$

where $\mathbf{P} = \{p_{ij}\} = \mathbf{AB}$. This performance index $\mathcal{I}$ measures the difference of $\mathbf{P}$ from a permutation matrix. A permutation matrix is defined so that on its each row and column, one and only one of its elements is 1, while the rest of their elements are zero. If the matrix $\mathbf{P}$ is a perfect permutation matrix, $\mathcal{I} = 0$. The farther the matrix $\mathbf{P}$ is from the permutation matrix, the higher is the value of the performance index $\mathcal{I}$.

While the global rejection index provides a good measure of the separation process, by using the notion of *gain matrices* [182] the gain in SIR can be defined. Assuming a model as in (4.1), with negligible noise ν and, the pre-processed data as defined in Section 4.1.6, the gain matrix can be defined as [182],

$$\mathbf{G} = \mathbf{W}^T \mathbf{C}^{-1/2} \mathbf{A} \mathbf{D}^{1/2}, \tag{4.45}$$

where $\mathbf{W}$ is the orthogonal de-mixing matrix, $\mathbf{C}$ is the covariance matrix of the observed mixtures, $\mathbf{A}$ is the mixing matrix and $\mathbf{D}$ is the diagonal matrix of the variances of the sources. The sources are assumed to be centered. The normalized sources $\mathbf{S}^\dagger$ are connected to the original sources by the following relation,

$$\mathbf{S}^\dagger = \mathbf{D}^{-1/2} \mathbf{S}, \tag{4.46}$$

where the following relations hold

$$\mathbf{D} = \text{diag}[\hat{\sigma_1}^2, \ldots, \hat{\sigma_k}^2], \tag{4.47}$$

$$\hat{\sigma}_k^2 = \mathbf{s}_k \mathbf{s}_k^T / N. \tag{4.48}$$

Thus the gain matrix effectively connects the normalized sources and the estimated sources as,

$$\begin{aligned} \hat{\mathbf{S}}^\dagger &= \mathbf{B}\mathbf{X}_{orig}, \\ &= \mathbf{W}^T \mathbf{V} \mathbf{A} \mathbf{S}, \\ &= \mathbf{W}^T \mathbf{V} \mathbf{A} \mathbf{D}^{1/2} \mathbf{S}^\dagger, \\ &= \mathbf{W}^T \mathbf{C}^{-1/2} \mathbf{A} \mathbf{D}^{1/2} \mathbf{S}^\dagger, \\ &= \mathbf{G}\mathbf{S}^\dagger. \end{aligned} \tag{4.49}$$

Note that if $\mathbf{A} = \mathbf{D} = \mathbf{I}$, then the gain matrix and the de-mixing matrix are one and the same.

The gain matrix quantifies the presence of the kth source signal in the estimated ith source signal. This is now the (i,k)th entry of $\mathbf{G}$ and is denoted as G_{ik}. Then, the total signal-to-interference of the kth source signal is defined as

$$\text{SIR}_k = \frac{E\{\mathbf{G}_{kk}^2\}}{E\{\sum_{k=1, k\neq i}^{K} \mathbf{G}_{ik}^2\}}. \tag{4.50}$$

The gain in SIR can be seen in Fig. 4.39. This gain is calculated experimentally by mixing a source with the desired source. The other source was considered

"interference" and its strength varied, while keeping the desired source strength constant. After estimating the sources with FastICA and the obtained gain in SIR is calculated using (4.50). This is the enhancement of the SIR after processing by FastICA. According to [182] the simulated gain is around 36 dB for a BPSK

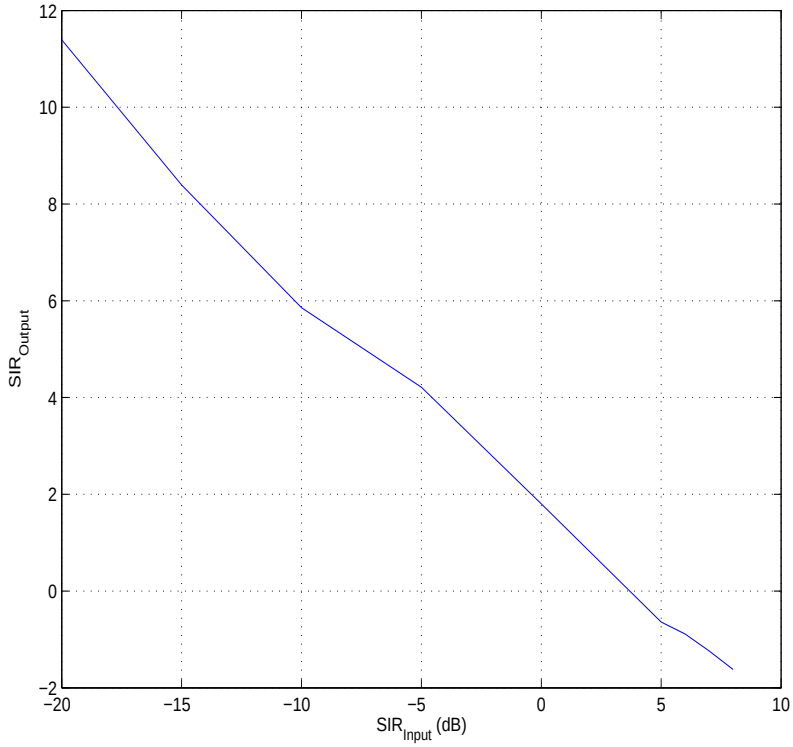


Figure 4.39: *Signal-to-interference (SIR) improvement with FastICA [75]. This experiment involves two mixed sources. The power of the desired source is kept constant while that of the interfering source is varied. The SIR after estimating the sources with FastICA is calculated using* (4.50)*. When the SIR is low, FastICA separates both the source and the interference signals, and gain in SIR is significant. The gain in SIR is not very high when the input SIR is close to* 0 *dB; i.e, all the signals are approximately of the same power. The estimates are an average of* 10 *independent simulations with the FastICA algorithm.*

signal. The Cramér-Rao lower bound (derived in [182]) for these signals is ∞,

and is a case when the gain matrix **G** is close to an identity matrix. In the above figure, it can be seen that the SIR enhancement has an inverse linear relation to the input SIR. When the SIR is very high, the interference signal is nearly absent, in which case, the gain induced by ICA is nearly nothing. This is clearly seen in the above figure. In this case, if ICA manages to extract the "interference", there is a reverse gain induced on the interfering signal.

4.10.3 Performance of the plain ICA-RAKE receiver

Starting from considerations presented in (4.40), the performance of the plain ICA-RAKE receiver can be derived in a semi-analytical manner. This first term of this equation can be interpreted as the performance of an unjammed system, while the second term is the loss in performance due to the jammer. Thus the above equation can be rewritten as

$$\overline{P_b} = (1-\rho)P_{b1} + \rho P_{b2}, \tag{4.51}$$

where, P_{b1} and P_{b2} are $Q(\gamma_b)$ and $Q\left(\sqrt{\frac{E_b}{N_o+N_J/\rho}}\right)$ respectively. P_{b2} is the term that describes the performance loss due to the presence of the jammer. Dividing the ratio in E_b results in,

$$P_{b2} = \rho Q\left(\sqrt{\frac{1}{N_o/E_b + (N_J/E_b)(1/\rho)}}\right). \tag{4.52}$$

Since E_b/N_o represents the signal-to-noise ratio and E_b/N_J is the signal-to-jammer ratio, (4.52) can be written as,

$$P_{b2} = \rho Q\left(\sqrt{\frac{1}{SNR^{-1} + SJR^{-1}/\rho}}\right). \tag{4.53}$$

Since the basic ICA algorithm separates the sources, the quality of separation being assessed by the SIR given by (4.50) is the improvement in the SJR after ICA has been performed. This ICA induced gain can be represented as Γ_{ICA}. Hence, after ICA, the performance is given by

$$\overline{P}_{b2} = \rho Q\left(\sqrt{\frac{1}{\gamma_b^{-1} + \Gamma_{ICA}^{-1}/\rho}}\right), \tag{4.54}$$

where γ_b is the SNR per bit and Γ_{ICA} is the ICA induced gain in SJR per bit. Substituting (4.54) in (4.40), the performance of the plain ICA-RAKE receiver can be derived as,

$$\overline{P_b} = (1-\rho)Q\left(\gamma_b\right) + \rho Q\left(\sqrt{\frac{1}{\gamma_b^{-1} + \Gamma_{ICA}^{-1}/\rho}}\right). \tag{4.55}$$

Fig. 4.40 shows the effect of enhanced SIR on the system performance. With

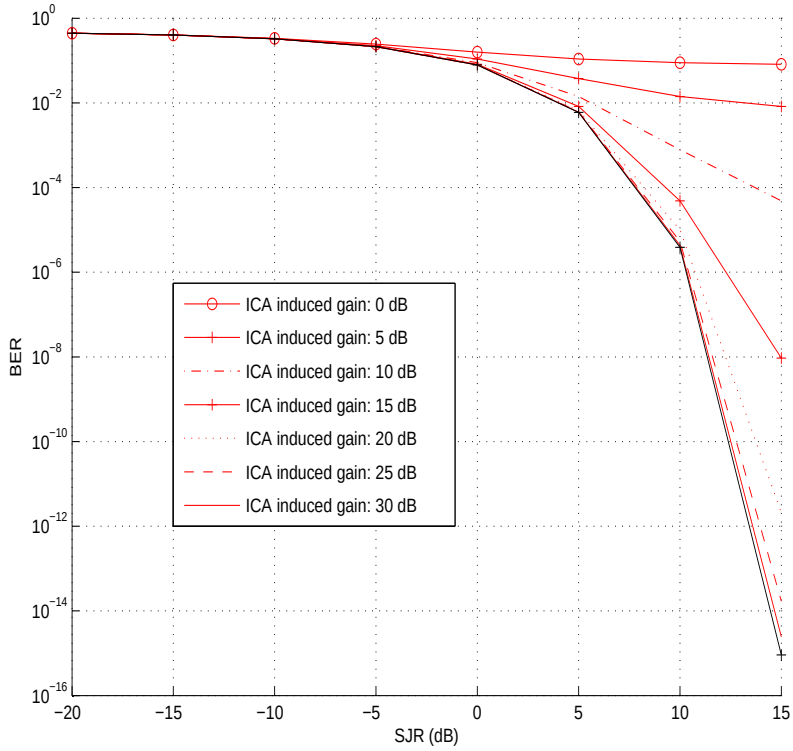


Figure 4.40: *Performance curves for different values of enhanced SIR's. With no enhancements in SIR, the theoretical performance is poor. With* 20*dB enhancement, the influence of the jammer is nearly non-existent.*

no gain, the results are quite poor. An enhancement of around 10 dB already produces a significant improvement in performance. From Fig. 4.39, assuming a

gain of 25 dB, the effect of jammer is nearly non-existent now.

4.10.4 Performance analysis of the switching process

The ICA-RAKE Pre- and Post-Switch, introduced in Sec. 4.6 and Sec. 4.7 rely on estimating the presence of the jammer and switching between the ICA portion and the MRC portion of the receiver. Based on the principles of hypothesis testing [123] for a binary detection problem, the following hypothesis can be postulated,

$$H_o : \text{jammer is absent} \tag{4.56}$$
$$H_1 : \text{jammer is present} \tag{4.57}$$

H_o is called the *null hypothesis* and H_1 is called the *alternative* hypothesis. The job of the switch is now to detect which of these hypothesis is true. For this binary hypothesis test, the decision space consists of two points (do not switch, switch), corresponding to accepting the hypothesis H_o and H_1.

Based on the estimation of the training symbols, a decision is made on the hypothesis to be accepted. Denoting this by P_b^*, when $P_b^* < P_b^\dagger$, the null hypothesis is accepted and no switching takes place, else the alternative hypothesis is accepted and ICA is applied before conventional detection. $P_b^\dagger$ is the threshold.

There is no obvious way to calculate this threshold $P_b^\dagger$. $P_b^\dagger$ is clearly related to the size of the training sequence, and the distributions of the signals in question. Assuming that jamming occurs if 10% of the estimated training symbols are false, this value can be set to $P_b^\dagger = 0.1$. Now the hypothesis can be postulated as[7],

$$H_o : \text{if} P_b^* < 0.1 \tag{4.58}$$
$$H_1 : \text{if} P_b^* \geq 0.1 \tag{4.59}$$

The aim now is to find the corresponding SJR threshold beyond which it is assumed that the jammer is present. Examining the worst-case jamming plot

[7]This threshold is set in a semi-analytical manner, after examining the results of several experiments. A more concrete threshold can be set after a rigorous statistical analysis such as the t-test or the z-test. Also, assuming that training sequences form $2-5\%$ (in all the experiments above) of the entire data, the probability of jamming is determined by examining a very small portion of the actual block.

(left hand plot) in Fig. 4.37, this occurs when the SJR is around 3 dB. Setting this as the threshold the performance of the switching algorithms can be written as,

$$\overline{P}_b = \begin{cases} P_b(\text{MRC}) & \text{SJR} > 3 \text{ dB} , \\ P_b(\text{ICA-RAKE}) & \text{SJR} \leq 3 \text{ dB}. \end{cases} \tag{4.60}$$

So, the ICA-RAKE algorithms switch when the SJR $\leq$ 3dB. For jammer less than this threshold, the ICA-RAKE portion is bypassed there by gaining speed and reducing complexity of the receiver chain. For stronger jammers/sources of interference, the ICA-RAKE chain is involved in pre-processing and separating the jammer, there by improving the estimates to the conventional detector.

Chapter 5

DSS Based Interference Cancellation

Blind detectors of category (C2) (Sec. 3.5) belong to a class that requires no additional assistance information. These receivers know only the spreading waveform of the desired user. Denoising source separation (DSS) is a general framework that combines the separation of non-Gaussian signals and signals with differing time structures. This framework allows for the construction of source separation algorithms around denoising (temporal filtering) functions. This principle is illustrated in Fig. 5.1. The procedure involves iterative denoising and demixing *i.e.*, separation around a denoising function, which is either linear or non-linear. This chapter presents a technique for separating and suppressing the interference based on the above principles.

5.1 A brief overview of DSS

Denoising source separation (DSS) is a novel framework for constructing source separation algorithms with a denoising feature [166]. These denoising algorithms vary by nature of their denoising function. DSS is introduced by illustrating linear separation in the section below (The notations used throughout this section are similar to the ones used in Sec. 4.1 of Chapter 4. The connections between the variables are as laid out in Sec. 4.2 of Chapter 4).

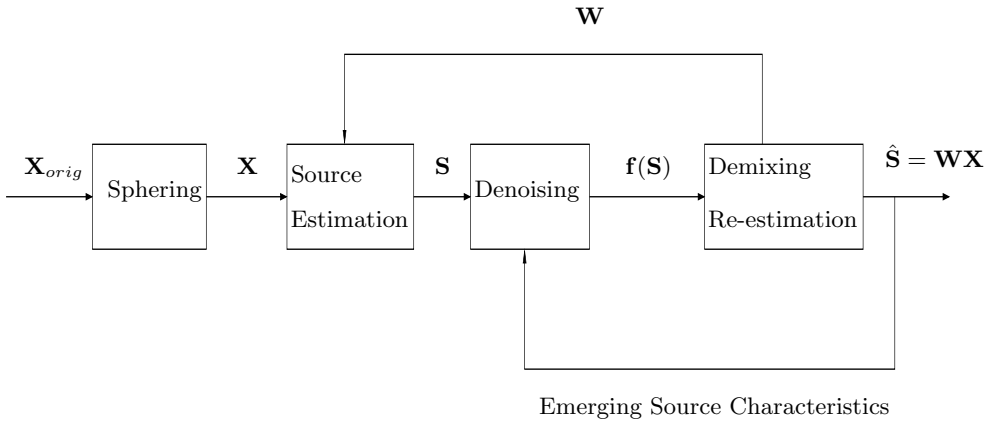


Figure 5.1: *Schematic illustration of Denoising Source Separation (DSS) [166].* $\mathbf{W}$ *is the separating matrix, while* $\mathbf{f}(\mathbf{S})$ *is the denoising function around which separation is done.* $\hat{\mathbf{S}}$ *is the new estimate of the sources.*

5.1.1 One-unit algorithm for source separation

The *expectation-maximization* (EM) algorithm [32] is a method for performing maximum likelihood estimation when a part of the data is missing. For linear mixing models as in (4.1), the source separation problem can be viewed as an EM estimation with the sources missing. The job of the EM algorithm is now to estimate the mixing matrix. The paper [14] describes an EM algorithm for

source separation and in [75] a one-unit version of the same is derived.

Considering a linear mixing model as in (4.1) and assuming that the data $\mathbf{X}$ is pre-whitened by the whitening transformation (refer to subsection 4.1.6), the EM algorithm for separating the components one by one is (Sec. 2.1 of [166] contains a detailed review of this algorithm):

$$\begin{aligned} \mathbf{s} &= \mathbf{w}^T\mathbf{X}, && (5.1)\\ \mathbf{s}^+ &= \mathbf{f}(\mathbf{s}), && (5.2)\\ \mathbf{w}^+ &= \mathbf{X}\mathbf{s}^{+T}, && (5.3)\\ \mathbf{w}_{\text{new}} &= \text{orth}(\mathbf{w}^+). && (5.4) \end{aligned}$$

In this algorithm, the first step (5.1), calculates a noisy estimate of one source. $\mathbf{w}$ is the separating vector as before. The second step (5.2) can be considered as denoising based on the model of the sources. Here, $\mathbf{f}(\mathbf{s})$ is a row-vector-valued function of a row-vector argument. The re-estimation step in (5.3) calculates a new estimate of the separating vector. This maximum likelihood estimate corresponds to the maximization stage in the classical EM algorithm. Step (5.4) is the normalization step, that prevents the norm of the separating vector from diverging. Although this algorithm separates only one component, it has been shown that the original sources correspond to stable fixed points of the algorithm under quite general conditions (see Theorem 8.1 of [75]), provided that the independent source model holds. For extracting multiple sources, this step is an orthonormalization step.

The paper [166] interprets step (5.2) as a denoising step. This interpretation allows for the development of new algorithms that are not derived starting from generative models. The steps (5.1) — (5.4) form the core of a denoising source separation (DSS) algorithm.

5.1.2 Linear denoising

Considering sources with Gaussian distributions, a denoising procedure is now derived using the Bayes principle [136]. A linear denoising function can be derived by examining the posterior distribution of the sources. This distribution is derived by multiplying the likelihood of the mixture with the Gaussian prior.

Consider a Gaussian source with a autocovariance matrix of $\mathbf{\Sigma_{ss}}$, its probability

density function is given by

$$p(\mathbf{s}) = \frac{1}{\sqrt{(2\pi)^K |\mathbf{\Sigma_{ss}}|}} \exp\left(-\frac{1}{2}\mathbf{s}\mathbf{\Sigma}_{\mathbf{ss}}^{-1}\mathbf{s}^T\right), \tag{5.5}$$

where $|\mathbf{\Sigma_{ss}}|$ denotes the determinant. Furthermore, the likelihood of the unnormalized Gaussian with a diagonal cross-correlation $\mathbf{\Sigma_{s,x}}$ is [166]

$$L(\mathbf{s}) = \exp\left(-\frac{1}{2}(\mathbf{s} - \mathbf{w}^T\mathbf{X})\mathbf{\Sigma}_{\mathbf{s,x}}^{-1}(\mathbf{s} - \mathbf{w}^T\mathbf{X})^T\right). \tag{5.6}$$

The posterior can then be written as:

$$q(\mathbf{s}) = \frac{1}{\sqrt{(2\pi)^K |\mathbf{\Sigma}|}} \exp\left(-\frac{1}{2}(\mathbf{s} - \mu)\mathbf{\Sigma}^{-1}(\mathbf{s} - \mu)^T\right), \tag{5.7}$$

with the mean $\mu = \mathbf{w}^T\mathbf{X}(\mathbf{I} + \sigma_x^2\mathbf{\Sigma}_{\mathbf{ss}}^{-1})^{-1}$, and the covariance matrix $\mathbf{\Sigma}^{-1} = \frac{1}{\sigma_x^2} + \mathbf{\Sigma}_{\mathbf{ss}}^{-1}$. Hence, the denoising step in (5.2) becomes,

$$\mathbf{s}^+ = \mathbf{f}(\mathbf{s}) = \mathbf{s}(\mathbf{I} + \sigma_x^2\mathbf{\Sigma}_{\mathbf{ss}}^{-1})^{-1} = \mathbf{s}\mathbf{D}, \tag{5.8}$$

which corresponds to linear denoising. The denoising step in the DSS algorithm $\mathbf{s}^+ = \mathbf{f}(\mathbf{s})$ is thus equivalent to multiplying the current source estimate $\mathbf{s}$ with a constant matrix $\mathbf{D}$.

