Blind Multiuser Detection Algorithms for Time-Varying Channels with Diversity Antenna Arrays

by

Jing Wang

M.A.Sc. (EE), Simon Fraser University, 1990

M.A.Sc. (Physics), Case Western Reserve University, 1986

B.A.Sc. (Physics), University of Science and Technology of China, 1984

a thesis submitted in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
in the School
of
Engineering Science

SIMON FRASER UNIVERSITY
April 2001

APPROVAL

Name: Jing Wang

Degree: Doctor of Philosophy

Title of thesis: Blind Multiuser Detection Algorithms for Time-Varying Channels with Diversity Antenna Arrays

Examining Committee: Dr. Mehrdad Saif
Chair

_____

Dr. James K. Cavers
Senior Supervisor

_____

Dr. Shawn Stapleton
Supervisor

_____

Dr. Ljiljana Trajkovic
Supervisor

_____

Dr. Aaron Gulliver
Internal Examiner, University of Victoria

_____

Dr. Max Wong
External Examiner, McMaster University

Date Approved: _____

Abstract

Wireless communications have witnessed an explosive growth in the past fifteen years in the number of subscribers, the type of services and the range of coverage. An important challenge for wireless communication today is an ever increasing demand for more services and relatively limited spectrum resources. Another important aspect of the wireless communication systems is that channels suffer multipath fading, which may be time-varying due to user mobility. Antenna arrays offer a promising solution for increasing the spectral efficiency of multiple access networks over wireless communication channels. The spatial diversity provided by antenna arrays allows multiple users to share the same channel. In this thesis, we focus on developing blind space-time algorithms for detecting multiple cochannel users signals received at a diversity antenna array over a time and frequency dispersive fading channel.

We assume that cochannel signals are digitally modulated with the same symbol periods and alphabet. The signals arrive at the array through a time-varying multipath propagation channel. The array output is modeled as a linear combination of signals from all the users, and additive noise. Our objective in blind multiuser detection is to detect each of the user sequences without training sequences. We achieve this goal by i) exploiting the finite alphabet (FA) property of the signals and ii) modeling the channel dynamics by a low order polynomial or as a stochastic process of known statistics.

The blind estimation problem is studied in stages. We propose two general approaches - the blind multiuser maximum likelihood (BMML) detection and blind multiuser Bayes (BMB) detection. We first develop the BMML detector for memoryless channels with negligible delay spread and for linearly and nonlinearly modulated signals. We then develope the BMB detector for linearly modulated signals. We study the uniqueness of these detection

methods and propose suboptimal but computationally e±cient algorithms for these detectors. E®ective initialization routines are also developed for robust detection. Performance is evaluated through numeric simulation.

We next consider multipath channels with nonzero delay spread. In this case, each antenna output is temporally oversampled at a rate higher than the symbol rate. A blind multiuser detection algorithm is proposed to tackle this most general channel scenario - time varying multipath channels. By exploiting the structure of the temporally oversampled and stacked input matrix by subspace projection and by exploiting the FA property of the data signals, the algorithm equalizes the multipath channels and detects multiple users using an extended BMML detection. The e®ectiveness of this algorithm is studied by simulation.

Acknowledgements

I am extremely grateful to my senior supervisor Professor Jim Cavers, for introducing and motivating me to the area of antenna array processing and blind signal detection for wireless communication systems, and for providing me with continuous feedback and encouragement during the entire Ph.D. program. His patience and optimism undoubtedly helped me to overcome many tough times. Also his generous ¯nancial support was integral to this work. I also thank him for many fruitful discussions and suggestion pertinent to this work.

I would like to thank Professors Shawn Stapleton, Ljiljana Trajkovic, Aaron Gulliver of University of Victoria, Max Wong of McMaster and Mehrdad Saif for taking the time to serve on my defense committee.

I would like to thank all my friends at Simon Fraser University for making my stay here an enjoyable one.

I would like to acknowledge partial support of this research from Glenayre RD Inc.

Finally, I would like to thank my wife, Hua, for her constant support and her willingness to always listen and understanding. I am indebted to my parents, for their emotional support and for giving their time and energy.

Dedication

To my parents.

Contents

List of Tables

List of Figures

List of Abbreviations

AR	Autoregressive (process)
ARMA	Autoregressive moving average (process)
BER	Bit error rate
BIR	Blind initialization routine
BMB	Blind multiuser Bayes (detection)
BMDF	Blind multiuser decision feedback (detection) (variant of BMB)
BMML	Blind multiuser ML (detection)
BMPSP	Blind multiuser PSP (detection) (variant of BMB)
BMVA	Blind multiuser VA (detection) (variant of BMB)
CCI	Cochannel interference
CDMA	Code division multiple access
CMA	Constant Modulus Algorithm or Complex Multiplication and Addition
CPFSK	Continuous phase FSK
CPM	Continuous phase modulation
CS	cyclostationarity
DF	Decision feedback (detection)
DOA	Direction of arrival
DQPSK	¼=4-shift di®erentially encoded PSK
DS	Data structure (condition)
FA	Finite Alphabet
FDMA	Frequency division multiple access
FIR	Finite impulse response
FSK	Frequency shift keying modulation
GSM	Global system of mobile communication
HOS	High order statistics

IDA	I.D. assisted (initialization)
ILSE	Iterative least-squares with enumeration (detection)
ISI	Intersymbol interference
MA	Moving average (process)
PAM	Phase-amplitude modulation
MAP	Maximum a posteriori
MCD	Multiuser channel deconvolution
MI	Multiple input
MLSE	Maximum-likelihood sequence estimate
ML	Maximum likelihood (detection)
MO	Multiple output
MRDF	Memory reduced decision feedback (detection) (variant of BMB)
PCS	Personal communication systems
PDA	Personal digital assistant
pdf	Probability density function
PSK	Phase shift keying modulation
QAM	Quadrature amplitude modulation
RBD	Random with backward detection
RIR	Random initialization routine
PSP	Per survivor processing
SBJD	Subspace based joint detection
SBJD/S	Static channel SBJD
SBSD	Subblock by subblock detection (variant of BMML)
SI	Single input
SNR	Signal to noise ratio
SO	Single output
SR	Sample rate (condition)
SWFD	Slide window feedback detection (variant of BMML)
TDMA	Time division multiple access
TVMCD	Time-varying MCD
TVILSE	Time-varying ILSE
UM	Unique mapping (condition)

VA	Viterbi algorithm
WLAN	Wireless local area network
WSSUS	Wide sense uncorrelated scattering

List of Symbols

$a_{jil}(t)$	Channel gain for the lth path of the ith user to the jth antenna at time t
$a_{ji}(n)$	Channel gain for the ith user to the jth antenna at time t
$a_{¢il}(n)$	Channel gain column vector of size m; $a_{¢il}(n) = (a_{1il}(n); :::; a_{mil}(n))^T$
a_{jil}	Channel gain row vector of size n; $a_{jil} = (a_{jil}(1); :::; a_{jil}(n))$
$a_{j¢l}(n)$	Channel gain row vector of size d; $a_{j¢l}(n) = (a_{j1l}(n); :::; a_{jdl}(n))$
a_j	Channel gain row vector of size dn; $a_j = (a_{j1}; :::; a_{jd})$
$A_l(n)$	Channel gain matrix of size m-by-d; $A_l(n) = (a_{¢1l}(n); :::; a_{¢dl}(n))$
A_{aug}	Augmented channel gain matrix of size m-by-dp; $A_{aug} = (A_0; :::; A_p)$
A_k	kth order polynomial channel model coe±cient matrix of size m-by-d
$\hat{A}_{poly}$	Polynomial channel gain estimate of size m-by-d
d	Number of multiple cochannel users
D(v)	Diagonal matrix of the vector v of size n; D(v) =diag($v_1; v_2; :::; v_n$)
E_E	Equivalence transformation matrix
E_p	Permutation matrix
E_r	Phase rotation matrix
f_dT	Fade rate
f(n)	Predictor ¯lter of size M
g(t)	Pulse shape function
h	Modulation index
$h_{jil}(t; t')$	Time-varying channel impulse response for path l of user i to antenna j
$h_i(n; n')$	Vector channel impulse response of size m
$h_i^k(n')$	kth polynomial expansion of the channel impulse response vector
$H^k(n')$	kth polynomial expansion of the channel impulse response matrix
$H_{aug}(n')$	Augmented channel impulse response matrix

$H_{aug}(K)$	Smoothed augmented channel matrix H_{aug} by K
$J_0(\cdot)$	Zero-th order Bessel function of the first kind
K	Smoothing factor
L_i	Total number of paths for the ith user
L_{pulse}	Duration of the pulse shape in symbol periods
m	Number of antennas
M	VA memory length
M_{VA}	PSP state set
M_{surv}	Survivor length
N	block size in symbols
N_{sub}	Sub-block size in symbols
p	Polynomial order
P	Sample rate
q(t)	Frequency shape pulse
Q	Data alphabet size
r_{MLS}	Mean square value of LS estimation error
$R_{a_{ji}}$	Channel autocorrelation (from user i to antenna j) matrix of size n-by-n
$R_j(S)$	Conditional input correlation matrix at antenna j of size m-by-m
$s_i(n)$	Transmitted nth symbol from ith user
s(n)	Data column vector of size d, $s(n) = (s_1(n); \cdots; s_d(n))^T$
s_i	Data row vector of size n, $s_i = (s_i(1); \cdots; s_i(n))$
$s_{aug}(n)$	Column vector of $S_{aug}(n)$, size d(p + 1)
S(n)	Data matrix of size d-by-n, $S(n) = (s(1); \cdots; s(n))$
$S_{aug}(n)$	Augmented data matrix of size d(p + 1)-by-n
$S_k(n)$	kth submatrix of $S_{aug}(n)$, size d-by-n, $S_k(n) = \begin{bmatrix} s(1); 2^k s(2); \cdots; n^k s(n) \end{bmatrix}$
$S_{aug}(K)$	Smoothed augmented data matrix S_{aug} by K
tr (·)	Trace of a matrix
$w_j(t)$	Equivalent complex baseband additive white Gaussian noise
x(n)	Input column vector of size m; $x(n) = (x_1(n); \cdots; x_m(n))^T$
$x_j(t)$	Received complex baseband signal at antenna j
x_j	Input row vector of size n, $x_j = (x_j(1); \cdots; x_j(n))$
X(K)	Smoothed input matrix X by K

$X(n)$	Input matrix of size m-by-n, $X(n) = (x(1); :::; x(n))$
$y_i(t)$	Transmitted complex baseband signal waveform of user j
¿	Discrete sample time within a symbol period
$¿_{jil}(t)$	Path delay for path l of user i to antenna j at time t
T_i	Delay spread for user i
T	Maximum delay spread
$¾_w^2$	Noise variance
$(¢)^¤$	Complex conjugate
$(¢)^T$	Transpose
$(¢)^H$	Hermitian transpose
$(¢)^\#$	Moore-Penrose pseudo-inverse
$j¢j^2$	Square norm or scalar product of a vector
$k¢k_F$	Frobenius norm of a matrix

Chapter 1

Introduction

1.1 General Review of Wireless Communication Networks

Telecommunication has been experiencing an explosive growth in the number of subscribers, coverage and type of services in the last fifteen years and the trend is expected to continue for at least the next ten years. At the end of 1998, there were about 305 million mobile subscribers throughout the world, and predictions show that there will be over 700 million users by the end of 2003 and over 1 billion wireless telephone users world-wide by the year 2005 [1]. Cellular communication systems have also evolved rapidly over a very short period of time: from the first (analog) generation technology (1G) in the mid 80's to the second (digital) generation (2G) in the early 90's to the third (wideband digital) generation (3G) beginning next year. While the 1G and 2G (including PCS) provide voice and limited data services, 3G will provide much extended data services, including multimedia via wireless Internet access. The types of hand held devices range from simple phone to phone with wireless Internet access and personal digital assistant (PDA) with wireless web-browsing capability. To accommodate these demands (both the number of users and type of services) a large frequency bandwidth is required. However, frequency spectrum allocated for wireless services is a limited physical resource. A challenge for wireless communication systems is to meet the increase in demand within the available spectrum. Efficient spectrum utilization and spectrum sharing are two key approaches to solve the spectrum drought problem. Techniques that increase spectrum efficiency/sharing are therefore critical. Innovations that have played critical roles in today's wireless communications include spectrally efficient digital modulation and cellular telephony technology.

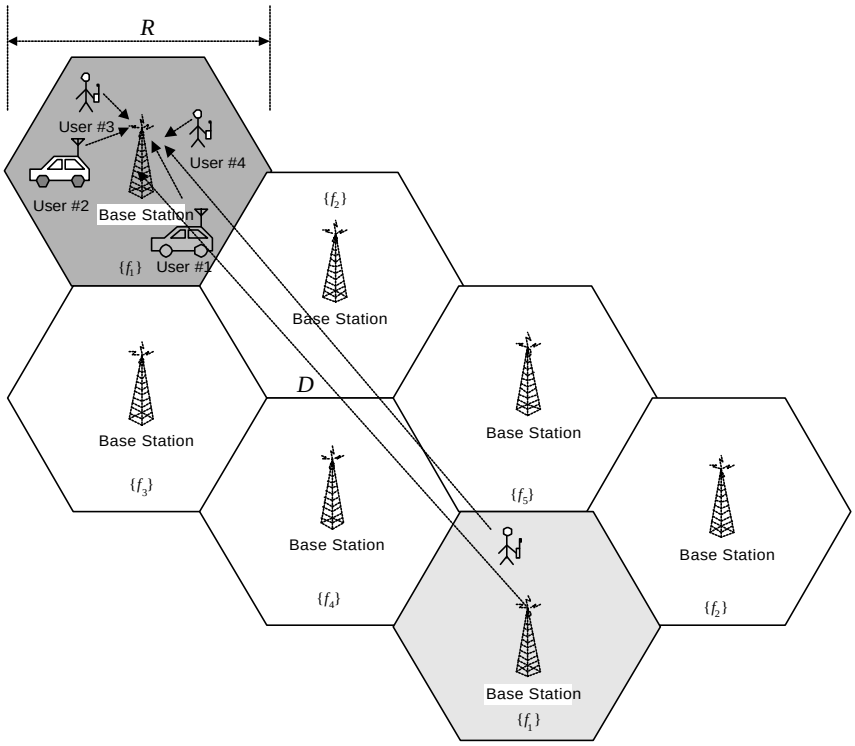

Figure 1.1: An idealized cellular network with hexagonal cells.

The cellular concept is to a divide large geographical region into small areas, each called a cell. At the center of each cell is the base station that, together with other base stations, forms a wireless network. To achieve spectral e±ciency, a set of frequencies designated to one cell is also reused in other cells, as illustrated in Figure. 1.1. In the ¯gure, a set of frequencies ff_1g is shown to be used by two cells, separated by a distance D. The cell size R may vary from a few tens of meters (pico-cell) to a few kilometers (macro-cell). The frequency reuse is possible because the intercell interference is small, as the terrestrial propagation power loss grows approximately as the fourth power of the distance (D) due to signal re°ection and de°ection. However, each user within a cell is assigned to a di®erent frequency channel or di®erent time slot to avoid (intracell) cochannel interference (CCI), as in time or frequency division multiple access (TDMA or FDMA) systems. The main examples of TDMA are the 2G cellular telephony systems, such as IS-136 [2] and GSM [3].

Spectrally e±cient digital modulation is another major technique to increase capacity

used in many present systems, including IS-136 and GSM systems.

Antenna array processing has provided yet another powerful avenue to increase a system's performance and is compatible with basically all the wireless systems via space-time processing.

Antenna arrays improve a system's performance primarily in three ways. First, they increase received signal to noise ratio (SNR). An increased SNR improves error performance. Conversely, with the same error performance the transmission rate is improved, (as seen from Shannon's channel capacity theorem [4],) and coverage is increased. Second, antenna arrays provide spatial diversity, the most effective way to fight channel fading.

The third way of improving system performance with antenna arrays is the subject of this thesis: they increase system capacity. As the user density reaches designed system's capacity and more users require service, there is a need to increase capacity. One way to increase channel capacity is to reduce cell size so that user density is increased. However, there are two problems associated with this approach: increased CCI and high cost. The first problem is an increased intercell CCI due to the reduced distance between neighboring cells sharing the same frequencies. The second one involves frequency replanning and increasing the number of base stations, which may not be economically feasible. Another approach is to increase the frequency reuse rate by reducing the cluster size instead of reducing the cell size and maintain the number of base stations unchanged. A cluster is a group of cells in which the same block of a frequency bandwidth is used. Typical cell cluster grouping is shown in Fig. 1.2 with the most common cluster size $K = 7$, based on a \good" voice reception [5]. As the cluster size is reduced, the number of users per cell is increased because the same number of users (or frequency bandwidth) in each cluster are now served by fewer cells. This approach resolves the second problem of the previous approach, but an increased intercell CCI still remains. (There are more interfering cells than those shown in Fig. 1.2, however, since they are farther away from the home cell and their effect to the home cell can be ignored.) The third approach to increase channel capacity is to increase the number of cochannel users within each cell while keeping the current cell or cluster structure unchanged. This approach may create very strong intracell CCI. If the intracell CCI can be eliminated or greatly suppressed, the system's capacity can be increased dramatically without modifying the cell or cluster layout. Antenna arrays can be used here to achieve this goal by suppressing cochannel interference or detecting cochannel signals jointly. This is the approach we will study in this thesis.

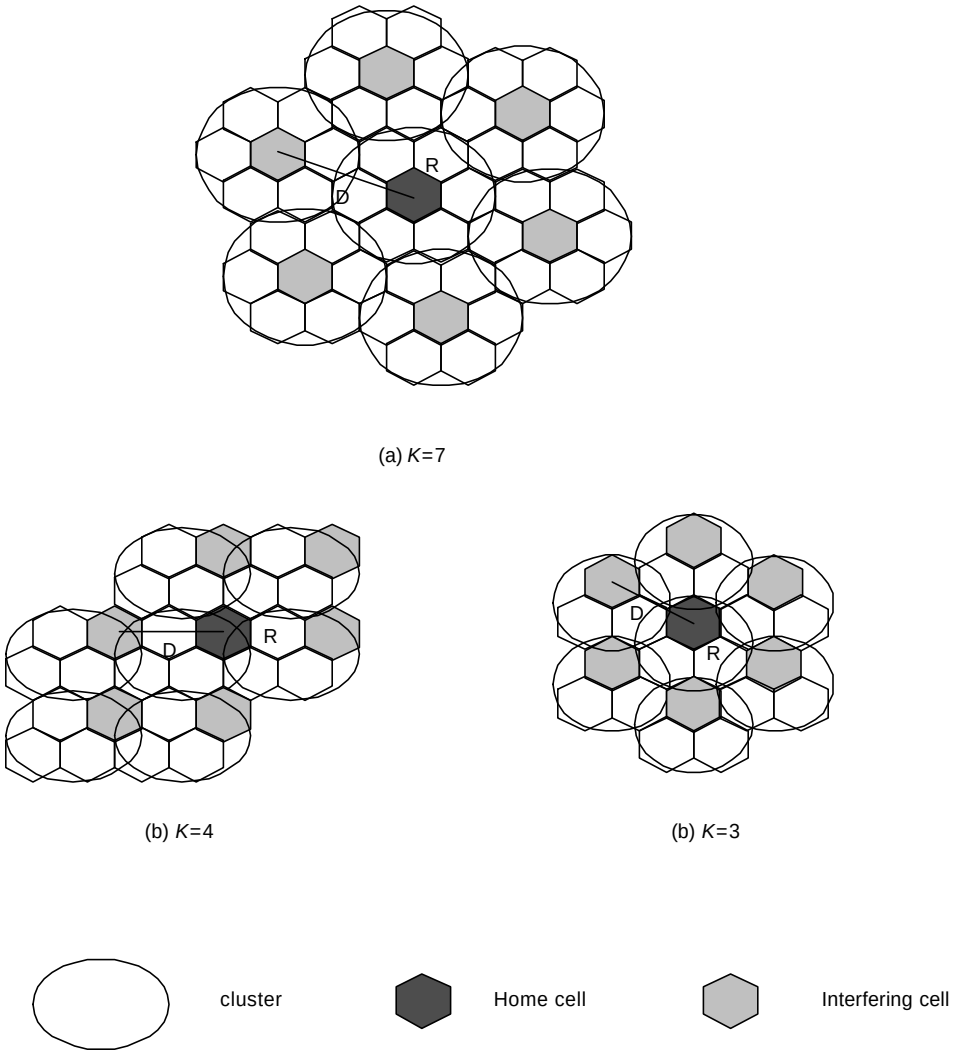

Figure 1.2: The most common cell cluster grouping. D is separation of the cells sharing the same set of frequencies, R is the cell size and K is the cluster size. Note that only the first tier of interfering cells are shown for clarity.

1.2 Motivation for Blind Detection

This thesis deals with a class of algorithms that increases capacity through blind detection with an antenna array. In this section and Section 1.3, we explain and justify our approach. Blind detection can be de¯ned as a process of channel or signal estimation based on prior knowledge of structural or spatial signal or channel properties, either statistically or deterministically, without access to the original input.

In non-blind detection, a training sequence is accessible that can be used to estimate the unknown wireless channel so that estimated channel values are available for the rest of input. The estimated channel can then be used to detect the unknown user sequence. If the channel is time-varying, then some known (training or pilot) sequences must be inserted into the user's data periodically. The training based detection algorithms can be used to combat not only channel fading but also ISI and CCI by means of pilot sequences [6], modulation [7] or coding [8].

In blind detection, on the other hand, the channel or signal is estimated based on prior knowledge of temporal or spatial signal or channel properties, either statistically or deterministically, without access to either the actual input or the channel values. This makes the problem intrinsically more di±cult than the \non-blind" or \informed" one.

Blind detection is itself an old subject of enquiry. However, in signal processing research and communication community, a number of signi¯cant developments began in the mid-1970's. Many blind algorithms have been proposed in the context of equalization and signal detection in static or very slow fading communication channels. Terms such as self-recovering equalization, blind channel equalization/identi¯cation/deconvolution, or blind detection, generally all refer to the same detection problem { blind detection. The use of training sequences is probably the most robust way to estimate the channel or signal. However, blind detection has the following advantages over the non-blind one:

- Bandwidth is conserved by eliminating or reducing training or pilot sequences.
- In multi-point communications, collision can be avoided or reduced for two or more simultaneously transmitted signals, whereas in non-blind detection training sequences in each signal can not be extracted easily due to CCI.
- No change is necessary to present air interface or protocol, and therefore the method is applicable directly to many existing TDMA systems. (Although TDMA systems have

training sequences, they are not sufficient to separate multiple signals for non-blind detections.)

- Training sequences may be unavailable or unreliable in cases of severe channel distortion, e.g., deep (time or frequency selective) fades during training.
- Training sequences for interference (cochannel users) are often inaccessible.
- Blind detection is applicable to special purpose communications where training sequences are not known.

A major drawback for all blind algorithms is the computational complexity. This is not surprising since information embedded in the signals is harder to extract than with the non-blind counterpart. However, computational complexity can be regarded as a \soft" limit in that the continued increase in the speed and decrease in the cost of digital signal processors will mitigate this problem. On the other hand, the bandwidth is a \hard" limit that is typically regulated by a standard that should not be violated.

Generally, whenever the training sequences are available in the system, the non-blind detection algorithm should be considered first, since any available side information about the signal or communication channel is helpful in improving the performance and reducing system's complexity. However, even in this case, the blind detection still has its value since it needs less frequent or shorter training sequences, and therefore bandwidth is saved. As a reference, the bandwidth for the training sequences is about 20% for GSM, and 13% for IS-136, respectively (assuming the synchronization and digital verification color code are known and can be used as a training signal).

Even with a significant portion of bandwidth allocated for training sequences, many existing communication systems are not designed to handle cochannel signals. Unfortunately, these systems include the widely deployed GSM and IS-136 systems for telephone services and ReFLEX systems for two-way paging service. In these systems, cochannel signals themselves do not contain sufficient side information that is needed for the non-blind single or multiuser detection algorithms. One obvious way to improve system capacity is to design the new air interface embedding training sequences so that multiple users can be easily separated via space-time processing. Unfortunately, protocol standardization is both expensive and time consuming. An alternative and immediate solution is to perform multiuser detection without changing the air interface. In this case, the blind detection algorithms that do

not require this embedded side information are essential.

1.3 Classi¯cation of Blind Detection Algorithms

A large number of blind detection algorithms is available in the literature, which are motivated partially by the rapid advance of communication systems. In a communication system, the number of input signals (desired users and interferers) may be one (single input or SI) or many (many input or MI). The number of output signals, e.g., the number of antennas, may also be one (single output or SO) or many (many output or MO). Note MO can also be provided by temporal oversampling (with respect to symbol rate). To simplify the notation, we will not distinguish temporal or spatial oversampling (antenna array) when no confusion is raised. Based on the number of input signals, blind detection methods could then be broadly classi¯ed as single user detection and multiuser detection, respectively. Based on the number of antennas, detection algorithms could be further divided as one antenna or multiple antennas (antenna array). However, this classi¯cation allows similar algorithms to appear in di®erent groups so it does not tell the whole story.

All blind detection processors exploit certain known properties of the signals and/or the channels to fully utilize any information leading to the detection. The most exploited properties include the temporal signal structure and temporal channel properties. For antenna array processing, spatial channel properties are also exploited. Most algorithms are developed by utilizing one or more of these properties. It is therefore constructive to classify the blind detection algorithms based on \property exploitation." We will use this classi¯cation to review some of the relevant blind detection algorithms.

1.3.1 Structural Signal Property Based Blind Detection

Structural signal properties include the waveform properties such as the ¯nite alphabet (FA) and the constant modulus (CM), and statistical properties of the waveform such as nonGaussianity, high order statistics (HOS) or cyclostationarity (CS).

Signal Statistical Property Based Detection

- High Order Statistics Algorithm (HOSA)

HOSA explicitly exploits higher (greater than two) order statistical properties of the signal. The motivation of HOSA is that second order statistics, such as autocorrelation or power spectrum, are not sufficient to identify the channels because the phase information is removed in the second order statistics. For a non-minimum phase system, as is commonly the case for multipath communication channels, there is a need for HOSA to provide this phase information. For signals to possess HOS, they cannot be Gaussian (since Gaussian sources have only the first two non-zero cumulants). This is not a serious problem since most communication signals today are digitally modulated and, therefore, are non-Gaussian, even with pulse shaping. Practical realizations rely on estimation of empirical statistics. Unfortunately they generally require much larger data size (~1000 symbol periods at least) than the second order moment based algorithms, which makes them unsuitable for time-varying mobile channels, since over such a time scale the mobile channel has probably changed significantly.

- SISO Case:

SISO blind channel identification is perhaps the first application of HOSA in communications. Many works have been done on SISO HOSA, as can be found in [41] { [44] and the references therein. However, HOSA's require a longer output data length to perform as effectively as second order based cyclostationarity algorithms, which will be described next.

- MIMO Case:

Many MIMO HOSA's have also been proposed for separating multiple sources of different statistics [39], [52], [39], or same statistics [53], [54] in which the identifiability problem was also considered. In [54] the global convergence is discussed.

² Cyclostationarity Algorithm (CSA)

Gardner [45], [46] discovered that many signals exhibit an interesting property { cyclostationarity { that statistical properties, such as moments, are periodic. Tong et al., [56] were the first to show that non-minimum phase systems identifiable only by HOSA previously, can now be identified using only the second-order cyclo-statistics. Compared with HOSA, the required data size is much smaller for CSA, since they generally only need second order cyclo-statistic quantities. Subsequently, the concept has been successfully applied in detection, filtering, parameter extraction, direction finding, system identification and other various related signal processing problems.

- SISO Case:

The CS period is determined by signals themselves. For digitally modulated signals of constant symbol rate the period of the cyclostationarity is a submultiple of that of the symbol. However, the received signal will exhibit this self-induced CS property only if it is oversampled with respect to the constant symbol rate, to avoid aliasing e®ects. If the received signal does not have an excess bandwidth (with respect to the Nyquist sample rate), oversampling is ine®ective.