To analyze the denoising matrix $\mathbf{D}$ further by considering its eigenvalue decomposition, it is seen that $\mathbf{D}$ and $\mathbf{\Sigma_{ss}}$ have the same eigenvectors and eigenvalue decompositions [166],

$$\mathbf{\Sigma}_{ss} = \mathbf{V}\mathbf{\Lambda}_\Sigma\mathbf{V}^T \tag{5.9}$$

$$\mathbf{D} = \mathbf{V}\mathbf{\Lambda}_D\mathbf{V}^T, \tag{5.10}$$

where, $\mathbf{V}$ is an orthonormal matrix with eigenvectors as the columns and $\mathbf{\Lambda}_D$ is a diagonal matrix with the following eigenvalues on the diagonal,

$$\lambda_{D,i} = \frac{1}{1 + \sigma_x^2/\lambda_{\Sigma,i}}, \tag{5.11}$$

with $\lambda_{\Sigma,i}$ is the i-th largest eigenvalue of $\mathbf{\Lambda}_\Sigma$.

Note that $\lambda_{D,i}$ is a monotonically increasing function of $\lambda_{\Sigma,i}$. Directions of $\mathbf{s}$ that have the smallest variances according to the prior model of $\mathbf{s}$ are suppressed the most.

Packing the different phases: (5.1), (5.8) and (5.3) together, yields

$$\mathbf{w}^+ = \mathbf{X}\mathbf{s}^{+T} = \mathbf{X}\mathbf{D}\mathbf{s}^T = \mathbf{X}\mathbf{D}\mathbf{X}^T\mathbf{w}. \tag{5.12}$$

Here the transpose is dropped from the above equation as $\mathbf{D}$ is symmetric. By writing $\mathbf{\Lambda}_D = \mathbf{\Lambda}_D^{1/2}\left(\mathbf{\Lambda}_D^{1/2}\right)^T = \mathbf{\Lambda}^*\mathbf{\Lambda}^{*T}$ and noting that $\mathbf{V}^T\mathbf{V} = \mathbf{I}$, the denoising matrix can be split as:

$$\mathbf{D} = \mathbf{D}^*\mathbf{D}^{*T}, \tag{5.13}$$

where $\mathbf{D}^* = \mathbf{V}\mathbf{\Lambda}^*\mathbf{V}^T$. Further, denoting $\mathbf{Z} = \mathbf{X}\mathbf{D}^*$, the DSS algorithm for estimating one separating vector is of the form:

$$\mathbf{w}^+ = \mathbf{Z}\mathbf{Z}^T\mathbf{w}. \tag{5.14}$$

This is the classical *power method* [204] implementation for principal component analysis (PCA). Note that in (5.14), $\mathbf{Z}\mathbf{Z}^T$ is the unnormalized covariance matrix. This algorithm converges to the *fixed point* $\mathbf{w}^*$ satisfying,

$$\lambda\mathbf{w}^* = \left(\mathbf{Z}\mathbf{Z}^T/T\right)\mathbf{w}^*, \tag{5.15}$$

where λ corresponds to the principal eigenvalue of the covariance matrix $\mathbf{Z}\mathbf{Z}^T/T$ and $\mathbf{w}^*$ is the principal direction. The asterisk — $*$ — is used to emphasize that the estimate is at the fixed point.

Fig. 5.2 shows the effect of the linear DSS algorithm on a set of mixed sources. Figure 5.2a shows the scatter plot of two mixed signals. The directions of the columns of the mixing matrix are plotted by the dashed lines. The curve shows the standard deviation of the data projected onto different directions. The principal eigenvector, depicted by the solid line does not separate any of the sources. For separation, two things are required: (a) the mixing vectors should be made orthogonal and (b) the eigenvalues should differ. After sphering in Fig. 5.2b, the basis and the sphered mixing vectors are roughly orthogonal. However, any unit-length projection yields unit variance, and PCA cannot separate the sources. Fig. 5.2c shows the effect of low-pass filtering. Now the principal component gives the approximate direction of the original source signal. The sources are now recovered by $\mathbf{s} = \mathbf{w}^T\mathbf{X}$. Fig. 5.2d shows one of the original sources, a noisy mixed signal and the denoised estimate corresponding to the source.

5.1.3 Denoising functions in practice

DSS is a general framework for designing source separation algorithms. The idea is that the denoising function $\mathbf{f}(\mathbf{s})$ differentiate the algorithms, while the core

remains the same. Denoising is practical and useful in real world applications. It has been adequately addressed in [6]. Usually, some prior knowledge about the signals and the applications is all that is required to perform denoising. Even in the case of signals estimated via a blind algorithm, *e.g.*, ICA, the estimates require significant denoising [196]. In DSS, these denoising methods constructed with some prior knowledge can be directly applied to source separation. This invariably leads to better results than pure BSS [166]. The denoising functions in practice can range from simple but powerful linear ones to sophisticated non-linear ones. These functions can be used in both cases: when some knowledge

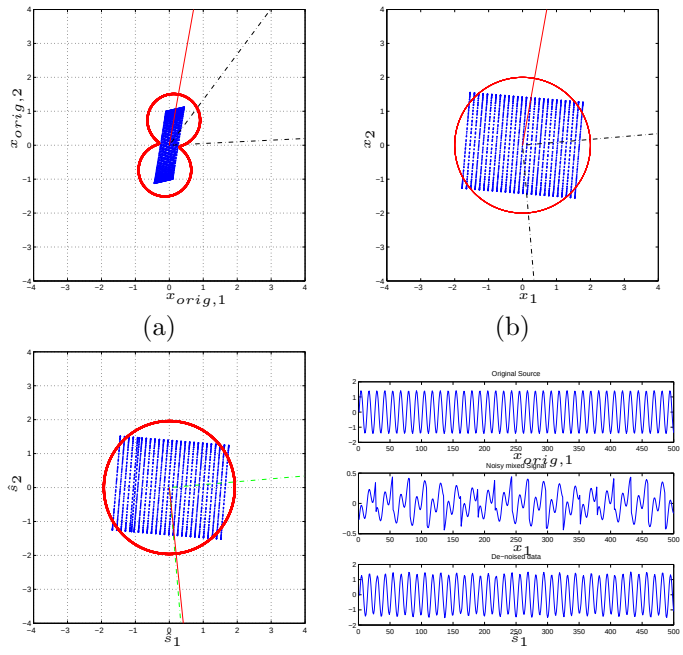


Figure 5.2: *Effect of a denoising procedure from [166]. (a) Scatter plot of two mixed signals with the solid lines denoting the principal component and dotted lines denoting mixing coefficients of the original signals. The circular boundaries represent the standard deviation of the projection of the data in different directions. (b) Data after sphering. (c) Data after low-pass filtering. Now, the principal component gives the approximate direction of the original source signal. (d) The original source, the mixture, and the estimated source after low-pass filtering.*

about the signals are known or when the separation is absolutely blind.

It should be noted, though, that DSS is not very useful if the main objective of the algorithm is to remove noise precisely rather than to separate the sources [166]. Fortunately, this is usually not the case, and it is enough for the denoising function $\mathbf{f}(\mathbf{s})$ to remove a significant portion of the noise (*ie.,* remove more noise than the signal) [75], assuming that the independence of the sources holds. This is because the estimation steps in (5.3) and (5.4) constrain the source $\mathbf{s}$ to the subspace spanned by the data. Even if the denoising discards parts of the signal or creates nonexistent signals, then subsequent re-estimation steps restore them [166].

If no detailed information about the characteristics of the signal is available, it is possible to start from a relatively general function and then tune this function by analyzing the structure of the noisy signals extracted in the first phase. This is called *bootstrapping*. In fact, some of the non-linear DSS algorithms can be interpreted as linear DSS algorithms where a linear denoising function is adapted to the sources, leading to non-linear denoising.

Linear denoising functions

This section considers several simple but powerful linear denoising schemes. Examining the EVD of the denoising matrix $\mathbf{D}$ in (5.10) shows that any denoising in linear DSS can be implemented as an orthonormal rotation in the $T-$dimensional space of signals, followed by point-wise scaling of the sample and rotation back to the original space. Thus EVD also offers good intuitive insight into the denoising function as well as practical means for its implementation.

On-Off Denoising. On-Off denoising refers to signals that have periods of activity and periods of non-activity *e.g.*, in the fields of psychophysics or biomedicine. Then denoising can be simply implemented by

$$\mathbf{D} = \mathrm{diag}(m_1, \ldots, m_T), \tag{5.16}$$

where $\mathbf{D}$ refers to the linear denoising matrix and,

$$m_i = \begin{cases} 1, & \text{for the active times } i, \\ 0, & \text{for the inactive times } i. \end{cases} \tag{5.17}$$

This amounts to multiplying the source estimate by a binary mask, where ones represent active parts and zeros non-active parts[1]. Masking usually satisfies the relation $\mathbf{D} = \mathbf{D}^*\mathbf{D}^{*T}$. This means that DSS is equivalent to the PCA applied to the denoised $\mathbf{Z} = \mathbf{XD}$ even with exactly the same filtering. In practice, this can be implemented by PCA applied to the active parts of the data with the sphering stage involving the whole data.

Denoising based on frequency content. If, on the other hand, the signals are characterized by having certain frequency components, the signals can be transformed to the frequency space, then the spectrum can be masked, *e.g.*, with a binary mask, and transformed back to the original space to denoise the signal,

$$\mathbf{D} = \mathbf{V}\mathbf{\Lambda}_D\mathbf{V}^T, \tag{5.18}$$

where $\mathbf{V}$ is the transform, $\mathbf{\Lambda}_D$ is the mask on its diagonal and $\mathbf{V}^T$ is the inverse transform. In practice, $\mathbf{V}$ can be implemented by Fourier transform or *discrete cosine transform* (DCT)[2]. After the transform, the signal can be filtered using the diagonal matrix $\mathbf{\Lambda}_D$, *i.e.*, by point-wise scaling of the frequency bins. Finally, the signal can be inverse transformed using $\mathbf{V}^T$. In the case of *linear time-invariant* (LTI) filtering, the filtering matrix has a Toeplitz structure and the denoising characteristics are manifested only in the diagonal matrix $\mathbf{\Lambda}_D$, while the transforming matrix $\mathbf{V}$ represents rotation. Under such cases, the algorithm can be further simplified by imposing the transformation on the sphered data $\mathbf{X}$. Then the iteration can be performed on the transformed basis.

Other functions. Other functions can be constructed by looking at the time-frequency plot of the signals. This is called *spectrogram denoising*. When the signal $\mathbf{s}$ has a repetitive structure, and the average repetition rate is known, denoising can be based on the *quasi-periodicity* of the signal. The paper [166] considers these cases in a detailed manner.

[1] This is a point-wise multiplication of a signal or transformation of a signal.

[2] Fourier transform as opposed to EVD is a complex transformation. Real Fourier transform can be considered where the sine and cosine terms have been separated in different elements. This makes it simple.

Denoising based on estimating signal value

In the linear denoising schemes the data was assumed to be unknown. It is also possible to estimate the denoising specifications from the data itself. This makes the denoising non-linear or adaptive.

Thresholding denoising. A form closely related to the on-off denoising is the *thresholding* operation. This implies that signal values below a constant are damped while signals above the constant are retained. This can be implemented via (5.16) and using,

$$m_i = \begin{cases} 1, & \text{if} \quad s(i) \geq \alpha, \\ 0, & \text{if} \quad s(i) < \alpha, \end{cases} \tag{5.19}$$

where α is the thresholding constant. If the signal is known to have binary values and have equiprobable $(0, 1)$ states[3] then, $\alpha = 0.5$ is a natural choice for α . This is also the case when the signals are known to have a super-Gaussian distribution (relatively small values but many zeros). Thresholding is closely related to the so-called "shrinkage denoising" principle [74] [4].

Denoising based on estimating signal variance

The variance of the signal σ_s^2 offers a naïve estimate for the signal power. For zero mean signals, this estimate is approximate. Estimating the signal variance and the noise variance leads to better separation of the signal and the noise. Several denoising functions can be constructed using variance estimation and refinement as an objective. Some of them are examined below.

Kurtosis-based DSS. Signal separation (Sec. 4.1) can be achieved by optimization of the estimated kurtosis of the sources. There the objective function

[3] BPSK signals tend to have a bi-modal distribution, hence values $0.5 \pm \delta$ are mostly due to noise.

[4] Shrinkage denoising is seemingly similar to null-zone decision function for interference cancellation. Though it might appear that they both are same, shrinkage denoising is a Bayesian techniques and is not necessarily restricted to the region around the null zone.

is

$$g(\mathbf{s}) = \sum_{t=1}^{T} s^4(t)/T - 3(\sum_{t=1}^{T} s^2(t)/T)^2. \tag{5.20}$$

Since the source variance is normalized to unity, this can be simplified as $g(\mathbf{s}) = \sum_{t=1}^{T} s^4(t)/T$, and the denoising function can be derived by gradient ascent. This yields $\Delta_{\mathbf{s}} g(\mathbf{s}) = (4/T)\mathbf{s}^3$, where $\mathbf{s}^3 = [s^3(1) \quad s^3(2) \quad \dots]$. Hence, the denoising function can be approximated as

$$\mathbf{f}(\mathbf{s}) = \mathbf{s}^3. \tag{5.21}$$

This is the ICA algorithm with non-linear denoising. On examining (5.21), $\mathbf{s}^3$ can be interpreted as being $\mathbf{s}$ masked by $\mathbf{s}^2$, the later being a somewhat naïve estimate for the signal power.

Refining variance estimates. Kurtosis can be regarded as a naïve or approximate estimate for the square of the variance of the signal. Assuming that $\mathbf{s}$ is composed of Gaussian noise with a constant variance σ_n^2, and of a Gaussian signal with non-stationary variance $\sigma_s^2(t)$, the denoising step can be formulated as,

$$s^+(t) = s(t)\frac{\sigma_s^2(t)}{\sigma_{tot}^2(t)}, \tag{5.22}$$

where $\sigma_{tot}^2(t) = \sigma_s^2(t) + \sigma_n^2$ is the total variance of the observation. This is also the *maximum-a-posteriori* (MAP) estimate [166].

The kurtosis-based DSS in (5.21) can be obtained from this MAP estimate if the signal variance is assumed to be far smaller than the total variance. In that case it is reasonable to assume σ_{tot}^2 to be constant and the signal variance $\sigma_s^2(t)$ can be estimated by $s^2(t) - \sigma_n^2$. Several methods for refining the estimate are shown in [166] and [187].

Tanh non-linearity based estimates. A popular replacement of the kurtosis-based non-linearity (5.21) is the hyperbolic tangent $\tanh(\mathbf{s})$, operating point-wise on the sources. This non-linearity is generally considered to be robust to outliers. The denoising function can be written as

$$\mathbf{f}_t(\mathbf{s}) = s(t) - \tanh[s(t)] = s(t)(1 - \frac{\tanh[s(t)]}{s(t)}). \tag{5.23}$$

Now the term $1 - \frac{\tanh[s(t)]}{s(t)}$ can be viewed as a mask that is related to the SNR. Unlike the naïve $s^2(t)$ mask from kurtosis, this mask saturates more slowly.

Other denoising functions. There are usually cases where the system itself suggests some denoising schemes. One such example is the case of CDMA signals to be considered in Sec. 5.2. Another example is source separation with a microphone array combined with speech recognition.

On-line *i.e.*, real-time source separation is often required. On-line sphering algorithms [38, 129] aid in the development of real-time DSS. One simple case of on-line denoising is presented by moving-average filters. However, these on-line filters are typically non-symmetric and thus have no definite objective functions [6], resulting in potentially unstable DSS. The eigenvalues of the denoising matrix $\mathbf{D}$ may be complex numbers. The resulting DSS algorithm converges in the 2-D space corresponding to the eigenvalues with the largest absolute magnitude, but fails to converge within the subspace.

It is also possible to combine various denoising functions when the sources are characterized by more than one type of structure. The combination order might be crucial, due to the fact that, in general $\mathbf{f}_i(\mathbf{f}_j(s)) \neq \mathbf{f}_j(\mathbf{f}_i(s))$, where $\mathbf{f}_i, \mathbf{f}_j$ represents two different denoising functions, either linear or non-linear. For example, the combination of the on-off mask (5.16) and (5.17) and the non-linear variance-based mask (5.22), is helpful only when the on-off masking is performed after the non-linear denoising.

5.1.4 Convergence of DSS algorithm

Given that DSS is a framework for source separation algorithms, the convergence of these algorithms has to be analyzed. The linear DSS algorithm is derived from the power method and hence the convergence properties of the power method [204] apply here. This analysis can be extended to the non-linear case, assuming that the mixing model holds and there are infinite data.

In the linear case, DSS is equivalent to the power method. The convergence is governed by the eigenvalues λ_i corresponding to the fixed points $\mathbf{w}_i^*$. If some of the eigenvalues are equal ($\lambda_i = \lambda_j, i \neq j$), the fixed points degenerate and there are subspaces of fixed points. In any case, it is possible to choose an orthonormal basis spanned by $\mathbf{w}_i^*$. This implies that any $\mathbf{w}$ can be expressed as

$$\mathbf{w} = \sum_i c_i \mathbf{w}_i^*, \tag{5.24}$$

where $c_i = \mathbf{w}^T \mathbf{w}_i^*$. With the linear denoising function $\mathbf{f}_{lin}$, the unnormalized

estimate of $\mathbf{w}^+$ is

$$\begin{aligned}
\mathbf{w}^+ &= \mathbf{X}\mathbf{s}^{+T}, \\
&= \mathbf{X}\mathbf{f}_{lin}^T(\mathbf{w}^T\mathbf{X}), \\
&= \mathbf{X}\mathbf{f}_{lin}^T(\sum_i c_i \mathbf{w}_i^{*T}\mathbf{X}), \\
&= \mathbf{X}\sum_i c_i \mathbf{f}_{lin}^T(\mathbf{w}_i^{*T}\mathbf{X}), \\
&= \mathbf{X}\sum_i c_i \mathbf{f}_{lin}^T(\mathbf{s}_i^*), \\
&= \sum_i c_i \mathbf{X}\mathbf{f}_{lin}^T(\mathbf{s}_i^*), \\
&= \sum_i c_i \lambda_i \mathbf{w}_i^*,
\end{aligned} \tag{5.25}$$

where λ_i is the ith eigenvalue corresponding to $\mathbf{w}_i^*$ and $\mathbf{s}_i^* = \mathbf{w}_i^{*T}\mathbf{X}$. The normalization step in (5.4) changes the contribution to the fixed points by equal fractions. After n iterations, the relative contributions of the fixed points thus change from $\frac{c_i}{c_j}$ to $\frac{c_i\lambda_i^n}{c_j\lambda_j^n}$ [166].

Linear DSS cannot separate sources if fixed points have equal eigenvalues *i.e.*, $\mathbf{w}_i^*$ and $\mathbf{w}_j^*$ have identical eigenvalues $\lambda_i = \lambda_j$. Hence, Gaussian sources having identical autocovariance matrices $\Sigma_{\mathbf{s}_i\mathbf{s}_i} = \Sigma_{\mathbf{s}_j\mathbf{s}_j}$ cannot be separated. In other words, the sources should have distinct time structures. As long as $c_i \neq 0$, DSS should converge globally to the source having the largest eigenvalue.

The speed of the convergence in the linear case depends linearly on the log-ratio of the largest (absolute) eigenvalues $\log(|\lambda_1|/|\lambda_2|)$ where $|\lambda_1| \geq |\lambda_2| \geq |\lambda_i|, i = 3, \ldots N$. The absolute values have been used since negative eigenvalues may exist *e.g.*, in complex sources.

The above analysis for linear denoising makes no assumptions about the data generating process. As such it does not extend to the non-linear case. Here, the number of fixed points can be larger or smaller than the dimensionality of the data, and these fixed points $\mathbf{w}_i^*$ need not be orthogonal, in general. Hence, it is assumed that the data are generated by sources that are independent, and that the mixing vectors are orthogonal after whitening. With these assumptions, the orthonormal basis spanned by the mixing vectors corresponds to the fixed points.

This is because the following condition holds for independent sources $\mathbf{s}_i$ [166]:

$$\lim_{T\to\infty} \frac{1}{T} \sum_{t=1}^{T} s_j(t) f_j(s_i(t)) = 0, \quad \text{when} \quad i \neq j. \tag{5.26}$$

In the linear power method, eigenvalues λ_i govern the rate of relative changes of the contributions of individual basis vectors in the estimate. Here, *local eigenvalues* $\lambda_i(\mathbf{s})$ [166] govern the rate of convergence. These local eigenvalues play a similar role as their linear counterparts, but unlike them the value of the local eigenvalues depends on the current source estimate. Formally, they can be defined as follows: Assume that the current weight vector and the subsequent unnormalized new weight vector are as follows:

$$\mathbf{w} = \textstyle\sum_i c_i(\mathbf{s}) \mathbf{w}_i^*, \tag{5.27}$$

$$\mathbf{w}^+ = \textstyle\sum_i \gamma_i(\mathbf{s}) \mathbf{w}_i^*. \tag{5.28}$$

The local eigenvalue can be defined to be the relative change in the contribution,

$$\gamma_i(\mathbf{s}) = T c_i(\mathbf{s}) \lambda_i(\mathbf{s}) \Leftrightarrow \lambda_i(\mathbf{s}) = \frac{\gamma_i(\mathbf{s})}{T c_i(\mathbf{s})}. \tag{5.29}$$

The idea of the DSS framework is that the user can tailor the denoising function to the task at hand. The purpose of defining local eigenvalues is then to draw attention to the factors influencing separation quality and convergence speed.

Practically, the important issue to consider is whether the algorithm converges at all. Non-linear denoising can be considered as constantly adapting linear denoising. This implies that different sources can have locally the largest eigenvalue. If the adaptation is consistent, *i.e.*, $\lambda_i(\mathbf{s})$ grows monotonically with c_i, all stable fixed points correspond to the original sources. In general, the best separation quality and the fastest convergence is achieved when $\lambda_i(\mathbf{s})$ is very large compared to $\lambda_j(\mathbf{s})$ with $j \neq i$ in the vicinity of $\mathbf{s}_i^*$ [166].

Sometimes it may be sufficient to separate a signal subspace. Then it is enough for the denoising function to make the eigenvalues corresponding to this subspace large compared to the rest but the eigenvalues do not need to differ within the subspace.

If the linear mixing model as in (4.1) holds and there is an infinite amount of data, the sources can usually be separated even in the linear case, because small differences in the eigenvalues of the sources are sufficient for separation. In practice though, data is often insufficient and the model assumptions do not strictly

hold. Conceptually, it can be thought that finite sample size introduces noise to the eigenvalues and leakage between mixing vectors [166]. Hence separation quality in practice is better if the local eigenvalues differ significantly around the fixed points and this is often easiest to achieve with non-linear denoising which utilizes a lot of prior information.

5.2 DSS based interference cancellation

Based on the principles of denoising source separation described in the previous section, the *IC-DSS* receiver - Interference Canceling DSS receiver is now introduced. The relation between the linear mixing model in (4.1) and the interference models in (2.21), (2.34), (2.38), described in Sec. 4.2 holds.

The IC-DSS algorithm belongs to the category (C2). The knowledge the algorithm has is the spreading code of the desired user k, and the modulation scheme. Based on the information regarding the code and the modulation scheme, the algorithm estimates the user's symbols and the mixing matrix, which can either hold the estimate of the channel or the estimate of the array steering vector $\boldsymbol{\theta}$, depending on the interference model (It is assumed that the signal is whitened and the matrix $\mathbf{Y}$ contains the signals for all the elements of the antenna array, *ie.,* there are N_a rows in $\mathbf{Y}$.). Each iteration of the DSS receiver on the received signal $\mathbf{y}$ consists of the following steps

(1) Initialize a random separating vector	:	$\mathbf{w}_k,$
(2) Separate the sources	:	$\mathbf{x}_k = \mathbf{w}_k^T \mathbf{Y},$
(3) Denoising phase	:	$\hat{\mathbf{b}}_k = \mathbf{f}(\mathbf{x}_k),$
(4) Reconstruction	:	$\hat{\mathbf{x}}_k = \mathbf{s}_k \hat{\mathbf{b}}_k,$
(5) Update phase	:	$\mathbf{w}_k^+ = \mathbf{Y}\hat{\mathbf{x}}_k,$
		$\mathbf{w}_k = \frac{\mathbf{w}_k^+}{\|\mathbf{w}_k^+\|}.$

Step 3 is the denoising phase. This is a two step process here: linear denoising consists of spreading the signal, and non-linear denoising is performed by thresholding the values using the sign function. This non-linearity is based on the modulation scheme, and restricts the possible set of values the symbols can

have. This step can be described as

$$f(\mathbf{x}_k) = \text{sign}(\mathbf{s}_k^T \mathbf{x}_k). \tag{5.30}$$

The schematic diagram is shown in Fig. 5.3. The present structure assumes an

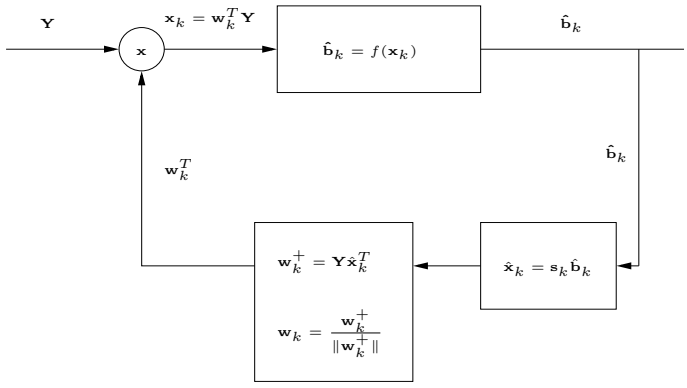


Figure 5.3: *Basic DSS scheme. Denoising is performed by* $f(\mathbf{x}_k)$ *which involves despreading and thresholding. Estimating the mixing structure from* $\mathbf{w}$ *is equivalent to estimating the channel for user* k*.*

antenna array and the model in (2.34) or (2.38). IC-DSS can also be applied to a data model that does not consider antenna arrays. In this case, IC-DSS estimates with minimal changes in addition to the user k the corresponding channel for the user. The channel estimation is blind and is embedded into the process.

The pseudo code for implementing the IC-DSS algorithm for complex signals is provided in **Algorithm 4**.

The algorithm is implemented as a two phase procedure. The initial phase - called *pre-conditioning or burn-in* - occurs when the error Δ is high. During this phase, a noisy estimate for $\hat{b}_k$ is obtained. Here, the normalization step can be replaced by another non-linearity *e.g.*, $\tanh(\hat{b}_k)$, if some additional information about the source distribution is known. This non-linearity is more robust to spiky noise. Replacing the non-linearity by $\hat{b}_k^3$ results in the independent component analysis (ICA) algorithm based on kurtosis. The **for**, **end for** portion is the burn-in phase in the algorithm.

Algorithm 4 IC-DSS Receiver

$w_k \Leftarrow \texttt{rand}(N_a, 1)$
$\hat{w}_{kold} \Leftarrow \texttt{rand}(N_a, 1)$
$\hat{b}_k \Leftarrow \texttt{rand}(1, M_c)$
$N_{iters} \Leftarrow 10$
for $i = 1$ to N_{iters} **do**
  $\hat{x}_k \Leftarrow s_k * \hat{b}_k$ {# — Burn-in phase — #}
  $w_k^+ \Leftarrow \hat{x}_k * y^T$
  $x_k \Leftarrow w_k * y$
  $\hat{b}_k \Leftarrow s_k^T * x_k / C$
  $\hat{b}_k \Leftarrow \hat{b}_k / \texttt{norm}(\hat{b_k}) * \sqrt{T} + j * \texttt{sign}(\hat{b}_k)$
  $\hat{b_k} \Leftarrow \texttt{sign}(\hat{b}_k)$
end for
$\epsilon \Leftarrow 10^{-6}$
$\Delta = \texttt{norm}(w_k - w_{kold}) + \epsilon$
while $\Delta \geq \epsilon$ **do**
  $w_{kold} \Leftarrow w_k$ {# — DSS phase — #}
  $\hat{x}_k \Leftarrow s_k * \hat{b}_k$
  $w_k^+ \Leftarrow y * \hat{x}_k^T$
  $w_k = w_k^+ / \texttt{norm}(w_k^+)$
  $x_k \Leftarrow w_k^T * y$
  $\hat{b}_k \Leftarrow s_k^T * x_k / C$
  $\hat{b}_k \Leftarrow \Delta * \hat{b}_k / \texttt{norm}(\hat{b_k}) + (1 - \Delta) * \texttt{sign}(\hat{b}_k))$
  $\Delta = \texttt{norm}(w_k - w_{kold})$
end while

The second phase is the *non-linear DSS* phase. Now, the error Δ is low, and effect of the normalization is minimal. In this phase, the noisy estimate is refined. It is possible to perform IC-DSS without the pre-conditioning, however, then convergence is longer.

5.3 Examples of interference cancellation

Examples of interference cancellation with the IC-DSS receiver are based on **Scenario 9**. The interference structure is a bit-pulsed jammer with a frequency offset. The block size is assumed to be $M = 5000$ symbols.

In Fig. 5.4, the effect of interference cancellation is seen. The IC-DSS method is better than MRC up to about 5 dB SJR. Another benefit of IC-DSS is its ability to suppress multipath interference. While MRC receivers perform poorly in this respect, IC-DSS is relatively unaffected. Fig. 5.5 compares IC-DSS with the Post-Switch described in Chapter 4. The Post-Switch detector is better than IC-DSS detector. This is because Post-Switch receivers obtain channel information using training symbols. IC-DSS on the other hand, does not rely on any training sequence, the only information is the spreading code. As the number of paths increases, Post-Switch's performance deteriorates while IC-DSS copes well.

The distributions in Fig. 5.6 provide another view of the above results. For a single path channel, IC-DSS is slightly inferior to Post-Switch, because there are blocks to the left of the peak. Such blocks are not present in the Post-Switch ICA-RAKE case. As the number of paths increase, the drop in performance is not drastic. Even with $L = 3$ paths, the distribution is skewed to the right. While Post-Switch has similar characteristics, the variance is far higher in this case. MRC on the other hand has significant degradation as the number of paths increase. As in the case of Post-Switch, the block size affects the performance of the IC-DSS receiver.

Both these methods are subspace based and are dependent on the quality of the covariance matrix. This effect is seen in Fig. 5.7. The BLER as a function of block size can be observed here. From the figures, a block size of $M = 1000$ seems appropriate for the scenario concerned.

Fig. 5.8 shows the processing time as a measure of computational complexity. MRC does not depend on the block size, which is evident from the figures. IC-

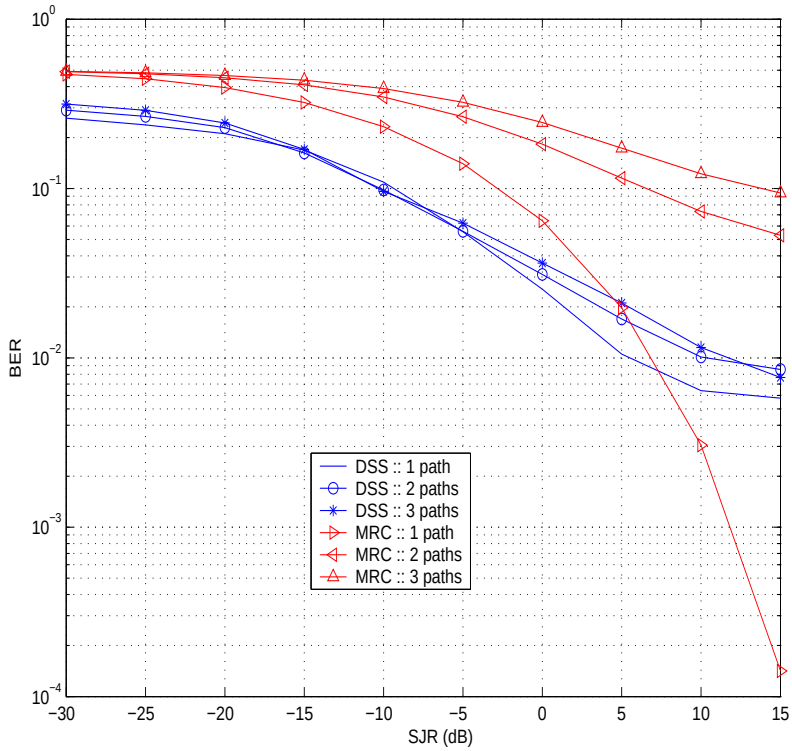


Figure 5.4: *The effect of interference cancellation with IC-DSS detection. The interference is a bit-pulsed jammer as in* **Scenario 9**. *Multiple paths seem to affect IC-DSS the least.*

DSS seems to have a linear complexity. The complexity of Post-Switch increases roughly linearly, too, but at a much smaller rate than that of IC-DSS[5]. With a block size of $M \approx 1500$, the complexities of both the methods are comparable. For lower block sizes (< 1000), IC-DSS is less complex than Post-Switch, with the cut-off around a block size of $M = 1000$.

[5]The convergence requirements for the DSS algorithm were stricter than the Post-Switch ICA-RAKE algorithm. DSS algorithm stopped when the change in weight $\mathbf{w}_k$, was less than $\epsilon = 10^{-6}$, while the Post-Switch did not enforce such conditions, and stopped after a fixed number of iterations

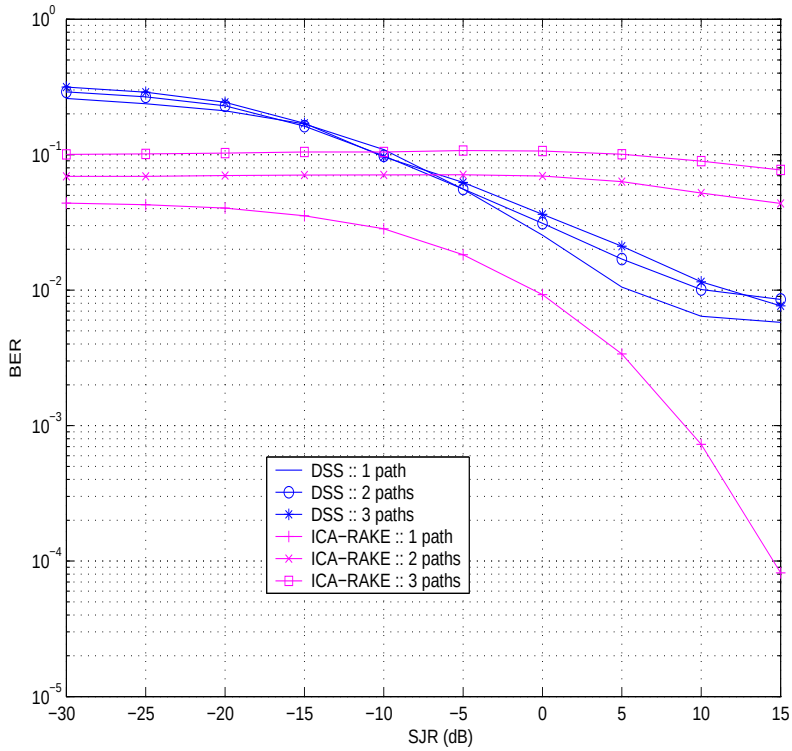


Figure 5.5: *The effect of interference cancellation with IC-DSS detection. The interference is a bit-pulsed jammer as in* **Scenario 9**. *Multiple paths seem to affect IC-DSS the least.*

5.4 Computational considerations

The computational complexity of the IC-DSS algorithm is now expressed as a function of the number of multiply and add operations. As in Section 4.9, the number of users is K, and the length of the data vector before despreading is $CM = C \times M$ (C being the code length and M the block size), and L is the number of correlated paths of the channel. DSS requires that the data is whitened before the filtering stage, and hence the complexity of whitening by EVD or SVD of the data correlation matrix is usually $\mathcal{O}(CM^3)$ [52]. Efficient algorithms [206]

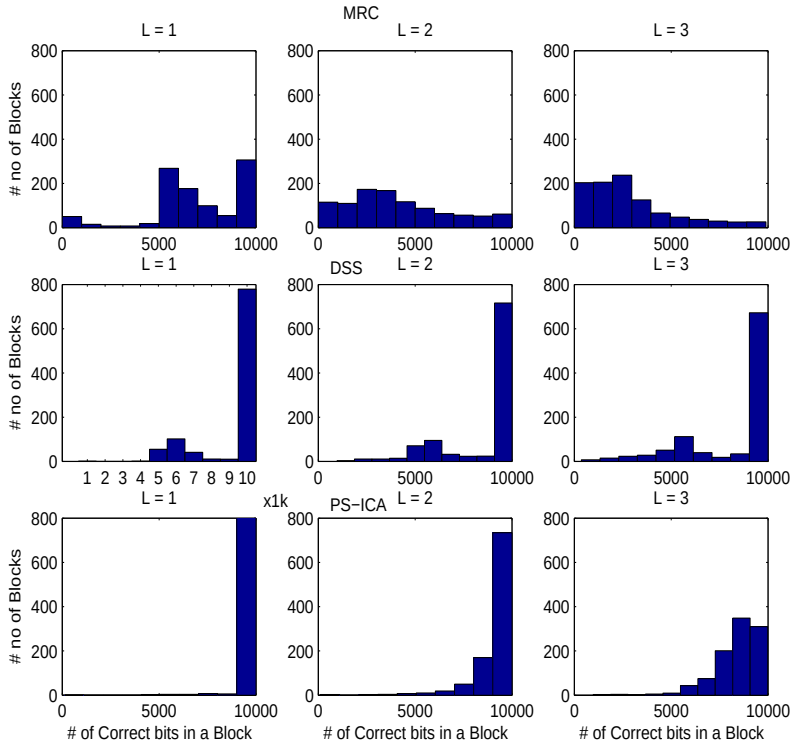


Figure 5.6: *Distribution of correct bits. As the number of paths increases, the distributions tend to move towards the left. The variance of IC-DSS though is far less than for the other detectors as the number of paths increases.*

can reduce this complexity to $\mathcal{O}(CMK)$, if only K principal eigenvectors are computed, at the expense of lower accuracy of the estimates.