- SIMO Case:

If the received signal has an excess bandwidth, oversampling will enforce CS property, and the oversampled output is transferred to the vector output of multichannel system by stacking output samples within each symbol period. Tong et al. [56] { [58] showed that oversampled equalizers are theoretically capable of obtaining perfect equalization with a ¯nite impulse response ¯lter. This is a useful result with important implications for practical equalizer design. His work also paved the way to the development of the popular class of subspace based detection algorithms, see [49], [51] and references therein for more details. Note that a multichannel output can also be obtained via antenna array. In this sense, antenna arrays can be interpreted as oversampling. Of course, antenna arrays are more powerful than CSAs because ¯rst they are not limited by the excess bandwidth of oversampling and second they provide spatial diversity gain.

- MIMO Case:

Antenna arrays have been used with CSA to detect more than one signal. A class of algorithms, the Self-COherence REstoral (SCORE) algorithms [61], [62] have been proposed based on the assumption that interferers do not have the same cyclo-frequencies as the signal of interest. For sources with similar or identical statistics SCORE algorithms do not work well without additional processing such as beamforming or FAA. FAA will be discussed in the next section and beamformers in Section 1.3.2.

Signal Waveform Property Based Detection

The signal waveform property based detection algorithms exploit the instantaneous structural signal property. The most utilized waveform properties are that of the constant modulus (CM) { signals with constant envelope and ¯nite alphabet (FA) { size of signal alphabet or constellation is ¯nite.

² Constant Modulus Algorithm (CMA)

The idea behind CMA is that any channel distortion or interference will destroy the CM property, thus by restoring the CM property of the transmitted signals channel distortion or interference is eliminated or reduced.

- SISO Case:

Sato [11] in 1975 pioneered the blind channel equalization in the context of pulse amplitude modulation (PAM) system by decomposing the PAM signal as a strong and several weak binary signals. Then by using a simple LMS-like adaptation algorithm with a nonlinear binary detection criterion, to enforce the CM property of received 'binary' data, channels can be equalized blindly. This is the earliest known CMA. The reference signal in the non-blind zero-forcing based equalizer is replaced by a nonlinear estimate of the channel input. (The nonlinearity is designed to minimize a cost function that is implicitly based on HOS.) Godard [12] extended this idea of 'CM property restoral' to a more complicated signal constellation - quadrature amplitude modulation (QAM) and proposed a more general class of nonlinear detection criteria. Variations of Sato or Godard CMA have also been proposed, as in [13], [14]. One problem with these algorithms is the slowness of convergence as the step size in LMS-like adaptation is extremely small since the variance of residual error near convergence is large, caused by enforcing the CM property on non CM signals (e.g., PAM, QAM). Nonetheless, these CMA converge globally. On the other hand, when such a CMA is used for signals with CM property, which exist for a large class of signals, such as FM, PSK, FSK, exact signal property restoral is achieved and the speed of convergence is much improved [15]. Unfortunately, the convergence speed is still not fast enough for tracking fast time-varying channels. More on the CMA convergence can be found, in [16, 17] and references therein.

- MIMO Case:

CMA has been extended to multiuser CMA or MIMO-CMA by Agee [18] and Van der Veen et al. [19] through incorporating the CMA into antenna arrays. In Agee's approach, a bank of parallel SISO-CMAs was used. To avoid them converging to the same source a soft-orthogonalization procedure is performed. Van der Veen transformed the optimization problem in CMA into a generalized eigenvalue problem at a higher computational cost but the algorithm is non-iterative.

² Finite Alphabet Algorithm (FAA)

When signals of interest are digitally modulated, their waveforms also possess another important property{¯nite alphabet property. Compared with CM, FA property is a stronger property since it further restrains the signals to a ¯nite set of constellation points. However, a FAA usually is more complicated than CMA, since more information has to be processed. FAA is also highly nonlinear which, like CMA, has its root in HOS. A minor limitation of FA property is that it applies only to digitally modulated signals.

- SISO Case:

An algebraic approach was proposed to identify time-invariant FIR systems by transforming the original channel identi¯cation problem into one of solving a set of nonlinear equations by using the FA property of signals [22]. It was shown that if the input and system impulse response satisfy a set of conditions then the system can be uniquely determined within a trivial phase ambiguity. Lodge et al. [25] proposed a Maximum likelihood Sequence Estimation (MLSE) for continuous phase modulation (CPM) signals in a time-varying channel whose statistics are assumed known. This work was later extended to frequency selective time-varying channels and to large constellations [26]. The MLSE is implemented in practice as the Viterbi Algorithm (VA) that was proposed ¯rst as an e±cient method to decode a convolutional code, and later, was shown to be useful for more general MLSE problems [27]. Seshadri in [23] { [24] proposed a joint data and channel estimation method using a generalized VA (GVA) without assuming known channel statistics. Rather, channels are estimated using a least squares (LS) method at each step of GVA. As in [22], the channel is assumed static.

- MIMO Case:

Talwar et al. [29] proposed a MIMO-FAA by iteratively performing data detection and channel estimation of QAM signals. Two implementations were developed to detect multiple users in a static channel, which alternate between using a least squares beamformer and either projecting the resulting signal estimate onto the nearest symbol (ILSP), or using enumeration (ILSE) to approximate the maximum likelihood solution, respectively. In particular, ILSE (iterative least-squares with enumeration) detect user sequences and estimate channel gains separately: it alternately detects multiple user sequences with enumeration with given channel estimate from the beamformer or some initialization and estimates the static channel gains with a least-squares technique. This two-step process continuous until a local convergence is reached. Their work was later extended to multipath channels [31]. Similar to [22], these algorithms require channels to be time-invariant. Time-varying

channel FAAs were developed by Wang and Cavers by modeling the channels as low-order polynomials [33], or as stochastic processes with known statistics [36, 37]. A more di±-cult problem, the asynchronous multiuser detection over time-varying channels, was solved in [35]. The algorithm in [35] can be extended easily to the more general time-varying, multipath channel environment.

- MISO Case:

This class of signal detection algorithms can be regarded as a special case of multiuser detection. Previously, blind multiuser detection with one antenna was considered as a ill-posed problem [20], [50], [29] since the output subspace does not span the input subspace fully, and therefore, the subspace based detection algorithms do not work. It was shown in [33] that equal-rank between the input signal and output subspaces is not a fundamental requirement of a blind MLSE detector. As a consequence, it does not impose a minimum antenna requirement.

1.3.2 Spatial Channel Property Based Blind Detection

Spatial channel properties are controlled by antenna spacing. When the spacing between antenna sensors is short, e.g., half of the wavelength, the spatial correlation between the channels "seen" by any two sensors is very high and the array is referred to as spatially coherent array. An important property of the half wavelength spacing sensor array is that it does not have spatial aliasing { the direction (angle) of arrival (DOA), between 0 and 2¼, of a point source and the array response are in one-to-one correspondence. On the other hand, when array sensors are more than several wavelengths apart, the spatial channel correlation is weak and the array is referred to as partially independent array. Partially coherent channels can be decomposed as the coherent and independent components. For this reason, we only consider the two extremes { fully coherent array with half wavelength spacing sensors and independent antenna arrays. In the following we concentrate on the MIMO case and treat SIMO as a special case.

Spatially Coherent Antenna Arrays

Coherent antenna arrays, also called beamformers, have been used widely to detect one or more radiating point-sources in radar, sonar, communication, and spectral estimation related applications. Rich references in this area can be found in [63], [64], [65], [79], [10]. Signal

detection is typically done by first estimating the DOAs of the signals, then calculating the beamformer weights based on some optimization criteria.

Pisarenko [66] was one of the first to apply eigenanalysis to the problem of extracting signal information from an estimated data covariance matrix. In contrast to the classical spectrum estimation methods, such as periodogram and other methods [79], the eigenvector based (also referred to as \super-resolution") approaches can theoretically produce unbiased frequency estimates with infinite resolution regardless of noise level. The super-resolution based beamformers in fact are spatial filters that are designed to pass the desired signal while nulling others (interferers) of different DOA based on some criterion. The most common DOA estimation methods, including MUSIC [67] and ESPRIT [68], have been developed for this purpose. Although elegant and analytically tractable, these super-resolution methods suffer from some practical problems. MUSIC, for example, requires accurate knowledge of the array manifold (a mapping between DOA and the array's response) to operate, which limits its applications in systems where such data is too difficult to obtain (for array calibration), while ESPRIT imposes a structural constraint (two or more arrays located close to each other so that their array manifolds are related by a simple translation or rotation) on the antenna array that can be difficult to satisfy in some practical systems, resulting in a loss in the degrees of freedom (null-steering capability). Moreover, the number of signals to be detected is limited by the number of array elements. In cases of multipath channels, the number of signals plus multipaths must be less than the number of antenna sensors [69]. This greatly restricts their applications in some mobile channel environment.

Furthermore, for signals with many local scatterers, DOA has some spread and sharp nulls of super-resolution based beamformers cannot be effectively placed onto it, resulting in a reduced cochannel suppression. On the other hand, even if an array is perfectly calibrated, the DOA estimation still has some statistical error due to finite data record length, which can have a large effect in certain applications [70]. For example, in TDMA, the interferers are strong in power but small in numbers. In this case unbiased sharp nulls are required to suppress the interferers effectively.

To reduce the sensitivity of super-resolution methods to the estimation error of correlation matrix based on limited data, many techniques have been developed by combining the basic beamformer and temporal based processing which exploits the temporal properties of signals within coherent arrays. For example, CM arrays devised in [74], [75] do not require

array calibration and have a low sensitivity to array estimation error. FA and CS properties have also been exploited in coherent arrays to provide robust estimation and lower the computational complexity. For example, by incorporating FA into coherent array, Talwar et al. [29] proposed two similar deterministic multiuser detection techniques, explained in Section 1.3.1. Two other related FA arrays, proposed by Swindlehurst et al. in [71], [72], iterate the following two steps: calculate the MMSE/LS estimate of the beamformer weight vector based on an estimated signal having a FA property and then demodulate-remodulate the estimated user symbol sequences using the estimated beamformer's weight vector. The advantages of these FA array algorithms are that i) the need for DOA estimation is alleviated and hence the need for array calibration data; ii) signal resolution is not limited by both spatial multipaths (multiple coherent sources of different DOAs) and temporal multipaths, an important distinction from others. CS arrays [61], [62], [76] also do not require array calibration and the correlation knowledge of the noise and interference, but a DOA estimate is needed. If users have similar CS properties (in the case of cochannel signals), DOA information is used to separate them.

In spite of all these efforts, coherent arrays still have some fundamental limitations in the time-varying mobile environment. One of them is the simultaneous deep fading. Specifically, since the antenna sensors are close together, it is very likely that a deep fade is experienced by all the sensors, resulting in a much reduced SNR and degraded performance. Almost all the coherent array based detections assume that the channel is static or quasi-static over a observation window (at least a few tens of symbol periods [55]). For fast fading channels, unfortunately, channels cannot be modeled effectively as quasi-static even for a small observation window of 20 symbol periods, as shown in [33]. The first problem can be solved by setting array sensors further apart. However, for the second problem, new techniques are needed, which must incorporate time-variation of the channel into the algorithms explicitly. This is the focus of this thesis.

Spatially Independent Antenna Arrays

Spatially independent arrays provide one of the most effective ways to fight time-varying channel fading [77]. As the channel connected to each antenna sensor fades independently, a simultaneous deep fade across the entire array is very unlikely [73]. Salz et al. [78] showed that the diversity gain lost to spatial channel correlation is insignificant even if the correlation is as high as 0.5. Spatially independent antenna arrays are also referred to as

diversity arrays due to their gain against independent channel fading. However, due to spatial aliasing and lack of spatial correlation, the DOA information is not available and diversity arrays must exploit other information, such as temporal properties of the signals, to \make up" the missing DOA information existing in the coherent arrays.

When fading channels are static or quasi-static, most of the blind multiuser detection algorithms in Section 1.3.1 can be incorporated into diversity arrays. HOS arrays ([53], [54]), the CM array ([19]) and the FA array ([29]) are good examples. It is interesting to note, however, that CS arrays by themselves cannot be used to separate multiple users even if there is no multipath, thus, other properties such as FA have to be utilized [50].

When channels are fast fading, however, most of the aforementioned algorithms do not work well. First, since HOS array requires accurate estimate of high order statistical properties, they are not suitable for time-varying channels due to a large data requirement. In the FA array ([29]), to satisfy some data structure condition, the required data size cannot be too small (a few tens of symbol periods at least) and channels are required to be constant, which may not be true for fast fading channels. Fading channels also make the convergence check more di±cult, and consequently, a®ect the performance. Since the CM array ([19]) is a subspace based algorithm, the determination of the dimension of subspaces becomes di±cult for fading channels, and an incorrectly determined subspace will likely result in a degraded performance.

Note that most of above array detectors have a very undesirable property { they require the number of signals to be detected to be no more than the number of antennas. In other words, it sets the limit of signal resolution for given antenna array.

In summary, we have seen that there is a variety of blind detectors, which can be well suited for speci¯c applications. In other words, there isn't one algorithm that is suitable to all applications. The choice of a particular technique depends on the signal, channel, and receiver structure. The objective of the technique may be to minimize inter-symbol interference (ISI) caused by multipath fading, or to minimize cochannel interference (CCI) caused by cochannel users/interferers, or to combat time-varying channel fading, or to minimize all above degradations. Since blind multiuser detections, especially those that minimize multiple degradation sources, can be computationally quite complicated, complexity reduction is also an important subject. Robustness, global convergence, and sensitivity to the observation errors are also important aspects of the detectors.

We have also seen that most of the blind detection algorithms focus on static or quasi-static channels and the blind detection for fast fading channels has received much less attention. Is time-varying channel scenario a real problem? Depending on the applications, it can be. For example, in IS-136 TDMA or paging systems, the channel fading can be very severe (up to a few percent of the symbol period). Thus, there is a real need for developing blind techniques that are suitable for time-varying fading channels as well as other impaired channels.

1.4 Thesis Overview

1.4.1 Subject of the Thesis

In this thesis we focus on the most general blind detection problem { blind MIMO detection. The signals are transmitted over unknown multipath and fast time-varying channels over the same frequency bandwidth at the same time. We assume all signals have the same symbol rate and symbol alphabet, both are known a priori. The data modulation can be linear or nonlinear. Our objective is to estimate each user sequence, which is corrupted by ISI, CCI and fast channel fading. We exploit two fundamental properties of digitally modulated signals, the ¯nite alphabet (FA) and cyclo-stationarity (CS) properties to separate, detect and equalize multiple users and channels (within the constellation). We do not assume that the array has certain physical geometry nor the minimum number of sensors. Towards our goal, we ¯rst study the detection problem in synchronous and memoryless channels in detail, and then consider extensions to asynchronous and delay spread channels.

1.4.2 Organization of the Thesis

The thesis is organized as follows. In this ¯rst chapter, we have motivated our work in the context of wireless networks and blind estimation algorithms. In Chapter 2, we describe the data model for multiple digital signals received by an antenna array at a base station. We consider both the memoryless data model with synchronous transmission and multipath of negligible delay spread, and delay spread data model for asynchronous signals and multipath with large delay spread. We consider both linear and nonlinear modulation schemes.

Optimum blind multiuser detectors are described in Chapter 3. To build up blind multiuser detection schemes, we ¯rst brie°y describe the classic optimum detectors { Bayes,

maximum a posteriori (MAP) and the maximum likelihood (ML) detectors. Then we extend them to multiuser detection. In this thesis, we focus on two of them { ML and Bayes detectors, which will be studied in detail in Chapters 4 and 5, respectively.

In Chapter 4, we describe a blind multiuser ML (BMML) detection algorithm for linear and nonlinear modulation. The time-variation of channels is transferred into an augmented data matrix by modeling channels as low order polynomials over properly selected detector windows. By exploiting the FA property more fully we show that BMML is able to simultaneously distinguish the cochannel signals jointly and estimate the channels (within a phase factor) with an array of diversity antennas of arbitrary number of elements over time-varying but memoryless channels. We study signal detectability in the noise case as well as the noise-free case and show a minimum su±cient block size that depends upon the system parameters algebraically. We also propose an e®ective initialization strategy for BMML. Two sub-block based algorithms are also presented, and simulation results demonstrate a promising performance.

In Chapter 5, we develop a blind multiuser Bayes (BMB) detection algorithm. In contrast to Chapter 4, we assume that the unknown time-varying channels be modeled as stochastic processes of zero mean and known or estimated second order statistics. The ambiguities inherent to blind detection are shown to be limited to simple equivalences under some easily satis¯ed conditions. The memory size of the detector is determined based on an estimate of the prediction error variance, which is found to depend not only on the order of the fading process but also on data{a distinct di®erence from similar MLSE detectors for single users. The memory size or state set is further reduced by incorporating the per-survivor processing technique, resulting in a practically feasible detector { BMPSP. The initialization or restart of BMPSP is also considered and a random state initialization routine turns out to be quite e®ective. BMPSP is highly e®ective for fast fading channels, as shown in numerical simulation.

In Chapter 6, we consider joint detection of asynchronous cochannel signals in multipath channels with an antenna array under Rayleigh channel fading. By exploiting the eigen-structure of the array matrix, we associate high-order modes with time-varying channels of di®erent polynomial orders. When the time variation of the channels is piecewise linear over an appropriately selected detector window, our new algorithm - Subspace Blind Joint Detection, or SBJD - is able to simultaneously distinguish the asynchronous cochannel signals and estimate the Rayleigh channels (within a phase factor). Finally in Chapter 7, we

conclude our work with a summary of the thesis and directions for future research.

Several of the main ideas in this thesis have been published, submitted or prepared for publication during the course of this research. BMML for PSK modulation were introduced in [32] and a complete description of the work is in [33]. Its extension to FSK modulation was given in [34] and a complete treatment of CPM modulation is submitted in [38]. The extension of this approach for asynchronous case was reported in [35]. BMB was introduced in [36] and a complete description is in preparation [37].

1.5 Thesis Contributions

This thesis focuses on the blind multiuser detection for fast time-varying channels with an array of diversity antennas. Detection algorithms are developed for this purpose. It seems that these detection algorithms are the first to tackle the more complicated time-varying channels, as opposed to previous detection algorithms which are only good for quasi-static channels. The thesis work consists of algorithm development, theoretical analysis, and verification by computer simulation. The key contributions of this research include the following:

- Introduced time-variation of fading channels explicitly into channel modeling. Previous works assumed channels are static or quasi-static over some observation periods (~100 symbols), a property that does not hold for some mobile channel environments. Proposed detection algorithms in this thesis remove this unrealistic assumption.
- Used the FA property to separate signals even when the channel is fast fading. The utilization of FA property in multiuser detection is not a new concept [29]. Unlike [29], however, this thesis shows that a full exploitation of the FA property enables the multiuser detection algorithms to function without requiring a minimum number of antennas.
- Developed a blind multiuser ML (BMML) detection algorithm. The main contributions are the following:
  - Modeled time-variations of the channels as low order polynomials.
  - Proved uniqueness of signal estimates.
  - Showed there is no minimum antenna limit.

- { Developed an e®ective blind initialization routine.
- { Developed a sub-block detection variant for fast fading channels.

² Developed a blind multiuser Bayes (BMB) detection algorithm. The main contributions are the following:

- { Proved uniqueness of signal estimates.
- { Showed there is no minimum antenna limit.
- { Performed memory size estimation for a blind multiuser VA detector.
- { Proposed complexity reduction techniques.

² Extended BMML to CPM modulated signals. The above detection algorithms are designed for linear modulations, due to their simplicity in analysis. The extension to blind multiuser time-varying CPM modulation is given in this work. The main contributions are the following:

- { Applied FA property directly to separate CPM signals without ¯rst using linear modulation approximation for CPM signals.
- { Proved uniqueness of signal estimates for CPM signals.
- { Exploited both FA property and temporal oversampling to avoid detection ambiguity.

² Extended BMML results to asynchronous multiple signals under multipath channels of delay spread. The main contributions are:

- { Associated the time-variations of the channels to the subordinate singular values of a structured input data matrix.
- { Applied a two-step process of multichannel equalization and multiuser detection, while the ¯rst step exploits the CS property by temporal oversampling to equalize the multipath channels and the second step utilizes FA property to separate multiple users.

Chapter 2

Wireless Communication Channels and Data Model

2.1 Wireless Communication Channels

The wireless communication channel presents one of the most complicated channel scenarios in that transmitted signals su®er distortions from not only terrain settings, which cause transmitted waves to be re°ected, di®racted and refracted before reaching the receiver, but also from other interfering signals, such as cochannel interferers (CCI). Moreover, the statistics of these channel impairments can also be time-varying due to user's mobility. Hence, reliable data transmission over the wireless communication channel is a challenging task. A good understanding and e®ective modeling of wireless channel is important in designing e®ective detection techniques for signal reception and transmission.

Wireless signal propagation can be characterized by three key e®ects: path loss, slow fading and fast fading. Path loss represents the average power loss due to a macroscopic terrain condition and is a function of the distance between the base station and mobile. For a two ray model, the path loss is inversely proportional to the fourth power of the distance. Slow fading is caused by shadowing, which represents an additional signal loss from large obstacles in the vicinity of the mobile receiver, e.g., from surrounding hills, trees and large buildings. The signal loss varies with the location and changes signi¯cantly over the length of obstacles. In other words, slow fading is more local than the path loss. Slow fading has a log-normal pdf with the mean equal to the path loss.

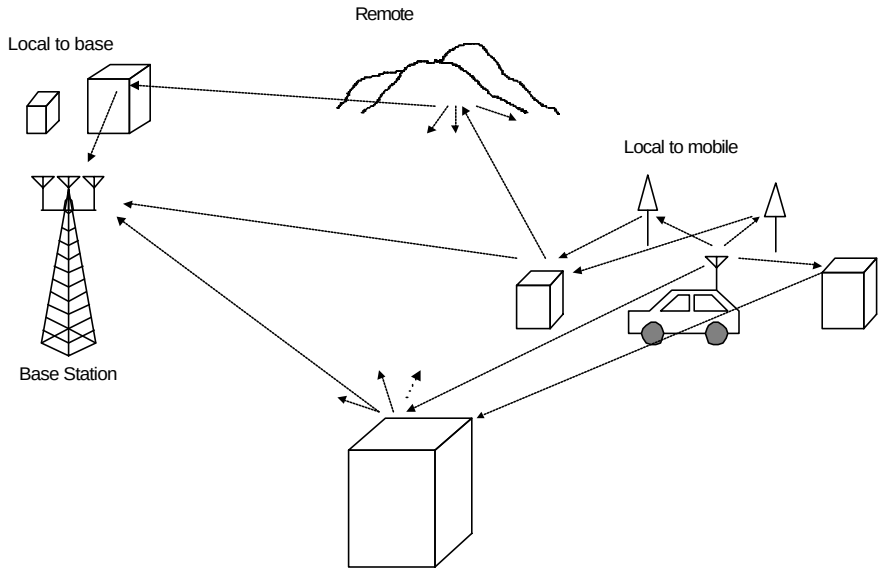

Figure 2.1: A wireless propagation scenario. There are scatterers that are local to mobile and to base, remote to mobille and base.

Fast fading is caused by multipath. The transmitted signal is scattered by many physical objects before reaching the receiver. Since each of scattered paths has some attenuation and propagation delay and from a di®erent DOA, a superposition of these signal paths forms interference at the receiving antenna. Figure 2.1 illustrates a wireless propagation scenario. The mobile illuminates its local scatterers. Some of the scattered paths reach to the base station directly while others are scattered again by remote dominant scatterers before reaching the base station. Some waves are further scattered by scatterers local to the base then arrive to the base station. Fast fading is even more local than slow fading, since the interference pattern changes very rapidly with the distance (order of a wavelength). Figure. 2.2 sketches the relative "speed" of the signal loss due to path loss, slow fading and fast fading. While the ¯rst two are considered during the initial layout phase or subsequent redesign of cellular systems, the fast fading can only be mitigated e®ectively in the receiver because it is a local phenomenon and changes rapidly over a small distance. Thus, fast fading plays a key role in determining the nature of the wireless channel and consequently determining the structure and complexity of the detector, such as whether channel equalization is needed or not, or whether the Doppler e®ect needs to be considered.

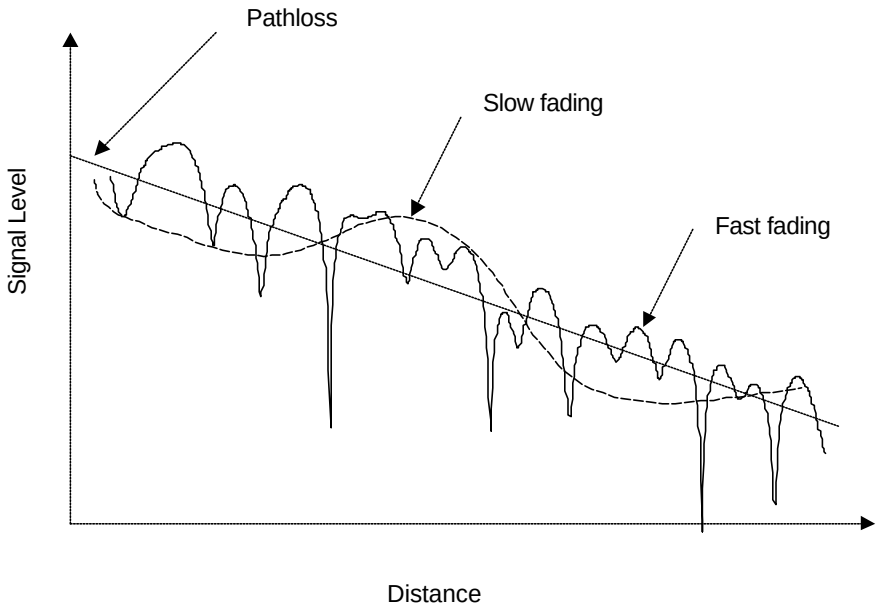

Figure 2.2: Signal loss due to path loss, slow fading and fast fading.

Fading has been studied intensively for several decades, and good sources of references on channel modeling and characterization can be found in [83], [84] and references therein. Note that when the transmitter is moving, the path length is also time-varying, resulting in an additional change in the interference pattern due to an induced phase shift.

In addition to multipath distortion, the cochannel interference (CCI) imposes another major difficulty for reliable wireless communications. As we have seen in Chapter 1, CCI is caused by users in the same cell as well as neighboring cells of different clusters that use same frequency at the same time. While the multipath propagation of cochannel signals can be mitigated by multiuser channel equalization techniques [31], [20], signals of interest can be extracted from CCI by utilizing various signal and channel properties. In the next section, we describe the data model for multiple sources, which are transmitted to a wireless channel and received by an antenna array at a base station.

2.2 Signal and Channel Model

In this thesis we consider narrowband data model. Narrowband signals are de¯ned as those whose signal bandwidth B is much smaller than the carrier (angular) frequency $!_c$. To build more complicated channel models, we ¯rst consider an ideal (excluding fast fading and user's movement) channel over which d synchronous cochannel signals transmit to an array of m sensors. In this thesis, we adopt the equivalent complex baseband representation of the real passband signal. The received complex baseband signal at the jth antenna sensor, $x_j(t)$, can be written as

$$x_j(t) = \sum_{i=1}^{d} a_{ji} y_i(t) + w_j(t) \qquad j = 1; ¢¢¢ ; m \tag{2.1}$$

where a_{ji} denotes the channel gain which combines the transmitter and receiver gain and the channel attenuation, connecting the ith signal of complex baseband signal waveform $y_i(t)$ to the jth antenna, and $w_j(t)$ is the equivalent complex baseband additive noise. Now we examine each term in (2.1) in more detail.