The core of the IC-DSS algorithm consists of the following steps:

1. $\mathbf{x}_k = \mathbf{w}_k^T \mathbf{Y}$: The separation of the sources is performed by estimating the separating matrix $\mathbf{W}$ of dimensions $K \times L$. This requires $\mathcal{O}(K)$ additions and $\mathcal{O}(LCM)$ multiplications. Since $K < LCM$, this order is $\mathcal{O}(LCM)$.
2. $\hat{\mathbf{b}}_k = sign(\mathbf{s}_k^T \mathbf{x}_k)$: This step denoises the sources. The single user IC-DSS algorithm denoises only the user of interest K. With a block size of M, this

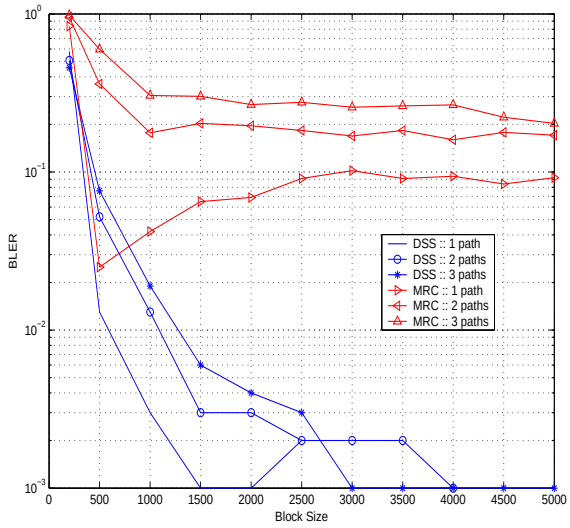


Figure 5.7: *Block size sensitivity of the algorithms. MRC is not dependent on the block size, while IC-DSS and Post-Switch's performance are directly dependent.*

complexity is $\mathcal{O}(CM)$.

3. $\hat{\mathbf{x}}_k = \mathbf{s}_k \hat{\mathbf{b}}_k$: Reconstruction of the sources involves respreading the sources. The complexity of this step is $\mathcal{O}(CM)$.
4. $\mathbf{w}_k^+ = \mathbf{Y}\hat{\mathbf{x}}_k$: Updating the separating matrix. The complexity of this matrix multiplication is of the order $\mathcal{O}(LCM)$.

Adding the separate complexity terms together, the final computational complexity of the IC-DSS algorithm can be summarized in Table 5.1.

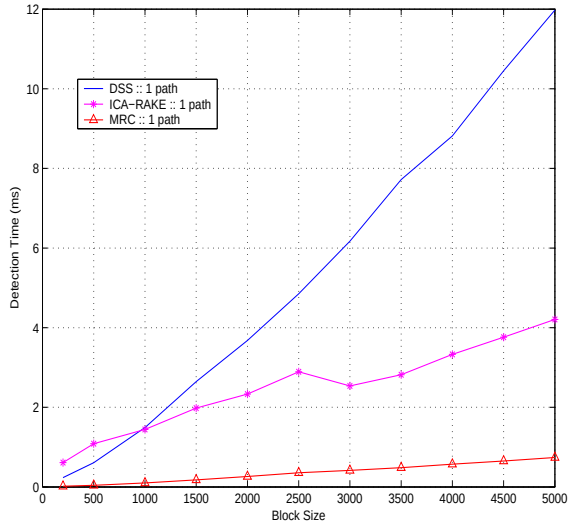


Figure 5.8: *Detection time as a function of block size. From Fig. 5.7, the block size can be estimated to be around* $M = 1000$. *For this size, the complexities of both the Post-Switch and IC-DSS algorithms are nearly equal.*

Step	Complexity
Whitening	$\mathcal{O}(CMK)$
Separation	$\mathcal{O}(LCM)$
Denoising	$\mathcal{O}(CM)$
Reconstruction	$\mathcal{O}(CM)$
Updating $\mathbf{W}$	$\mathcal{O}(LCM)$
Overall	$\approx \mathcal{O}(CMK + CM + LCM)$

Table 5.1: Computational complexity of IC-DSS method. The constants have been omitted. Here, C denotes the code length, K is the number of users, M is the block size, and L is the number of correlated paths.

Chapter 6

Discussion

Semi-blind and blind interference cancellation techniques, introduced in Chapters 4 and 5, offer promising alternatives to conventional techniques. They make optimal use of the available information while at the same time refrain from actively seeking more information. It is generally desirable to utilize all the "prior" knowledge effectively. This can help in improving the computational speed of the algorithms and in addition decrease the error rate. Adding this information adds a flavor of semi-blindness to blind algorithms, and hence notions of blindness have to be clearly defined. Additionally, equivalent algorithms exist in different fields. ICA has often been compared to beamforming [22], while conceptually, the IC-DSS technique is closely related to the despread-respread technique [162]. This chapter tries to place the methods and algorithms in a wider context.

6.1 Notions of blind signal processing

Blindness is an often sought feature in algorithms ever since computational capabilities have grown. This popularity has grown to the extent that many traditional algorithms are cast within a blind framework. This leads to misrepresentations of these algorithms as several "convenient" definitions of blind are used within different scientific communities. The Oxford American dictionary defines blind as[1],

blind: *without having all the relevant information*

With this definition, the algorithm to solve the model

$$\mathbf{x}(t) = \mathbf{A}\mathbf{s}(t) + \boldsymbol{\nu}(t), \tag{6.1}$$

is blind, as long as it has very little knowledge on the mixing structure $\mathbf{A}$ or the source signals $\mathbf{s}(t)$. Independent component analysis, a technique to solve this problem, is blind in spite of the strong assumption of independence made on the nature of the sources. This is the view taken by the signal processing community, and fits with the definition above.

By contrast, a communication system has much prior information embedded in the system. Typical prior information includes the knowledge of the modulation scheme, the codes used, the spreading code of the user[2], *etc.* This information can be considered basic. Conventional methods utilize this information to decode the transmitted message. Additionally, several methods seek other essential information *e.g.*, channel characteristics to improve the estimation. This is relevant to the method in question. Most conventional algorithms utilize both the basic part (modulation scheme, the codes, spreading code *etc.*), and this additional information to effectively demodulate and decode the transmitted message. Any algorithm that utilizes all the basic information, without additionally seeking the other relevant part, effectively is blind. In short, the algorithm *does not have all the relevant* information required traditionally. With this view, the definition of blind systems can be rephrased as,

[1]The first definition found in the dictionary for blind is usually *unable to see*. Since this thesis talks about blind as in separation, the word is used as an adverb, and hence this definition. "Unable to see" is the definition of blind when used as an adjective.

[2]There is some research in the area of *modulation unaware* systems, known as automatic modulation recognition systems.

blind: *without actively seeking all the relevant information*

This classification means that most schemes that do not require channel state characteristics fall in the category of blind systems. With this justification, systems in Category (C2) and (C3) (See Sec. 3.5) are blind. Systems in category (C1) additionally assume something about the timing — though their requirement of this is less than that of conventional methods — and hence they can be termed as *semi-blind*.

The plain ICA-RAKE receiver, Pre-Switch and Post-Switch separate the users in a blind manner. Due to inherent ambiguities in the separation algorithm, these algorithms require some amount of prior knowledge for identifying the separated sources. This is done by the use of very short portions of training data (typically 2% of the block length). By combining blind separation and training based identification, these methods fall into category (C1).

The IC-DSS receiver utilizes the spreading code of the desired user, along with all other basic prior information. The algorithm does not actively seek any further information and belongs to the category (C2), and is hence blind.

Even though the popularity of blind methods is increasing, their widespread application to many practical problems is limited due to increased complexity and computational costs. Semi-blind methods offer the way forward. They tend to use all the relevant information efficiently at reasonable computational cost and are more flexible with respect to on-line implementations.

6.2 ICA-RAKE Pre- and Post-Switch

ICA-RAKE Pre- and Post-Switch methods have been described in Sections 4.6 and 4.7 respectively, in Chapter 4.

ICA-RAKE Pre-Switch is a first-order structure that processes the received signal before interference cancellation. The estimated training sequence can be corrupted by the interfering jammer, or the multiple access interference, or both. Thus the effectiveness of the estimated training sequence can affect significantly the quality of the estimated source. The switching threshold δ_{sw} is set as a fraction of the length N_t of the training sequence. Moreover, this threshold is ad

hoc, and a rigorous statistical test can be applied to determine it.

ICA-RAKE Post-Switch is a correlation based detection structure. The received signal is processed independently by ICA and RAKE, and by MRC simultaneously. The estimated training sequence is now free of all interference, and is hence more accurate. The correlation between the true training sequence and the estimated training sequence leads to a better choice, and hence improved performance.

6.3 Role of ICA in communications

Blind source separation (BSS) is a general concept of solving the problem in (6.1), where the matrix $\mathbf{A}$ is assumed to mix the source signals $\mathbf{s}(t)$. This model can also be viewed as a blind identification problem in the context of narrow-band array processing, where N_a sensors receive waves emitted by K narrow-band sources. Each vector in the array $\mathbf{a}_i$ can be interpreted as the *directional vector* associated with the ith source since it depends on the location specific direction of the incoming source.

Beamformers [191] are essentially spatial filters designed to form pencil beams, in order to receive a signal radiated from a specific location and attenuate signals from all other locations. The beamformer should also account for the other signals and the noise, in order to be optimal filters. By assuming that the signal $s_i(t)$ is independent of the other contributions, the highest SNR at the filter output is obtained by choosing a filter $\mathbf{w}_i$ as

$$\mathbf{w}_i = \mathbf{R}_x^{-1}\mathbf{a}_i \quad \text{with} \quad \mathbf{R}_x = E\{\mathbf{x}(t)\mathbf{x}^T(t)\}. \tag{6.2}$$

This filter is the *minimum variance distortionless response* (MVDR) filter [191].

If it is assumed that the direction vectors of the other coherent sources are known, the spatial filter can be constrained to cancel these interfering sources. This leads to a *linear constrained minimum variance* (LCMV) filter [191]. For mutually independent sources, and spatially white noise, the filter $\mathbf{w}_i$ is of the form,

$$\mathbf{w}_i = \mathbf{R}^{\#}\mathbf{a}_i, \tag{6.3}$$

where $\mathbf{R}$ denotes a generic matrix that computes the filter co-efficient, when $\mathbf{R} = \mathbf{R}_x^{-1}$, this leads to the MVDR filter, and # denotes the pseudoinverse.

Both these approaches assume that the direction vector associated with the desired signal is known, and they are sensitive to errors in these vectors. These errors can either be deformation of the array, drift in the hardware (calibration errors) or multiple paths and/or reflections in the vicinity of the array (modeling errors). These approaches are also limited by the sample statistics used to estimate the matrix $\mathbf{R}_x$, which might not be the true estimate due to finite sample size.

If the direction vectors are unknown, the problem is blind, in which case the direction vectors have to be *estimated*. Independent component analysis, a popular technique to solve the BSS problem, can be viewed as blind identification technique that estimates these direction vectors *without knowing the array manifold i.e.*, without physical modeling of the propagation or array calibration. The benefit now is that the technique is insensitive to errors in the model.

In the presence of several coherent sources – which is usually the case – the "blind" approach typically yields the direction vectors associated with *all* the sources. When only some sources are of interest, then it is necessary to process the estimated directional vectors and to select the ones associated with the sources of interest. This identification can be performed by using other non-spatial information like spectral content, modulation, *etc.* Other approaches include choosing the estimated vector that is as close as possible to the modeled one.

ICA assumes that the sources are mutually independent at a given time. Although this is a very strong assumption, it is often plausible and is quite useful and is met in practice when the emitters are physically separated. In a multipath channel, the single transmitter contributes several correlated signals. When these paths have similar delays the sources are coherent *i.e.*, $s_1(t) = \beta s_2(t)$, where β is a complex number. The combination of these signals at the array output is $s_1(t)\mathbf{a}_1 + s_2(t)\mathbf{a}_2 = s_1(t)(\mathbf{a}_1 + \beta\mathbf{a}_2)$. This can be viewed as a single source with a composite "direction" vector $(\mathbf{a}_1 + \beta\mathbf{a}_2)$. Such composite direction vectors would be a problem in parametric array processing, which associate each independent component with one specific direction. Blind approaches do not make this assumption as they do not deal with directions-of-arrival (DOA). Hence, the blind array sees only one source with a unique composite direction vector.

Thus, ICA can be viewed as a blind beamforming technique that estimates these "composite direction vectors". Since the term "direction vector" is misleading in the blind context, it is not used. These vectors are instead called "mixing vectors" in order to emphasize the generic nature of the method.

6.4 Is IC-DSS a despread-respread concept?

The concept of the *despread-respread* (DR) algorithm is to generate a reference signal by despreading the received signal. The received signal is usually the output of the array. A bit decision on the possible transmitted sequence is formed with this generated reference signal. The bit estimate is then respread and used as the reference signal. The DR algorithm requires PN code synchronization, and hence its performance depends on the performance of the synchronization circuits. A example of a DR algorithm used in a multi-target constant modulus array is presented in [162].

The *least-squares despread-respread multi-target constant modulus array* (LS-DRMTCMA) shown in Fig. 6.1, is an adaptive algorithm that combines constant modulus adaptation and the despread-respread adaptation. During each iteration, the output from the beamformer $\mathbf{y}_k$ is despread to form the bit decision, which is subsequently respread with the kth user code. Adaptation of the beamformer is performed by comparing the respread signal with constant modulus constrained array output signal.

The *interfering canceling-denoising source separation* algorithm (IC-DSS) is similar to the above algorithm. The output from the array $\mathbf{y}_k$ is despread-adapt-respread. During the adaptation phase, the bit decision is modified based on the changes in the kth user's unmixing vector $\mathbf{w}_k$. Gradually, the IC-DSS algorithm mimics the mixing structure — the channel or the array — as the "fitness" of the bit decision increases.

The major difference between the above algorithms is that IC-DSS adapts the bit-decision based on the change in weights $\mathbf{w}$, while the LS-DRMTCMA algorithm adapts the weights based on the error between the respread and the output of the beamformer. The adaptation process of the IC-DSS algorithm lends itself to various improvements like compensation for Doppler effects and adaptation to fading channels *etc.*

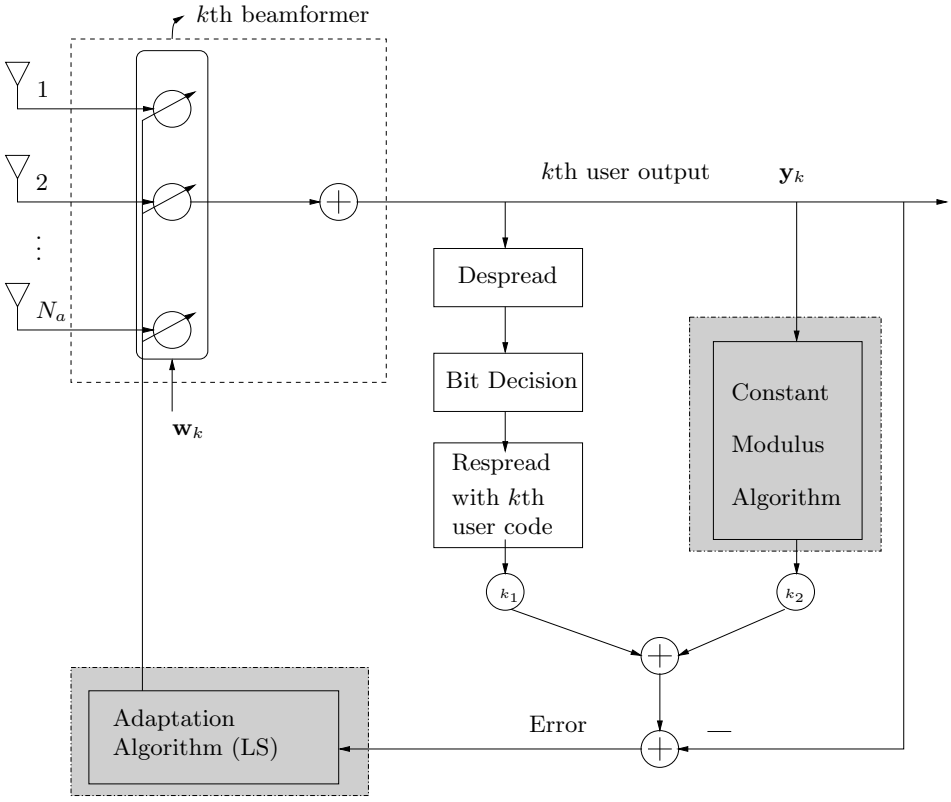


Figure 6.1: *Schematic illustration of the Least-Squares Despread-Respread (LS-DR) algorithm (Adapted from [162]) and the IC-DSS algorithm [152]. Both algorithms modify the weight vector* $\mathbf{w}_k$ *taking into account the respread signal and the original received signal. The shaded portions of the figure represents blocks exclusive to the LS-DR algorithm.*

Chapter 7

Conclusions

7.1 Summary

This thesis considers the issue of interference cancellation and suppression in downlink CDMA systems. The methods developed either augment existing methods by providing a suppression block based on the blind source separation principle, or introduce a new structure that requires only the spreading codes of the desired user to separate and suppress the interference.

The beauty of blind source separation is that it allows for a solution to a seemingly unsolvable problem: Assume that the observed data are generated by some unknown mixing of unknown variables. Using only the observed data, the aim is to find both the mixing structure and the latent variables. Under certain assumptions, described in Sec. 4.1.1, this problem is indeed solvable [27, 75]. Such problems occur in many application domains *e.g.*, localizing areas of activity of the brain from the measured electroencephalogram traces. Blind methods provide a practical means for solving these problems.

In communications, blind methods have a wider appeal. Given that available bandwidth is both finite and scarce, and the need for higher data rates is incessant, blind methods offer promising solutions that require no training data. This provides a means to increase the data rate by allowing transmission of user data in place of training data.

Independent component analysis (ICA) [75] is a popular method for solving the BSS problem. By assuming that the unknown variables are statistically independent, ICA tries to estimate both the mixing matrix and the original sources. ICA is a proven and popular method in the BSS community. Several ICA/BSS algorithms [5, 8, 21, 73, 76, 214] solve the BSS problem by maximizing some measure of independence.

The appeal of applying these methods to communication provides the motivation for this work. The aim is to cancel and suppress *unknown* sources of interference, as they tend to have an adverse impact on performance [119]. The first extension is the plain ICA-RAKE receiver structure (Sec. 4.3), wherein the interference is estimated by an ICA algorithm, but identified and suppressed by the conventional RAKE receiver. Several experiments with different settings demonstrate the performance of this structure.

The fact that the plain ICA-RAKE receiver under-performs when the interference level is low, led to the development of the ICA-RAKE Pre-Switch algorithm. The ICA-RAKE Pre-Switch is a first order distance based receiver, that estimates the "raw" interference level, and switches between conventional and ICA based detection (Sec. 4.6). A more robust structure ICA-RAKE Post-Switch was later developed (Sec. 4.7). This switch operates on the post-processed data and is based on correlating the estimated and true training data. This structure provides effective interference cancellation for all levels of interference. Both these structures are evaluated under different circumstances *e.g.*, changing nature of interference structure, multipath propagation environment *etc.* Additionally, several ICA algorithms are evaluated in this framework.

The fact that all three structures considered are semi-blind led to the development of the fourth structure — IC-DSS, a denoising source separation based algorithm, described in Sec. 5.2. This algorithm separates the interference by repeated despreading-respreading of the data with the desired user's spreading code, while denoising the data simultaneously. Without additional information, the IC-DSS algorithm tries to construct a possible transmitted sequence that is as close as possible to the received bit stream, building a "possible" channel matrix additionally. The fact that this method does not require training sequences means it is blind.

Given the amount of prior information available in modern communication, it would be futile to completely disregard this data and search for truly blind solutions. Moreover, many blind methods suffer from slow convergence and may

have some or other ambiguities. Prior information in this context is valuable. Appendix A examines a certain class of priors termed as "conjugate priors" in this setting. The elegance of this approach, is that these priors can be applied to existing algorithms without major modifications. Earlier examples from image extraction, as well as examples from the CDMA problem setting help to illustrate this point.

Finally, being a monograph, this thesis provides a brief overview of direct-sequence CDMA (DS-CDMA), and constructs the models that assist in the application of ICA (Chapter 2). Chapter 3 contains a literature survey outlining the existing algorithms for interference cancellation. Similarities between existing methods are discussed in Chapter 6.

7.2 The Future

There are several enhancements to the methods presented here, from both the semi-blind and blind viewpoints. The ICA based algorithms involve the use of conventional detection mainly for identification and timing extraction. Though the models in Sec. 2.2.2 or Sec. 2.2.3 are convolutional due the channel, the array presents a means for application of instantaneous algorithms. The obvious extension would be to apply blind deconvolution methods [56, 57, 60] or the several extensions of convolutional ICA algorithms (See [26], Chapter 23 of [75] and [59].).

There are several routes for extending the IC-DSS algorithm, some of which are outlined below:

1. performance improvement (on-line adaptation) of the algorithm;
2. adaptation to fading channels;
3. explicit channel estimation;
4. extension to the multiuser domain;
5. adaptation to include prior information.

Possible extensions using prior information can be in the domain of overloaded systems and in adaptation and evaluating a class of channel specific priors, that help in the development of on-line versions of these algorithms. Clearly, there remain good prospects for further progress based on extending the work presented here.

Bibliography

[1] Bluetooth specification. Version 1.0B. `http://www.bluetooth.com`.

[2] B. Aazhang, B.-P. Paris, and G. Orsak. Neural networks for multi-user detection in CDMA communications. *IEEE Transactions on Communications*, 40:1212–1222, 1992.

[3] P. D. Alexander, L. K. Rasmussen, and C. Schlegel. A linear receiver for coded multiuser CDMA. *IEEE Transactions on Communications*, 45(5):605–610, 1997.

[4] S.-I. Amari. Natural gradient works effciently in learning. *Neural Computation*, 10(2):251–276, 1998.

[5] S.-I. Amari, A. Cichocki, and H. H. Yang. A new learning algorithm for blind source separation. *Advances in Neural Information Processing*, 8:757–763, 1996.

[6] B. D. Anderson and J. B. Moore. *Optimal filtering*. Prentice-Hall, 1979.

[7] A. N. Barbosa and S. L. Miller. Adaptive detection of DS/CDMA signals in fading channels. *IEEE Transactions on Communications*, 46(1):115–124, 1998.

[8] A. J. Bell and T. J. Sejnowski. An information-maximization approach to blind separation and blind deconvolution. *Neural Computation*, 7:1129–1159, 1995.

[9] A. J. Bell and T. J. Sejnowski. The "independent components" of natural scenes are edge filters. *Vision Research*, 37:3327–3338, 1997.

[10] A. Belouchrani, K. Abed-Meriam, J.-F. Cardoso, and É. Moulines. A blind source separation technique based on second order statistics. *IEEE Trans. on Signal Processing*, 45(2):434–44, 1997.

[11] A. Belouchrani and M. Amin. A two-sensor array blind beamformer for direct sequence spread spectrum communications. *IEEE Transactions on Signal Processing*, 47:2191–2199, 1999.

[12] A. Belouchrani and M. Amin. Jammer mitigation in spread spectrum communications using blind source separation. *Signal Processing*, 80:723–729, 2000.

[13] S. Bensley and B. Aazhang. Subspace-based channel estimation for code division multiple access communication systems. *IEEE Transactions on Communications*, 44(8):115–124, 1996.

[14] O. Bermond and J.-F. Cardoso. Approximate likelihood for noisy mixtures. In *Proc. of the 1999 Int. Workshop on Independent Component Analysis (ICA 1999)*, pages 325–330, Aussois, France, January 1999.

[15] E. Bingham and A. Hyvärinen. A fast fixed-point algorithm for independent component analysis of complex valued signals. *Int. Jo. of Neural Systems*, 10(1):1–8, 2000.

[16] E. Bingham and H. Mannila. Random projection in dimensionality reduction: Applications to image and text data. In *Proc. of the 7th ACM SIGKDD Int. Conference on Knowledge Discovery and Data Mining*, pages 245–250, San Francisco, USA, August 2001.

[17] J. Boer. Direct sequence spread spectrum: Physical layer specifications - IEEE 802.11b. *IEEE WLAN Specification: IEEE P802.11b-96/49E*, 1996. `http://grouper.ieee.org/groups/802/11/Tutorial/ds.pdf`.

[18] F. N. Brannstrom, T. M. Aulin, and L. K. Rasmussen. Iterative multiuser detection of trellis code multiple access using posteriori probabilities. In *Proc. of the IEEE Int. Conference on Communications*, pages 11–15, Helsinki, Finland, June 2001.

[19] J.-F. Cardoso. Source separation using higher order moments. In *Proc. of the IEEE Int. Conf on Acoustics, Speech and Signal Processing (ICASSP 1989)*, pages 2109–2112, Glasgow, UK, May 1989.

[20] J.-F. Cardoso. Blind signal separation: statistical principles. *Proceedings of the IEEE. Special issue on blind identification and estimation*, 9(10):2009–2025, 1998.

[21] J.-F. Cardoso and A. Souloumiac. Blind beamforming for non gaussian signals. *IEE Proceedings-F*, 140(6):362–370, 1993.

[22] J.-F. Cardoso and A. Souloumiac. Jacobi angles for simultaneous diagonalization. *SIAM J. Mat. Anal. Appl.*, 17(1):161–164, 1996.

[23] A. B. Carlson. *Communication Systems - An Introduction to Signals and Noise in Electrical Communication*. McGraw-Hill, third edition, 1986.

[24] D. S. Chen and S. Roy. An adaptive multiuser receiver for CDMA systems. *IEEE Journal on Selected Areas in Communications*, 12(6):808–816, 1994.

[25] L.-C. Chu and U. Mitra. Performance analysis of an improved MMSE multiuser receiver for mismatched delay channels. *IEEE Transactions on Communications*, 46(10):1369–1380, 1998.

[26] A. Cichocki and S.-I. Amari. *Adaptive Blind Signal and Image Processing*. Wiley & Sons, 2002.

[27] P. Comon. Independent component analysis - a new concept. *Signal Processing*, 36:287–314, 1994.

[28] T. Cover and J. Thomas. *Elements of Information Theory*. John Wiley & Sons Inc., 1991.

[29] R. Cristescu, T. Ristaniemi, J. Joutsensalo, and J. Karhunen. CDMA delay estimation using a FastICA algorithm. In *Proc. of the IEEE International Symposium on Personal, Indoor and Mobile Communications*, pages 1117–1120, London, UK, September 2000.

[30] R. Cristescu, T. Ristaniemi, J. Joutsensalo, and J. Karhunen. Delay estimation in CDMA communications using a FastICA algorithm. In *Proc. of the 2nd International Workshop on Independent Component Analysis*, pages 105–110, Espoo, Finland, June 2000.

[31] S. Davidovici and E. G. Kanterakis. Narrow-band interference rejection using real-time Fourier transforms. *IEEE Transactions on Communications*, 37(7):713–722, 1998.

[32] P. Dempster, N. M. Laird, and D. B. Rubin. Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society, Series B (Methodological)*, 39(1):1–38, 1977.

[33] A. J. Van der Veen. Analytical method for blind binary signal separation. *IEEE Transactions on Signal Processing*, 45(4):1078–1082, 1997.

[34] A. J. Van der Veen and A. Paulraj. An analytical constant modulus algorithm. *IEEE Transactions on Signal Processing*, 44(5):1136–1155, 1996.

[35] A. J. Van der Veen and J. Tol. Separation of zero/constant modulus signals. In *Proc. of the IEEE International Conference on Acoustics, Speech and Signal Processing*, pages 3445–3448, Munich, Germany, April 1997.

[36] K. I. Diamantaras and S. Y. Kung. *Principal Component Analysis: Theory and Applications*. Wiley & Sons, 1996.

[37] R. C. DiPietro. An FFT based technique for suppressing narrowband interference in PN spread spectrum communication systems. In *Proc. of the IEEE Int. Conference on Acoustics, Speech and Signal Processing*, volume 2, pages 1360–1363, Glasgow, UK, May 1989.

[38] S. C. Douglas and A. Cichocki. Neural networks for blind decorrelation of signals. *IEEE Transactions on Signal Processing*, 45(11):2829–2842, 1997.

[39] A. Duel-Hallen. Decorrelating decision-feedback multiuser detectors for synchronous code-division multiple access channels. *IEEE Transactions on Communications*, 41(2):285–290, 1993.

[40] A. Duel-Hallen. A family of multiuser decision-feedback detectors for asynchronous code-division multiple access channels. *IEEE Transactions on Communications*, 43(2/3/4):421–434, 1995.

[41] A. Duel-Hallen, J. Holtzman, and Z. Zvonar. Multiuser detection for CDMA systems. *IEEE Personal Communications*, 2(2):46–58, 1995.

[42] L. Fang and L. B. Milstein. Successive interference cancellation in multicarrier DS/CDMA. *IEEE Transactions on Communications*, 48(9):1530–1540, 2000.

[43] F. H. P. Fitzek. *Quality of Service in Wireless Multi-Code CDMA Systems*. Technical University of Berlin, 2002.

[44] G. D. Forney. The Viterbi algorithm. *Proc. IEEE*, 61(3):268–278, 1973.

[45] K. Fukunaga and D. R. Olsen. An algorithm for finding intrinsic dimensionality of data. *IEEE Transactions on Computers*, 202:176–183, 1971.

[46] M. Funaro, E. Oja, and H. Valpola. Independent component analysis for artefact separation in astrophysical images. *Neural Networks*, 16(3-4):469–478, 2003.

[47] G. Giannakis, Y. Hua, P. Stoica, and L. Tong, editors. *Signal Processing Advances in Wireless and Mobile Communications, Vol. 1: Trends in Channel Estimation and Equalization*. Prentice-Hall, 2000.

[48] M. Girolami. *Self-Organizing Neural Networks - Independent Component Analysis*. Springer-Verlag, 1999.

[49] S. Glisic. *Advanced Wireless Communications*. John Wiley & Sons, 2004.

[50] D. N. Godard. Self-recovering equalization and carrier tracking in two dimensional data communication systems. *IEEE Transactions on Communications*, 28(11):1867–1875, 1980.

[51] A. Goldsmith. *Wireless Communications*. Cambridge University Press, 2005.

[52] G. H. Golub and C. F. Van Loan. *Matrix Computations*. The John Hopkins University Press, 1996.

[53] E. Gönen and J. M. Mendel. Applications of cumulants to array processing - part III: Blind beamforming for coherent signals. *IEEE Transactions on Signal Processing*, 45(9):2252–2264, 1997.

[54] W. Haifeng, J. Lilleberg, and K. Rikkinen. A new sub-optimal multiuser detection approach for CDMA systems in Rayleigh fading channel. In *Proc. Conference on Information Sciences and Systems (CISS)*, volume 1, pages 276–280, The John Hopkins University, Baltimore, USA, March 1997.

[55] A. Haimovich and A. Vadhri. Rejection of narrow-band interferences in PN spread spectrum systems using an eigenanalysis approach. In *Proc. IEEE Workshop on Statistical Signal and Array Processing*, pages 383–386, Quebec, Canada, June 1994.

[56] S. Haykin. *Blind Deconvolution*. Prentice Hall, 1994.

[57] S. Haykin. *Adaptive Filter Theory*. Prentice Hall, 3rd edition, 1996.

[58] S. Haykin. *Neural Networks: A Comprehensive Foundation*. Prentice Hall, 1999.

[59] S. Haykin. *Unsupervised Adaptive Filter, Vol 1: Blind Source Separation*. Wiley, 2nd edition, 2000.

[60] S. Haykin. *Unsupervised Adaptive Filter, Vol 2: Blind Deconvolution*. Wiley, 2000.

[61] F. Herrmann and A. K. Nandi. Maximisation of squared cumulants for blind source separation. *Electronics Letters*, 36(19):1664–1665, 1996.

[62] H. Holma, A. Toskala, and A. Hottinen. Performance of CDMA multiuser detection with antenna diversity and closed loop power control. In *Proc. of the IEEE Vehicular Technology Conference*, pages 362–366, Atlanta, USA, 1996.

[63] J. M. Holtzman. DS/CDMA successive interference cancellation. In *Proc. of the IEEE Int. Symposium on Spread Spectrum Techniques and Applications*, pages 69–78, Oulu, Finland, July 1994.

[64] M. L. Honig, U. Madhow, and S. Verdú. Blind adaptive interference suppression for near-far resistant CDMA. In *Proc. of the IEEE Globecom*, pages 379–384, San Francisco, USA, Nov-Dec 1994.

[65] M. L. Honig, U. Madhow, and S. Verdú. Blind adaptive multiuser detection. *IEEE Transactions on Information Theory*, 41:944–960, 1995.

[66] K. Hooli. *Equalization in WCDMA Terminals*. Doctoral Thesis, Acta. Univ. Ouluensis, C-192, 2003.

[67] H. Hotelling. Analysis of a complex of statistical variables into principal components. *Journal of Educational Psychology*, 24:417–441, 498–520, 1933.

[68] A. Hottinen, H. Holma, and A. Toskala. Performance of multistage multiuser detection in a fading multipath channel. In *Proc. of the IEEE Int. Symposium on Personal, Indoor, Mobile, Radio, Communications*, pages 960–964, Toronto, Canada, September 1995.

[69] F. M. Hsu and A. A. Giordano. Digital whitening techniques for improving spread-spectrum communications performance in the presence of narrowband jamming interference. *IEEE Transactions on Communications*, COM-26:209–216, 1978.

[70] T. Huovinen and T. Ristaniemi. Comparison of nonlinear interference cancellation and blind source separation techniques in the DS-CDMA uplink. In *Proc. of the 2003 IEEE Int. Symposium on Personal, Indoor, Mobile, Radio, Communications (PIMRC 2003)*, Beijing, China, September 2003.

[71] T. Huovinen and T. Ristaniemi. DS-CDMA capacity enhancement using blind source separation group-wise successive interference calcellation. In *Proc. of the IEEE International Workshop on Signal Processing Advances for Wireless Communications*, Lisbon, Portugal, July 2004.

[72] T. Huovinen and T. Ristaniemi. Private communication. 2005.

[73] A. Hyvärinen. Fast and robust fixed-point algorithms for independent component analysis. *IEEE Transactions on Neural Networks*, 10(3):626–634, 1999.

[74] A. Hyvärinen. Sparse code shrinkage: Denoising of nongaussian data by maximum likelihood estimation. *Neural Computation*, 11:1739–1768, 1999.

[75] A. Hyvärinen, J. Karhunen, and E. Oja. *Independent Component Analysis*. John Wiley & Sons Inc., 2001.

[76] A. Hyvärinen and E. Oja. A fast fixed-point algorithm for independent component analysis. *Neural Computation*, 9:1483–1492, 1997.

[77] A. Hyvärinen and E. Oja. Independent component analysis - algorithms and applications. *Neural Networks*, 13(4-5):411–430, 2000.

[78] A. Hyvärinen and K. Raju. Imposing sparsity on the mixing matrix in independent component analysis. *Neurocomputing*, 49:151–162, 2002.

[79] A. Hyvärinen, J. Särelä, and R. Vigário. Spikes and bumps: Artefacts generated by independent component analysis with insufficient sample size. In *Proc. of the 1999 Int. Workshop on Independent Component Analysis and Signal Separation (ICA 1999)*, pages 425–429, Aussois, France, January 1999.

[80] J. Igual and L. Vergara. Prior information about mixing matrix in BSS-ICA formulation. In *Proc. of the 2000 Int. Workshop on Independent Component Analysis and Signal Separation (ICA 2000)*, pages 123–128, Helsinki, Finland, June 2000.

[81] J. Iinatti. Mean acquisition time of DS code synchronization in fixed multipath channel. In *Proc. of the 1998 IEEE International Symposium on Spread Spectrum Techniques and Applications (ISSSTA 1998)*, volume 1, pages 116–120, Sun City, South Africa, September 1998.

[82] J. Iinatti and M. Latva-aho. Matched filter acquisition in fixed multipath channel. In *Proc. of the 1998 IEEE International Symposium on Personal, Indoor and Mobile Radio Communications (PIMRC 1998)*, volume 3, pages 1501–1505, Boston, USA, September 1998.

[83] R. A Iltis and L. B. Milstein. Performance analysis of narrowband interference in DS spread spectrum signals using transversal filters. *IEEE Transactions on Communications*, COM-32:1169–1177, 1984.

[84] R. A Iltis and L. B. Milstein. An approximate statistical analysis of the Widrow LMS algorithm with application to narrowband interference rejection. *IEEE Transactions on Communications*, COM-33:121–130, 1985.

[85] D. H. Johnson and D. E. Dudgeon. *Array Signal Processing: Concepts and Techniques*. Prentice-Hall, 1993.

[86] W. W. Jones and K. R. Jones. Narrowband interference suppression using filter-bank analysis/synthesis technqiues. In *Proc. of the IEEE MILCOM Conference*, San Diego, USA, October 1992.

[87] C. R. Johnson Jr., P. Schnitter, T. J. Endres, J. D. Behm, D. R. Brown, and R. Casas. Blind equalization using the constant modulus criterion: A review. *Proc. IEEE*, 86(10):1927–1950, 1998.

[88] M. Juntti. *Multiuser Demodulation for DS-CDMA Systems in Fading Chanels*. Doctoral Thesis, Acta. Univ. Ouluensis, C-106, 1997.

[89] M. J. Juntti, B. Aazhang, and J. O. Lilleberg. Iterative implementation of linear multiuser detection for dynamic asynchronous CDMA systems. *IEEE Transactions on Communications*, 46(4):503–508, 1998.

[90] C. Jutten and J. Herault. Blind separation of sources, part I: An adaptive algorithm based on neuromimetic architecture. *Signal Processing*, 24:1–10, 1991.

[91] M. Katz, J. Iinatti, and S. Gilsic. Two-dimensional code acquisition in fixed multipath channels. In *Proc. of the 2000 IEEE International Vehicular Technology Conference (VTC Fall 2000)*, volume 5, pages 2317–2324, Boston, USA, September 2000.

[92] S. M. Kay. *Fundamentals of Statistical Signal Processing: Estimation Theory*. Prentice-Hall, 1993.

[93] M. Kendall and A. Stuart. *The Advanced Theory of Statistics*. Charles Griffin & Company, 1958.

[94] J. Ketchum and J. G. Proakis. Adaptive algorithms for estimating and suppressing narrowband interference in PN spread spectrum systems. *IEEE Transactions on Communications*, COM-30:913–924, 1982.

[95] H. Knuth. A Bayesian approach to source separation. In *Proc. of the 1999 Int. Workshop on Independent Component Analysis and Signal Separation (ICA 1999)*, pages 283–288, Aussois, France, January 1999.

[96] W. Krattenthaler and F. Hlawatsch. Time-frequency design and processing of signals via smoothed Wigner distributions. *IEEE Transactions on Signal Processing*, 41(1):278–287, 1993.

[97] S. Kullback and R. Leibler. On information and sufficiency. *Ann. Math. Stat*, 22:79–86, 1951.

[98] J. Laster and J. Reed. Interference rejection in digital wireless communications. *IEEE Signal Processing Magazine*, 14(3):37–62, 1997.

[99] M. Latva-Aho. *Advanced Receivers for Wideband CDMA Systems*. Doctoral Thesis, Acta. Univ. Ouluensis, C-125, 1998.

[100] M. Latva-aho and J. Lilleberg. Parallel interference cancellation in multiuser detection. In *Proc. of the 1996 IEEE Int. Symposium on Spread Spectrum Techniques and Applications (ISSSTA 1996)*, pages 1151–1155, Mainz, Germany, September 1996.