- Channel gain a_{ji}: The complex channel gain represents the signal loss due to path loss and slow fading for a ¯xed transmitter and receiver gain. In coherent arrays and under the ideal (Gaussian) channel condition, a_{ji}'s are related and, in fact, they are proportional to the array manifold { a set of normalized gains across the antenna array { which is a function of array geometry, sensor spacing, and local scatterers in the vicinity of the base station. If the array manifold can be calibrated accurately, then the class of super-resolution methods, discussed in Chapter 1, can be used for multiuser detection.

- Digital signal waveform $y_i(t)$: As mentioned in Chapter 1, spectrally e±cient digital modulation is one of the keys for capacity improvement. The digital signal can be created by modulating the amplitude and/or phase of the carrier frequency by data sequences either in a linear fashion, that is, the modulated waveform can be expressed linearly with nonmodulated user sequences, or a nonlinear fashion otherwise [90]. Examples of linear modulation include pulse amplitude modulation (PAM), quadrature amplitude modulation (QAM), and phase shift keying (PSK). Nonlinear modulations include frequency shift keying (FSK), continuous phase FSK (CPFSK) and continuous phase modulation (CPM).

For linear modulations, the digitally modulated signal can be expressed as the convolution of the user's symbol sequence and a pulse shape,

$$y_i(t) = \sum_{n=-1}^{\infty} s_i(n)g(t - nT); \quad i = 1,\ldots,d \tag{2.2}$$

where $\{s_i(\cdot)\}$ is the ith user's symbol sequence. For PSK modulation, $s_i(n) \in Q = \{q_k = e^{j2\pi(k-1)/Q} | k = 1,\ldots,Q\}$, T is the symbol period and $g(\cdot)$ is the pulse shape. The purpose of the pulse shape function $g(\cdot)$ is twofold: first, it controls the spectrum of the transmitted waveform $y_i(t)$; and second, it removes ISI at sampling instants nT by satisfying the Nyquist sampling criterion. While a large class of Nyquist pulses is available, the most popular ones are the so called raised-cosine pulses. Usually the full raised-cosine pulse is distributed between the transmitter (shape pulse) and the receiver (matched filter), and each has the form of

$$g(t) = \begin{cases} \frac{\beta}{2}\left[\left(1+\frac{2}{\pi}\right)\sin\left(\frac{\pi}{4\beta}\right) + \left(1-\frac{2}{\pi}\right)\cos\left(\frac{\pi}{4\beta}\right)\right] & t = \pm\frac{1}{4\beta} \\ \dfrac{\sin(\pi(1-\beta)\frac{t}{T}) + 4\beta\frac{t}{T}\cos(\pi(1+\beta)\frac{t}{T})}{\pi\frac{t}{T}\left[1-(4\beta\frac{t}{T})^2\right]} & \text{otherwise} \end{cases} \tag{2.3}$$

which is band limited to $|f| \le (1+\beta)\pi/T$. The roll-off factor β controls the excess bandwidth, $\beta\pi/T$. The advantage of using a larger roll-off β is that it allows $g(t)$ to decay with $|t|$ rapidly and hence reduces ISI due to sample timing errors. Another advantage is that it allows oversampling to be more effective (larger excess bandwidth), which is essential to some of the CSA detection algorithms that are based on self-induced cyclostationarity. See [85] for more discussion on the resolution limit for oversampling band limited signals. A fine balance of controlled excess bandwidth and ISI is very important to a system design.

[2] For many nonlinear modulations, such as CPM, the transmitted waveform can be expressed as

$$y_i(t) = \exp\{j\theta(t; s_i(n))\}, \; nT \le t < (n+1)T, \; i = 1,\ldots,d \tag{2.4}$$

where

$$\theta(t; s_i(n)) = \pi h \sum_{n=-1}^{\infty} s_i(n)q(t - nT); \quad i = 1,\ldots,d \tag{2.5}$$

where h is the modulation index, $s_i(n) = (s_i(1); s_i(2); :::; s_i(n))$, $s_i(k) \in Q = \{\pm 1; \pm 3; :::; \pm(Q-1)\}$ is the kth transmitted symbol from the ith user, and $q(t)$ is the phase function,

$$q(t) = \int_{-\infty}^{t} g(\tau)d\tau \tag{2.6}$$

where $g(t)$ is the frequency shape pulse with $q(1) = 1$. In GSM systems, $g(t)$ has a Gaussian shape

$$g(t) = \frac{1}{2T}\left[Q\left(2\pi B_b \sqrt{\frac{t-\frac{T}{2}}{\ln 2}}\right) - Q\left(2\pi B_b \sqrt{\frac{t+\frac{T}{2}}{\ln 2}}\right)\right], \quad 0 \le B_b T \le 1 \tag{2.7}$$

where $Q(t) = \int_t^{\infty} \frac{1}{\sqrt{2\pi}} \exp\left\{-\frac{\tau^2}{2}\right\} d\tau$. B_b is the 3-dB bandwidth of the Fourier transform of $g(t)$.

- Noise $w_j(t)$: It is modeled as a complex zero-mean Gaussian random process that is both temporally and spatially white, that is, $\frac{1}{2}E[w_j(t)w_{j'}^*(t')] = \sigma_w^2 \delta_{j,j'} \delta(t-t')$, where σ_w^2 is the variance of the Gaussian noise, and $\delta_{(\cdot)}$ and $\delta(\cdot)$ are the Kronecker and Dirac delta function, respectively. Furthermore, $w_j(t)$ is assumed to be circularly symmetric. That is, $\frac{1}{2}E[w_j(t)w_{j'}(t')] = 0$. In other words, $w_j(t)$ can be decomposed as two independent (real and imaginary) Gaussian processes of identical variance.

Now, we add local scatters to our data model. We have seen in Section 2.1 that a major effect of fast fading is the multipath that may be time-varying. Under the narrowband assumption, Equation 2.1 can be easily extended to include fast fading as

$$\begin{aligned} x_j(t) &= \sum_{i=1}^{m} \sum_{l=1}^{L_i} a_{j,i,l} e^{j\omega_c \tau_{j,i,l}(t)} y_i(t - \tau_{j,i,l}(t)) + w_j(t) \\ &= \sum_{i=1}^{m} \sum_{l=1}^{L_i} a_{j,i,l}(t) y_i(t - \tau_{j,i,l}(t)) + w_j(t); \qquad j = 1, \cdots, m \end{aligned} \tag{2.8}$$

where $\tau_{j,i,l}(t)$ and $a_{j,i,l}$ are the path delay and the channel gain, respectively, for the lth path from signal i to antenna j. L_i is the number of paths for user i. Note in above equation the phase factor $e^{j\omega_c \tau_{j,i,l}(t)}$ is due to the passband and baseband signal transformation: a delay in passband signal waveform result in a phase change in its equivalent baseband signal. More on the passband and baseband transformation is given in [84]. Since the time-variation of

$\zeta_{jil}(t)$ is insignificant compared to the symbol period over the entire observation period (up to a few hundreds of symbol periods), we may regard it as constant,

$$\zeta_{jil}(t) \approx \zeta_{jil}.$$

Also since the relative path delay between antenna sensors for each signal is very small compared to symbol period over the observation period, we may regard ζ_{jil} to be independent of antenna sensors, even if antenna sensors are separated by several wavelengths as in diversity arrays. Based on these two approximations, we have

$$\zeta_{jil}(t) = \zeta_{il},$$

and consequently,

$$x_j(t) = \sum_{i=1}^{K} \sum_{l=1}^{N} a_{jil}(t) y_i(t - \zeta_{il}) + w_j(t), \qquad j = 1, \cdots, m. \tag{2.9}$$

To measure the extent of the multipath, the delay spread is defined as

$$T_i = \max_l \zeta_{il} - \min_l \zeta_{il} \tag{2.10}$$

for each user i. We consider two different data models in this thesis, depending on the maximum delay spread caused by the wireless channel, $T = \max_i T_i$.

2.2.1 Memoryless Channel Model

In this subsection, we consider channels with negligible delay spreads. In addition, we assume that multiple users are synchronous. (We will treat the asynchronous multiuser channel in section 2.2.2 as a special case of multipath channel with delay.) That is, the maximum delay spread $T \ll T$, where T is a symbol period. For typical urban and rural cells, the delay spread is about 0.5μs and 5μs [96], respectively. In IS-136 systems, the symbol period is 41.6μs and therefore the delay spread could be ignored for these channel scenarios. However, to improve the detection, an equalizer may be desired. For some hilly terrains, however, the delay can be as large as 15μs and the delay spread must be considered. In GSM systems, on the other hand, the symbol period is 3.7μs, and delay spread should

be considered in most channel scenarios. Thus, if $T \gg T$, (2.9) becomes

$$x_j(t) = \sum_{i=1}^{d} \sum_{l=1}^{L_i} a_{jil}(t) e^{j\omega_c \tau_{il}} y_i(t) + w_j(t)$$
$$= \sum_{i=1}^{d} a_{ji}(t) y_i(t) + w_j(t) \qquad j = 1, \cdots, m \tag{2.11}$$

where $a_{ji}(t) = \sum_{l=1}^{L_i} a_{jil}(t) e^{j\omega_c \tau_{il}}$. Comparing with (2.9), (2.11) linearly combines all multiple signal paths from each user into a single term, which effectively removes the spatial multipath effect.

We need a discrete representation of (2.11) for digital signal processing (DSP). For linear modulations, we obtain it by first filtering the array output with a matched filter, $g^*(-t)$, which is matched to the transmit pulse. If the transmit filter $g(t)$ is a square-root Nyquist filter, then the output of the matched filter (convolution of the transmit and matched filter) satisfies Nyquist criterion. Then, by symbol rate sampling, i.e., at instants $t = nT$, we obtain

$$x_j(n) = \sum_{i=1}^{d} a_{ji}(n) s_i(n) + w_j(n) \qquad j = 1, \cdots, m \tag{2.12}$$

where $x_j(n) = x_j(t) * g^*(-t)|_{t=nT}$; $w_j(n) = w_j(t) * g^*(-t)|_{t=nT}$ and $a_{ji}(n) s_i(n) = [a_{ji}(t) y_i(t)] * g^*(-t)|_{t=nT}$, where $*$ denotes convolution. Note that to avoid introducing more notations, we abuse some of them in an understandable way. For example, the sampled convolution of $x_j(t)$ with pulse shape function is denoted as $x_j(n)$ in the above equation. In the last equation, we have assumed that a change in channel gain over the duration of the transmit pulse (typically a few symbol periods) is negligible. Also for display clarity, we normalize the symbol period to one ($T = 1$). It is clear that the filtered and sampled noise $w_j(n)$ is a discrete Gaussian random process with zero-mean and with autocorrelation $\frac{1}{2} E[w_j(n) w_{j'}^*(n')] = \sigma_w^2 \delta_{jj'} \delta_{nn'}$, where σ_w^2 is the noise variance. It can be shown that $x_j(n)$; $j = 1, \cdots, m$, is a sufficient statistic for estimating $s_i(n)$, $i = 1, \cdots, d$, for linear modulations. It is convenient to write (2.12) in its vector form,

$$x(n) = A(n) s(n) + w(n) \tag{2.13}$$

where $x(n)$ is antenna output from all m antennas, defined as $x(n) = (x_1(n), x_2(n), \cdots, x_m(n))^T$, $s(n)$ is the transmitted data vector from all d users at symbol time n, defined as

$s(n) = (s_1(n); s_2(n); :::; s_d(n)^T$; and $A(n)$ is a m-by-d channel gain matrix with $(A(n))_{ji} = a_{ji}(n)$: $w(n)$ is de¯ned similarly.

For nonlinear modulations, the set of samples $fx_j(n)g$ produced by a bank of matched ¯lters is not a su±cient statistic for estimating $s_i(n)$ because of the modulation memory (the phase of the waveform in each symbol period also depends on data of preceding symbol period). However, the set $fx_j(n)g$ produced by a bank of matched ¯lters, each is matched to one of Q^d possible waveforms, is a su±cient statistic for estimating the sequence $fs_i(n)g$ and thus, the detection problem becomes one of sequence detection problems (since each data vector (all user symbols at time nT) cannot be detected independently). For a single source the problem is well studied and can be solved by a MLSE or suboptimal symbol-by-symbol detection based on the observation of several symbol periods [86]. For multiple sources, however, we will show in Chapter 4 that temporally oversampled observations also form a set of su±cient statistics for sequence detection. Thus, to reduce the complexity of detection, we will perform oversampling rather than matched ¯ltering for nonlinear modulations. The required oversample rate will be determined by the number of signals. Detailed discussion is given in Chapter 4.

To obtain (2.13) we have used three assumptions. First, we have assumed that all d signals are synchronous with respect to the symbol period. In practice, however, each user may begin to transmit with an arbitrary delay so that each user overlaps with two or more symbols of every other user. A blind detection approach for estimating asynchronous signals is presented in Chapter 6. Second, by restricting the multipath to small time delays, we have ensured that there is no intersymbol interference (ISI). In the presence of ISI, the problem again inherently requires sequence detection, and is also considered in Chapter 6. Third, the time-variations of the channel over the duration of the transmit pulse is negligible. Suppose the pulse duration is L_{pulse} times of the symbol period. Let f_d denote the maximum Doppler shift. The assumption then implies $L_{pulse} f_d T \ll 1$ which is easily satis¯ed for most of practical systems. For example, in IS-136 systems, $f_d = 83Hz$ (corresponding to the vehicular speed of 100km/hr), $T = 41.6\mu s$; and $L_{pulse} \simeq 7$ (for roll-o® factor of 0.35), yielding $L_{pulse} f_d T < 0.03$. Even more so for GSM systems due to the shorter symbol period.

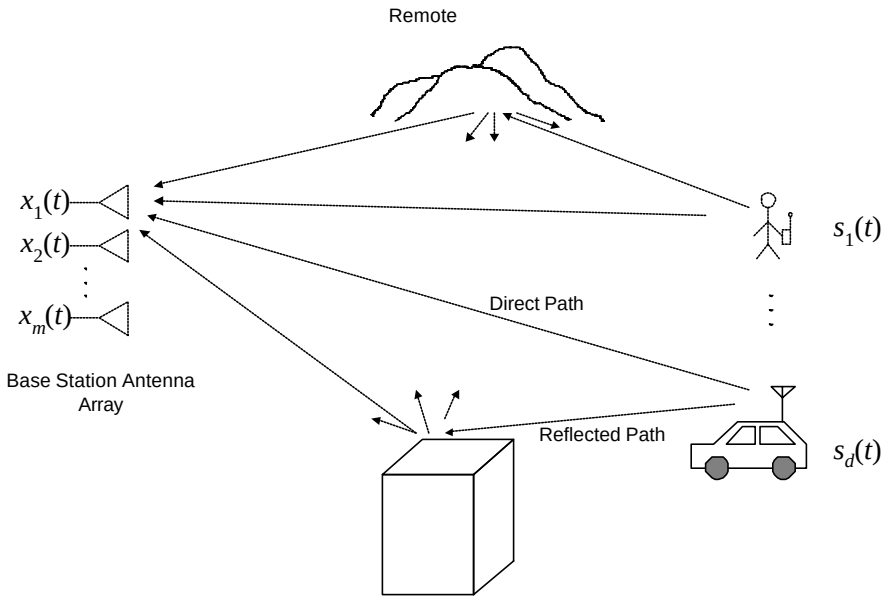

Figure 2.3: A multiuser signal propagation scenario. d multiusers transmit to an antenna array of m sensors at the base station.

2.2.2 Multipath Channel Model

In this section, we consider the more general data model given by including delay spread. In addition, the users may not be synchronized with each other. The waveform received at the array consists of multiple scattered paths from each signal with different delays and attenuations, as shown in Figure 2.3.

We assume that signal paths and transmission delays are on the order of a symbol period. For multipath delay channels or asynchronous transmissions, the sampled matched filter output is not a sufficient statistic for detecting user symbols even for linear modulations because of intersymbol interferences (referring to Figure 2.4). That is, the sampled filter output does not satisfies the Nyquist criterion. Verdú [87] indicates in a CDMA system of multiple asynchronous users that the sequence of sampled output from a bank of matched filters is a sufficient statistic for sequence estimation, but the result is not suitable here because i) we do not assume known user delays in blind multiuser detection and ii) we will need many more filters (d filters in CDMA vs. Q^d filters in TDMA). Therefore, instead of matched filtering we will adopt oversampling for multipath channels with delays.

For simplicity of the discussion, we only consider linear modulations. We oversample the array output P times per symbol period, where $P \geq N$ is known as the oversample rate. At the sample instant $t = n + \tau$, where $n \in Z$, and $\tau = k/P$ with integer $k = 1, \ldots, P$, the jth antenna output is simply given by (2.9) taken at $t = n + \tau$;

$$\begin{aligned} x_j(n+\tau) &= \sum_{i=1}^{N} \sum_{l=1}^{L_i} a_{jil}(n+\tau) y_i(n+\tau - \tau_{il}) + w_j(n+\tau) \\ &= \sum_{i=1}^{N} \sum_{l=1}^{L_i} a_{jil}(n+\tau) \sum_{m'=-1}^{1} s_i(m') g(n+\tau-\tau_{il}-m') + w_j(n+\tau) \\ &= \sum_{i=1}^{N} \sum_{m'=-1}^{1} \sum_{l=1}^{L_i} a_{jil}(n+\tau) g(m'+\tau-\tau_{il}) s_i(n-m') + w_j(n+\tau) \\ &= \sum_{i=1}^{N} \sum_{m'=-1}^{1} h_{ji}(n+\tau; m'+\tau) s_i(n-m') + w_j(n+\tau) \end{aligned} \tag{2.14}$$

where

$$h_{ji}(t;t') = \sum_{l=1}^{L_i} a_{jil}(t) g(t'-\tau_{il}). \tag{2.15}$$

Equation (2.15) is the total time-varying channel (impulse) response at time t, from all L_i paths of delays τ_{il}, for the ith signal to the jth antenna.

When asynchronous transmission is the only source of intersymbol interference (no multipath delays), (2.14) holds with T_i equal to either L_{pulse} or $L_{pulse} + 1$, depending on the relative delays of users and the start of sample points. A particular two asynchronous users channel is shown in Fig. 2.4. In this example the pulse width $L_{pulse} = 2$ and the oversample rate $P = 2$. As can be seen from the figure, if we start to count the sample at $x_j(n)$, then $T_i = 2$ for user 1 and $T_i = 3$ for user 2. If, on the other hand, we start sampling at $t = n + 1/2$, then $T_i = 3$ for user 1 and $T_i = 2$ for user 2. In any case, the asynchronous multiuser channel is a special case of more general multiuser channel with a delay spread with a well defined T_i.

2.2.3 Time-Varying Channel Model

Notice that (2.9) is quite general since no restriction on $a_{jil}(t)$ is specified. The reason for the restriction via parameter modeling is twofold: First, the dynamics of $a_{jil}(t)$ follow some physical laws (e.g. the nature of scattering and the Doppler effect), and therefore

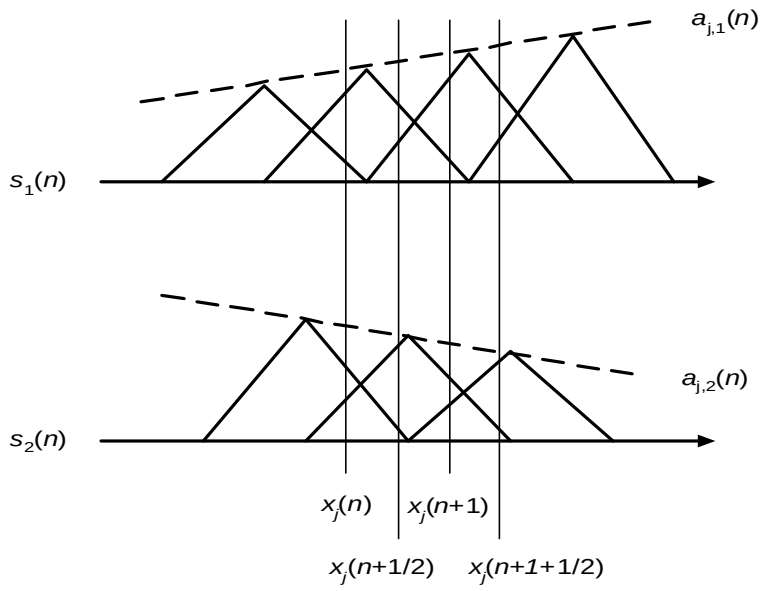

Figure 2.4: Asynchronous PSK signals with a triangular pulse shape of a span of 2 symbol periods and P = 2.

some information is lost if $a_{j\,i\,l}(t)$ is allowed to vary arbitrarily. Second, we cannot, without further assumption, determine the sequence of parameters $fs_i(n)g$ uniquely by (2.9) from a sequence of measurements $fx_j(t)g$, even in the noise-free case. To avoid this di±culty, assumptions or modeling must be introduced on the dynamics or the statistical description of $a_{j\,i\,l}(t)$. In other words, modeling helps to reduce uncertainty. Overparameterization will cause the detection algorithm to be more susceptible to noise and other interferences in addition to a high computational cost.

In communication area, there are two general approaches to model the time-varying parameters. One is to express $a_{j\,i\,l}(t)$ over some time span with a series expansion of known functions and the other is to describe the dynamics of $a_{j\,i\,l}(t)$ by stochastic models. We refer to the ¯rst approach as the deterministic and second one as stochastic due to the nature of the modeling. We exploit both approaches in this thesis.

Over a time interval of interest, $a_{j\,i\,l}(t)$ could be assumed to evolve as a linear combination of deterministic basis functions, such as the harmonic basis expansion [80], eigenfunction expansion [10], wavelet expansion [81] or polynomial basis expansion (e.g. Legendre

polynomials [92]). With the polynomial basis function, $a_{jil}(t)$ is expressed, around $t = t_0$; as

$$a_{jil}(t) = a^0_{jil}(t_0) + a^1_{jil}(t_0)(t - t_0) + a^2_{jil}(t_0)(t - t_0)^2 + \cdots + a^p_{jil}(t_0)(t - t_0)^p \tag{2.16}$$

where $\{a^k_{jil}\}$ are the expansion coefficients and p is the polynomial order determined by the observation window (in symbol periods), symbol duration and fade rate, defined as the maximum Doppler frequency shift f_d. We adopt the polynomial basis because it simplifies the signal detection algorithm as well as the detectability study, as will be seen in Chapter 4. However, their use requires a little more justification than simple expediency. Here, we address the accuracy and the required order of the polynomial model.

Consider two additional alternatives for the basis set on which to represent the gains: the Fourier series and the set of eigenvectors of the correlation matrix of the gain processes. (Note that the latter is for comparison only; our blind method does not require detailed knowledge of channel statistics.) For a given dimensionality of an approximate representation, it is well known that the eigenvector basis provides the smallest relative mean square error (MSR) [65],

$$E\left\{\frac{\sum_{t=1}^{N} |a_{jil}(t) - a_{jil;est}(t)|^2}{\sum_{t=1}^{N} |a_{jil}(t)|^2}\right\} \tag{2.17}$$

where $E(\cdot)$ denotes the ensemble average and $a_{jil}(t)$ and $a_{jil;est}(t)$; respectively, are the channel gain and its estimate for the lth path of ith user from the jth antenna.

It is interesting to see how the polynomial and Fourier expansions compare to the eigenbasis expansion. In the following example we compute (2.17) for these expansions for a Rayleigh fading model with Jakes' power spectrum (e.g., see [73]). (The total received signal from many scatterers can be considered as Gaussian by the central limit theorem. Under an isotropic scattering models, the power spectrum of the complex channel gain has a well known U-shape, which is often referred to as the Jakes' power spectrum [84]). The corresponding autocorrelation of this channel gain is given by $r(k) = r(0)J_0(2\pi f_d k/P)$ where $J_0(\cdot)$ denotes the zero-th order Bessel function of the first kind. As a fair comparison, we set p = 1 for the polynomial (thus two terms) and keep only the first two eigenfunctions for the eigenfunction decomposition and also the first two terms in the Fourier basis expansion. The results for $f_d = 0.1\%$, 1%; and data window size $N_w = 20$; 100 symbols are listed

	f_d = 0:1%		f_d = 1%	
	N_w = 20	N_w = 100	N_w = 20	N_w = 100
Eigenfunction	4:3 £ 10^{-8}	2:7 £ 10^{-5}	4:3 £ 10^{-4}	0:18
Polynomial	1:3 £ 10^{-7}	8 £ 10^{-5}	1:2 £ 10^{-3}	0:32
Fourier	2:6 £ 10^{-4}	6:3 £ 10^{-3}	2:4 £ 10^{-2}	5:8 £ 10^{-2}

Table 2.1: Ensemble average of the relative MSR for the polynomial, eigenfunction and Fourier basis expansion.

in Table 2.1. The eigenbasis is best, as expected. However, the truncated Fourier basis expansion has a signi¯cantly larger ¯tting error than both the polynomial and eigenbasis over a wide range of f_dT and N_w. We can understand the relatively good accuracy of the polynomial basis simply by observing the eigenvectors in Fig. 2.5. For small f_dT and/or N_w, the eigenvectors resemble the ¯rst monomials (i.e., the ¯rst two eigen-functions in the upper two and lower left sub-graphs of Fig. 2.5 resemble the ¯rst two monomials), which implies polynomial basis are good approximation for the channel dynamics. The same is not true of the Fourier basis. For high fade rate and longer data block, on the other hand, eigenfunctions are more like sinusoidal (e.g., see the lower right sub-graph of Fig. 2.5), which implies that Fourier series is better, as was con¯rmed in Table 2.1. We conclude that, for $f_dT \cdot 1\%$ or $N_w \gg 20$, we can approximate the channels well with a polynomial of p = 1. Conversely, if p and f_dT are given we can estimate the optimal window size. In Chapter 4 more discussion will be given on how to choose p in practical scenarios.

In the stochastic approach, $a_{j,i,l}(t)$ can be described by linear time-invariant stochastic models, such as the autoregressive (AR), the moving average (MA) and the autoregressive-moving average (ARMA), of proper orders. It represents either a priori information about time variations, known, estimated or design assumptions. These models are in fact quite general and can be used to describe a large variety of parameter dynamics. For example, time-variations are modeled as AR processes in some innovation based blind detection algorithms [25], [26]. Based on the well known WSSUS (wide sense stationary, uncorrelated scattering) channel model, $a_{j,i,l}(t)$ can also be described as a zero mean stochastic random process with autocorrelation [84]

$$\frac{1}{2}E[a_{j,i,l}(t)a_{j',i',l'}(t')] = \delta_{j,j'}\delta_{i,i'}\delta_{l,l'}\sigma^2_{j,i,l}J_0\left(2\pi f_d(t-t')\right) \tag{2.18}$$

where $\sigma^2_{j,i,l}$ is the average power of the lth scattered waveform from signal i to antenna sensor j; and $J_0(\cdot)$ is the zero-th order Bessel function of the ¯rst kind. Although obtained

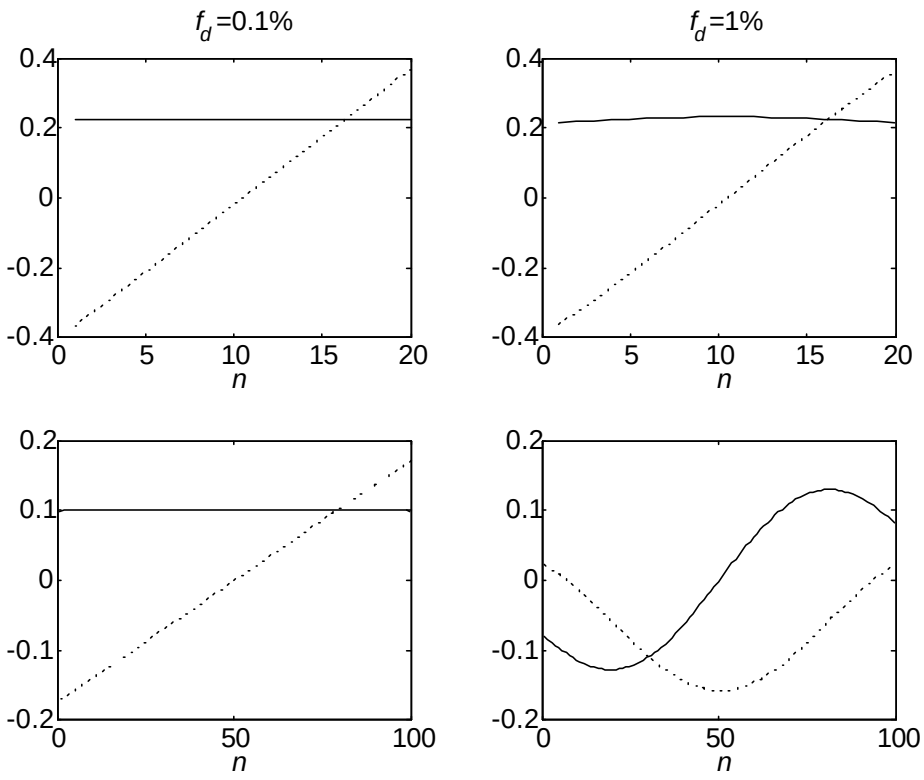

Figure 2.5: Eigen functions in decreasing eigenvalue order: ¯rst (|), second (¢¢¢). The upper and lower two sub-plots are for the block size N = 20 and N = 100, respectively. At low fade rate (f_d = 0:1%) or smaller block size (N = 20); the ¯rst two eigen functions resemble the polynomial basis function of order p = 1

from a physical model, the autocorrelation function (2.18) is not always convenient to work with. For example, the autocorrelation matrix created from (2.18) is near singular and thus non-invertible. Also, MLSE (maximum likelihood sequence estimation) type detection algorithms would require very long memory size (very large state set) with the Bessel autocorrelation function [36] since the Bessel function decays rather slowly over its argument. Note, however, that the near-singularity problem mentioned above is not due to Bessel function in particular, it is a result of Doppler spectrum being band-limited ($|f| \cdot f_d$), so any band limited Doppler spectrum has the same problem.