[101] M. Latva-Aho and J. O. Lilleberg. Parallel interference cancellation in multiuser CDMA channel estimation. *Wireless Personal Communications, Kluwer Academic Publishers, Special Issue on CDMA for Universal Communication Systems*, 7(2/3):171–195, 1998.

[102] W. Lee. *Mobile Cellular Communications*. McGraw Hill, New York, 1989.

[103] L. Li and L. B. Milstein. Rejection of narrowband interference in PN spread spectrum signals using transversal filters. *IEEE Transactions on Communications*, COM-30:925–928, 1982.

[104] L. Li and L. B. Milstein. Rejection of pulsed CW interference in PN spread spectrum signals using complex adaptive filters. *IEEE Transactions on Communications*, COM-31:10–20, 1983.

[105] M. M. Loéve. *Probability Theory*. Van Nostrand, Princeton, 1955.

[106] R. Lupas and S. Verdú. Linear multiuser detectors for synchronous code-division multiple-access channel. *IEEE Transactions on Information Theory*, IT-35(1):123–136, 1989.

[107] D. J. C. Mackay. *Information Theory, Inference and Learning Algorithms*. Cambridge University Press, 2003.

[108] U. Madhow. Blind adaptive interference suppression for the near-far resistant acquisition and demodulation of direct-sequence CDMA systems. *IEEE Transactions on Signal Processing*, 45(1):124–136, 1997.

[109] U. Madhow. Blind adaptive interference suppression for direct-sequence CDMA. *Proc. IEEE*, 86(10):2049–2069, 1998.

[110] U. Madhow. MMSE interference suppression for timing acquisition and demodulation in direct-sequence CDMA system. *IEEE Transactions on Communications*, 46(8):1065–1075, 1998.

[111] U. Madhow and M. L. Honig. MMSE interference suppression for direct-sequence spread-spectrum CDMA. *IEEE Transactions on Communications*, 42(12):3178–3188, 1994.

[112] U. Madhow and M. L. Honig. On the average near-far resistance for MMSE detection of direct-sequence CDMA signals with random spreading. *IEEE Transactions on Information Theory*, 45(6):2039–2045, 1999.

[113] E. Masry. Closed form analytical results for the rejection of narrowband interference in PN spread spectrum signals - Part I: Linear prediction filters. *IEEE Transactions on Communications*, COM-32:888–896, 1984.

[114] E. Masry. Closed-form analytical results for the rejection of narrowband interference in PN spread system signals - Part II: Linear interpolation filters. *IEEE Transactions on Communications*, COM-33:10–19, 1985.

[115] E. Masry and L. B. Milstein. Performance of DS spread-spectrum receiver employing interference-suppression filters under a worst-case jamming condition. *IEEE Transactions on Communications*, COM-34:13–21, 1986.

[116] E. Masry and L. B. Milstein. Interference suppression to aid acquisition in direct-sequence spread-spectrum communications. *IEEE Transactions on Communications*, 36(11):1200–1202, 1988.

[117] J. M. Mendel. *Lessons in Estimation Theory for Signal Processing, Communications and Control*. Prentice-Hall, 1995.

[118] A. Mertins. *Signal Analysis: Wavelets, Filter Banks, Time-Frequency Transforms and Applications*. John Wiley & Sons, 1999.

[119] L. B. Milstein. Interference rejection techniques in spread spectrum communications. *Proc. of the IEEE*, 66:657–671, 1988.

[120] L. B. Milstein and P. K. Das. Analysis of a real-time transform domain filtering digital communication system -Part I: Narrowband interference rejection. *IEEE Transactions on Communications*, COM-28(6):816–824, 1980.

[121] U. Mitra and H. V. Poor. Adaptive receiver algorithms for near-far resistant CDMA. *IEEE Communications Magazine*, 43(2-4):1713–1724, 1995.

[122] L. Molgedey and H. G. Schuster. Separation of a mixture of independent signals using time delayed correlations. *Physical Review Letters*, 72:3634–3636, 1994.

[123] T. K. Moon and W. C. Stirling. *Mathematical Methods and Algorithms for Signal Processing*. Prentice-Hall, 1999.

[124] S. Moshavi. Multi-user detection for DS-CDMA communications. *IEEE Communications*, 34(10):124–137, 1996.

[125] R. S. Mowbray, R. D. Pringle, and P. M. Grant. Increased CDMA system capacity through adaptive cochannel interference regeneration and cancellation. *IEE Proceedings*, I-139(5):515–524, 1992.

[126] A. K. Nandi. *Blind Estimation using Higher-Order Statistics*. Kluver Academic Publishers, 1999.

[127] A. K. Nandi and F. Herrmann. Fourth-order cumulant based estimator for independent component analysis. *Electronic Letters*, 37(7):469–470, 2001.

[128] A. K. Nandi and V. Zarzoso. Fourth-order cumulant based blind source separation. *IEEE Signal Processing Letters*, 3(12):312–314, 1996.

[129] E. Oja. Principal components, minor components, and linear neural networks. *Neural Networks*, 5:927–935, 1992.

[130] E. Oja. Convergence of the symmetrical FastICA algorithm. In *Proc. of the 9th Int. Conf. on Neural Information Processing (ICONIP 2002)*, Singapore, November 2002.

[131] T. Ojanperä and R. Prasad. *Wideband CDMA for Third Generation Systems*. Artech House, 1998.

[132] B. A. Olshausen and D. J. Field. Emergence of simple-cell receptive field properties by learning a sparse code for natural images. *Nature*, 381:607–609, 1996.

[133] T. B. Oon, R. Steele, and Y. Li. Performance of an adaptive successive serial-parallel CDMA cancellation scheme in flat Rayleigh fading channels. In *Proc. of the IEEE Vehicular Technology Conference*, pages 193–197, Phoenix, USA, May 1997.

[134] P. Pajunen. Blind source separation using algorithmic information theory. *Neurocomputing*, 22:35–48, 1998.

[135] A. Papoulis. *Probability, Random Variables and Stochastic Processes*. McGraw-Hill, 3rd edition, 1991.

[136] A. Papoulis and S. Unnikrishan Pillai. *Probability, Random Variables and Stochastic Processes*. Mc-Graw Hill, 2002.

[137] S. Parkvall. *Near-Far Resistant DS-CDMA Systems*. Ph.D. Thesis, Royal Institute of Technology, Sweden, October, 1996.

[138] P. Patel and J. Holtzman. Analysis of successive interference cancellation in m-ary orthogonal DS-CDMA system. *IEEE Journal on Selected Areas in Communication*, 12(5):796–807, 1994.

[139] A. Paulraj and H. Papadias. Space-time processing for wireless communications. *IEEE Signal Processing Magazine*, 14(6):49–83, 1997.

[140] R. L. Peterson, R. E. Ziemer, and D. E. Broth. *Introduction to Spread Spectrum Communications*. Prentice Hall Internation, 1995.

[141] R. L. Pickholtz, L. B. Milstein, and D. L. Schilling. Spread spectrum for mobile communications. *IEEE Transactions on Vehicular Technology*, 40(2):313–322, 1991.

[142] Li Ping, Lihai Liu, and W. K. Leung. A simple approach to near-optimal multiuser detection - interleave-division multiple-access. In *Proc. of 2003 Wireless Communications and Networking Conference*, volume 1, pages 391–396, New Orleans, USA, March 2003.

[143] Li Ping, Lihai Liu, K. Y. Wu, and W. K. Leung. On Interleave-Division Multiple-Access. In *Proc. of 2004 Int. Conf. on Communications*, volume 5, pages 2869–2873, Paris, France, June 2004.

[144] Li Ping, K. Y. Wu, Lihai Liu, and W. K. Leung. A simple unified approach to nearly optimal multiuser detection and space-time coding. In *Proc. of IEEE 2002 Information Theory Workshop*, volume 1, pages 53–56, Bangalore, India, October 2002.

[145] R. Price and P. E. Green. A communication technique for multipath channels. *Proceedings of the IRE*, 46:555–570, 1958.

[146] J. G. Proakis. *Digital Communications*. McGraw-Hill, third edition, 1995.

[147] K. Raju and T. Ristaniemi. ICA-RAKE-switch for jammer cancellation in DS-CDMA array systems. In *Proc. of the 2002 IEEE Int. Symposium on Spread Spectrum Techniques and Applications (ISSSTA 2002)*, pages 638 – 642, Prague, Czech Republic, September 02 - 05, 2002.

[148] K. Raju and T. Ristaniemi. Exploiting interference to cancel interferences due to adjacent cells in a DS -CDMA cellular system. In *Proc. of the 2003 IEEE Int. Symposium on Personal, Indoor, Mobile, Radio, Communications (PIMRC 2003)*, Beijing, China, September 07-10, 2003.

[149] K. Raju, T. Ristaniemi, and J. Karhunen. Semi-blind interference suppression on coherent multipath environments. In *Proc. of the First Int. Symposium of Control, Communications and Signal Processing (ISCCSP 2004)*, pages 283–286, Hammamet, Tunisia, March 21-24, 2004.

[150] K. Raju, T. Ristaniemi, J. Karhunen, and E. Oja. Suppression of bit-pulsed jammer signals in a DS-CDMA array system using independent component analysis. In *Proc. of the 2002 IEEE Int. Symposium on Circuts and Systems (ISCAS 2002)*, pages I–189/I–192, Phoenix, USA, May 26 - 29, 2002.

[151] K. Raju, T. Ristaniemi, J. Karhunen, and E. Oja. Jammer cancellation in DS-CDMA array systems using independent component analysis. *IEEE Transactions on Wireless Communication*, 5(1):77–82, January 2006.

[152] K. Raju and J. Särelä. A denoising source separation based approach to interference cancellation in DS-CDMA array systems. In *Proc. of the 38th Asilomar Conference on Signals, Systems and Computers*, pages 1111–1114, Pacific Grove, November, 2004.

[153] K. Raju and B. Phani Sudheer. Blind source separation for interference cancellation - a comparison of several spatial and temporal statistics based techniques. In *Proc. of the 3rd Workshop on the Internet, Telecommunications and Signal Processing*, Adelaide, Australia, Dec 20-22, 2004.

[154] L. K. Rasmussen, T. J. Lim, and A.-L. Johansson. A matrix-algebra approach to successive interference cancellation in CDMA. *IEEE Transactions on Communications*, 48(1):145–151, 2000.

[155] V. U. Reddy, A. Paulraj, and T. Kailath. Performance analysis of the optimum beamformer in the presence of correlated sources and its behavior under spatial smoothing. *IEEE Transactions on Acoustics, Speech and Signal Processing*, ASSP-35:927–936, 1987.

[156] R. Rick and L. B. Milstein. Parallel acquisition in mobile DS-CDMA systems. *IEEE Transactions on Communications*, 45(11):1466–1476, 1997.

[157] T. Ristaniemi. *Synchronization and Blind Signal Processing in CDMA Systems*. Doctoral Thesis, University of Jyväskylä, 2000.

[158] T. Ristaniemi and J. Joutsensalo. Advanced ICA-based receivers for block fading DS-CDMA channels. *Signal Processing*, 82:417–431, 2002.

[159] T. Ristaniemi, K. Raju, and J. Karhunen. Jammer mitigation in DS-CDMA array system using independent component analysis. In *Proc. of the IEEE Int. Conference on Communications (ICC 2002)*, pages 232–236, New York, USA, April 28 - May 02, 2002.

[160] T. Ristaniemi, K. Raju, J. Karhunen, and E. Oja. Jammer cancellation in DS-CDMA array systems: Pre and post switching of ICA and RAKE. In *Proc. of the 2002 IEEE Int. Symposium on Neural Networks for Signal Processing (NNSP 2002)*, pages 495–504, Martingy, Switzerland, September 04 - 06, 2002.

[161] T. Ristaniemi, K. Raju, J. Karhunen, and E. Oja. Inter-cell interference cancellation in DS-CDMA array systems by independent component analysis. In *Proc. of the Fourth Int. Symp. on Independent Component Analysis and Blind Signal Separation*, pages 739–744, Nara, Japan, April, 2003.

[162] Z. Rong, P. Petrus, T. S. Rappaport, and J. H. Reed. Despread-respread multi-target constant modulus array for CDMA systems. *IEEE Communication Letters*, 1(4):114–116, 1997.

[163] R. H. Roy. Space division multiple access technology and its applications to wireless communication systems. In *Proc. of the 1997 IEEE Int. Vehicular Technology Conference*, volume 2, pages 730–734, Rhodes, Greece, May 1997.

[164] J. J. K. Ruanaidh and W. J. Fitzgerald. *Numerical Bayesian Methods Applied to Signal Processing*. Springer-Verlag, 1996.

[165] Y. Sanada and Q. Wang. A co-channel interference cancellation technique using orthogonal convolutional codes. *IEEE Transactions on Communications*, 44(5):549–556, 1996.

[166] J. Särelä and H. Valpola. Denoising source separation. *Journal of Machine Learning Research*, 6:233–272, 2005.

[167] G. I. Saulnier, P. Das, and L. B. Milstein. Suppression of narrowband interference in a PN spread-spectrum receiver using a CTD-based adaptive filter. *IEEE Transactions on Communications*, COM-32(11):1227–1232, 1984.

[168] G. I. Saulnier, P. Das, and L. B. Milstein. An adaptive suppression filter for direct-sequence spread-spectrum communications. *IEEE Journal on Selected Areas in Communications*, SAC-3(5):676–686, 1985.

[169] B. Sayadi and S. Marcos. A hybrid SIC/PIC detector based on reduced network of Kalman filters for DS-CDMA systems over multipath fading channels. In *Proc. of the IEEE Int. Conference on Acoustics, Speech and Signal Processing*, volume IV, pages 977–980, Sun City, South Africa, September 2004.

[170] R. A. Scholtz. The origins of spread-spectrum communications. *IEEE Transactions on Communications*, 30(5):822–854, May, 1982.

[171] T. J. Shan and T. Kailath. Adaptive beamforming for coherent signals and interference. *IEEE Transactions on Acoustics, Speech and Signal Processing*, ASSP-33:527–536, 1985.

[172] C. Shannon. A mathematical theory of communication. *Bell Sys. Tech. Journal*, 27:379–423,623–656, 1948.

[173] M. K. Simon, J. K. Omura, R. A. Scholtz, and B. K. Levitt. *Spread Spectrum Communications Handbook*. McGraw-Hill, revised edition, 1994.

[174] R. Steele. *Mobile Radio Communications*. Pantech Press, London, 1992.

[175] G. Strang and T. Nguyen. *Wavelets and Filter Banks*. Wellesley College, 1996.

[176] E. Ström, S. Parkvall, S. Miller, and B. Ottersten. Propagation delay estimation in asynchronous direct-sequence code division multiple access system. *IEEE Transactions on Communications*, 44(1):84–93, 1996.

[177] P. H. Tan and L. K. Rasmussen. Linear interference cancellation in CDMA based on iterative techniques for linear equation systems. *IEEE Transactions on Communications*, 48(12):2099–2018, 2000.

[178] A. S. Tannenbaum. *Computer Networks*. Prentice Hall, 4th edition, 2003.

[179] W. Teich and M. Seidl. Multiuser Detection based on a recurrent neural network structure. In *Proc. of the IEEE Int. Symposium on Spread Spectrum Techniques and Applications*, volume 3, pages 979–984, Mainz, Germany, September 1996.

[180] S. Theodosidis, N. Kalouptaidis, J. Proakis, and G. Koyes. Interference rejection in PN spread-spectrum systems with LS linear-phase FIR filters. *IEEE Transactions on Communications*, 37(9):991–994, 1989.

[181] TIA/EIA-95-B. *Mobile Station - Base Station Compatibility for Dual Mode Wideband Spread Spectrum Systems*. Telecommunication Industry Association, Washington, D.C.

[182] P. Tichavský, Z. Koldovský, and E. Oja. Performance Analysis of the FastICA Algorithm and Cramér-Rao Bounds for Linear Independent Component Analysis. *IEEE Transactions on Signal Processing*, 2006. in press.

[183] L. Tong and S. Perreau. Multichannel blind identification: From subspace to maximum likelihood methods. *Proc. IEEE*, 86(10):1951–1968, 1998.

[184] K. Torkkola. Blind separation for audio signals: are we there yet? In *Proc. Int. Workshop on Independent Component Analysis and Blind Separation of Signals (ICA 1999)*, pages 239–244, Aussois, France, January 1999.

[185] J. R. Treichler and M. G. Larimore. New processing techniques based on constant modulus algorithm. *IEEE Transactions on Acoustics, Speech and Signal Processing*, 33(2):420–431, 1985.

[186] C. Ung and R. H. Johnston. A Space Division Multiple Access Receiver. In *Proc. of the 2001 IEEE Int. Symposium of Antennas and Propagation Society*, volume 1, pages 422–425, Boston, USA, July 2001.

[187] H. Valpola and J. Särelä. Accurate, fast and stable denoising source separation algorithms. In *Proc. 5th Int. Conference on Independent Component Analysis and Blind Signal Separation (ICA'04)*, pages 65–72, Granada, Spain, September 2004.

[188] F. Van der Wijk, G. M. J. Janssen, and R. Prasad. Groupwise successive interference cancellation in a DS/CDMA system. In *Proc. of the IEEE Int. Symposium on Personal, Indoor, Mobile, Radio, Communications*, volume 2, pages 742–746, Toronto, Canada, September 1995.

[189] M. K. Varanasi and B. Aazhang. Multistage detection in asynchronous code division mulitple-access communications. *IEEE Transactions on Communications*, 38(4):509–519, 1990.

[190] M. K. Varanasi and B. Aazhang. Near-optimum detection in synchronous code division mulitple-access systems. *IEEE Transactions on Communications*, 39(5):725–736, 1991.

[191] B. D. Van Veen and K. M. Buckley. Beamforming: A versatile approach to spatial filtering. *IEEE ASSP Magazine*, (4):4–24, 1988.

[192] S. Verdú. *Multiuser Detection*. Cambridge University Press, 1998.

[193] N. Veselinovic. *Iterative Receivers for Interference Cancellation and Suppression in Wireless Communications*. Acta. Univ. Ouluensis, C-215, 2004.

[194] R. Vigário, V. Jousmäki, M. Hämäläinen, R. Hari, and E. Oja. Independent component analysis for identification of artifacts in magnetoencephalographic recordings. *Advances in Neural Information Processing Systems*, 10:229–235, 1998.

[195] R. Vigário, J. Särelä, V. Jousmäki, M. Hämäläinen, and E. Oja. Independent component approach to the analysis of EEG and MEG recordings. *IEEE Transactions on Biomedical Engineering*, 47(5):589–593, 2000.

[196] V. Vigneron, A. Paraschiv-Ionescu, A. Azancot, O. Sibony, and C. Jutten. Fetal electrocardiogram extraction based on non-stationary ICA and wavelet denoising. In *Proc. Int. Symposium on Signal Processing and its Applications (ISSPA'03)*, Paris, France, July 2003.

[197] A. J. Viterbi. Error bounds for convolutional codes and an asymptotically optimum decoding algorithm. *IEEE Transactions on Information Theory*, 13:260–269, 1967.

[198] A. J. Viterbi. *CDMA: Principles of Spread Spectrum Communication*. Addison Wesley, 1995.

[199] K. Waheed and F. M. Salem. Blind Information-Theoretic Multi-User Detection Algorithms for DS-CDMA and WCDMA Downlink Systems. *IEEE Transactions on Neural Networks*, 16(4):937–948, July 2005.

[200] X. Wang and H. V. Poor. Blind equalization and multiuser detection in dispersive CDMA channels. *IEEE Transactions on Communications*, 46(1):91–103, 1998.

[201] X. Wang and H. V. Poor. Blind multiuser detection: A subspace approach. *IEEE Transactions on Information Theory*, 44(2):677–690, 1998.