2.3 Convensions

For convenience,bold lettered lower-case and upper-case symbols represent either row or column vectors and matrixes, respectively. A block is de¯ned as a record of data, usually a data frame or slot in the protocol, which contains an I.D. sequence. Denote $\{v(n_1), \cdots, v(n_2)\}$ as $v(n_1 : n_2)$.

For clarity of presentation, we will drop some indices when no confusion is caused.

2.4 Summary

In this chapter we have described the data models used in this thesis. These data models include time-varying memoryless multipath channels and multipath channels with delay spread. The time-variations of the channel are modeled as deterministic polynomial basis expansion or stochastic random processes with known, estimated or designed parameters. In linear modulations, user symbols at any time instant can be detected independently whereas in nonlinear modulations the detection problem becomes the sequence detection problem which usually employs the VA type of MLSE detection algorithms with an increased detection complexity. Generally, matched ¯ltering is used for synchronous multiple users while oversampling is used for asynchronous multiple users. For nonlinear modulations, we oversample the array output even for synchronous users to reduce complexity. In the next chapters, we will use these signal and channel models to develop novel detection algorithms.

Chapter 3

Optimum Blind Multiuser Detection

3.1 Introduction

In this chapter we first describe some relevant classic optimum detectors and then generalize them to the context of blind multiuser detection. There exist different optimum detectors based on different optimum criteria. The advantage of optimum detectors is two fold: First, they have the best achievable detection performance under some criteria and therefore, they can be used as benchmark detectors. Second, they provide detector structures. In blind multiuser detection, it may be necessary to implement these algorithms or their complexity reduced variants because other type of simpler detectors may not work satisfactorily to separate multiple users. One example of simple detectors is the successive interference cancelation multiuser detector. The disadvantage of optimum detectors, on the other hand, is that they are usually more computationally expensive than simpler ones.

3.2 Optimum Detectors

In this section we review some classic optimum detectors and the discussion of the blind multiuser detection will be given in the next section.

Signal detection can be cast as one of hypothesis testing problems. More specifically, the multiuser detection problem can be regarded as a composite hypothesis problem in which the unknown parameters are the channel gains, and hypotheses are a collection of all user

sequences. The goal of the hypothesis testing is to decide, among these hypotheses, which one is the best description of the measurements. In signal detection, the best decision is made so that detection errors, measured in some way, are minimized. Optimum detection criteria may be, for example, to minimize the probability of errors or the average cost of errors, or to maximize the probability of correct decisions or the average cost of correct decisions. Based on the availability of some side information about the signals, e.g., cost for different type decision errors or prior statistical information about the transmitted signals, different optimum detection criteria can be used in hypothesis testing. See [91] or other sources on general signal detection. In digital signal detection, the most widely used optimum detection criteria include the Bayes criterion, the maximum a posteriori (MAP) criterion and the maximum likelihood (ML) criterion.

3.2.1 Bayes Criterion

This detection criterion was originated by Rev. Thomas Bayes [94]. It says, if the cost or risk associated with detection errors and prior probabilities of the parameters about the signals are available, then the optimum detection criterion is the one that minimizes the average cost or risk of the detection errors [91].

We now study Bayes criterion in the context of digital signal detection by one antenna. Let the signal to be detected be a PSK signal, that is, $s(n) \in Q$. Let $H_k(n)$ denote that "the kth signal is transmitted during the nth symbol interval," $k = 1, \ldots, Q$. To simplify the discussion, we will drop the time index n. It can be shown that if the cost for any incorrect decision is the same, Bayes criterion is equivalent to maximizing the posterior probability $Pr(H_k|X)$; the probability of the hypothesis H_k given the observation $X = \{x(t);\ nT < t \le (n+1)T\}$. Bayes decision rule then chooses H_k if $Pr(H_k|X) > Pr(H_{k'}|X)$; for all $k' \neq k$.

If signal has some parameters, then composite hypothesis testing has to be used, with posterior probability $Pr(H_k, \mu|X)$ given by

$$Pr(H_k, \mu|X) = p(X|H_k, \mu)\, Pr(H_k, \mu)/p(X) \tag{3.1}$$

where $p(X|H_k, \mu)$ is the conditional probability density function (pdf) of X for given H_k and μ and also referred to as the likelihood function, $Pr(H_k, \mu)$ is the joint prior probability of H_k and μ. In this example, the parameter μ is the channel gain or the amplitude of the received signal (assuming the amplitude of the transmitted signal is normalized to one).

Note in (3.1) sometimes $p(X)$ is left out since it is a constant that is independent of H_k and μ.

In a non-blind signal detection, the value of μ is available so the problem becomes a simple hypothesis testing problem as just described. In blind signal detection, however, the knowledge of the channel is not available, but its prior distribution may be known or modeled based on some physical channel modeling, such as the ones introduced in Chapter 2. Since μ and H_k are physically independent, $Pr(H_k;\mu) = Pr(H_k)p(\mu)$. If the prior pdf $p(\mu)$ is known, the parameter μ can be averaged out of (3.1) as

$$Pr(H_k|X) = Pr(H_k) \int p(X|H_k;\mu)p(\mu)d\mu = p(X); \tag{3.2}$$

leaving only the hypothesis parameter H_k. Bayes detection can be expressed as a process of deciding on $\hat{H}_k$ such that

$$\hat{H}_k = \max_{H_k} Pr(H_k|X): \qquad \text{(Bayes)} \tag{3.3}$$

3.2.2 Joint Maximum A Posteriori Criterion

One problem with Bayes detection is that to get $Pr(H_k|X)$ one needs sometimes to carry out the integration, which makes the detection computationally expensive. Instead of computing (3.2), one can use alternative criterion to select $\hat{H}_k$ by maximizing the posterior probability $Pr(H_k;\mu|X)$ in (3.1) with respect to the parameters μ and H_k. The estimated pair $\{\hat{\mu};\hat{H}_k\}$ is referred to as the maximum a posteriori (MAP) estimate and the detector is referred to as the MAP detector, which can be formally defined as

$$\{\hat{\mu};\hat{H}_k\} = \max_{H_k;\mu} Pr(H_k;\mu|X): \qquad \text{(MAP)} \tag{3.4}$$

MAP criterion is also referred to as the unconditional maximum likelihood (UML) criterion [91] by viewing the posterior probability $p(X|H_k;\mu)\,Pr(H_k;\mu)$ as the unconditional (with respect to μ) likelihood function.

Note that there is no fundamental difference between Bayes and MAP criterion, since they both rely on maximizing the posterior probability, and the only difference is the treatment of parameter μ: in Bayes μ is averaged out and in MAP a maximizing value $\hat{\mu}$ is used for μ.

3.2.3 Maximum Likelihood Criterion

When $p(\mu)$ is unknown or not precisely known, one may adopt the so called maximum likelihood (ML) criterion. To see why maximum likelihood detection, we start from (3.1) again. Suppose first the prior $p(\mu)$ is known. Then we have (3.2). Now let the knowledge of μ become less and less precise and eventually is only vaguely known, which means the dependence of $p(\mu)$ on μ is weak. If data X is to be of any use in gaining the knowledge in making decision on H_k; the posterior $Pr(H_k; \mu | X)$ must be much narrower in μ than the prior $p(\mu)$: Hence, in carrying out the integration in (3.2), the most dominant contribution from the integrand for given observation X is for some μ at which the likelihood function $p(X|H_k; \mu)$ attains its maximum value. Thus, maximizing $Pr(H_k|X)$ can be replaced by maximizing $p(X|H_k; \mu)$, which is the ML criterion. Formally, this can be expressed as

$$\left\{\hat{\mu}; \hat{H}_k\right\} = \max_{H_k;\mu} p(X|H_k; \mu): \qquad \text{(ML)} \tag{3.5}$$

This joint optimization problem yields another approach for multiuser detection. We can see that in ML, like in MAP, the signal and the parameter are estimated simultaneously. In applications where both $H_k; \mu$ are desired, ML and MAP are more suitable than the Bayes since they can be estimated simultaneously.

Note that although (3.5) seems to be derived from (3.2) they are viewed quite differently in statistics community. In Baysian statistics viewpoint, the parameter μ is the variable and therefore integration in (3.2) makes sense, whereas ML has its root in classical statistics, which regards the parameter μ as a fixed value, although unknown. Based on this viewpoint, we may refer to ML detector as a deterministic approach since μ is treated as a fixed but unknown value whereas Bayes detector is a stochastic approach.

Some comparisons of Bayes (3.3), MAP (3.4) and ML (3.5) detectors are in order:

- ML detectors do not require the prior knowledge of the parameter μ and therefore they can be applied to more general scenarios whereas Bayes and MAP detectors require the priors of μ. In our case, this implies that if the time-variation of the channel cannot be accurately modeled then ML detectors is more appropriate than the others. If, on the other hand, we can obtain the stochastic description of the channel, then Bayes or MAP is superior since more information is utilized.

- All three detectors are systematic approaches and give the optimum receiver structures automatically, although they may not be easy to realize.

² All three detectors are quite expensive, especially Bayes. In practice, therefore, some kind of suboptimal forms must be considered.

3.3 Optimum Blind Multiuser Sequence Detectors

In this section, we relate the optimum detectors described in the last section to the blind multiuser sequence detection problem. The solutions of these detectors are the goal of this thesis and will be studied in detail in the remaining of the thesis. As in the last section, we study the Bayes, MAP and ML detectors separately. To simplify the discussion, we assume in this section that channel is memoryless so that the multipath e®ects are not considered. Detection of multiple asynchronous users under multipath channels will be discussed in Chapter 6. Note that in section we refer to the following optimum multiuser detectors as blind optimum multiuser detectors to remind ourself that no training sequences are utilized in them, although the non-blind optimum multiuser detectors have similar forms.

3.3.1 Blind Multiuser Bayes Detector

Suppose now we have d user sequences, each of length N symbols. Let $S = \{S(N)\}$ denote a collection of all Q^{dN} possible user sequences, of PSK signals of the same constellation. If each of the di®erent sequences in S is labeled by k, $k = 1, \ldots, Q^{dN}$, we can de¯ne H_k as "the kth signal sequence is transmitted". Observation X is de¯ned as $X = \{x_j(t); 0 < t < NT;$ $j = 1, \ldots, m\}$: Since we only concern the memoryless channel in this section the path index l is removed from all the channel gain vectors for notational convenience. Thus, a_{ji} is the gain at antenna j from user i at time n. Further, with $a_{ji} = (a_{ji}(1), \ldots, a_{ji}(n))$; a row vector of size n, and $a_j = (a_{j1}, \ldots, a_{jd})$, a vector of size dn.

From the signal model in (2.13), we can express the likelihood function, or posterior probability, at antenna j as

$$
\begin{aligned}
& L_j^u(S; a_j | x_j) \\
&= p(x_j; a_j; S) \\
&= p(x_j | a_j; S) p(a_j; S) \\
&= p(x_j | a_j; S) p(a_j) \Pr(S) \\
&= C \exp\left(-\frac{\| x_j - \sum_{i=1}^{d} a_{ji} D(s_i) \|^2}{2\sigma_w^2} - \frac{\sum_{i=1}^{d} a_{ji}^{*} R_{a_{ji}}^{-1} a_{ji}^{T}}{2} \right)
\end{aligned}
\tag{3.6}
$$

where $|x|^2$ applied to a column vector denotes $x^H x$; σ_w^2 is the noise variance, and $R_{a_{ji}}$, $\frac{1}{2}E[a_{ji}^T a_{ji}^*]$,

$$p(x_j|a_j;S) = \frac{1}{(2\pi\sigma_w^2)^N} \exp\left(-\frac{|x_j - \sum_{i=1}^{d} a_{ji} D(s_j)|^2}{2\sigma_w^2}\right) \tag{3.7}$$

and

$$p(a_j) = \prod_{i=1}^{r} p(a_{ji}) \tag{3.8}$$

with

$$p(a_{ji}) = \frac{1}{(2\pi)^N \sqrt{R_{a_{ji}}}} \exp\left(-\frac{\sum_{i=1}^{d} a_{ji}^* R_{a_{ji}}^{-1} a_{ji}^T}{2}\right) : \tag{3.9}$$

Note that in the above we regard a_{ji} to be Gaussian distributed, this is justified in light of central limit theorem (see Section 2.2.3 for more discussion). Since $R_{a_{ji}}$ is independent of antenna index j by our far field assumption, we abbreviate it as R_{a_i}.We also refer to (3.6) as the unconditional likelihood function in the sense that the likelihood function is not conditioned on the parameters a_j. In obtaining above equations we have used fact that S is a discrete variable and all possible S are assumed to be equally probable and therefore its contribution to (3.6) is included in the constant C which combines all proportionality terms of the pdfs that do not depend upon the variables interested.

As stated in the last section, Bayes detector requires one to compute the posterior probability

$$Pr(S|X) = \prod_{j=1}^{r} Pr(S|x_j) \tag{3.10}$$

with

$$Pr(S|x_j) = \frac{p(x_j|S)\,Pr(S)}{p(x_j)}: \tag{3.11}$$

Note in (3.10) we have used the assumption that antennas are fully uncorrelated. Since $p(x_j)$ is independent of S and by assumption, Pr(S) is independent of S, we have

$$Pr(S|X) \propto \prod_{j=1}^{r} p(x_j|S) = p(X|S): \tag{3.12}$$

Hence Bayes criterion becomes one of maximizing the marginal likelihood function $p(X|S)$ with

$$p(x_j|S) = \frac{\int_{a_j \in C^{dN}} p(x_j; a_j; S)da_j}{Pr(S)} = \int_{a_j \in C^{dN}} p(x_j|a_j; S)p(a_j)da_j \quad (3.13)$$

where $da_j = da_{j1} \cdots da_{jd}$ with $da_{ji} = da_{ji}(1) \cdots da_{ji}(N)$: With given $p(x_j|a_j; S)$ and $p(a_j)$; as in (3.7) – (3.9), (3.13) can be computed. Detailed computation is tedious but possible, as shown in Appendix A. However a much simpler way can be used to obtain the same result since x_j, conditioned on S, is known to be Gaussian with zero mean. Its conditional autocorrelation $R(S)$ is defined as

$$R_j(S) = \frac{1}{2}E[x_j^T x_j^*|S] = \sum_{i=1}^{K} D(s_i)R_{a_{ji}}D(s_i^*) + \sigma_w^2 I \quad (3.14)$$

where I is the identity matrix of proper dimension, and the channel covariance matrix $R_{a_{ji}}$ connecting user i and antenna j is defined as

$$R_{a_{ji}} = \frac{1}{2}E[a_{i;j}^T a_{i;j}^*]: \quad (3.15)$$

Thus, we have

$$p(x_j|S) = \frac{1}{(2\pi)^N |R_j(S)|} \exp\left(-\frac{x_j^* R_j(S)^{-1} x_j^T}{2}\right): \quad (3.16)$$

Since the channel responses between antennas are assumed to be independent, the overall marginal likelihood function $p(X|S)$

$$p(X|S) = \prod_{j=1}^{N} p(x_j|S) \quad (3.17)$$

and Bayes criterion for blind multiuser detection is

$$\hat{S} = \arg \max_{S(N) \in \{F\}^N} p(X|S) \quad (3.18)$$

where $\{F\}$ is the set of all possible values of $\{s_1(n); \ldots; s_d(n)\}$. We will study this blind multiuser Bayes (BMB) detector in Chapter 5.

3.3.2 Blind Multiuser MAP Detector

In MAP a particular variable set $\{\hat{S}, \hat{\mathbf{a}}_j\}$ is chosen so that (3.6) is maximized. An alternative but suboptimal approach is to maximize $L_j^u(S, \mathbf{a}_j | \mathbf{x}_j)$ over $\mathbf{a}_j$ and then S in two separate steps. Thus, this two-step approach may yield different estimates $\{\hat{S}, \hat{\mathbf{a}}_j\}$ as the single-step one does. To maximize $L_j^u(S, \mathbf{a}_j | \mathbf{x}_j)$ over $\mathbf{a}_j$, we take the derivative of (3.6) with respect to $\mathbf{a}_j$ and set it to zero,

$$(S_{diag}^H S_{diag} + \sigma_w^2 R_{a_j}^{-1}) \hat{\mathbf{a}}_j^T = S_{diag}^H \mathbf{x}_j^T \tag{3.19}$$

$$\text{or } \hat{\mathbf{a}}_j^T = (S_{diag}^H S_{diag} + \sigma_w^2 R_a^{-1})^{-1} S_{diag}^H \mathbf{x}_j^T. \tag{3.20}$$

where $S_{diag}(n) = (D(s_1(n)), \ldots, D(s_d(n)))$ and $R_{a_j} = \frac{1}{2} E[\mathbf{a}_j^T \mathbf{a}_j^*]$. With (3.14), (3.19) becomes

$$S_{diag} \hat{\mathbf{a}}_j^T = \left[I - \sigma_w^2 R_j^{-1}(S) \right] \mathbf{x}_j^T \tag{3.21}$$

By applying the matrix inversion lemma to (3.20) and using the result and (3.21) in (3.6) we have, after some cancellations,

$$\Lambda_j^u(S|\mathbf{x}_j) = L_j^u(S, \hat{\mathbf{a}}_j | \mathbf{x}_j) = \exp\left(- \frac{\mathbf{x}_j^* R_j^{-1}(S) \mathbf{x}_j^T}{2} \right). \tag{3.22}$$

Again since the channel responses between antennas are assumed to be independent, the overall likelihood function $\Lambda^u(S|\mathbf{X})$ from all antennas is simply $\Lambda^u(S|\mathbf{X}) = \prod_{j=1}^m \Lambda_j^u(S|\mathbf{x}_j)$. Thus, MAP criterion for blind multiuser detection is given as

$$\hat{S} = \arg \max_{S(N) \in \{F\}^N} \Lambda^u(S|\mathbf{X}). \tag{3.23}$$

By comparing $\Lambda_j^u(S|\mathbf{X})$ (3.22) with $p(\mathbf{x}_j|S)$ (3.16) we see that the only difference is in a "hypothesis dependent" scalar. However, (3.23) only guarantees local optimal solution because S sought in this two-step solution is not globally optimal. Nonetheless, MAP is slightly simpler than Bayes detector since $|R(S)|$ need not be calculated. We will not study MAP detector in this thesis work as the Bayes detector can be easily changed to the MAP detector and the performance of the latter is expected to be slightly worse than Bayes since the linear term $\frac{1}{|R_j(S)|}$ in (3.16) is less sensitive to different decisions than the exponential term in $\Lambda^u(S|\mathbf{X})$ and $p(S|\mathbf{X})$.

3.3.3 Blind Multiuser ML Detector

In both Bayes or MAP detectors, we have assumed priors for both S and $\{a_j\}$ are known. While the assumption on the prior or S seems reasonable, that is, all users with all possible transmitted sequences are equally probable, the knowledge on the prior of $\{a_j\}$ is less obvious. If an accurate modeling of the time-varying channels is not possible, then it is desirable to adopt detection criteria that do not depend on priors. ML criterion is one of such decision rules. From (3.6) we see the likelihood pdf is

$$L_j^c(S, a_j | x_j) = p(x_j | a_j, S) = C \exp\left(- \frac{|x_j - \sum_{i=1}^{d} a_{ji} D(s_i)|^2}{2\sigma_w^2} \right) : \tag{3.24}$$

This likelihood pdf is also referred to as the "conditional" likelihood pdf [91] since it depends on the channel parameter a_j. Then the ML criterion for blind multiuser detection is

$$\{\hat{S}, \{\hat{a}_j\}\} = \arg \max_{S(N) \in \{F\}^N; a_j \in C^{dN}} L^c(S|X; \{a_j\}) \tag{3.25}$$

with $L^c(S|X; \{a_j\}) = \prod_{j=1}^{m} L_j^c(S|x_j; a_j)$ assuming the noises are spatially independent. In Chapter 4 we will study this blind multiuser ML (BMML) detector.

3.4 Summary

In this section, we have introduced the blind multiuser Bayes, MAP and ML detectors. Two of them, the Bayes and ML detectors, will be studied in detail in the next two chapters.

An important aspect of the blind multiuser detection is not whether there is a solution, but whether the solution is unique. If solutions are not unique, we need to know if there exists a simple way to identify which is the correct one.

Although optimum under different criteria, optimum detectors are prohibitively expensive computationally. For these optimum blind multiuser detectors, detection complexity rises exponentially with the number of users d and the sequence length N. Thus, a direct solution of $\hat{S}$ is prohibitively high because of its immense search space $\{F\}^N$. Therefore, suboptimal variations of above detectors will have to be developed. Good initialization is also critical for blind multiuser detection.

These problems will also be discussed in the following chapters.

Chapter 4

Blind Multiuser ML Detector

4.1 Introduction

In this chapter we develop the blind multiuser maximum likelihood (BMML) algorithm, in which the time-variations of the channels are modeled as polynomials, described in detail in Chapter 2. As mentioned in Chapter 3 this approach is deterministic in nature in that time-varying channels are treated as unknown but ¯xed values, rather than stochastic random processes. It does not require the statistical knowledge of the fading channels, and hence to learn or estimate it. To simplify the discussion, we will focus on solving the problem of memoryless channels with synchronous users. More general cases will be discussed in Chapter 6.

Due to the memory in non-linear signal modulations, the detection algorithms for linear and non-linear modulation are di®erent. We ¯rst present the deterministic multiuser detection for linearly modulated signals, then for nonlinearly modulated signals.

4.2 Application to PSK Modulation

In this section, we study the detection algorithm for multiple PSK modulated signals with the same symbol rate. The result can be extended to other linear modulations, such as PAM and QAM. However, details for the latter are not presented in this thesis.

The work presented in this section is an extension of an earlier one [32] in which time varying channels are considered as piecewise linear over (sub-) block intervals. Through a simple transformation the explicit time-dependence of the channels is shifted into an

augmented signal matrix, resulting in a detection algorithm with much reduced complexity. Based on the unique mapping condition rather than the full rank condition of the channel matrix, the algorithm is able to separate cochannel signals by exploiting the finite alphabet (FA) property of signals. It requires only minimal information of a single I.D. sequence per user per block in order to distinguish the detected streams (i.e., which one corresponds to user #1, etc.).

The implementation of BMML relates to but differs from the iterative least-squares with enumeration (ILSE) [29], a technique that iteratively detects the signals and channels. ILSE assumes the channel is static and requires the number of antennas to be no less than the number of users. See Section 1.3.1 for more description on ILSE or the original paper by Talwar et al. [29]. BMML, on the other hand, allows the channel to be time-varying and does not impose a limit on the number of antennas, as will be seen in this chapter. The detection order ambiguity in each sub-block is much reduced by using the condition that channels (or their estimates) vary continuously across the sub-block boundaries. Two different sub-block algorithms are designed. On the other hand, if the channel variation is very slow, as in some applications, sub-block detection is not necessary and replaced by block detection. We study signal detectability and derive the unique detection conditions, and show that the number of signals to be detected is not limited by the number of antenna elements for reliable detection.

This section is organized as follows. We first establish ideal BMML and its iterative implementation for multi-user detection. Then we study the solutions of BMML and show that under some conditions the signals can be identified uniquely with the side information. Moreover, we provide a proper selection rule for the sub-block size and the order of polynomials for a given fade rate. Furthermore, we propose an effective initialization strategy for the iterative BMML and estimate a threshold for detecting plausible solutions. Finally, two sub-block algorithms are presented and their performance is studied by numerical simulations.

4.2.1 Blind Multiuser Maximum Likelihood (BMML) Scheme

We construct the matrix form of the channel model proposed in Chapter 2. In matrix form, the channel model (2.13) becomes

$$x(n) = A(n)s(n) + w(n) \tag{4.1}$$

where $x(n)$, $s(n)$, and $A(n)$ are defined in Chapter 2. For clarity of the presentation at this point we assume that the channels can be adequately characterized by low order polynomials of degree p over full block intervals, of length N symbols each. Every block contains an I.D. sequence per user. In contrast, a sub-block, a shorter data record, may not contain an I.D. sequence. Sub-block data detection will be discussed separately in Section 4.2.1.

Ideal BMML

Although there are many choices of channel basis expansions, we here choose the polynomial basis because it simplifies the signal detectability study, as will be seen in Section 4.2.2.

From the polynomial channel model (2.16), $A(n)$ can now be expressed as a polynomial expansion of order p,

$$A(n) = A_0 + nA_1 + n^2A_2 + \cdots + n^pA_p. \tag{4.2}$$

(4.1) can then be formulated alternatively as

$$\begin{aligned} X(n) &= \begin{bmatrix} A_0 & A_1 & \cdots & A_p \end{bmatrix} \begin{bmatrix} S_0(n) \\ S_1(n) \\ \vdots \\ S_p(n) \end{bmatrix} + W(n) \\ &= A_{aug}S_{aug}(n) + W(n) \end{aligned} \tag{4.3}$$

where $X(n) = (x(1); x(2); \ldots; x(n))$, $W(n) = (w(1); w(2); \ldots; w(n))$, and

$$S_k(n) = \begin{bmatrix} s(1); 2^k s(2); \ldots; n^k s(n) \end{bmatrix}; \; k = 0 \ldots p. \tag{4.4}$$

For convenience, denote $S_0(n)$ as $S(n)$. The advantage of (4.3) over (4.1) is that the explicit time dependence of channels $A(n)$ is transferred to $S_{aug}(n)$ since coefficient matrices A_k are now time-invariant. As a result, it allows each signal vector $s(n)$ to be detected independently, as explained next.

Since each column of $X(n)$ in (4.3), when conditioned on A_{aug} and S_{aug}, is Gaussian and independent of others, the ML estimation of $S(n)$ and A_{aug} (3.25) reduces to a least-squares (LS) estimate:

$$\theta_{LS} = \min_{A_{aug};\, S_{aug}(N)} \|X(N) - A_{aug}S_{aug}(N)\|_F^2 = \min_{A_{aug}} \sum_{n=1}^{N} \min_{s_{aug}(n)} |x(n) - A_{aug}s_{aug}(n)|^2. \tag{4.5}$$

where $s_{aug}(n) = (s(n)^T; ns(n)^T; :::; n^p s(n)^T)^T$ and $\frac{1}{2}_{LS}$ is the residual. S_{aug} could be solved in principle by enumeration: after expressing A_{aug} as an LS estimate ([10])

$$A_{aug} = XS^{\#}_{aug} \tag{4.6}$$

where $(\cdot)^{\#}$ denotes the pseudoinverse [82], S_{aug} can be detected by enumerating all Q^{dN} possible value in (4.5). However, this method is highly impractical due to its computational complexity.