[202] X. Wang and H. V. Poor. *Wireless Communication Systems: Advanced Techniques for Signal Reception*. Printice-Hall, 2004.

[203] Y.-C. Wang and L. B. Milstein. Rejection of multiple narrowband interference in both BPSK and QPSK DS spread-spectrum systems. *IEEE Transactions on Communications*, COM-36(2):195–204, 1988.

[204] J. H. Wilkinson. *The algebraic eigenvalue problem*. Monographs on numerical analysis. Clarendon Press, London, 1965.

[205] Z. Xie, C. K. Rushforth, and R. T. Short. A family of suboptimum detectors for coherent multiuser communications. *IEEE Journal on Selected Areas in Communication*, 8(4):685–690, 1990.

[206] B. Yang. Projection approximation subspace tracking. *IEEE Transactions on Signal Processing*, 44(1):95–107, 1995.

[207] V. Zarzoso and A. K. Nandi. Blind separation of independent sources for virtually any source probability density function. *IEEE Transactions on Signal Processing*, 47(9):2419–2432, 1999.

[208] V. Zarzoso and A. K. Nandi. Adaptive blind source separation for virtually any source probability density function. *IEEE Transactions on Signal Processing*, 48(2):477–488, 2000.

[209] V. Zarzoso and A. K. Nandi. Non-invasive fetal electrocardiogram extraction: blind separation versus adaptive noise cancellation. *IEEE Transactions on Biomedical Engineering*, 48(1):12–18, 2001.

[210] V. Zarzoso, A. K. Nandi, F. Herrmann, and J. Millet-Roig. Combined estimation scheme for blind source separation with arbitary source pdfs. *Electronics Letters*, 37(2):132–133, 2001.

[211] H. H. Zeng, L. Tong, and C. R. Johnson Jr. Relations between the constant modulus and Wiener receivers. *IEEE Transactions on Information Theory*, 44(4):1523–1538, 1998.

[212] Y. Zhang, M. G. Amin, and A. R. Lindsey. Anti-jamming GPS receivers based on bilinear signal distributions. In *Proc. of the IEEE MILCOM Conference*, volume 2, pages 1070–1074, McLean, USA, October 2001.

[213] L. Zhao, M. G. Amin, and A. R. Lindsey. Subspace projection techniques for anti-FM jamming GPS receivers. In *Proc. of the IEEE Workshop on Statistical Signal and Array Processing*, pages 529–533, Pocono Manor, USA, August 2000.

[214] A. Ziehe and K.-R. Müller. TDSEP - an efficient algorithm for blind separation using time structure. In *Proc. Int. Conference on Artifical Neural Networks (ICANN'1998)*, pages 675–680, Skövde, Sweden, September 1998.

[215] Z. Zvonar and D. Brady. Multiuser detection in single-path fading channels. *IEEE Transactions on Communications*, 42(2/3/4):123–136, 1994.

Appendix A

Using Prior information for Interference Cancellation

The linear model in (4.1) expresses the observed data as a transformation of non-Gaussian and mutually independent latent variables. The noiseless case can be expressed as

$$\mathbf{X} = \mathbf{AS}, \tag{A.1}$$

where $\mathbf{X}$ is the matrix of observed random variables, $\mathbf{S}$ is the matrix of the latent variables, and $\mathbf{A}$ is an unknown constant matrix, called the mixing matrix. ICA assumes that the latent variables are independent. The fundamental conditions for the identifiably of this model are given in Chapter 4 and have been given in [27]. The most fundamental condition is that the independent components are non-Gaussian.

The only prior knowledge that this model assumes is on the distribution of the independent components which are assumed to be non-Gaussian (Of course, they are also assumed to be statistically independent.). Non-Gaussian variables can be roughly divided into two groups: super-Gaussian and sub-Gaussian variables, although slightly different definitions exist. In many cases, assumptions are usually made on this nature of the independent components. Such assumptions are made, for example, in image feature extraction [9, 74, 132], in which the components are assumed to be super-Gaussian, or sparse. This assumption is not arbitrary, but usually holds for independent components from image data.

Other than non-Gaussianity, assumptions of time correlations of the components S_i [10] can also lead to alternative estimation methods for the model.

Instead of prior knowledge on the components, prior knowledge on the mixing matrix [80] can be used in the basic ICA model (A.1). Use of priors on the mixing matrix gives the model a great generality. This knowledge on the mixing matrix is usually application specific. The mixing matrix can be assumed to be sparse in CDMA when the system is overloaded *i.e.*, the number of users is greater than the length of the spreading code. Using prior information on the mixing matrix is likely to give better estimates for a given sample size. This helps bring down the computational costs of ICA especially in situations where the computational costs of the estimation are so high that they severely restrict the amount of data that can be used, or in situations where the amount of data is restricted due to the nature of the application as in CDMA systems (due to fixed sizes of the block length).

This situation can be compared to that found in regression, where over learning is a very general phenomenon. The classical way of avoiding over learning in regression, *i.e.*, over fitting, is to use regularizing priors, which typically penalize regression functions that have large curvatures, *i.e.*, many "wiggles". This makes it possible to use regression methods even when the number of parameters in the model is very large compared with the number of observed data points. In the extreme theoretical case, the number of parameters is infinite, but the model can still be estimated from finite amounts of data by using prior information. Thus suitable priors can reduce over-learning [79].

Another example of using prior knowledge is in the literature of beamforming (see the discussion in [21]) where a very specific form of the mixing matrix is represented by a small number of parameters. Application of ICA to magnetoencephalography [194], results in independent components that can be modeled by the classic dipole model. This information can be used to constrain the form of the mixing coefficients [95]. However, the issue of using prior information is too application specific and may be applicable to a few data sets only.

A.1 Priors and modeling

The model in (A.1) can be rewritten as,

$$\mathbf{X} = \mathbf{f}(\mathbf{S}, \mathbf{A}), \tag{A.2}$$

where the function $\mathbf{f}$ relates the parameters $\mathbf{A}$ to the variables $\mathbf{S}$ and $\mathbf{X}$. This model is linear in (A.1), but this is not a pre-requisite.

The model can be estimated by the *m*aximum-a-posteriori principle (MAP). Using the Bayes' theorem [107], the parameters of the model can be related as

$$p(\mathbf{A}|\mathbf{X}) = \frac{p(\mathbf{A})p(\mathbf{X}|\mathbf{A})}{p(\mathbf{X})}, \tag{A.3}$$

where,

- $p(\mathbf{A})$ is the *prior* probability that summarizes all the knowledge of the values of the parameters prior to observing the data,
- $p(\mathbf{X}|\mathbf{A})$ is the *likelihood*, the probability of realizing the data given the value of the parameters $\mathbf{A}$,
- $p(\mathbf{X})$ the normalizing term called *evidence*
- $p(\mathbf{A}|\mathbf{X})$ is the *posterior* probability, that summarizes the state of knowledge about the value of the parameters after the data has been observed.

The MAP estimation provides a method for refining the estimation of the model. The likelihood is now complemented by the prior knowledge. This helps in the prevention of maximum likelihood from over learning. Thus the use of the Bayes' theorem provides a compact method for describing the estimation process.

A.1.1 Priors

In the following, it is assumed that the estimator $\mathbf{B}$ of the inverse of the mixing matrix $\mathbf{A}$ is constrained so that the estimates of the independent components $\mathbf{y} = \mathbf{Bx}$ are *white*, *i.e.*, decorrelated and of unit variance: $E\{\mathbf{y}\mathbf{y}^T\} = \mathbf{I}$. This

restriction facilitates greatly the analysis. For its justification, see *e.g.*, [27, 76]. The aim is to formulate priors for $\mathbf{B} = \mathbf{A}^{-1}$. Completely analogue results hold for priors on $\mathbf{A}$.

Flat Priors

Flat priors indicate a state of ignorance. They indicate no knowledge about the value of the parameter prior to observing the data. In this case, the priors can have a very flat, broad density function. An example of this kind of a prior is the *uniform* density function,

$$p(\mathbf{B}) = k, \tag{A.4}$$

where k is a constant. The estimates obtained by the use of such priors are identical to those obtained by maximum likelihood [164].

Jeffreys' prior

The classical prior in Bayesian inference is Jeffreys' prior. It is considered a maximally uninformative prior, which already indicates that it is probably not useful for our purpose.

Indeed, it was shown in [134] that Jeffreys' prior has the form,

$$p(\mathbf{B}) \propto |\det \mathbf{B}^{-1}|. \tag{A.5}$$

Now, the constraint of whiteness of the $\mathbf{y} = \mathbf{Bx}$ means that $\mathbf{B}$ can be expressed as $\mathbf{B} = \mathbf{UV}$, where $\mathbf{V}$ is a constant matrix, and $\mathbf{U}$ is restricted to be orthogonal. However, $\det \mathbf{B} = \det \mathbf{U} \det \mathbf{V} = \det \mathbf{V}$, which implies that Jeffreys's prior is constant in the space of allowed estimators (*i.e.*, decorrelating $\mathbf{B}$). Thus we see that Jeffreys' prior has no effect on the estimator, and therefore it cannot reduce over learning.

Quadratic priors

In regression, the use of quadratic regularizing priors is very common. The same concept can be used in ICA. This would require that the columns of $\mathbf{A}$ are smooth.

In other words, every column of $\mathbf{A}$ is considered to be a discrete approximation of a smooth function. The prior is chosen to impose smoothness for the underlying continuous function. Similar arguments hold for the priors on the rows of $\mathbf{B}$, *i.e.*, the corresponding separating matrix.

Quadratic priors are a simple class of regularizing priors. Consider the priors of the form,

$$\log p(\mathbf{B}) = \sum_{i=1}^{n} \mathbf{b}_i^T \mathbf{M} \mathbf{b}_i + const., \tag{A.6}$$

where the $\mathbf{b}_i^T$ are the rows of $\mathbf{B} = \mathbf{A}^{-1}$, and $\mathbf{M}$ is a matrix that defines the quadratic prior. For example, for $\mathbf{M} = \mathbf{I}$ there is a "weight decay" prior $\log p(\mathbf{B}) = \sum_i \|\mathbf{b}_i\|^2$.

Alternatively, it is possible to include in $\mathbf{M}$ some differential operators so that the prior would measure the "smoothness" of the $\mathbf{b}_i$, in the sense explained above. The prior can be manipulated algebraically to yield,

$$\sum_{i=1}^{n} \mathbf{b}_i^T \mathbf{M} \mathbf{b}_i = \sum_{i=1}^{n} \mathrm{tr}(\mathbf{M} \mathbf{b}_i \mathbf{b}_i^T) = \mathrm{tr}(\mathbf{M} \mathbf{B}^T \mathbf{B}). \tag{A.7}$$

Quadratic priors have little significance in ICA estimation, however. To see this, constraining the estimates of the independent components to be white, yields,

$$E\{\mathbf{y}\mathbf{y}^T\} = E\{\mathbf{B}\mathbf{x}\mathbf{x}^T\mathbf{B}^T\} = \mathbf{B}\mathbf{C}\mathbf{B}^T = \mathbf{I}, \tag{A.8}$$

in the space of allowed estimates, which gives after some algebraic manipulations $\mathbf{B}^T\mathbf{B} = \mathbf{C}^{-1}$. Now,

$$\sum_{i=1}^{n} \mathbf{b}_i^T \mathbf{M} \mathbf{b}_i = \mathrm{tr}(\mathbf{M}\mathbf{C}^{-1}) = const. \tag{A.9}$$

In other words, the quadratic prior is constant. The same result can be proven for a quadratic prior on $\mathbf{A}$. Thus, quadratic priors are of little interest in ICA.

A.1.2 Sparse priors on the mixing matrix

Another class of priors that could be more useful are priors that are sparse. In this case, the information is that most of the elements on each row of $\mathbf{B}$ are zero. Motivations for considering sparse priors are both empirical and algorithmic.

Empirically, in the field of feature extraction of images, the obtained filters tend to be localized in space. This means that the distribution of the elements of b_{ij} of the filters $\mathbf{b}_i$ are sparse *i.e.*, most elements are practically zero. This usually happens when the source signal is captured by a limited number of sensors as in magnetoencephalography or CDMA array processing. This is due to the spatial localization of the sources and the sensors.

The algorithmic appeal of sparsifying priors, on the other hand, is based on the fact that sparse priors can be made to be conjugate priors. This is a special class of priors, and means that estimation of the model using this prior requires only very simple modifications in ordinary ICA algorithms [78].

Measuring sparsity of mixing matrix

Sparsity of a random variable, say s, can be measured by expectations of the form $E\{G(s)\}$ [74], where G is a non-quadratic function, for example the following,

$$G(s) = -|s|. \tag{A.10}$$

The use of such measures requires that the variance of s is normalized to a fixed value, and the mean of s is zero.

Assuming that the data $\mathbf{x}_{orig}$ is *whitened* as a preprocessing step, the data is linearly transformed into $\mathbf{x} = \mathbf{V}\mathbf{x}_{orig}$ so that the covariance matrix of the whitened data equals identity: $E\{\mathbf{x}\mathbf{x}^T\} = \mathbf{I}$. $\mathbf{W}$ is used as the separating matrix applied to the whitened data.

Now, constraining the estimates $\hat{\mathbf{s}} = \mathbf{W}\mathbf{x}$ of the independent components to be white implies that $\mathbf{W}$ is orthogonal, which implies that the sum of the squares of the elements $\sum_j w_{ij}^2$ is equal to one for every i. The elements of each row of $\mathbf{W}$ can be then considered a realization of as a random variable of zero mean and unit variance. This means the sparsities of the rows of $\mathbf{W}$ can be measured using a sparsity measure of the form (A.10).

Thus, a sparse prior can be of the form,

$$\log p(\mathbf{W}) = \sum_{i=1}^{n} \sum_{j=1}^{n} G(w_{ij}) + const., \tag{A.11}$$

where G is the logarithm of some super-Gaussian density function (up to some additive constant), and again $\mathbf{w}_i^T = (w_{i1}, ..., w_{in})$ are the rows of $\mathbf{A}^{-1}$. The function G in (A.10) is such log-density, that measures the sparsity of $\mathbf{w}_i$ [78].

The prior in (A.11) has the convenient property of being a conjugate prior. Assume that the *independent components are super-Gaussian*, for simplicity, and further assume that they have identical distributions, with log-density G and the same log-density as the log-prior density G in (A.11). Now, the prior can be written in the form,

$$\log p(\mathbf{W}) = \sum_{i=1}^{n} \sum_{j=1}^{n} G(\mathbf{w}_i^T \mathbf{e}_j) + const., \tag{A.12}$$

where $\mathbf{e}_i$ are the canonical basis vectors, *i.e.*, the ith element of $\mathbf{e}_i$ is equal to one, and all the other elements are zero.

Assume now that there are T whitened observations $\mathbf{x}(t), t = 1, ..., T$. The likelihood of $\mathbf{W}$ that is constrained to be orthogonal is simply given by [77, 78],

$$\log L(\mathbf{W}) = \sum_{t=1}^{T} \sum_{i=1}^{n} G(\mathbf{w}_i^T \mathbf{x}(t)) + const. \tag{A.13}$$

Thus the posterior distribution has the form

$$\log p(\mathbf{W}|\mathbf{x}) = \sum_{i=1}^{n} [\sum_{t=1}^{T} G(\mathbf{w}_i^T \mathbf{x}(t)) + \sum_{j=1}^{n} G(\mathbf{w}_i^T \mathbf{e}_j)] + const. \tag{A.14}$$

This form shows that the posterior distribution has the same form as the prior distribution (and, in fact, the original likelihood). Priors with this property are called conjugate priors in Bayesian theory. The usefulness of conjugate priors resides in the property that the prior can be considered to correspond to a "virtual" sample. The posterior distribution in (A.14) has the same form as the likelihood of a sample of size $T + n$ which consists of both the observed $\mathbf{x}(t)$ and the canonical basis vectors $\mathbf{e}_i$. In other words, the posterior in (A.14) is the likelihood of the augmented (whitened) data sample,

$$\mathbf{x}^*(t) = \begin{cases} \mathbf{x}(t), & \text{if } 1 \leq t \leq T, \\ \mathbf{e}_{t-T}, & \text{if } T < t \leq T + n. \end{cases} \tag{A.15}$$

Thus, using conjugate priors has the additional benefit that exactly the same algorithm for maximization of the posterior as in ordinary maximum likelihood

estimation of ICA can be used. All that is needed is to add the virtual sample to the data; the virtual sample is of same size n as the dimension of the data.

There could be sub-Gaussian (anti-sparse) conjugate priors just in the same way as sparse priors. The property of being conjugate is more general concept [164]. However, the utility of non-sparse conjugate priors is not as obvious as the utility of sparse priors and often has to be justified by the area of application.

Modifying prior strength

The conjugate priors given above can be generalized by considering a family of super-Gaussian priors given by

$$\log p(\mathbf{W}) = \sum_{i=1}^{n} \sum_{j=1}^{n} \alpha G(\mathbf{w}_i^T \mathbf{e}_j) + const. \tag{A.16}$$

Using this kind of prior means that the virtual sample points are weighted by some parameter α. This parameter expresses the degree of belief on the prior. A large α means that the belief in the prior is strong. Also, the parameter α could be different for different i, but this seems less useful here. The posterior distribution has then the form [78]

$$\log p(\mathbf{W}|\mathbf{x}) = \sum_{i=1}^{n} [\sum_{t=1}^{T} G(\mathbf{w}_i^T \mathbf{x}(t)) + \sum_{j=1}^{n} \alpha G(\mathbf{w}_i^T \mathbf{e}_j)] + const. \tag{A.17}$$

The above expression can be further simplified in the case where the assumed density of the independent components is Laplacian, *i.e.*, $G(y) = -|y|$. In this case, the α can multiply the $\mathbf{e}_j$ as

$$\log p(\mathbf{W}|\mathbf{x}) = \sum_{i=1}^{n} [\sum_{t=1}^{T} |\mathbf{w}_i^T \mathbf{x}(t)| - \sum_{j=1}^{n} |\mathbf{w}_i^T (\alpha \mathbf{e}_j)|] + const., \tag{A.18}$$

which is simpler than (A.17) from the algorithmic viewpoint: it amounts to the addition of just n virtual data vectors of the form $\alpha \mathbf{e}_j$ to the data. This avoids all complications due to the differential weighting of sample points in (A.17), and ensures that any conventional ICA algorithm can be used by simply adding the virtual sample to the data. In fact, the Laplacian prior is most often used in ordinary ICA algorithms, sometimes in the form of the log cosh function that can be considered as a smoother approximation of the absolute value function.

The estimation of a suitable α is a further problem. A trial and error approach is typically needed.

Priors and whitening

When the data are whitened, the effect of the sparse prior is dependent on the whitening matrix. This is because sparseness is imposed on the separating matrix of the whitened data, and the value of this matrix depends on the whitening matrix. There is an infinity of whitening matrices, so imposing sparseness on the whitened separating matrix may have different meanings.

On the other hand, it is not necessary to whiten the data. The above framework can be used for non-white data as well. If the data are not whitened, the meaning of the sparse prior is somewhat different, though. This is because every row of $\mathbf{b}_i$ is not constrained to have unit norm for general data. Thus the measure of sparsity does not any more measure the sparsities of each $\mathbf{b}_i$. On the other hand, the developments of the preceding section show that the sum of squares of the whole matrix $\sum_{ij} b_{ij}^2$ does stay constant. This means that the sparsity measure is now measuring the global sparsity of $\mathbf{B}$, instead of the sparsity of each individual row.

In practice, whitening the data is done for technical reasons. Then the problems arises of how to impose the sparseness on the original separating matrix even when the data used in the estimation algorithm needs to be whitened. The above framework can be easily modified so that the sparseness is imposed on the original separating matrix. Denote by $\mathbf{V}$ the whitening matrix and by $\mathbf{B}$ the separating matrix for original data. Thus, $\mathbf{WV} = \mathbf{B}$ and $\mathbf{x} = \mathbf{Vx}$ by definition. Now, the prior (A.12) can be expressed as [78]

$$\log p(\mathbf{B}) = \sum_{i=1}^{n} \sum_{j=1}^{n} G(\mathbf{b}_i^T \mathbf{e}_j) + const. = \sum_{i=1}^{n} \sum_{j=1}^{n} G(\mathbf{w}_i^T (\mathbf{V}\mathbf{e}_j)) + const. \quad \text{(A.19)}$$

Thus, the virtual sample added to $\mathbf{x}(t)$ now consists of the columns of the whitening matrix, instead of the identity matrix.

Notice that a similar manipulation of (A.12) shows how to put the prior on the original mixing matrix instead of the separating matrix. Since $\mathbf{VA} = (\mathbf{W})^{-1} = \mathbf{W}^T$, then $\mathbf{a}_i^T \mathbf{e_j} = \mathbf{a}_i^T \mathbf{V}^T (\mathbf{V}^{-1})^T \mathbf{e_j} = \mathbf{w}_i^T (\mathbf{V}^{-1})^T \mathbf{e}_j$. This shows that imposing a sparse prior on $\mathbf{A}$ is done by using the virtual sample given by the rows of

the inverse of the whitening matrix. (Note that for whitened data, the mixing matrix is the transpose of the separating matrix, so the fourth logical possibility of formulating prior for the whitened mixing matrix is not different from using a prior on the whitened separating matrix.)

In practice, the problems implied by whitening can often be solved by using a whitening matrix that is sparse itself. Then imposing sparseness on the whitened separating matrix is meaningful. In the context of image feature extraction, a sparse whitening matrix is obtained by the zero-phase whitening matrix, for example (see [9] for discussion). Then it is natural to impose the sparseness for the whitened separating matrix, and the complications discussed in this subsection can be ignored.

A.1.3 Estimation using priors

An example of the effect of priors on image feature extraction is examined here. The essential idea is as in [9, 74, 132]. The data was obtained by taking 20×20 pixel image patches at random locations from monochrome photographs depicting wild-life scenes (animals, meadows, forests, *etc*,.). The patches are normalized to unit norm. The data are whitened by the zero-phase whitening filter, which means multiplying the data by $\mathbf{C}^{-1/2}$, where $\mathbf{C}$ is the covariance of the data. (see [9]). In the results shown above, the inverse of these preprocessing steps is performed.

The sample size is fixed at 20 000. This is insufficient for such a large window size. The estimated basis vectors are shown in Fig. A.1 (For reasons of space, only 200 of the 400 basis vectors are shown; these are randomly selected). Fixing the prior strength parameter α to 25, a much better bases is obtained as seen in Fig. A.2. Visually, the features are much better. The features are oriented, as well as localized in space and in frequency. The aspect of multiresolution seems to be less developed than in results with sufficient data however [9, 132].

A.2 Using priors for interference cancellation

Given the fact that prior knowledge helps in the estimation of both the mixing matrix and the sources, it is natural to examine ways of incorporating prior

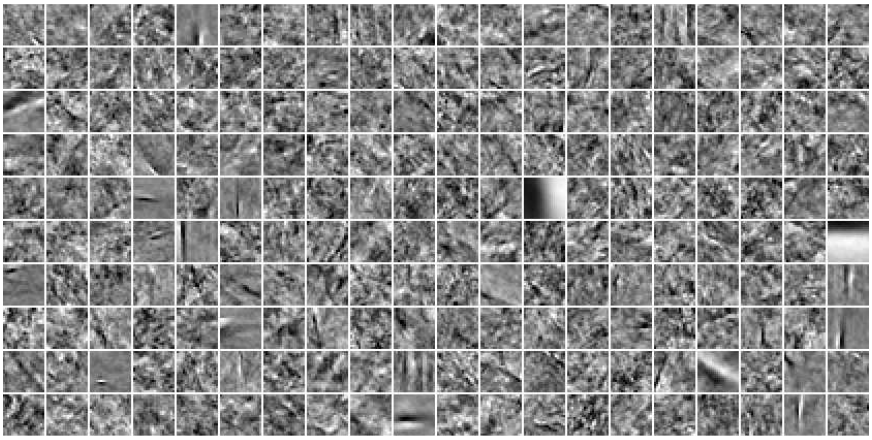


Figure A.1: *Estimation of the image features with no prior information. The sample size is insufficient to give useful estimates.*

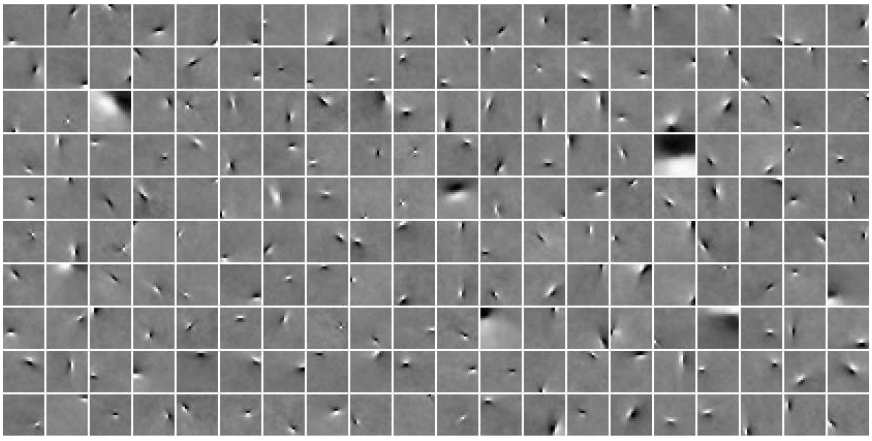


Figure A.2: *Estimation of the image features with suitable prior information. The estimation is successful even with this small sample size.*

knowledge in the interference cancellation algorithms described in this thesis. The main motivation for the use of prior knowledge are the following:

- enhanced interference suppression;
- requirement of smaller sample sizes for a target BER;
- faster convergence of the ICA algorithm.

Enhanced interference suppression would naturally be beneficial. ICA-RAKE Post-Switch already offers good interference suppression capabilities as seen in Fig 4.27. Examining the figure closely, it can be seen that the change in BER is not drastic from about -30 dB SJR to -10 dB. This is in spite of the fact that ICA-RAKE Post-Switch offers the best solution among the detectors considered. Improvement of the performance further will be a stronger motivation for practical consideration of semi-blind algorithms.

Again referring to Fig. 4.27, the gain at very low SJR comes at the cost of increasing the block size. From Figs. A.1 and A.2, it can be seen that the addition of "sparse" priors to the data facilitated in estimation of basis functions with a sample size of 20000, a size insufficient for estimation without the aid of priors. The same effect should be visible in CDMA systems, too. Reduction of the block size from $M = 5000$ or $M = 1500$ (Fig. 4.32) to sizes $M < 500$ should aid in the adaptation of the algorithm to on-line implementations.

The third motivation for use of priors is faster convergence. From the experiments conducted on images it is found that the use of priors increases the speed of convergence. At this stage this claim is somewhat heuristic, and further investigation of the convergence is needed. In any case, increase in the speed of convergence should again aid on-line implementations.

A.2.1 Maximum likelihood based ICA

The ICA method described in Sec. 4.1.7 is based on kurtosis. The approach, though practical, and justifiable from the statistical point of view, is still heuristic. The addition of prior information, as developed in the previous sections require a maximum likelihood based approach.

The fixed point algorithm (4.17) in Sec. 4.1.7, can now be modified for maximum likelihood estimation as [75],

$$\mathbf{W} \leftarrow \mathbf{W} + \mathrm{diag}(\beta_i)[\mathrm{diag}(\gamma_i) + E\{\mathbf{g}(\mathbf{s})\mathbf{s}^T\}], \tag{A.20}$$

where,

$$\begin{aligned} \beta_i &= -E\{s_i g(s_i)\}, \quad \text{for} \quad i = 1, \ldots, N, & \text{(A.21)} \\ \gamma_i &= -1/(\beta_i + E\{g'(s_i)\}), \quad \text{for} \quad i = 1, \ldots, N, & \text{(A.22)} \end{aligned}$$

assuming $\mathbf{s} = \mathbf{W}\mathbf{x}$. The non-linearity g is typically chosen as

$$\begin{aligned} g(s) &= \tanh(as), & \text{(A.23)} \\ g'(s) &= a(1 - \tanh^2(as)). & \text{(A.24)} \end{aligned}$$

Hence, with a small modification, the algorithm in (4.17) can perform ML estimation, and can be used along with priors.

A.2.2 Choosing suitable priors

Finding the correct prior is a non-trivial matter. In the previous application, the prior is assumed to be sparse, due to the fact that the estimated filters are localized in space, and the distribution of the elements of the filter is sparse. This is not the case in the model considered for interference cancellation. In this case possible priors need to be constructed for the specific problem.

A simple way of using priors is to use the whitening matrix as a possible prior. The motivation for the use of this prior is the fact that the new mixing matrix is now orthogonal. If $\mathbf{x}$ is the whitened data, then it follows,

$$E\{\mathbf{x}\mathbf{x}^T\} = \tilde{\mathbf{A}} E\{\mathbf{s}\mathbf{s}^T\} \tilde{\mathbf{A}} = \tilde{\mathbf{A}} \tilde{\mathbf{A}}^T = \mathbf{I}, \tag{A.25}$$

where $\tilde{\mathbf{A}} = \mathbf{V}\mathbf{A}$. This prior essentially restricts the search for a mixing matrix to the space of orthogonal matrices.

On the other hand, if some information is available on the channel *e.g.*, fading characteristics, this information can be used as priors. Assuming that $\mathbf{G}$ contains some raw estimate of the channel, the canonical basis vectors that form the virtual sample can be chosen as,

$$\mathbf{e}_t = \Delta^{-1/2} \mathbf{E}^T \mathbf{G}, \tag{A.26}$$

where Δ and $\mathbf{E}$ are the matrices containing the eigenvalues and the eigenvectors of the covariance matrix of the data $\mathbf{x}_{orig}$. This is the whitened (prior) channel state matrix. Crude information on the array characteristics can also be incorporated into the prior using the same technique.

Sparse priors of the form (A.15) can also be used in interference cancellation problems. This is possible when there is a strong belief in the sparseness of the mixing matrix. This is usually the case in overloaded systems [72]. This is a case where the number of independent components (*i.e.*, users) exceeds an upper bound $\frac{2}{3}C$, the maximum number of allowable users from the sampled single antenna DS-CDMA model. Preliminary experiments show promising results with the use of priors.

A.2.3 Examples of interference cancellation with priors

With the whitening matrix as a possible prior, preliminary simulations try to illustrate the use of priors. The task helps in understanding the role of weight α and slope a in the separation of the unknown jammer.

Scenario 11: Variation of α, a in interference cancellation. The setup is similar to **Scenario 3**. For each value of the slope $a \in [0.1, 0.2, \ldots, 11.4]$, the prior strength α is varied in order to find the optimal values. The block size $M = 5000$ and the detection algorithm is the Post-Switch algorithm.

In Fig. A.3, the BER for all values of α and a is shown. It can be seen that lower values of α are preferred, while the slope seem to have no influence on the results. Keeping $\alpha = 0.1$, the BER obtained for different values of the slope is shown in Fig. A.4. It can be seen that $a = 0.6$ is the ideal value, though the scale is too fine.

The motivation for using 0.6 as the slope for the contrast function stems from the fact that it is neither too fast nor too slow a function as seen in right hand plot of Fig. A.4, which plots the values of the non-linearity against different slopes a. Very strong values of α are not preferred usually. With a very strong value for α, the roles of the data and the prior are often interchanged with the data acting as a prior on the virtual sample. This leads to the estimation of the virtual sample, which is often not desired. Now, the priors do not regulate over-learning.

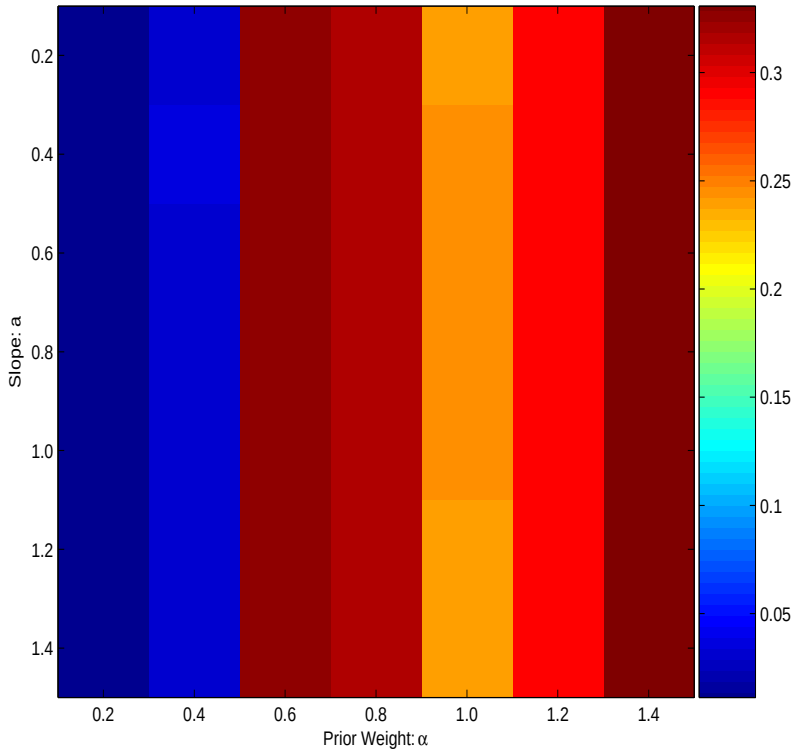


Figure A.3: *Variation of the prior strength and the slope of the contrast function. For this application, smaller prior strengths seem to provide the best results while the slope does not seem to affect the results.*

A.2.4 Where to use priors in CDMA?

Use of priors is always suited when the amount of the data is insufficient (compared with the number of independent variables) so that it leads to over-learning. In the above example with interference cancellation, the number of users is low (max $C = 31$). This is because short spreading codes are used to spread the data. These issues become more interesting when long spreading codes (*i.e.*, $C = 128$) are used. In this case, naturally, the system can support more users, and given that the block sizes are fixed and small for reasonable estimation of the statistical

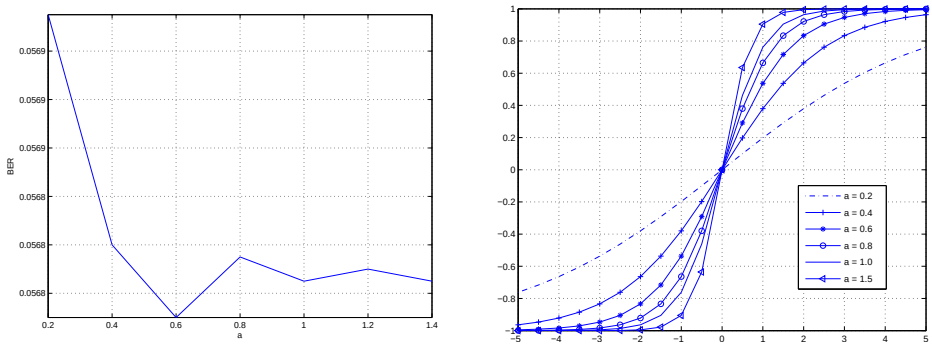


Figure A.4: *Variation of slope. The left hand plot shows the BER as a function of the slope of* tanh*. While the scale is fine, it can still be observed that* $a = 0.6$ *represents the minimum in the BER curves. On the right hand side, the value of the non-linearity is plotted against different values of the slope.* $a = 0.6$ *represents a curve that is neither too fast or too slow.*

parameters, priors could have a considerable impact on the performance of the detectors.

Appendix B

Simulation Details

All the simulations in this thesis are performed using a simulator developed for this purpose. The simulator is written in *Matlab* and *Python*. The simulator consists of three essential parts:

1. transmitter,
2. channel,
3. receiver.

The transmitter modules consist of a source generator, a modulator, and a spread spectrum generator. Vector modulation schemes are used in the modulator. *Quadrature phase shift keying* (QPSK) is the modulation mostly used in all the simulations. The number of users K, and the block size M are the variable parameters here. The spread spectrum generator generates Gold codes of length $C = 31$. Other codes such as Walsh or Hadamarad can also be generated and used.

The channel simulator modules generate different kinds of channels *e.g.*, an AWGN, or a Rayleigh, or a Rician channel of L significant paths. The interference structure and its strength (SJR or SIR) and the noise level (SNR) are the free parameters. Other parameters like the speed of the terminal, and the frequency of operation can also be varied to generate the appropriate fading co-efficients. The channel simulator is a generic block, and many of its functionalities have not been used in the thesis (*e.g.*, fading channels, *etc.*).

The receiver forms the test bed for all the algorithms discussed. They are classified as: fixed; adaptive; semi-blind, and blind. Fixed receivers are the conventional sign detector, MF, decorrelating detector and the MMSE receiver. Adaptive receivers are based on the LMS principle. The semi-blind detectors are the ICA-RAKE, ICA-RAKE Pre-Switch, ICA-RAKE Post-Switch, and the IC-DSS detectors. These detectors form the core of this thesis and are discussed in Secs. 4.3, 4.6, and 4.7 of Chapter 4. The blind detector is the IC-DSS detector described in Sec. 5.2 of Chapter 5.

B.1 Scenarios

This thesis presents experiments in the form of scenarios. Each scenario assumes that a certain parameter is varied, while all the other parameters remain constant. Most of the experiments consist of varying the SJR or the SNR. Based on the number of parameters, the following scenarios can be summarized:

1. varying SJR, all others being constant,
2. varying the SJR with a different interferer (continuous jamming),
3. varying the SJR with a different interferer (bit-pulsed jamming),
4. varying SNR, all others being constant,
5. varying the nature of the interference (interference due to adjacent cells),
6. varying the cell load, all others being constant,
7. varying the block size M, all others being constant,
8. varying the number of coherent paths L, all others being constant,
9. varying the source separation algorithm,
10. determining the optimal block size M, with all other parameters constant.

HELSINKI UNIVERSITY OF TECHNOLOGY
DISSERTATIONS IN COMPUTER AND INFORMATION SCIENCE

Report D1	Koskela, M., Interactive Image Retrieval Using Self-Organizing Maps, 2003.
Report D2	Sinkkonen, J., Learning Metrics and Discriminative Clustering, 2003.
Report D3	Hurri, J., Computational Models Relating Properties of Visual Neurons to Natural Stimulus Statistics, 2003.
Report D4	Bingham, E., Advances in Independent Component Analysis with Applications to Data Mining, 2003.
Report D5	Himberg, J., From Insights to Innovations: Data Mining, Visualization and User Interfaces, 2004.
Report D6	Särelä, J., Exploratory Source Separation in Biomedical Systems, 2004.
Report D7	Peltonen, J., Data Exploration with Learning Metrics, 2004.
Report D8	Könönen, V., Multiagent Reinforcement Learning in Markov Games: Asymmetric and Symmetric Approaches, 2004.
Report D9	Inki, M., Extensions of Independent Component Analysis for Natural Image Data, 2004.
Report D10	Honkela, A., Advances in Variational Bayesian Nonlinear Blind Source Separation, 2005.
Report D11	Nikkilä, J., Exploratory Cluster Analysis of Genomic High-Throughput Data Sets and their Dependencies, 2005.
Report D12	Seppänen, J. K., Using and Extending Itemsets in Data Mining: Query Approximation, Dense Itemsets, and Tiles, 2006.
Report D13	Creutz, M., Induction of the Morphology of Natural Language: Unsupervised Morpheme Segmentation with Application to Automatic Speech Recognition

ISBN 951-22-8260-7

ISSN 1459-7020