Note that there are multiple solutions to (4.5). Lacking channel reference information, BMML has rotational and order-exchange symmetries in data and channel estimation. That is, in a solution S_{aug}, if any user sequence is given a phase rotation from the constellation set and/or the order of the users is permuted, and corresponding changes are made to A_{aug}, the newly modified user sequences and channel $\{\mathbf{S}_{aug}, \mathbf{A}_{aug}\}$ are also solutions since they result in the same residual in (4.5). Obviously the combination of the two transformations is also a solution. For easy references, we give the following definition.

Definition: A transformation of S by the equivalent translation matrix T_E, $S_E = T_E S$ belongs to the equivalent set $S_E = \{S_E \mid T_E$: all possible translations defined below$\}$ and any $S \in S_E$ is called an equivalent solution. The equivalent translation matrix T_E is a $d \times d$ matrix with only one non-zero element $\varphi(q)$ in each row and column, which can be factored as $T_E = E_r E_p$ where E_r is a diagonal matrix with the kth element $\varphi(q_k)$ (it rotates the kth user sequence by a common phase factor $\varphi(q_k)$) and E_p is a matrix with one 1 in each column and row and other entries zero (it permutes the order of detected users).

We refer to the particular set $\{S_E\}$ containing the transmitted one as the correct equivalence set $\{S_{CE}\}$. If a solution $\mathbf{S} \in \{S_{CE}\}$ and there is additional information contained in each user sequence, the correct solution can be obtained. For example, differential encoding can remove the phase rotation ambiguity and a single embedded I.D. sequence for each user per block can remove the signal order ambiguity. Note that these requirements are reasonable, as they are fulfilled in most of the communication systems, e.g. in the TDMA IS-136 standard.

We would like to know under what conditions the solutions are in the same equivalent set. It is possible that without any restrictions on A_{aug}; not all solutions S_{aug} are equivalent. A necessary condition for A_{aug} to allow equivalent solutions is given below:

Condition 4.2.1 The channel A_{aug} maps any data signal s_{aug} uniquely onto its range

space, that is, for any $s_{aug;1} \neq s_{aug;2}$ $A_{aug}s_{aug;1} \neq A_{aug}s_{aug;2}$. This condition is referred to as the unique mapping (UM) condition.

To see this is a necessary condition, suppose that $S_{aug;1}$ is a solution. Suppose also that at time n, $s_{aug;1}(n) \neq s_{aug;2}(n)$ but $A_{aug}s_{aug;1}(n) = A_{aug}s_{aug;2}(n)$. Now construct $S_{aug;2}$ from $S_{aug;1}$ by substituting its nth column vector $s_{aug;1}(n)$ with $s_{aug;2}(n)$. Clearly data matrices $S_{aug;2}$ and $S_{aug;1}$ in general are not equivalent, however $A_{aug}S_{aug;1} = A_{aug}S_{aug;2}$: Thus we cannot distinguish between these two data matrices, and therefore Condition 4.2.1 is a necessary condition.

Iterative BMML

In practice, we find a sub-optimal solution of (4.5) by optimizing A_{aug} and $s_{aug}(n)$ alternately, as in [93], [29]. First, we fix A_{aug} at a previously estimated or initialized value and obtain $\hat{S}_{aug}$ by solving (4.5) over each $s_{aug}(n)$ in turn by enumerating Q^d possible values of $s_{aug}(n)$. Then we use the resulting $\hat{S}_{aug}$ to re-estimate the continuous variable A_{aug} via (4.6). The two-step process is repeated until a certain convergence criterion is reached. We term this approach "iterative BMML" (or simply BMML, for short), to distinguish it from the ideal BMML. In essence, BMML solves the following two equations iteratively:

$$r_k = \min_{\mathbf{b}_{aug}} \left\| X - \left(\hat{A}_{aug}\right)_{k-1} \left(\mathbf{b}_{aug}\right)_k \right\|_F^2 = \sum_{n=1}^{N} \min_{\mathbf{b}_{aug}(n)} \left\| x(n) - \left(\hat{A}_{aug}\right)_{k-1} \left(\mathbf{b}_{aug}(n)\right)_k \right\|^2 \tag{4.7}$$

$$\left(\hat{A}_{aug}\right)_k = X \left(\hat{S}^{\#}_{aug}\right)_k : \tag{4.8}$$

As indicated by the R-H side of (4.7), the minimization can be performed for each $s_{aug}(n)$ independently. Consequently, its complexity grows only linearly with the block size N and number of antenna elements m and exponentially with number of users d.

The complete BMML algorithm is summarized as follows:

1. Set polynomial order p based either on a priori information about the time variation of the channel or a design assumption.
2. At k = 0, set $(\hat{A}_{aug})_k$ by an initialization procedure (Section 4.2.1).
3. Increment k and solve $\hat{S}_{aug}$ from (4.7) by enumeration at each n.
4. Estimate $(\hat{A}_{aug})_k$ using (4.8).

5. Repeat step 3 and 4 until the algorithm converges locally, i.e., until $(r)_k - (r)_{k-1} \to 0$. The attained solution is referred to as a local solution of BMML.

6. Check the local solution for near-global convergence (defined below). If it is reached, output the corresponding solution, referred to as a plausible or near-global solution. Otherwise restart BMML at step 2 with a different $(\hat{A}_{aug})_0$, which may be selected randomly or based on some initialization routine, as suggested in Section 4.2.1. Finally the algorithm stops when either the near-global convergence is reached or the number of iterations exceeds a preset number. In latter case, the local solution that corresponds to the least r_k is chosen.

In step 4, $\hat{S}_{aug}$ is required to have full row rank. Lemma 4.2.4 in Section 4.2.2 shows that if $\hat{S}_{aug}$ contains certain data structures (DS) it has a full row rank. Even if $\hat{S}_{aug}$ is rank deficient, one can use the rank-deficient LS method to obtain $\hat{A}_{aug}$ [82]. However, the probability that S_{aug} is rank deficient diminishes exponentially with N, as can be seen in Section 4.2.2. Section 4.2.1 describes an alternative remedy in sub-block algorithms, if it should occur for a smaller N.

Near-global convergence is defined as local convergence with $r_k \le \mu$, where μ is a threshold which will be described next. In contrast, global convergence is defined as local convergence with $r_k = \aleph_{LS}$ and the corresponding solution is the global solution. In step 6, we seek plausible rather than global solutions for two reasons: (1) $\aleph_{LS}$ is not exactly known a priori and (2) global solutions may not be found for the given number of iterations. Although local solutions of BMML exist ([93] [29]), there is no guarantee that global or even plausible solutions can be found within a given number of iterations. Simulation results show, on the other hand, that plausible solutions can almost always be found if μ is set properly.

We now determine a suitable value for μ as follows. It is reasonable to set the threshold close to the mean residual, $E\|X(N) - A_{aug}S_{aug}(N)\|_F^2$, denoted as r_{MLS}, where $\hat{A}_{aug}$ is the LS solution corresponding to the correct data and the average is taken over the signal, channel and noise ensembles, which are assumed mutually independent. The polynomial model error and the channel noise all contribute to r_{MLS}. From (2.13), the estimation error vector

$$e(n) = x(n) - \hat{A}_{poly}(n)\hat{s}(n) = \left[A(n) - \hat{A}_{poly}(n)\right] s(n) + w(n)$$

where $\hat{A}_{poly}(n) = \sum_{i,j,k} n^k (\hat{a}_k)_{j,i}$ represents the estimated polynomial channel model to the actual channel $A(n)$ (which may deviate from the polynomial channel model) and the $(\hat{a}_k)_{j,i}$

here denotes the coe±cients of the polynomial channel approximation. Then

$$r_{MLS} = E\sum_{n=1} |e(n)|^2 = \sum_{j=1}\sum_{i=1} i\,\mathrm{tr}(R_{j;i}) - \mathrm{tr}\left\{\Sigma^{-1} V R_{j;i} V^T\right\} + mN\sigma_w^2$$
$$= r_{model} + r_{noise} \tag{4.9}$$

where $V_{l;m} = m^{l-1}$; $l = 1;\ldots;p+1$; $m = 1;\ldots;N$, $\Sigma = VV^T$ and $R_{j;i} = \frac{1}{2}E\{a_{j;i}a_{j;i}^*\}$ which can be calculated once the channel statistics are known. Note in obtaining (4.9), we have assumed that the estimation error is much smaller than the model error. One can show that if $f_dT \to 0$, $r_{MLS} \to r_{noise}$. However, if f_dT is not zero or if $p \to \infty$ r_{MLS} will be larger. Thus r_{noise} alone is inadequate to estimate the near-global convergence of BMML. Instead, one could use r_{MLS} or μ; de¯ned as

$$\mu = \alpha r_{model} + \beta r_{noise} \tag{4.10}$$

where $\alpha = 1.8$ and $\beta = 1$ are determined through simulation (see Section 4.2.3) to maximize the probability of near-global convergence while maintaining the computational load reasonable. The simulation covers a wide range of the noise and fade rate. See Fig. 4.4 and associated discussion for more detail.

Fig. 4.1 displays the computed r_{MLS} from (4.9) for channels (users) that have an identical Jakes fading channel model (all users have the same fade rate and average power level). SNR is de¯ned for user #1 (the ratio of signal power of user #1 to that of noise). As expected, increasing p leads to a smaller r_{MLS} for a ¯xed N and fade rate f_dT. However, a larger p also leads to poor convergence in BMML which results in a reduced performance. Hence, there is balance in choosing p for optimal BER performance, more on the e®ect on BER is given in Section 4.2.3. It also shows that below some fade rate the model error is insigni¯cant. Moreover, to minimize the model error a small N is desired, yet it con°icts with the data structure condition (introduced later), suggesting some optimal p and N should be used for a given fade rate.

It is also worth mentioning that the value of r_{MLS} or μ should not be used as a direct measure of the performance of the BMML as a lower μ will always produce a better result but at a much higher computational cost, as will be seen in Section 4.2.3.

BMML Initialization

The complexity of BMML in each iteration can be estimated straightforwardly, as in [29]. Since this class of algorithms is iterative in nature, the overall complexity is determined by

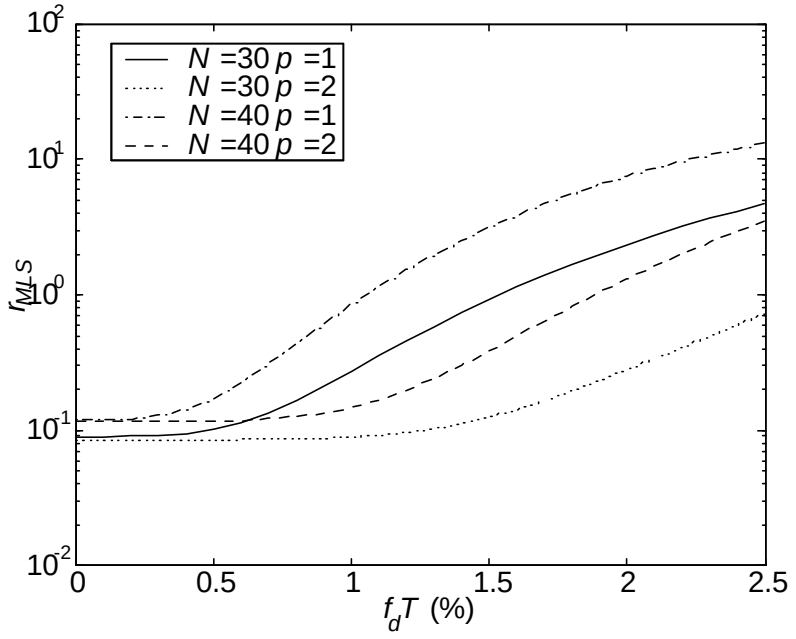

Figure 4.1: r_{MLS} vs. f_dT for different N and p. SNR=25dB.

the number of iterations needed to achieve plausible solutions. The number of iterations depends on the fade rate, noise level, number of antennas and the order of polynomials.

A reliable initialization strategy not only reduces the number of iterations but also improves performance. If $\mathbf{A}_{aug}$ has full column rank, the analytical constant modulus algorithm (e.g.,[19]), a noniterative blind multiuser detection method for constant modulus signals developed for time-invariant channels, may be used to initialize BMML. Unfortunately, $\mathbf{A}_{aug}$ may not have full column rank because the number of rows (m) may be less than the number of columns ((p + 1)d). In some cases, the embedded I.D. sequences in each block may be utilized to obtain $(\mathbf{A}_{aug})_0$ by (4.8) with $\mathbf{S}_{aug}$ replaced by the I.D. sequences if they form a full row rank signal matrix, resulting in a complexity lower than that of randomly started BMML. We refer to this I.D. assisted initialization strategy as the IDA initialization. However this approach may not be effective in fast fading channels, for reasons given in Section 4.2.3.

We propose here a blind initialization strategy which does not rely on the I.D. sequences or on the rank of $\mathbf{A}_{aug}$. We first obtain a very rough detection of the data at each antenna

separately. The detection is successive by user, as follows: assume a single user with unit channel gain at the antenna; detect the user's data, which should most closely resemble that of the strongest user; revise the user's channel gain estimate (assumed time invariant) by a trivial least squares; and subtract the reconstructed user signal (data times gain) from the antenna observation to reveal the remaining users. Repeat $d-1$ times to obtain estimates of the data for the remaining users. Note that there will be a different set of data decisions, with many decision errors and user order permutations, at each antenna. We choose one of those decisions sets as a $S_{aug,0}$ and from it obtain the corresponding $\hat{A}_{aug,0}$ by least squares (4.6). The set of $\hat{A}_{aug,0}$ is then used to start the regular BMML (4.7), (4.8). If the residual of the local solution produced by BMML is too large for it to be considered a plausible solution, we restart BMML by selecting an approximate decision set from a different antenna as new $S_{aug,0}$. One implicit assumption of this approach is that signals are roughly orthogonal, which can be justified if the signals are random and the block size is large enough. This blind initialization strategy is ad hoc and is not guaranteed to converge globally, however, as will be seen in the simulation, it works quite effectively for the scenarios considered.

Sub-Block Based Algorithms

In fast fading channels, the optimal block size may be less than the actual data frame/slot size. In this case we can apply BMML to sub-blocks of size N_{sub} each. We consider the difficult case in which only the first sub-block contains an I.D. sequence, either because the signal format is dictated by existing standards or because more frequent I.D. sequences would reduce capacity. If detection were applied independently from one sub-block to another, the ordering of the users in the recovered data could vary across the full block. Thus, some means of order restoration strategies are needed. Several strategies are possible, e.g.: the initial $\hat{A}_{aug}$ in a sub-block takes the value inherited from the last iteration on the previous sub-block, optionally with partial overlap of sub-blocks; operate backward from the next sub-block with an I.D. sequence if the algorithm fails to converge when operated in the forward direction. Here we detail the first two.

The first method is the sub-block by sub-block detection (SBSD) with no sub-block overlapping. More specifically, the channel estimate for the $(i-1)$th sub-block $\hat{A}^{i-1}$ is used

to initialize the channel estimate in the subsequent one as

$$\hat{\mathbf{A}}^{(i)}_{0} = \hat{\mathbf{A}}^{(i-1)} \tag{4.11}$$

where $\hat{\mathbf{A}}^{(i)}_{0}$ is the initial channel estimate of the $i$the sub-block from which $\hat{\mathbf{A}}^{(i)}_{aug\ 0}$ can be calculated.

The second method introduces sub-block overlapping. In each subsequent sub-block, only the last part of data needs to be detected with the channel again initialized via (4.11). In the extreme, they are overlapped such that each sub-block contains only the last signal vector that needs to be detected. In this case the channel estimate in the preceding sub-block is very close to that in the current one, so that overlapping provides a more robust way to estimate the channel gain in the current sub-block and consequently a more reliable way to join the data in sub-blocks through better gain initialization. This approach is referred to as the sliding window feedback detection (SWFD) since it is analogous to the classical decision feedback detector [90].

If any of the sub-blocks $\hat{\mathbf{S}}_{aug}$ is rank de¯cient, the algorithms fail to compute the $\hat{\mathbf{A}}_{aug}$ correctly through (4.6). One way to overcome this is to increase the sub-block size or combine adjacent sub-blocks to reduce the singularity problem, at the expense of causing larger channel model errors. For SWFD, the problem is somewhat relaxed as one can simply append the newly detected $\hat{\mathbf{b}}_{aug}(n)$ to $\hat{\mathbf{S}}_{aug}$ and defer the $\hat{\mathbf{A}}_{aug}$ update until the full rank condition is ful¯lled.

4.2.2 Selected Properties of BMML

Limited Ambiguity

We have seen in Section 4.2.1 that BMML solutions are not unique and have a multiplicity of at least the size of the equivalence set. The question is: Are there more solutions for the same residual?

Ideal BMML has multiple solutions. However, with probability one they are con¯ned to a single equivalent set. The proof will be in three parts:

(a) With probability one, all solutions have the same full row space; that is, any two solutions $\hat{\mathbf{S}}^{1}_{aug}$ and $\hat{\mathbf{S}}_{aug}$ are related by a nonsingular transformation matrix $\mathbf{T}_{aug}$ such that $\hat{\mathbf{S}}^{1}_{aug} = \mathbf{T}_{aug}\hat{\mathbf{S}}_{aug}$. (In the absence of noise, we require the true $\mathbf{S}_{aug}$ to have full row rank).

(b) If any two data sets $\hat{S}_{aug}$ and $\hat{S}_{aug}$ have the same full row space, and if any of them satisfies a certain data structure condition, then $\hat{S}$ and $\hat{S}$ have the same full row space.

(c) If any two data sets $\hat{S}$ and $\hat{S}$ have the same full row space, and if any of them satisfies a certain data structure condition, then $\hat{S}$ and $\hat{S}$ are equivalent.

First we show part (a) when noise is absent.

Lemma 4.2.1 In the absence of channel noise, if the true data matrix S_{aug} has full row rank, and true channel matrix A_{aug} satisfies the UM condition, then all solutions have the same row space.

Proof. Since the rows of S_{aug} are linearly independent by assumption, they span a space of dimension $d(p+1)$. Consider the most difficult case, with a single antenna. Since the elements of A_{aug} are drawn from a continuous ensemble, the row vector $X = A_{aug}S_{aug}$ has a component in the direction of each of the rows with probability one. Any other solution $X = \hat{A}_{aug}\hat{S}_{aug}$ must therefore span the same space at least, and because it also has $d(p+1)$ rows, the two row spaces are the same. ■

It can be shown that with probability one, UM condition is satisfied. Suppose A_{aug} maps two different signal vectors $s_{aug;1}$ and $s_{aug;1}$ to a common point, i.e., $s_{aug;1} \neq s_{aug;2}$ but $A_{aug}(s_{aug;1} - s_{aug;2}) = 0$. Since A_{aug} is a continuous random process, then with probability one, $A_{aug}(s_{aug;1} - s_{aug;2}) \neq 0$ unless $s_{aug;1} = s_{aug;2}$; which is a contradiction. Therefore with probability one, A_{aug} satisfies UM condition.

Then we show part (a) when noise is present.

Lemma 4.2.2 In the presence of channel noise, all solutions have full row rank with probability one.

Proof. Suppose there exists a solution pair ($\hat{A}_{aug}$ and $\hat{S}_{aug}$). It is clear from (4.3) that X is not confined to any subspace, since additive noise has no such constraint. We now prove the lemma by contradiction. If $\hat{S}_{aug}$ does not have full row rank, then construct another data matrix $\hat{S}'_{aug}$ by replacing a linearly dependent row with an arbitrarily selected signal sequence that is independent of the other rows. Then with probability one,

$$\min_{\hat{S}'_{aug};\,\hat{A}_{aug}} \left\| X - \hat{A}_{aug}\hat{S}'_{aug} \right\|_F^2 < \min_{\hat{S}_{aug};\,\hat{A}_{aug}} \left\| X - \hat{A}_{aug}\hat{S}_{aug} \right\|_F^2$$

since $\hat{S}'_{aug}$ has a higher dimension than $\hat{S}_{aug}$, more signal energy from X is projected onto the former signal row subspace and hence have a less residual. Thus, with probability one, all solutions have full row rank (to achieve the minimum residual). ■

Lemma 4.2.3 In the presence of the channel noise, all solutions have the same row space.

Proof. Suppose there exists a solution pair ($\tilde{A}_{aug}$ and $\tilde{S}_{aug}$) in addition to ($\hat{A}_{aug}$ and $\hat{S}_{aug}$). In this case $\tilde{A}_{aug}$ and $\tilde{S}_{aug}$ satisfy

$$\min_{\tilde{S}_{aug},\,\tilde{A}_{aug}} \left\| X - \tilde{A}_{aug}\tilde{S}_{aug} \right\|_F^2 = \min_{\hat{S}_{aug},\,\hat{A}_{aug}} \left\| X - \hat{A}_{aug}\hat{S}_{aug} \right\|_F^2 . \tag{4.12}$$

Substituting $\hat{A}_{aug}$ and $\tilde{A}_{aug}$ with their corresponding LS solutions in (4.12) results in

$$\begin{aligned} 0 &= \min_{\hat{S}_{aug}} \left\| X\hat{P} \right\|_F^2 - \min_{\tilde{S}_{aug}} \left\| X\tilde{P} \right\|_F^2 \\ &= \operatorname{tr}\left[X \left(\hat{P} - \tilde{P} \right) X^{y} \right] \end{aligned} \tag{4.13}$$

where $\hat{P}$ and $\tilde{P}$ are the orthogonal projection operators associated with $\hat{S}_{aug}$ and $\tilde{S}_{aug}$, respectively.

We now show $\Pr\left[\tilde{P} = \hat{P} \right] = 1$ if noise variance $\sigma_n^2 \neq 0$. If $\tilde{P} \neq \hat{P}$ (4.13) requires X to be either orthogonal to $\hat{P} - \tilde{P}$ or to be zero. Since $X = A_{aug}S_{aug} + W$ (4.3), either case requires W to lie in one or more of a finite number of subspaces of zero volume. Thus with probability one X cannot satisfy (4.13), unless $\tilde{P} = \hat{P}$. In other words, to satisfy (4.13) $\Pr\left[\tilde{P} = \hat{P} \right] = 1$.

Pre-multiplying $\tilde{P} = \hat{P}$ by $\tilde{S}_{aug}$, we have $\tilde{S}_{aug} = \tilde{S}_{aug}\hat{S}_{aug}^{y}(\hat{S}_{aug}\hat{S}_{aug}^{y})^{-1}\hat{S}_{aug}$ since $\hat{S}_{aug}$ has full rank from Lemma 4.2.2. Thus all solutions have the same row space. ■

Next we show part (b).

Condition 4.2.2 The signal S and its estimate contains (in any order) the following distinct column vectors once,

$$S_{sp} = \begin{cases} \begin{bmatrix} q_1 & q_2 & \cdots & q_Q & q_1 & \cdots & q_1 & \cdots & q_1 & \cdots & q_1 \\ q_1 & q_1 & \cdots & q_1 & q_2 & \cdots & q_Q & \cdots & \vdots & & \vdots \\ \vdots & \vdots & & \vdots & q_1 & \cdots & q_1 & & \vdots & & \vdots \\ \vdots & \vdots & & \vdots & \vdots & & \vdots & & q_1 & \cdots & q_1 \\ q_1 & q_1 & \cdots & q_1 & q_1 & \cdots & q_1 & \cdots & q_2 & \cdots & q_Q \end{bmatrix}; & \text{for } d > 2 \\ \begin{bmatrix} q_1 & q_2 & \cdots & q_Q \\ q_1 & q_1 & \cdots & q_1 \end{bmatrix}; & \text{for } d = 2 \end{cases} \tag{4.14}$$

where $q_i \neq q_j$ for $i \neq j$, and the following d vectors,

$$S_{sp_2} = \begin{cases} \begin{pmatrix} q_1^1 & q_1 & \cdots & q_1 \\ q_1 & q_1^1 & \cdots & \vdots \\ \vdots & q_1 & \ddots & q_1 \\ \vdots & \vdots & & q_1^1 \\ q_1 & q_1^1 & & \\ q_1 & q_1 & & \end{pmatrix}; & \text{for } d > 2 \\ !\ ; & \text{for } d = 2 \end{cases} \tag{4.15}$$

$pQ + 1$ times each, where $q_1^1 \neq q_1$. This condition is referred to as the data structure (DS) condition. (The straightforward case in which $d = 1$ is not considered here.)

Note in the above condition, $q_1; :::; q_Q$ merely denote distinct members of Q, they do not necessarily mean in a order such that q_i is related to q_{i+1} by a constant phase factor of $e^{j2\pi/Q}$. To give an example of Condition 4.2.2, let $d = 3$ and $Q = 2$, i.e., 3 BPSK signals. Then one possible S_{sp} and S_{sp_2} would be

$$S_{sp} = \begin{pmatrix} 1 & -1 & 1 & 1 \\ 1 & 1 & -1 & 1 \\ 1 & 1 & 1 & -1 \end{pmatrix}; \quad S_{sp_2} = \begin{pmatrix} -1 & 1 & 1 & 1 \\ 1 & -1 & 1 & 1 \\ 1 & 1 & -1 & 1 \end{pmatrix}:$$

Lemma 4.2.4 *If two data sets, $\mathbf{S}_{aug}$ and $\mathbf{S}^1_{aug}$ have the same row space, and any of them satisfies Condition 4.2.2, then $\mathbf{S}$ and $\mathbf{S}^1$ have the same full row space.*

Proof. If two data sets, $\mathbf{S}_{aug}$ and $\mathbf{S}^1_{aug}$ have the same row space, we have

$$\mathbf{S}^1_{aug} = T_{aug}\mathbf{S}_{aug}: \tag{4.16}$$

To reach our goal, we first simplify (4.16). By partitioning T_{aug} into $(p + 1) \times (p + 1)$ equal size blocks with its $i; j$th block $T_{i;j}$, (4.16) can be written as

$$\mathbf{S}^1_{i\,i\,1} = \sum_{j=1}^{p+1} T_{i;j}\mathbf{S}_{j\,i\,1} \quad i = 1:::p + 1 \tag{4.17}$$

Comparing the first row elements of $\mathbf{S}^1_i$ and denoting the first row of $T_{i;j}$ as $t_{i;j}$, we have

$$n^{i\,i\,1}q(n) = \sum_{j=1}^{p+1} t_{i;j}\mathbf{b}(n)n^{j\,i\,1} \quad i = 1:::p + 1;\ n = 1; :::; N$$

where $q(n) \in Q$ is a first-row element of $\mathbf{S}$.

If a particular $\mathbf{b}(n)$ appears $(p+1)$ times at $n = n_1, \ldots, n_{p+1}$ in $\mathbf{b}$, we have

$$\begin{pmatrix} n_1^i q \\ \vdots \\ n_{p+1}^i q \end{pmatrix} = \begin{pmatrix} 1 & \cdots & n_1^p \\ \vdots & \ddots & \vdots \\ 1 & \cdots & n_{p+1}^p \end{pmatrix} \begin{pmatrix} t_{i;1}\mathbf{b} \\ \vdots \\ t_{i;p+1}\mathbf{b} \end{pmatrix}. \tag{4.18}$$

To ensure that the set of equations (4.18) can be found at least for one q, it is sufficient for $\mathbf{b}$ to appear $Qp+1$ times in $\mathbf{b}$ (so that at least one of the Q possible q's is selected $p+1$ times in $Qp+1$ trials). Since the Vandermonde matrix in (4.18) is invertible, we have

$$t_{i;j}\mathbf{b} = \begin{cases} 0, & \text{for } j \neq i \\ q, & \text{for } j = i \end{cases} \tag{4.19}$$

which holds for any $\mathbf{b}$. By the DS condition, the collection of distinct vectors $\mathbf{b}$ in S_{sp_2}, each satisfying (4.18), is of full row rank. (See the proof of Lemma 4.2.4 in Appendix B.) It then follows that $t_{i;j} = 0$ for $j \neq i$. By applying the same argument to other rows of $T_{i;j}$ we have $T_{i;j} = 0$ for $j \neq i$. Equation (4.17) can then be separated as

$$\mathbf{S}_i = T_{i;i}\mathbf{b}_i, \quad i = 0, \ldots, p. \tag{4.20}$$

Defining $D^{(i)} = \mathrm{diag}(1^i, 2^i, \cdots, N^i)$ and substituting S_i with $SD^{(i)}$, (4.20) can be further simplified as

$$\mathbf{S} = T\mathbf{b} \tag{4.21}$$

where T denotes any diagonal sub-blocks $T_{i;i}$. That is, the diagonal sub-blocks are identical. ■

Finally we show part (c).

Lemma 4.2.5 If $\mathbf{b}$ and $\hat{\mathbf{S}}$ have the same full row space and any of them satisfies Condition 4.2.2, then $\mathbf{b}$ and $\hat{\mathbf{S}}$ are equivalent.

Proof. See appendix B.2. ■

We summarize the above results in the following theorem:

Theorem 4.2.1 With probability one and Condition 4.2.2, all solutions are equivalent.

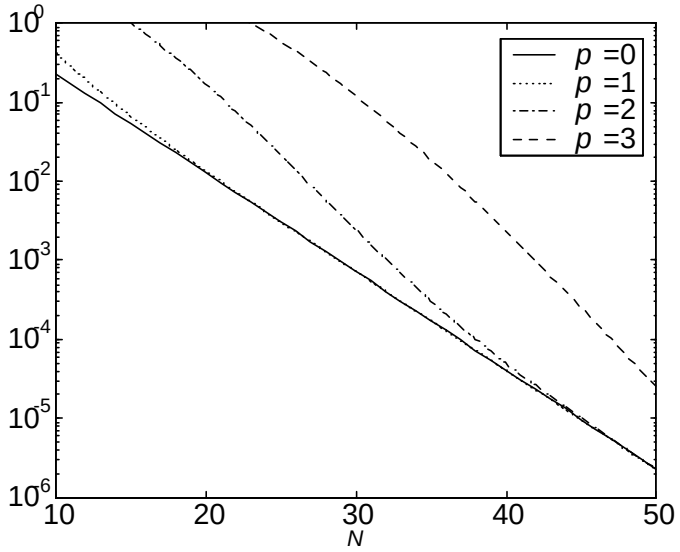

Figure 4.2: An upper bound on the probability of failing to meet the DS condition for p = 0; 1; 2 and 3.

The answer to our original question is that, with probability one, all BMML solutions are equivalent under Condition 4.2.2.

The probability for S not meeting the DS condition decreases exponentially with N, as can be illustrated from a calculated upper bound on the probability for d = 3 and Q = 2; in Fig. 4.2. See Appendix E for a derivation. We also see that as p increases the DS condition is less likely to be met. Notice that the results presented here are an upper bound on the probability of failing to meet a su±cient condition. In practice, however, better performance is achieved, as shown in simulation results later.

No Rank Requirement

The well-understood blind subspace based multi-signal detection algorithms, [29], [50], [31], [20], and many others, rest on the idea of the equivalence of signal and the array input subspaces, which requires the array input subspace to span the signal subspace fully, or equivalently, A_{aug} (and $\mathbf{h}_{aug}$) to have full row rank. In these algorithms the number of signals to be detected cannot be more than the number of antennas if processed at the symbol rate. Although fractional sampling eases the requirement to a certain extent [85],

the fundamental limitation still remains { the reconstructed channel matrix needs to be full rank and the number of antennas limits the number of signals to be detected.

In contrast, our approach is based on the unique projection or mapping of the signal subspace onto the input subspace that, depending on its dimension, may not fully span the signal subspace. By exploiting the FA property of signals more fully we have shown that with probability one, signals can be detected with ambiguity limited to interchanges and phase rotations. This result has both theoretical and practical importance. As a direct consequence, the signals are no longer limited by the number of antennas even when processed at the symbol rate, as we have shown. With this new insight it is clear that even ILSE (see Section 4.2) is able to detect multiple signals with an arbitrary number of antennas, since ILSE is a special case of BMML. (Prior to our work, it was not clear if ILSE would work with an arbitrary number of antennas.)

The UM condition is not only easier to meet but also more powerful than the full rank condition. For example, when applied to narrow beam antenna arrays, BMML is able to detect cochannel signals with identical directions of arrivals (DOA), as long as the UM condition is satis¯ed, e.g., if signals have di®erent fading dynamics or di®erent power. If channels are time invariant then ILSE will also be able to do the same. This prediction was indeed observed for a special case of two BPSK signals [30]. Unfortunately the discussion there was not generalized to more signals case or cases where $m \cdot d$ and no insight was given as to why it works since clearly the full rank condition is not ful¯lled in this case.

Note that with UM and DS conditions alone, signals can only be separated. To identify them, we need that the data (user streams) are di®erentially encoded and an I.D. sequence per user per block is inserted.

4.2.3 Numerical Examples

Simulation in this section and throughout the thesis is used to evaluate the performance of di®erent detection algorithms. The jth antenna output $x_j(n)$ is generated as follows. First, the time-varying Rayleigh fading channel $a_{ji}(n)$; connecting user i to the jth antenna, is generated by passing a complex Gaussian random process of some average power through a ¯lter which has the desired Jakes' U-shape spectrum (refer to Section 2.2.3 for more detail). The ith user's sequence $\{s_i(n)\}$, drawn randomly from a constellation set of size Q, is multiplied by $a_{ji}(n)$ to form the ith signal's contribution to $x_j(n)$, de¯ned as $x_j(n) = \sum_{i=1}^{d} a_{ji}(n)s_i(n) + w_j(n)$, i.e., (2.12), as shown in Fig. 4.3. Assume the data energy is

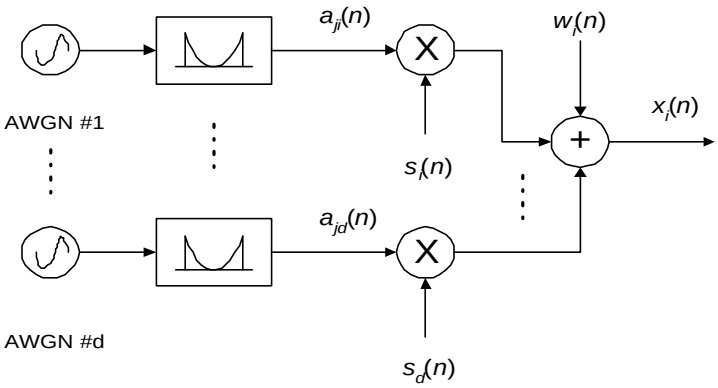

Figure 4.3: Signal $x_j(n)$ generation for the jth antenna.

normalized such that $|s_i(n)|^2 = 1$, and SNR is defined as $E[a_{11}(n)a_{11}^*(n)]=\sigma_w^2$, that is, the average signal energy of the first user to the first antenna with respect to the noise variance. For convenience, we also assume the any user has the same average signal energy across the antenna array, that is, $E[a_{j1}(n)a_{j1}^*(n)] = E[a_{11}(n)a_{11}^*(n)]$, $j > 1$.

The optimal channel model order p is determined based on the fade rate f_dT and the (sub-) block size (N_{sub}) N which is determined based on the DS condition. If we require the probability of not meeting the DS condition to be less than some value, then from Fig. 4.2, one can estimate N, the block size. From Fig. 4.1 and a given N; one can estimate p for a given fade rate f_dT. If f_dT is unknown, it can be estimated or its design value (e.g., f_dT = 1%) can be used to estimate p. Based on the estimated p, μ can be computed from (4.9). For a typical test, 1000 or more blocks of signals, each of size N symbols long, are generated. The initial channel gain $(\mathbf{h}_{aug})_0$ contains the coefficients of the polynomial channel model, and is either obtained from some initialization or selected randomly.

In this section, we demonstrate the performance of the proposed BMML by detecting three differential-encoded and equal-powered BPSK signals. For display clarity, only the averaged BER for three equal powered signals are presented. Most of the presented results are for $m = 6$ antennas; however, the effect of the number of antennas is explicitly evaluated further in this discussion. The errors are counted for each user assuming the order of the users is available, e.g., via I.D. sequences and SNR is defined for user #1, i.e., the ratio of signal power of user #1 to the channel noise variance.

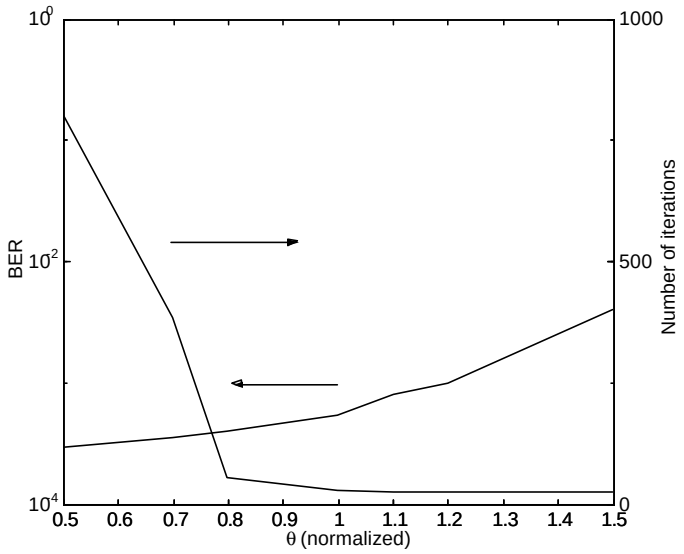

Figure 4.4: BER performance and computational load (measured in terms of number of iterations) of BMML vs. μ (normalized so that $\mu = 1$ corresponds to the value given by (4.10)). The number of antenna $m = 6$; the number of signals $d = 3$; block size $N = 20$ and SNR=15dB.

First, we would like to know how to choose the threshold μ. Without concerning the computational load, of course μ should be as small as possible. However, to balance the performance gain and computational complexity, there is a range of optimal values for μ. Fig. 4.4 shows the simulated performance and computational complex curves as a function of the normalized μ. As seen from the figure, as the normalized $\mu < 0{:}8$; the performance gain is diminishing but the computational load increases drastically. On the other hand, as the normalized $\mu > 1{:}1$, the computational complexity remains unchanged but the performance degrades significantly. From this figure, it seems the best value for the normalized μ should be around 1, which is the value suggested by (4.10).

Next, we test the robustness of BMML to channel model mismatch. More specifically, we test BMML of $p = 0; 1$ under a random static Gaussian channel. As can be seen from Fig. 4.5, BER improves quickly initially with increasing block size N and slowly for larger N, suggesting that the performance is dominated by the DS condition for small N and by noise for large N. Extensive simulations show that $\mathbf{S}_{\text{aug}}$ is almost never rank deficient but sometimes does not satisfy DS condition. As expected, BMML with $p = 0$

is better at this test because: i) it is a more constrained optimization problem (fewer parameters to estimate) and hence less detection noise; ii) channel is perfectly characterized (no channel order mismatch). Nonetheless, BMML with p = 1 shows good performance in static channels. The converse, BMML with p = 0, on the other hand, performs poorly in time varying channels, as can be seen in the next test.

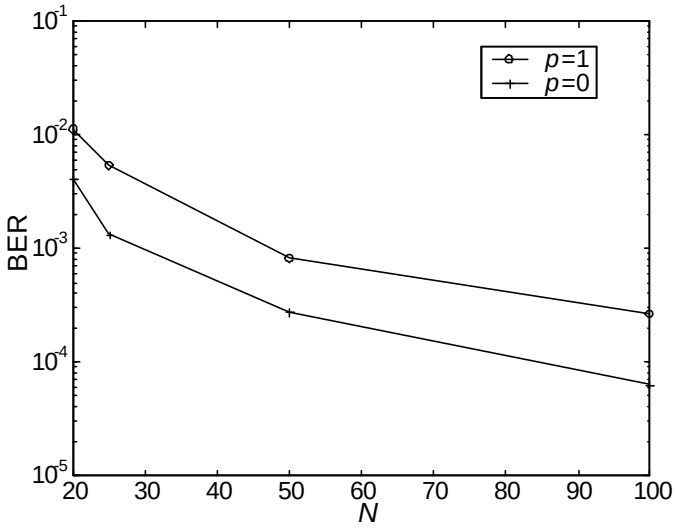

Figure 4.5: Simulated BMML performance for p = 0 and 1 on random static channels for di®erent block sizes for m = 6 and three equal power users each with SNR=5dB.

In the third test we study BMML for di®erent N; p at a ¯xed fade rate f_dT = 1%. In today's systems, this is considered fast fading. If we set the probability of not satisfying the DS condition 0.001, then N ¼ 30, as seen from Fig. 4.2. From Fig. 4.1, p should be less than two (p < 3) so that the LMS residual r_{LMS} is small. We now examine the test result. As seen from Fig. 4.6, the static channel model (corresponding to ILSE) is inadequate at this fade rate and the performance degrades with increasing N, indicating the performance is limited by model mismatch. The linear channel model performs very well for $20 \cdot N < 40$ and the quadratic channel model (p = 2) is well suited at this fade rate for N > 30. The channel model of order 3 produces a higher BER. At ¯rst this is somewhat a surprise, since a higher channel model has a reduced channel model error which would imply an improved performance. However, it is found from simulations that the local convergence for higher order channel models is harder to detect than the lower order counterpart, as the residual

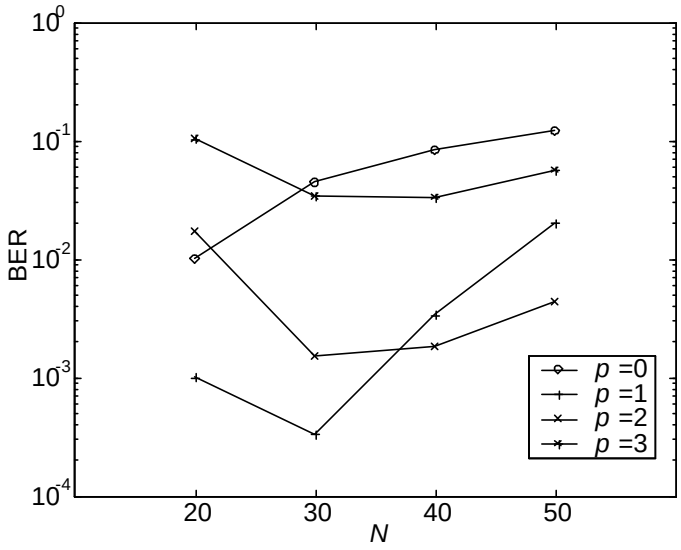

Figure 4.6: Performance of BMML for three equal power users at different block size N and polynomial order p: f_dT = 1%; m = 6 and each user with SNR=25dB.

difference between global and local convergence is much smaller for higher order models. As a result, the overall performance decreases due to lack of global convergence. As we can see, the simulation result confirms our estimate that p < 3. In the remaining experiments we select p = 1 and N · 30.

The fourth experiment evaluates BMML at different fade rates. Fig. 4.7 shows that BMML performs very well for fade rates up to f_dT = 0:01, and degrades only slightly with increasing fade rate. The results also show performance floors, a typical fading channel performance curve for one user case. The difference here is that the floor is caused by the combination of possibly failing to satisfy the DS condition and channel model error.

Two sub-block algorithms, SBSD and SWFD, are studied next. BER is defined as the number of errors per block with one known I.D. embedded for user order identification. Random initialization is used in this test. The block size N = 100 and sub-block size N_{sub} = 25. The results are shown in Fig. 4.8. It clearly demonstrates that the proposed sub-block detection algorithms are effective at fade rate f_dT · 1%. As expected, SWFD is better, at the same fade rate, than SBSD due to more robust channel estimation; however, the difference is much smaller at f_dT = 1%. This suggests that the channel model error

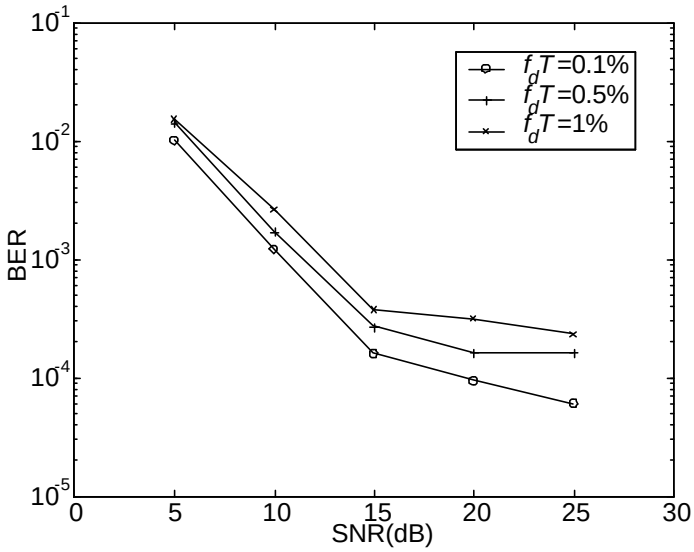

Figure 4.7: Simulated BMML performance for three equal power users at di®erent fade rate. p = 1; N = 25 and m = 6.

dominates the performance at higher fade rates.

Finally, we examine the e®ect of number of antennas m on BMML. Fig. 4.9 shows that BMML is able to detect more signals than the number of antennas. The performance improves with m due to antenna diversity and improved SNR. However, after m > 4 the improvement is marginal suggesting it is limited by other factors, such as failing to satisfy the DS condition and model errors. Fig. 4.10 clearly shows that BMML is able to detect multiple signals with only one antenna although the performance degrades as the number of the signals increases due primarily to reduced near-global convergence because of reduced diversity, as can be seen next. Note that when the number of antennas increases, the slope does not change, which suggested that all signals enjoy the same diversity gain, a result that has been observed in a non-blind multiuser ML detector [6].

BMML complexity is measured as product of the number of operations in each iteration and the average number of iterations and re-initialization. We will study the computational load of the initialization routines in this section while leaving the detailed estimate of the computational load of BMML in Appendix D. We assume that the ¯rst three symbols of each of the signals are known in each block and can be used for IDA initialization. The

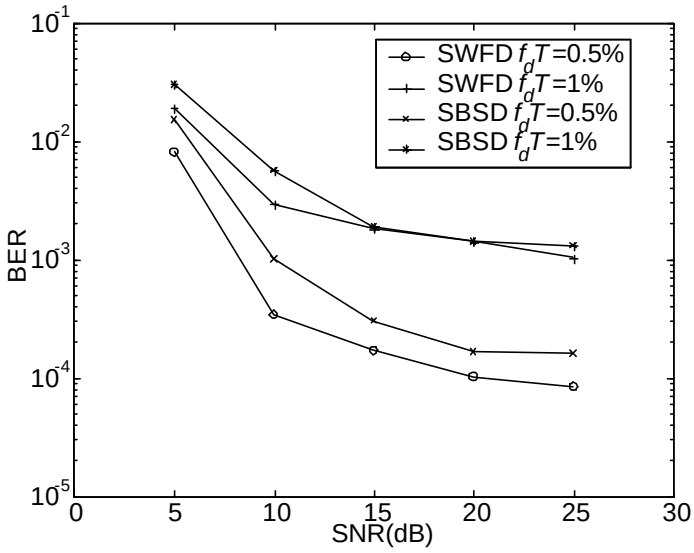

Figure 4.8: Simulated BER results of SWFD and SBSD for three equal power users at di®erent fade rate. p = 1; N_{sub} = 25; N = 100 and m = 6.

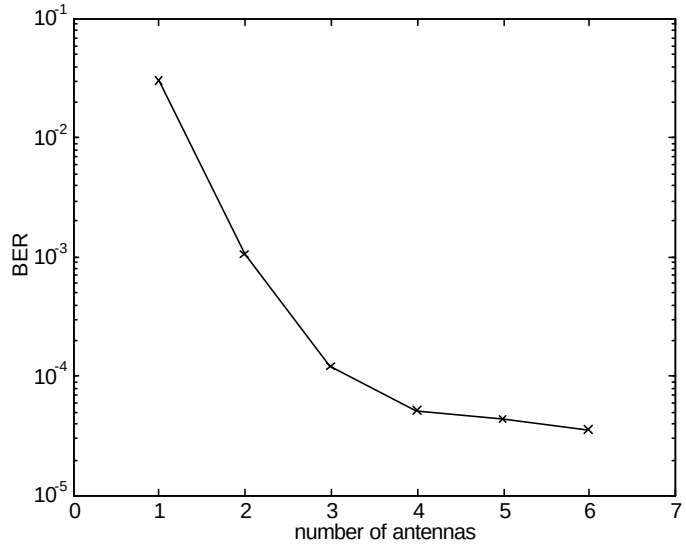

Figure 4.9: BER performance of BMML for three equal power users with di®erent number of antennas m. p = 1, N = 30, f_dT = 0:5% and SNR=25dB.

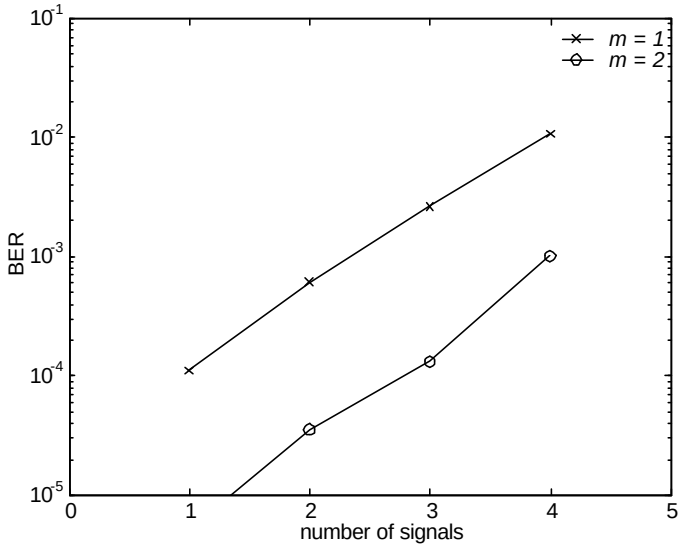

Figure 4.10: Result of BMML for multiple signals with a variable number of antenna. f_dT = 0:5%, N = 30 and SNR=45dB.

results are tabulated in Table 4.1. Both IDA and the blind strategy show a common result: the average number of iterations improves with the number of the antennas and degrades with the fade rate. We also found (not shown here) that at low SNRs the complexity is slightly higher but not as sensitive to noise as to other parameters. As expected, IDA is better at low fade rate than the blind initialization since the channel estimate is reliable. At high fade rate, however, IDA is worse as the channel estimate, based on I.D. sequences (3 symbols for 3 users), does not approximate well the overall channel variations in each (sub) block due to the channel model error. In contrast, the blind initialization utilizes all data in each (sub) block and the estimate is more reliable. The blind initialization is a preferred choice because it does not rely on the I.D. sequences which may not be used for channel estimation, since I.D. sequences are designed for signal identi¯cations and may be too short to have full d row rank and consequently, cannot be used for channel estimation. Moreover, the blind initialization is less sensitive to fade rate.

	IDA (N_{iter}=$N_{re:init}$)			Blind Initialization (N_{iter}=$N_{re:init}$)		
	f_dT = 0:1%	f_dT = 0:5%	f_dT = 1%	f_dT = 0:1%	f_dT = 0:5%	f_dT = 1%
m = 6	2.0/1.0	2.2/1.1	17.9/6.4	4.0/1.5	4.9/1.7	8.3/2.3
m = 4	2.8/1.4	4.0/1.83	35.7/12	4.4/1.6	6.0/1.9	12.2/3.2
m = 2	6.2/3.04	23.5/9.5	62/20	8.0/2.5	11.6/3.2	23.9/5.8

Table 4.1: Average number of iterations and re-initializations in BBJD for 3 equipower users. SNR=15DB and p=1.

4.2.4 Conclusion

A novel blind multiuser maximum likelihood detection for cochannel signals transmitted over time-varying fading channels is developed. It is based on the ideas of the unique mapping of the data subspace onto the input subspace and the low order polynomial basis expansion of the time-varying channels over appropriately selected data window. One attraction of BMML is that it does not require a minimum number of antennas, although increasing the number of antenna elements improves the performance due to spatial diversity. Signal detectability is studied and minimum su±cient conditions on signal data structure are found. A simple yet e®ective initialization strategy is proposed to improve convergence speed and reduce the complexity of BMML. Two sub-block detection algorithms are also presented, which can be used for moderate fade rates. Simulations demonstrate that these approaches are e®ective in detecting cochannel signals in time-varying channels. The optimal data window size and order of polynomials are determined by the fade rate. Although focused on PSK systems in this section, the entire work can be generalized to other linear modulation systems.

4.3 Application to CPM Modulation

In this section we study the BMML detector for continuous phase modulation (CPM) on time-varying channels. CPM and its variants have found wide applications in wireless communications, such as GSM and paging systems; the former uses Gaussian smoothed MSK modulation and the latter uses CPFSK modulation. The problem di®ers from PSK in some key respects: a di®erent DS condition is needed; CPM signals usually have a non-rectangular frequency pulse shape; and there is memory in CPM, which implies di®erent detection algorithms are needed.

This work appears to be first in 1) using BMML to detecting CPM signals directly (not through linear modulation approximations, as in [101]) and 2) in time-varying channels. This work is also an extension of [34]: i) the piecewise linear channel approximation is extended to polynomials and the optimal polynomial order p is determined based on the fade rate and block size; ii) CPFSK is generalized to CPM. In addition, an effective initialization routine is presented and the number of signals is estimated. By exploiting the data structure of signals alone, the algorithm is able to detect and separate cochannel signals without a hard minimum number of antennas. Just as in PSK modulation, it requires only minimal information, such as a single I.D. sequence per user per block, in order to identify the detected user streams.

This section is organized as follows. We first establish the BMML algorithm and then study the solution of BMML, which shows that all the BMML solutions are within a simple equivalence set with probability one. We next introduce an effective initialization routine and a simple way to estimate the number of signals within BMML. Similar to PSK BMML, a sub-block BMML that is appropriate for faster fading is also proposed. Finally, BMML performance is evaluated through simulation.

4.3.1 Blind Multiuser Maximum Likelihood (BMML) Scheme

For convenience we rewrite the transmitted waveform for the ith user (2.4) as

$$\begin{aligned} y_i(t; s_i(n)) &= \exp\{j\phi(t; s_i(n))\} \\ &= \exp\{j[\theta_{i;n} + \mu(t; s_i(n))]\}, \; nT \le t < (n+1)T, \; i = 1, \ldots, d \end{aligned} \tag{4.22}$$

where

$$\theta_{i;n} = \pi h \sum_{k=1}^{n-1} s_i(k) \tag{4.23}$$

and

$$\mu(t; s_i(n)) = \pi h s_i(n) q(t - nT), \; nT \le t < (n+1)T \tag{4.24}$$

where as before, h denotes the modulation index. To simplify the discussion we assume a full response pulse, so that the phase pulse shape, a non-decreasing function in t, $q(t) = 0$ for $t < 0$ and $q(t) = 1$ for $t > T$. After a rate P sampler, the system equation (2.11) is

$$x_j(t) = \sum_{i=1}^{d} a_{j;i}(t) y_i(t; s_i(n)) + w_j(t), \; j = 1, \ldots, m \tag{4.25}$$

sampled at time instant $t = n + \tau$ where n is an integer and $\tau = 1/P, \ldots, (P-1)/P$. $w_j(t)$ is the additive white Gaussian noise observed at the jth antenna. The function of the rate P sampler is two fold: for single user detection, it is used to remove phase transition ambiguity; for multiuser detection, it is used to remove the detection and phase transition ambiguities, as will be shown in Section 4.3.2. Note here we abuse the time index t in a understandable way, that is, we use t to represent a discrete time variable in this section to avoid introducing a new notation. In matrix form, (4.25) becomes

$$x(t) = A(t)y(t; S(n)) + w(t) \tag{4.26}$$

For clarity we assume in this subsection that polynomial channel model (2.16) in Chapter 2 holds over the entire block. Later in Section 4.3.1 we relax this assumption.

Under (2.16), (4.26) can be transformed into a more compact form

$$\begin{aligned} X(t) &= \begin{bmatrix} A_0 & A_1 & \cdots & A_p \end{bmatrix} \begin{bmatrix} Y_0(t; S(n)) \\ Y_1(t; S(n)) \\ \vdots \\ Y_p(t; S(n)) \end{bmatrix} + W(t) \\ &= A_{aug} Y_{aug}(t; S(n)) + W(t) \end{aligned} \tag{4.27}$$

where $X(t) = (x(1/P), x(2/P), \ldots, x(t))$, $W(t) = (w(1/P), w(2/P), \ldots, w(t))$, and $Y_k(t; S(n))$ is a matrix whose ith column is defined as

$$(i/P)^k y(i/P; S(n)), \quad n = \lfloor i/P \rfloor, \quad i = 1, \ldots, PN, \quad k = 0, \ldots, p \tag{4.28}$$

where $\lfloor \cdot \rfloor$ denotes the floor function (e.g., $\lfloor 3.4 \rfloor = 3$) and Y_0 is shortened as Y.

BMML Algorithm

For notational convenience, in the following we scale the time index up by a factor P so that the new time index is an integer, that is,

$$t = n + \tau, \quad n = \lfloor t/P \rfloor, \quad \tau = 0, \ldots, P-1. \tag{4.29}$$

From the assumption of Gaussian noise, $X(t)$ is a set of conditionally Gaussian processes and thus the ML criterion (3.25) in Chapter 3 is equivalent to the following least squares problem

$$\min_{S(N); A_{aug}} \| X(N) - A_{aug} Y_{aug}(S(N)) \|_F^2. \tag{4.30}$$

However, a direct solution of (4.30) is formidable due to the vast search space of S(N) (2 Z^{Nd}). Just like in Section 4.2, we takes a similar two-step approach that estimate S(N) and A_{aug} separately and iteratively:

$$\hat{S}(N) = \arg\min_{S(N)} \left\| X(N) - \hat{A}_{aug} Y_{aug}(S(N)) \right\|_F^2 \tag{4.31}$$

with $\hat{A}_{aug}$ ¯xed and

$$\hat{A}_{aug} = X \Phi_{aug}^H (\Phi_{aug} \Phi_{aug}^H)^{-1} = X \Phi_{aug}^{\#} \tag{4.32}$$

with Φ_{aug} ¯xed.

Since CPM has memory, each data vector s(n) in (4.31) cannot be detected independently, as in Section 4.2. Experience tells us that a Viterbi algorithm (VA) may be used to reduce its complexity. To use VA we need to de¯ne a state variable at time n. For this we write $y(t; S(n)) = y(\theta_n; \mu(t; s(n)))$ with $\theta_n = (\theta_{1;n}; :::; \theta_{d;n})^T$ and $\mu(t; s(n)) = (\mu(t; s_1(n)); :::; \mu(t; s_d(n)))^T$. With this substitution we may partition $Y_{aug}(Pn; S(n))$ as

$$Y_{aug}(Pn; S(n)) = [Y_{aug}(P(n-1); S(n-1)); Y_{aug}^P(\theta_n; s(n))] \tag{4.33}$$

where $Y_{aug}^P(\theta_n; s(n))$ denotes the last P columns of $Y_{aug}(Pn; S(n))$. From (4.33) it is clear that (4.31) can be solved by the standard VA if we de¯ne θ_n as the state variable with $\theta_{n+1} = \theta_n + \frac{1}{4}hs(n)$. Note that the branch metric in this case is a sum of sequences indicated in (4.31) taken over the P samples in $Y_{aug}^P(\theta_n; s(n)$. As well known, its complexity grows only linearly with block size N; the number of antennas m and the number of states.

For the matrix inverse in (4.32) to exist Φ_{aug} must have a full rank. Lemma 4.3.1 below shows that if Φ_{aug} contains a special data structure (Condition 4.3.1) it has a full rank. For clarity of presentation, let $\gamma_{i;n} = \exp\{j\theta_{i;n}\}$ and $\pm_{s_i(n)}(¿) = \exp\{j\mu(¿; s_i(n))\}$ with n and ¿ are de¯ned in (4.29). Then $y_i(t; s_i(n)) = \gamma_{i;n}\pm_{s_i(n)}(¿)$; $i = 1; :::; d$.

Condition 4.3.1 $\Phi(t)$ contains, in any symbol column order (de¯ned below), the following dP columns at time $¿ = 0; 1; :::; (P-1)=P$;

$$Y_{DS_1}(¿) = \begin{pmatrix} \gamma_{1;n_1}\pm_{q_1}(¿) & \gamma_{1;n_2}\pm_{q_2}(¿) & \cdots & \gamma_{1;n_d}\pm_{q_d}(¿) \\ \gamma_{2;n_1}\pm_{q_1}(¿) & \gamma_{2;n_2}\pm_{q_2}(¿) & & \vdots \\ \vdots & \vdots & \ddots & \gamma_{d-1;n_d}\pm_{q_d}(¿) \\ \gamma_{d;n_1}\pm_{q_1}(¿) & \gamma_{d;n_2}\pm_{q_2}(¿) & & \gamma_{d;n_d}\pm_{q_d}(¿) \end{pmatrix}$$

and the following dP columns at time $\tau = 0, 1, \ldots, (P-1)/P$;

$$Y_{DS_2}(\tau) = \begin{bmatrix} \alpha_{1,n_{d+1}} e^{j\theta_{q_{d+1}}(\tau)} & \cdots & \alpha_{1,n_{2d}} e^{j\theta_{q_{2d}}(\tau)} \\ \vdots & \ddots & \vdots \\ \alpha_{d,n_{d+1}} e^{j\theta_{q_{d+1}}(\tau)} & \cdots & \alpha_{d,n_{2d}} e^{j\theta_{q_{2d}}(\tau)} \end{bmatrix} = \left[\alpha_{n_{d+1}}, \ldots, \alpha_{n_{2d}} \right] \Phi(\tau).$$

where $q_i, q_{\bar{i}} \in Q$ but $q_i \neq q_{\bar{i}}$ and $\Phi(\tau) = \mathrm{diag}(e^{j\theta_{q_{d+1}}(\tau)}, \ldots, e^{j\theta_{q_{2d}}(\tau)})$. A symbol column is defined as any P consecutive columns in $Y(t)$ with time $t = n, \ldots, n + (P-1)/P$. In addition, $\det Y_{DS_2}(0) \neq 0$ and $\det(Y_{DS_1}(0) - \mathrm{diag}(Y_{DS_1}(0))) \neq 0$ where $\mathrm{diag}(\Phi)$ on a matrix denotes the diagonal of that matrix. These data structure requirements will be referred to as the data structure (DS) condition.

To see an example of $Y_{DS_1}(\tau)$ and $Y_{DS_2}(\tau)$, let $d = 2$; $h = 1/2$, $q(t) = t/T$ and $\alpha_{i,n_1} = 1$, $i = 1, 2$; then we have

$$Y_{DS_1}(\tau) = \begin{bmatrix} e^{j\pi\tau/T} & \alpha_{1,n_2} e^{j\pi\tau/T} \\ e^{j\pi\tau/T} & \alpha_{2,n_2} e^{j\pi\tau/T} \end{bmatrix} \quad \text{and}$$

$$Y_{DS_2}(\tau) = \begin{bmatrix} \alpha_{1,n_3} e^{j\pi\tau/T} & \cdots & \alpha_{1,n_4} e^{j\pi\tau/T} \\ \vdots & \ddots & \vdots \\ \alpha_{2,n_3} e^{j\pi\tau/T} & \cdots & \alpha_{2,n_4} e^{j\pi\tau/T} \end{bmatrix}$$

where α_{i,n_k}, $i = 1, 2$; are the initial phase factors of the waveform for user #1 or #2 at symbol time n_k.

Lemma 4.3.1 Ψ_{aug} has full rank under the DS condition.

Proof. Suppose there exist non-zero row vectors γ_k, $(k = 0, \ldots, p)$ that satisfy

$$\sum_{k=0}^{p} \gamma_k \Psi_k(t) = 0.$$

If, as in Y_{DS_2}, we take $s(n_1) = (q_1, \ldots, q_1)^T$, then the above equation can be expanded for samples $0 \leq \tau \leq p$ of symbol $s(n_1)$ as

$$\begin{bmatrix} 1 & Pn_1 & \cdots & (Pn_1)^p \\ 1 & Pn_1 + 1 & \cdots & (Pn_1 + 1)^p \\ \vdots & \vdots & \ddots & \vdots \\ 1 & Pn_1 + p & \cdots & (Pn_1 + p)^p \end{bmatrix} \begin{bmatrix} \gamma_0 \alpha_{n_1} \\ \gamma_1 \alpha_{n_1} \\ \vdots \\ \gamma_p \alpha_{n_1} \end{bmatrix} = 0 \tag{4.34}$$

where $\circ_{n_1} = (\circ_{1;n_1}; \circ_{2;n_1}; :::; \circ_{d;n_1})^T$. Since the Vandermonde matrix is non-singular, we have $\bar{}_k \circ_{n_1} = 0$ for $k = 0; :::; p$. As (4.34) holds for any $\circ_{n_1}$ it holds for $Y_{DS_2}(0)$, therefore, $\bar{}_k(\circ_{n_1}; :::; \circ_{n_d}) = 0$. By the DS condition $Y_{DS_2}(0)$ has full rank and hence, $\bar{}_k = 0$ which contradicts the assumption and therefore Lemma 4.3.1 follows. ■

Since (4.31) does not construct a '¯lter matrix' W such that $W^y X = Y$ in the noise-free case [20], BMML does not impose a minimum requirement on the number of antenna elements. As a result, the number of antenna can be arbitrary. Generally, however, increasing the number of antennas will improve the performance in fading conditions due to antenna diversity gain, as can be seen in Section 4.3.3.

For BMML to operate e®ectively the following implementation related questions need to be answered. How to choose the order of polynomial p? How to choose the initial value of $\hat{A}_{aug}{}_0$ or $\hat{\Gamma}_{aug}{}_0$? We will discuss each of them below.

Sub-block BMML

As just discussed, desirable block sizes are generally shorter as f_dT increases; conversely, shorter block sizes make is less likely that the users' data will satisfy the data structure condition of Section 4.3.1. If the optimum block size N_w so determined is less than the actual block size N we can divide the block into several sub-blocks of N_w symbols each and apply BMML in each sub-block. However, the independent application of BMML in each sub-block may result in permutations of the users from one sub-block to the next, because only the ¯rst sub-block carries the I.D. One remedy is to initialize each sub-block BMML with the converged channel estimate inherited from the previous sub-block. A detailed description of sub-block BMML is given in Section 4.3.3. This approach has been shown to improve the performance by reducing the detection-order ambiguity. It also increases the speed and accuracy of convergence since a good starting channel estimate leads to a faster convergence and a better solution. The same can be said of the original full-block BMML if the blocks are transmitted continuously.

BMML Initialization

The performance of an iterative algorithm is often limited by the initialization. For example, an inaccurate initial $\hat{A}_{aug}$ may cause BMML to converge slowly or to a poor solution. From Section 4.2.1, we know a good initial value of $\hat{A}_{aug}$ is the converged value from the previous

sub-block. However, for the first sub-block, or if BMML needs to restart within the block, we need a different initialization strategy. One possible way is to select the initial $\mathbf{H}_{aug}$ randomly. However, this approach is not very reliable, especially when the number of antennas is small and the fade rate is high. If $\mathbf{H}_{aug}$ has full row rank (which requires the number of antennas $m \geq (p+1)d$) then we speculate that the analytical constant modulus algorithm [19] could be used to initialize BMML. Unfortunately, if $m < (p+1)d$ this method fails.

We propose an alternative blind initialization strategy which does not rely on the rank of $\mathbf{H}_{aug}$, i.e., on the number of antennas. It was introduced in Section 4.2.1 and shown to be effective for PSK signals. Essentially it is a primitive successive interference cancellation detector. Though this detector is not on its own effective enough for detecting cochannel signals, it can be used to initialize BMML. In operation, the detection is performed successively by user and independently at each antenna, as follows: assume a single user with unit channel gain on antenna 1; detect the user's data, denoted as $\hat{\mathbf{b}}_1^1(n)$; revise the user's channel gain estimate (assumed time invariant) by a simple least-squares fitting, denoted as $\hat{\mathbf{a}}_1^1$; and subtract the reconstructed user signal $\hat{\mathbf{a}}_1^1\hat{\mathbf{b}}_1^1$ from the antenna observation to reveal the remaining users. Repeat $d-1$ times to obtain estimates of the remaining users and signal estimate $\hat{\boldsymbol{\gamma}}_{aug}^1$ for a given p. Repeat the above $m-1$ times for the remaining antennas and we obtain m estimates $\hat{\boldsymbol{\gamma}}_{aug}^i$, $i = 1, \ldots, m$, each of which can be used as a $\boldsymbol{\gamma}_{aug}(0)$ to start BMML. Note this initialization routine, like the one introduced in Section 4.2.1, does not guarantee the global convergence, and its effectiveness will be evaluated in Section 4.3.3.

Computational load of BMML

Since the operation of BMML is in essence an iterative VA processor (4.31) and LS solver (4.32), its computational load can then be estimated from these two processes.

The complexity of the VA is estimated as follows. If the modulation index h can be expressed as a fractional number $h = h_1/h_0$; the number state for each user is given as uh_0Q where $u = 1$ or 2 if h_1 is even or odd, respectively [90]. Thus for d users, the number is $(uh_0Q)^d$. For each state transition, there will be $m(p+1)P$ complex multiplications and additions (CMAs), which yields a total of $N\,m(p+1)P(uh_0Q)^d$ CMAs for detecting $S(N)$ in each iteration. For estimating channel gains, there will be about $2z^2(N - z/3) + mz^2 + Nmz$ CMAs (see Appendix D), where $z = (p+1)Pd$. Assume algorithm takes I iteration to reach

near-global convergence, we have the complexity of BMML $C_{BMML(CPM)}$ is

$$C_{BMML(CPM)} \approx [Nm(p+1)P(uh_0Q)^d + 2z^2(N - z/3) + mz^2 + Nmz]I \quad \text{CMAs}. \tag{4.35}$$

4.3.2 Solutions of BMML

One of the fundamental questions for a blind multi-user detection scheme is the uniqueness of the solution. Usually a blind detection system is under-determined and therefore the solution is not unique. In this case we would like know whether these multiple solutions are trivially related. This is the subject of the section.

To simplify the discussion, we first study the solutions of (4.30) and later that of BMML which solves (4.31) and (4.32) iteratively. Like linear modulations (Section 4.2), (4.30) has user-permutation and phase ambiguities; that is, the detected data sequences of any pair of users can be exchanged and/or the sequences given fixed phase shifts without affecting the residual of (4.30), provided the corresponding changes are made in A_{aug}. However, phase ambiguity in θ does not affect CPM signal detection for $\mathbf{b}$, since only the phase difference between symbol times is used in the detection.

Condition 4.3.2 The sample rate $P \geq \max\{p+2, 3\}$. This condition is referred to as the sample rate (SR) condition.

Lemma 4.3.2 If two solutions θ and $\tilde{\theta}$ are related by an invertible matrix T and either one satisfies the DS condition then T is an equivalence transformation.

Proof. See Appendix B.3. ■

Theorem 4.3.1 shows that solutions of (4.30) are within an equivalence set.

Theorem 4.3.1 All solutions of (4.30) belong to the same equivalence set with probability one, under conditions 4.3.1 and 4.3.2.

Proof. Suppose, in addition to h_{aug} and θ_{aug}, there exists another pair of simultaneous solutions ($\tilde{h}_{aug}$ and $\tilde{\theta}_{aug}$) such that $\tilde{\theta}_{aug} \neq \theta_{aug}$.

First we state that

$$\tilde{\theta}_{aug} = T_{aug}\theta_{aug} \tag{4.36}$$

with probability one, where T_{aug} is a $d(p+1) \times d(p+1)$ invertible matrix. The proof is similar to that of Theorem 4.2.1.

Next we show T_{aug} is a block diagonal matrix. Partitioning T_{aug} into $(p+1) \times (p+1)$ equal size blocks and denoting the first row of the i,jth block $T_{i,j}$ as $t_{i,j}$, a subset of (4.36) can be expanded as, for $i = 1:::p+1$; $\zeta = 0;1;:::;(p+1)$;

$$\pm_{s_1(n)}(\zeta)(Pn+\zeta)^{i-1}\circ_{1;n} = \sum_{j=1}^{p+1} \pm_{q_1}(\zeta)(Pn+\zeta)^{j-1} t_{i,j}\,\flat_n \tag{4.37}$$

if $\flat_k(n) = q_1$ for $k = 1;:::;d$. Expanding (4.37) for $\zeta = 0;1;:::;(p+1)$ (permitted by the SR condition) gives

$$\Phi v_i \circ_{1;n} = DVt_\flat \text{ or } Vt_\flat = \Re v_i \tag{4.38}$$

where

$$t_\flat = \begin{pmatrix} t_{i,1}\flat_n \\ t_{i,2}\flat_n \\ \vdots \\ t_{i,p+1}\flat_n \end{pmatrix}; \quad V = \begin{pmatrix} 1 & Pn & \cdots & (Pn)^p \\ 1 & Pn+1 & \cdots & (Pn+1)^p \\ \vdots & \vdots & \ddots & \vdots \\ 1 & Pn+p+1 & \cdots & (Pn+p+1)^p \end{pmatrix};$$

$D = \mathrm{diag}(1;:::;\pm_{q_1}(p+1))$; $\Phi = \mathrm{diag}(1;:::;\pm_{s_1(n)}(p+1))$; $\Re = \mathrm{diag}(\Re_0;\Re_1;...;\Re_{p+1}) = D^{-1}\Phi\circ_{1;n}$ with $\Re_0 = \circ_{1;n}$ and v_i is the ith column of V. As V is a $(p+2)$ by $(p+1)$ matrix, for linear equations (4.38) to have a unique solution $t_\flat$ we must have $\det[V;\Re v_i] = 0$ for $i = 1;:::;p+1$; or

$$\begin{vmatrix} 1 & Pn & \cdots & (Pn)^p & \Re_0 (Pn)^{i-1} \\ 1 & Pn+1 & \cdots & (Pn+1)^p & \Re_1 (Pn+1)^{i-1} \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 1 & Pn+p+1 & \cdots & (Pn+p+1)^p & \Re_{p+1}(Pn+p+1)^{i-1} \end{vmatrix} = 0; \tag{4.39}$$

We now show by induction on p that (4.39) implies all coefficients $\Re_k$ for $k = 1;:::;p+1$ are equal. For $p = 0$ it is obvious that $\Re_1 = \Re_0$. Suppose for $p = k$ the statement is also true. For $p = k+1$, by expanding (4.37) for $\zeta = 0;1:::;(k+2)$ and performing elementary operations on the columns of the matrix, (4.39) for $i = 1;:::;k+2$ can be reduced to

$$\begin{vmatrix} 1 & Pn & \cdots & (Pn)^k & -\frac{\Re_0 - \Re_{k+2}}{k+2}(Pn)^{i-1} \\ 1 & Pn+1 & \cdots & (Pn+1)^k & -\frac{\Re_1 - \Re_{k+2}}{k+1}(Pn+1)^{i-1} \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 1 & Pn+k+1 & \cdots & (Pn+k+1)^k & -\frac{\Re_{k+1} - \Re_{k+2}}{1}(Pn+k+1)^{i-1} \end{vmatrix} = 0: \tag{4.40}$$

By comparing (4.40) with (4.39) it follows that the coe±cients in last column of (4.40) must all be equal by assumption. If $i \in [1; k+1]$, by equating the last two coe±cients in (4.40), we obtain $\alpha_{k+1} = (\alpha_k + \alpha_{k+2})/2$. Since $|\alpha_k| = 1$ it follows that $\alpha_k = \alpha_{k+1} = \alpha_{k+2}$. Similarly all other α_k's are the same. If $i = k+2$, a similar argument leads to the same conclusion. Thus, all α_k's are the same for any p. From (4.38) this results in

$$t_{i,j}(\mathbf{b}_n) = \begin{cases} 0, & \text{for } j \neq i \\ \alpha_{1,n}, & \text{for } j = i \end{cases} \tag{4.41}$$

and $s_1(n) = q_1$. With the DS condition, we can find $Y_{DS_2}(\phi)$ such that $t_{i,j} Y_{DS_2}(0) = 0$: It follows then that $t_{i,j} = 0$ for $j \neq i$. Hence T_{aug} is a block diagonal matrix. Furthermore, T_{aug} contains identical diagonal sub-blocks, which we denote by T. Then (4.36) can be written as

$$\mathbf{r} = T\mathbf{b} \tag{4.42}$$

By Lemma 4.3.2, T is an equivalence transformation E_p. This completes the proof. ■

Comment: An interesting result in BMML is that, in order to satisfy the DS condition and restrain the solution within an equivalence set, the data record length cannot be arbitrarily small. We also know from Section 4.2.1 that for better channel approximation it cannot be too large. Thus a compromise must be made on the record length to minimize the detection error. Another interesting observation is that the sample rate P depends on the order of polynomial p which is, in turn, a function of the fade rate f_d. This result is expected but for a different reason { oversampling is utilized to reveal signal structures which limits the solution to an equivalence set in BMML. However, in subspace based detection algorithms (e.g. [20]) oversampling is used to reduce the rank requirement of the channel gain matrix.

Conditions 4.3.1 and 4.3.2 are easily met. The probability of not meeting the DS condition decreases exponentially with N if s(n) is random. For example, the probability of two random sequences belonging to the same equivalence set is very small. The sample rate condition is not di±cult to meet, since some degree of oversampling is common in modem implementations.

4.3.3 Numerical Examples

In this section we study the performance of BMML by simulation. The channel and data are generated in a similar way as that shown in Fig. 4.3 in Section 4.2.3, with two differences:

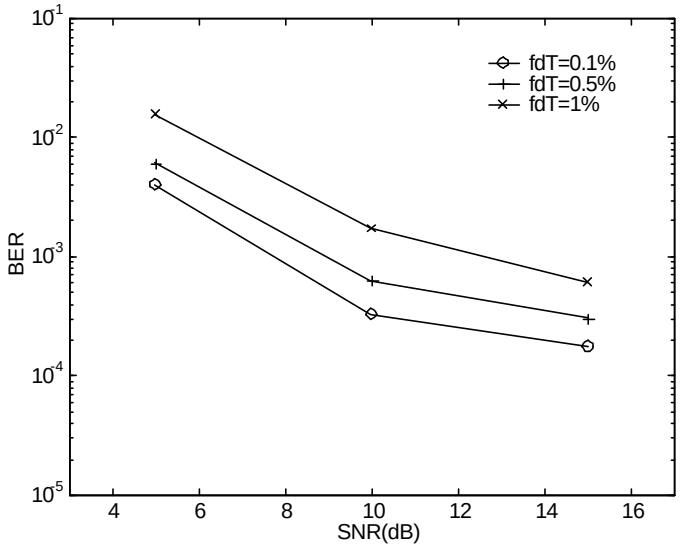

Figure 4.11: BMML for two equal power users with m = 3 antenna and block size N = 20.

1) the signal $s_i(n)$ is there replaced by the waveform $y_i(n)$ and 2) data is over-sampled with the rate P.

First we study BMML (i.e., (4.31), (4.32)) for small blocks, over which the time variation of the channel is approximately linear. (In Section 4.3.1 we showed how to make this assumption more accurate.) Then we examine the sub-block version of BMML for larger block size, using the same channels. Moreover, we study the e®ect of the number of antennas on BMML. Finally, the convergence of BMML is studied by measuring the average number of iterations required for a near-global solution. Errors are measured for each block assuming user order is known from the I.D. sequences.

For simplicity, we study two equipower cochannel CPM users under independent fading channels of identical fade rate. The channel is modeled as a Rayleigh process with Jakes' power spectrum. The signal alphabet size Q = 4; the modulation index h = 1=3 and the sample rate P = 5. The number of antennas is a variable. SNR is de¯ned as the ratio of user #1 power over the noise variance. Also for display clarity, only the averaged BER for the two equal powered signals are shown. In this section, the channel order p = 1 and block size N = 100; unless otherwise stated.

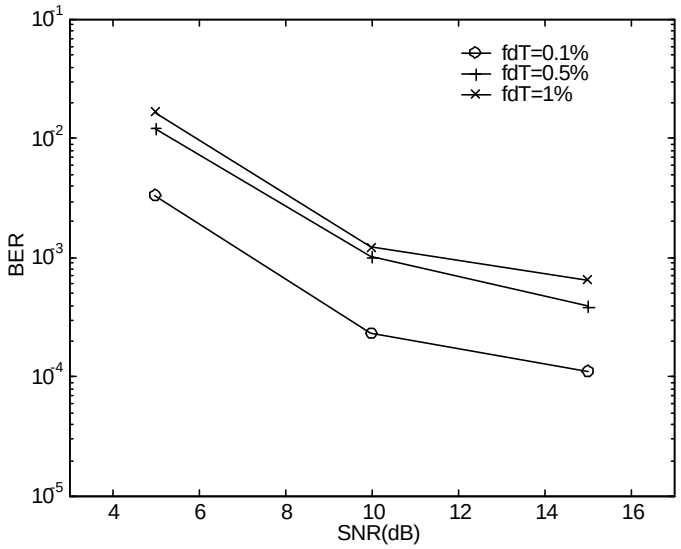

Figure 4.12: Sub-block BMML for two equal power users with antenna m = 3 and block size N = 100 with sub-block size N_{sub} = 20.

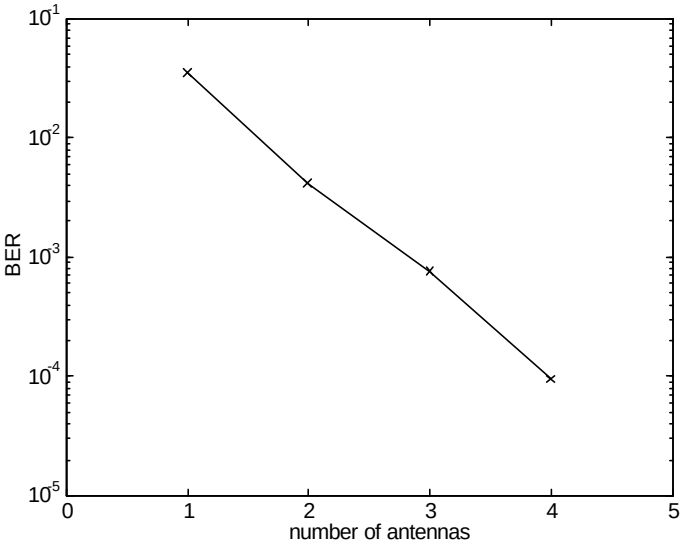

Figure 4.13: BMML of two equal power users for di®erent number of antennas with SNR = 15 dB, f_dT = 0:5% and N = 20.

	f_d = 0:1%	f_d = 0:5%	f_d = 1%
BIR	2.3	2.6	4.0
RIR	5.5	6.4	9.4

Table 4.2: The average number of iterations for near-global solutions

As seen from Fig. 4.11 BMML works quite effectively at moderate SNR's and its performance is better for lower fade rate since the time variation of the channels is more linear at lower fade rates. Fig. 4.12 depicts the results for sub-block BMML with N_{sub} = 1=5N. Despite the lack of I.D. sequences in all but the first sub-block, the performance of sub-block BMML is slightly better than that of BMML (comparing Fig. 4.12 with Fig. 4.11) because BMML is limited by the lack of convergence to near-global solutions. In sub-block BMML, on the other hand, the convergence is improved by the fact that in any of the subsequent sub-block detections its initial gain estimate (taken from the previous sub-block) is close to its final one so that the algorithm is more likely to converge to a near-global solution. Finally, Fig. 4.13 shows the effect of the number of antennas on BMML. It is apparent that the number of signals to be detected is not limited by the number of antennas, although increasing the number of antennas improves its performance. The near exponential relationship (linear on a semilog plot) between the number of antennas and BER clearly shows that the diversity gains of the antenna array, like for single user detection, is ~ $(SNR)^m$.

The effectiveness of the initialization routine is studied by measuring the number of iterations to achieve the near-global convergence of BMML. The result is listed in Table 4.2. As indicated in the table, the average number of iterations for the blind initialization routine (BIR) varies slightly from 2 to 4 for various fate rates as shown. This result clearly demonstrates BIR is highly effective. In comparison, the random initialization routine (RIR) shows a relatively poor performance. In addition to speed improvement, BIR also improves BER performance. For example, BIR gains about 2dB SNR compared with RIR at fade rate f_dT = 0:1% to 0:5%. However, the improvement is less at fade rate f_dT = 1%. It is also found (not shown here) that BMML converges slightly faster at higher SNR.

4.3.4 Conclusion

In this section a new block based blind detection scheme for cochannel CPM signals over time varying channels is presented. It is based on the idea that slow time-varying channel

can be approximated by low order polynomials by a appropriate selection of detector window size, which allows an e±cient VA based detection algorithm to be developed. Signal detectability is studied and conditions for unique detection within an equivalent class are derived. A variant of BMML, a sub-block based detection algorithm, used when the channel approximation is good for sub-block intervals, is also proposed. Moreover, an initialization routine is presented which is shown via simulation more e®ective than the random initialization routine. Simulations demonstrate that BMML can be applied e®ectively in detecting cochannel CPM signals in fading channels.

4.4 Summary

This chapter describes the blind multiuser maximum likelihood (BMML) detectors for linear and non-linear modulations. It shows that even when channel is time-varying and the number of signals is more than that of antenna sensors the signals can be detected satisfactorily. Detection is not unique because of unknown channels, but ML solutions are within a well de¯ned set. It is found in both cases that detection ambiguity can be easily overcome if transmitted user sequences contain some I.D. sequences periodically. Finite alphabet property of the data and some easily satis¯ed data structure conditions are su±cient to obtain an equivalent solution. An e®ective initialization routine is critical to the success of this method since it a®ects the computational complexity, and more importantly, the (BER) performance. The main advantages of BMML detectors are:

- No statistical model is assumed on the channel and therefore BMML is suitable for more general channel conditions.
- For multiple synchronous users and for linear modulations BMML is equivalent to the symbol vector-by-symbol vector detection.

The main disadvantage of BMML is the lack of guaranteed global convergence and its peak complexity can be quite high (it takes longer time to converge for some channel realizations). It is found in simulations that the number of iterations to achieve the near global convergence depends on fade rate very strongly. This fact limits its e®ectiveness for faster fading channels ($f_dT > 1\%$).

Chapter 5

Blind Multiuser Bayes Detector

5.1 Introduction

In this chapter, we study the second novel blind multiuser detector, the blind multiuser Bayes (BMB) Detector. We have seen in the last chapter that although BMML detectors can be quite effective in fading channels, the speed of the (near global) convergence can be slow, especially at higher fade rates and for a small number of antennas. From the last chapter we also know that to reduce channel model errors for faster fading channels the order of polynomials should be increased. Although increasing the number of parameters (channel gains) could reduce the channel model errors, it increases the chance of local solutions and hence detector errors. In addition, it makes the convergence even slower.

The BMB detector does not suffer from these problems. However, it requires prior pdfs of channel parameters. If (statistical) knowledge of the channels is not available, one can estimate it or design for the worst case scenarios, e.g. design for the fastest fade. The worst case design satisfies the so called Minimax criterion (see [91] for detail), and if the actual channel fade rate is lower than the designed value, the performance will not degrade further.

In this chapter, however, we assume that the statistics are known to simplify the analysis. Like BMML in the last chapter, BMB also requires a single I.D. sequence per user per block, in order to distinguish the detected streams. In comparison to BMML, BMB has the advantages of allowing faster fading channels, a fixed computational load (i.e., it is non-iterative and independent of fade rates of interest) and better performance.

This chapter is organized as follow. We first study the uniqueness of BMB detection. Then, we estimate the memory size of BMB and memory reduction techniques, which lead

to the development of the suboptimal but memory reduced forms of BMB, the blind multiuser per-survivor processing (BMPSP) and the blind multiuser decision feedback (MRDF) detectors. The performance of these detectors is demonstrated by numerical simulations. Finally, a performance/complexity comparison for the BMML and BMB detectors is given.

5.2 Solutions of BMB Detector

In this section we examine the uniqueness and correctness of BMB solutions, de¯ned in (3.18). To simplify the discussion, we assume in this chapter that the multipaths have delays much smaller than the symbol period and multiple users are synchronous so that the system can be well described by the memoryless channel model (2.13).

It is clear from (3.16) that if $\hat{\mathbf{b}}$ is a solution of (3.18), its equivalent solutions may also be solutions, depending on the statistics of the channels. To see this, notice that if $\mathbf{b}_i$ is a part of solution, so is $\mathbf{b}_i\,'(q)$, as rotation does not change (3.14) and consequently (3.18). Thus, if $\hat{\mathbf{b}}$ is a solution, so is $E_r\hat{\mathbf{b}}$. In cases where some of the $R_{a_{ji}}$ $(i = 1;:::;d;\ j = 1;:::;m)$ are the same, say, $R_{a_{ji}} = R_{a_{ji'}}$; for $i \neq i'$ and $j \neq j'$, exchanging $\mathbf{b}_i$ with $\mathbf{b}_{i'}$ will not a®ect (3.14) and consequently not (3.18). $E_p\hat{\mathbf{b}}$ is then also a solution. These kinds of non-uniqueness were discussed in the last chapter and can be solved by i) embedding a single I.D so that permutation ambiguity is resolved and ii) di®erential encoding so that rotation ambiguity is resolved. Both requirements are possible with many existing systems, e.g., in TDMA IS-136. To simplify the discussion, in this chapter we assume that all channels associated with any user have identical statistics (same fade rate and power level) for all antennas. In other words, channel correlation matrix $R_{a_{ji}}$ is independent of antenna index j, and is denoted as R_{a_i} for notational convenience.

In the following Theorem we make an interesting observation that relates di®erent solutions of (3.18) to their corresponding autocorrelation matrices.

Theorem 5.2.1 i) With probability one autocorrelation matrices for all solutions of (3.18) are the same for ¯nite m. ii) If $m \to 1$ autocorrelation matrices of all solutions are exactly the same and equal to that of S.

Proof. i) Let $\hat{\mathbf{b}}$ and $\tilde{\mathbf{b}}$ be any two solutions of (3.18) with $\hat{\mathbf{b}} \neq \tilde{\mathbf{b}}$. That is,

$$L(\hat{\mathbf{b}}|X) = L(\tilde{\mathbf{b}}|X): \tag{5.1}$$

Define a modified (geometric mean) likelihood function as

$$L_{mod} = \exp\left\{\frac{1}{m}\sum_{j=1}^{m} \ln[L_j(S|x_j)]\right\} \tag{5.2}$$

$$= \frac{1}{|R(\hat{S})|}\exp\left\{-\mathrm{tr}\left[R(\hat{S})^{-1}\frac{1}{2}\sum_{j=1}^{m}\frac{x_j x_j^H}{m}\right]\right\}$$

$$= \frac{1}{|R(\hat{S})|}\exp\left\{-\mathrm{tr}\left[R(\hat{S})^{-1}\hat{R}(S)\right]\right\}$$

where $\hat{R}(S) \triangleq \frac{1}{2}\sum_{j=1}^{m}\frac{x_j x_j^H}{m}$. From (5.1)

$$\ln\left[|R(\hat{S})|/|R(S)|\right] = \mathrm{tr}\left[R(S)^{-1} - R(\hat{S})^{-1}\hat{R}(S)\right]. \tag{5.3}$$

Notice that both $\ln\left[|R(\hat{S})|/|R(S)|\right]$ and $R(S)^{-1} - R(\hat{S})^{-1}$ are discrete functions of discrete variables $\hat{S}$ and S whereas $\hat{R}(S)$ is a function of continuous random noise and time-varying channel variables which both are continuous random variables, as seen from (2.13) and the definition of $\hat{R}(S)$. Thus, with probability one, (5.3) will not be satisfied unless $R(S)^{-1} - R(\hat{S})^{-1} = 0$ or $R(S) = R(\hat{S})$. That is, with probability one all solutions correspond to the same autocorrelation matrix.

ii) As $m \to \infty$, $\hat{R}(S)$ converges in the mean to $R(S)$ by the strong law of large number. Let

$$L_0 = \frac{1}{|R(S)|}\exp\{-\mathrm{tr}[I]\}. \tag{5.4}$$

We now show that if $R(\hat{S}) \neq R(S)$ then $L_{mod}/L_0 < 1$ and $L_{mod}/L_0 = 1$ iff $R(\hat{S}) = R(S)$: From (5.2) and (5.4),

$$L_{mod}/L_0 = |R(\hat{S})^{-1}R(S)|\exp\left\{-\mathrm{tr}\left[R(\hat{S})^{-1}R(S) - I\right]\right\}$$

Performing an eigenvalue decomposition on matrix $R(\hat{S})^{-1}R(S)$, we have

$$L_{mod}/L_0 = |\Lambda|\exp\{-\mathrm{tr}[\Lambda - I]\} = \prod_{k=1}^{N}\lambda_k \exp\{-(\lambda_k - 1)\} \tag{5.5}$$

where $\Lambda = \mathrm{diag}\{\lambda_k; k = 1:::N\}$ and all eigenvalues λ_k are positive (see Appendix C). Since (5.5) is convex in $\{\lambda_k\}$ it has an unique maximum, which occurs at $\lambda_k = 1; k = 1;:::;N$.

$R(\hat{S})^{i-1}R(S)$ can have all unity eigenvalues, e.g., when $\hat{S} = S$. A matrix with all eigenvalues equal to unity is the identity matrix, so $R(\hat{S})^{i-1}R(S) = I$, or $R(\hat{S}) = R(S)$. In other words, the autocorrelation matrices of all solutions are the same and equal to that of S as $m \to 1$. ■

In general, the solutions $\hat{S}$ and $\tilde{S}$ satisfying (5.1) may not have the same autocorrelation matrices. However, Theorem 5.2.1 tells us that, with probability one, autocorrelation matrices for different solutions are the same. Theorem 5.2.1 also allows us to study the uniqueness of solutions merely from their autocorrelation matrices. What we are interested is the opposite question: Are all the solutions with the same autocorrelation matrix equivalent? Theorem 5.2.2 below will answer this question.

For ease of derivation we partition d users into K groups of different fade rates. Let the number of users in each group be d_k with $\sum_{k=1}^{K} d_k = d$. Each group is further divided into two subgroups: subgroup 1 with users whose autocorrelation matrices are not linearly related by coefficients selected from $\alpha \in Q$ and subgroup 2 with users whose channel autocorrelation matrices are related linearly by $\alpha \in Q$. That is, for subgroup 1,

$$R_{a_{k;i}} \neq \sum_{l \in S_{k;i}} \alpha_l R_{a_{k;l}} \tag{5.6}$$

where $R_{a_{k;i}}$ denotes the ith channel autocorrelation matrix in group k with $i \le d_k^1$ and $d_k^1 \le d_k$ is the size of subgroup 1, $S_{k;i} = \{1, 2, \ldots, i-1, i+1, \ldots, d_k\}$ and $\alpha_l \in Q$. Notice in subgroup 2 there could be more than one linear relationship among channel autocorrelation matrices, all related by $\alpha \in Q$.

Theorem 5.2.2 Under the following sufficient conditions, solutions with the same autocorrelation matrix are equivalent.

i) the estimated data $\hat{S}$ contains, in any column order, the same data structure (DS) as that defined in (4.14),

$$S_{sp} = \begin{cases} \begin{bmatrix} \theta_1 & \theta_2 & \cdots & \theta_Q & \theta_1 & \cdots & \theta_1 & \cdots & \theta_1 & \cdots & \theta_1 \\ \theta_1 & \theta_1 & \cdots & \theta_1 & \theta_2 & \cdots & \theta_Q & \cdots & \vdots & & \vdots \\ \vdots & \vdots & & \vdots & \theta_1 & \cdots & \theta_1 & & \vdots & & \vdots \\ \vdots & \vdots & & \vdots & \vdots & & \vdots & & \theta_1 & \cdots & \theta_1 \end{bmatrix}; & \text{for } d > 2; \\ \begin{bmatrix} \theta_1 & \theta_1 & \cdots & \theta_1 & \theta_1 & \cdots & \theta_1 & \cdots & \theta_2 & \cdots & \theta_Q \\ \theta_1 & \theta_2 & \cdots & \theta_Q & & & & & & & \\ \theta_1 & \theta_1 & \cdots & \theta_1 & & & & & & & \end{bmatrix}; & \text{for } d = 2; \end{cases}$$

ii) no channel autocorrelation function in any subgroup 1 is linearly related to members of any other groups for some pair of adjacent time lags n^0 and $n^0 + 1$, i.e.,

$$r_{a_{k;i}}(n) \neq \sum_{l \in S_{k;i}} \bar{}_l r_{a_{k;j}}(n) + \sum_{\substack{k' \in \{1;:::;k-1;k+1;:::;K\} \\ l_{k'} \in \{1;:::;d_{k'}\}}} \beta_{k';l_{k'}} r_{a_{k';l_{k'}}}(n) \quad n = n^0; n^0 + 1 \tag{5.7}$$

where $i \in \{1;:::;d_k^1\}$, $\bar{}_l \in \{0;Q\}$ and $\beta_{k';l_{k'}} \in Q$:

iii) no channel autocorrelation function in any subgroup 2 is linearly related to members of any other subgroup 2 for at least time lags n^0 and $n^0 + 1$, i.e.,

$$r_{a_{k;i}}(n) \neq \sum_{\substack{k' \in \{1;:::;k-1;k+1;:::;K\} \\ l_{k'} \in \{d_{k'}^1+1;:::;d_{k'}\}}} \beta_{k';l_{k'}} r_{a_{k';l_{k'}}}(n) \quad n = n^0; n^0 + 1 \tag{5.8}$$

where $i \in \{d_k^1 + 1;:::;d_k\}$ and $\beta_{k';l_{k'}} \in Q$:

Proof. Assume, in addition to $\mathbf{b}$, there exists another solution $\mathbf{c}$ such that

$$R(\mathbf{b}) = R(\mathbf{c}): \tag{5.9}$$

Assume also that the leading symbols of all sequences to be one ($'(0)$) without losing generality. In fact, we have assumed $\mathbf{c}(1) = \mathbf{b}(1) = 1$. Our aim is to $\bar{}$nd all $\mathbf{c}$ satisfying (5.9) and we will achieve this by $\bar{}$rst determining all subgroup 1 users then separating all subgroup 2s and $\bar{}$nally determining users in each subgroup 2.

1) Detecting all group 1 users. The nth diagonal entries of (5.9) can be written as

$$\mathbf{b}(i;i+n) = \mathbf{c}(i;i+n) = \sum_{l=1}^{N} r_{a_{1;l}}(i;i+n)\mathbf{c}_l(i)\mathbf{c}_l^*(i+n); \; i = 1:::N - n - 1: \tag{5.10}$$

Let $n = n^0$, $n^0 + 1$: With (5.10), all subgroup 1 users at time n^0 and $n^0 + 1$; $\mathbf{c}_l(n^0 + 1)$ and $\mathbf{c}_l(n^0 + 2)$; for $l \in \{d_1^1;:::;d_K^1\}$; are readily solved since $\mathbf{c}_l(1)$ is known. Condition (5.7) and (5.6) guarantee the solution is unique. From (5.10) with $i = n^0 + 1$ and $n = -n^0 + 1$ we can solve $\mathbf{c}_l(2)$ uniquely, assuming channels are wide-sense stationary. In this fashion, all $\mathbf{c}_l(i)$; $i = 2:::N$ can be determined uniquely. For example, let $n^0 = 4$. Start from $\mathbf{c}_l(1)$ and use (5.10) we could determine $\mathbf{c}_l(5)$ and $\mathbf{c}_l(6)$. Then from $\mathbf{c}_l(6)$ and using (5.10) backward, we could get $\mathbf{c}_l(2)$ and so on.

2) Separating all subgroup 2s. What remained in (5.9) after removing the detected users are all the subgroup 2 users. By (5.8) all subgroup 2 users can be separated just as was

done above for solving subgroup 1 users. Each separated subgroup 2 forms a matrix equality (5.9) for users in that subgroup.

3) Detecting all users in each subgroup 2. Without losing generality, we assume all users are in one subgroup 2. To save notation, we assume that all d users are in one subgroup 2. All the users now have the same fade rate and channel autocorrelation matrices of all users are related by a scalar. First, we transform (5.9) into a simpler and equivalent form. Since the noise variance is assumed known, we remove it from (5.9) without a®ecting the equation. This operation is equivalent to set the noise to zero. Multiplying each entry of the noise removed equations in (5.9) by a scalar results in

$$R_0(\mathbf{c}) = R_0(\mathbf{b}) \tag{5.11}$$

where $R_0(\mathbf{c})$ is the same as $R(\mathbf{b})$ except the associated channel autocorrelation matrices (de¯ned in (3.15)) are modi¯ed. We choose each scalar such that the modi¯ed channel autocorrelation matrices have identical entries since all channel matrices are the same within a scalar. That is, the entry of the modi¯ed channel matrix $r_{a_{2;\,j}}(k) = r_{a_{2;\,j}}$ for $k = 1;:::;N$ and $j = 1;:::;d$. This operation is equivalent to set the fade rate to zero.

Expand the ith column of $R_0(\mathbf{b})$ as

$$\begin{aligned}
&\begin{pmatrix} R_0(\mathbf{b})_{1;i} \\ R_0(\mathbf{b})_{2;i} \\ \vdots \\ R_0(\mathbf{b})_{N;i} \end{pmatrix} = \begin{pmatrix} \sum_{j=1}^{d} r_{a_{2;\,j}} \mathbf{b}_j(1)\mathbf{b}_j^{*}(i) \\ \sum_{j=1}^{d} r_{a_{2;\,j}} \mathbf{b}_j(2)\mathbf{b}_j^{*}(i) \\ \vdots \\ \sum_{j=1}^{d} r_{a_{2;\,j}} \mathbf{b}_j(N)\mathbf{b}_j^{*}(i) \end{pmatrix} \\
&= \begin{pmatrix} \mathbf{b}_1(1) & \mathbf{b}_2(1) & \cdots & \mathbf{b}_d(1) \\ \mathbf{b}_1(2) & \mathbf{b}_2(2) & \cdots & \mathbf{b}_d(2) \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{b}_1(N) & \mathbf{b}_2(N) & \cdots & \mathbf{b}_d(N) \end{pmatrix} \begin{pmatrix} \mathbf{b}_1^{*}(i) r_{a_{2;\,1}} \\ \mathbf{b}_2^{*}(i) r_{a_{2;\,2}} \\ \vdots \\ \mathbf{b}_d^{*}(i) r_{a_{2;\,d}} \end{pmatrix} \\
&= \mathbf{b}^{T} \mathbf{b}_{0;i}; \quad i = 1;2;:::;N:
\end{aligned} \tag{5.12}$$

From (5.11) we have $\mathbf{b}^{T}\mathbf{b}_{0;i} = \mathbf{c}^{T}\mathbf{e}_{0;i}$. Since $\mathbf{c}$ satis¯es the DS condition it has a full rank

(see Lemma 4.2.4), we have

$$\begin{aligned} & \mathbf{e}_{0;i} = (\mathbf{S}^{\star}\mathbf{e}^{T})^{-1}\mathbf{S}^{\star}\mathbf{b}^{T}\mathbf{b}_{0;i} = T_2\mathbf{b}_{0;i}; i = 1:::N \\ \Leftrightarrow\ & \mathrm{diag}(r_{a_{2;1}}:::r_{a_{2;d}})\mathbf{e} = T_2\mathrm{diag}(r_{a_{2;1}}:::r_{a_{2;d}})\mathbf{b} \\ \Leftrightarrow\ & \mathbf{e} = \mathrm{diag}^{-1}(r_{a_{2;1}}:::r_{a_{2;d}})T_2\mathrm{diag}(r_{a_{2;1}}:::r_{a_{2;d}})\mathbf{b} \\ =\ & T\mathbf{b}: \end{aligned} \tag{5.13}$$

If $\mathbf{b}$ contains S_{sp} it can be shown that $T = E$ (Lemma 4.2.5) and thus all other solutions $\mathbf{e}$ are within the same equivalence class as $\mathbf{b}$.

Since users in subgroup 1 are uniquely determined and users in each subgroup 2 are determined within the same equivalence class, any $\mathbf{e}$ is within the same equivalence class. ■

Thus, from Theorem 5.2.1 and 5.2.2, all solutions are equivalent with probability one. Condition i) is needed in its entirety only if all users have the same autocorrelation matrix (same fade rate and power level.) A subset of it is needed if only some of users have identical fade rate and power level. This means that in practice that we may never require data matrix to satisfy the full DS condition. In other words, we have considered the worst case scenario. The probability that S satisfies condition i) increases exponentially with N: (See Section 4.2.2). Condition (ii) { (iii) are easily met since (5.7) and (5.8) are satisfied almost everywhere, as can be argued by virtue of Lebesgue measure. The implication of Theorem 5.2.1 and 5.2.2 are that with probability one, the detection ambiguity can be avoided.

5.3 Blind Multiuser Viterbi Algorithm (BMVA)

It is well-known that a practical implementation of a single user MLSE is to use of Viterbi algorithm (VA)[27, 25]. In this section we will incorporate VA into blind multiuser detection. An important parameter in VA is its memory size (state set). In general, the memory at time n extends back to time 1, and if that memory is truncated, we would expect a performance loss. A consequence of memory truncation is that DS condition in Theorem 5.2.2 may not be met, resulting in detection error. As reported in [36], an interesting type of detection error in multiuser case { signal swapping { was discovered when the memory-truncated BMVA was used. In particular, part of detected user sequences are mistakenly phase rotated and swapped. The \swapping error" of a memory-truncated BMVA is related to the convergence

of BMVA which will be discussed in Section 5.3.3. Assume that BMVA starts correctly, then diverges (from the correct state) due to detection error and then remerges. However, unlike a single user VA, the remerged state may not be the correct state since multiuser VA converges in terms of autocorrelation matrix only (Theorem 5.2.2) which corresponds to several possible equivalent solutions. As a result, the swapping error occurs. For a 2-user BMVA, the swapping error can be detected and restored easily through a post-detection-restoration technique [36]. However, for more users, the swap detection and restoration become more difficult.

In a single user blind VA detection, the detector memory is determined solely by the fading channel model memory [25, 26]. For example, if the time-varying channels plus noise is an AR processes of order L (symbol periods) then VA memory will also be L without compromising the performance. In the next subsection, we will check if this conclusion can be generalized to BMVA.

5.3.1 Estimate of BMVA Memory Size

To study the more complicated multiuser case it is constructive to review the one signal case first. We first study the non-modulated then the modulated channel case.

Case 1: One non-modulated channel. Suppose the channel $\{y(n)\}$ is modeled as a stationary AR process of order L

$$a_0 y(n) + a_1 y(n-1) + \cdots + a_L y(n-L) = x(n) \tag{5.14}$$

where the regressor $x(n)$ is a symmetric, complex, white noise of zero mean and unit variance and $\{a_i\}$ is the set of process parameters with $a_0, a_L \neq 0$. Multiplying (5.14) by $x^*(n+k)$ and $y^*(n+k)$, respectively, and taking the ensemble averages, we obtain

$$a_0 b_k + a_1 b_{k-1} + \cdots + a_L b_{k-L} = \delta_{k,0} \tag{5.15}$$

and

$$a_0 r_k + a_1 r_{k+1} + \cdots + a_L r_{k+L} = b_k^* \tag{5.16}$$

where $r_k = \frac{1}{2} E y(n) y^*(n+k)$ and $b_k = \frac{1}{2} E y(n) x^*(n-k) \neq 0$ for $k \geq 0$. These are the familiar Yule-Walker equations[10] but we are interested in their matrix form. Denote

$a = (a_0; :::; a_L)^T$ and

$$L = \begin{pmatrix} a_0 & & & 0 \\ \vdots & a_0 & & \\ a_L & \vdots & & \\ & a_L & \ddots & \\ & & \vdots & a_0 \\ 0 & & & \end{pmatrix}$$

With (5.15) and (5.16) it is readily shown that

$$R_{a_1} L L^H = I \tag{5.17}$$

where R_{a_1} is the channel autocorrelation matrix with Toeplitz structure for one user. Thus, the inverse of R_{a_1} is a band matrix of bandwidth L. Conversely, from the Toeplitz structure of R_{a_1} and uniqueness of the Cholesky decomposition [82] one can recover (5.15) and (5.16) from (5.17). We obtain an important observation: $(R_{a_1})^{i\,1}$ is a band matrix of bandwidth L i® the process fy(n)g is a stationary AR process of order L.

Case 2: One digitally modulated channel. If noise is zero then $R(S) = D(s_1)R_{a_1}D^H(s_1)$ (see (3.14)). We know that $R^{i\,1}(S) = D^{i\,H}(s_1)(R_{a_1})^{i\,1}D^{i\,1}(s_1)$ is also banded. With this result, we see that the log-likelihood function $L_j(Sjx_j)$ in (3.16) for one user can be expressed as a sum of terms each of which depends only on L past observations. The computationally e±cient VA with the memory size of L can be applied directly [25, 26]. In other words, if the channel can be modeled as a stationary AR process of order L then the memory size of MLSE can be reduced from N to L.

If the noise term in (3.14) is not zero but small ($¾_w^2 < r_0$), we have from (3.14) and [82]

$$\begin{aligned} R^{i\,1} &= \left(DR_{a_1}D^H + ¾_w^2 I\right)^{i\,1} \\ &= D(R_{a_1})^{i\,1}\left(I + ¾_w^2 (R_{a_1})^{i\,1}\right)^{i\,1} D^H \\ &= D\sum_{k=0}^{\infty} (-1)^k ¾_w^{2k} (R_{a_1})^{i(k+1)} D^H \end{aligned} \tag{5.18}$$

where we have used $D^{i\,1} = D^H$ for PSK signals. From this equation the bandwidth of $(R_{a_1})^{i(k+1)}$ is about (k + 1)L. For very small noise, (5.18) converges very quickly and can be approximated by the ¯rst K terms. In this case the bandwidth of $R^{i\,1}(S)$ is about KL.

Case 3: Multiple non-modulated channels. If we assume each individual channel $fy_i(n)g$, $i = 1; :::; d$, is modeled as a stationary AR process of order L_i, the composite channel fy(n)g

is then the sum of AR processes. We want to know if $fy(n)g$, under some conditions, is still an AR process so that the inverse of $R_a = \sum_{i=1}^{d} R_{a_i}(R^c)^{i-1}$ is banded. Assume the additive noise is zero-first. The z-transformation of $fy_i(n)g$ satisfies $Y_i(z) = \frac{X_i(z)}{F_i(z)}$ where $F_i(z)$ is a polynomial of z^{-i-1} of order L_i: Thus

$$Y(z) = \sum_{i=1}^{1} Y_i(z) = \sum_{i=1}^{1} \frac{X_i(z)}{F_i(z)} \tag{5.19}$$

We now show that $fy(n)g$ is a AR process iff all the $F_i(z)$'s have all their roots in common. The if part is straightforward. For the only if part, we need to show a white regressor $X(z)$ in $Y(z) = \frac{X(z)}{F(z)}$ can be found. Expressing (5.19) as

$$Y(z) = \frac{X_1(z) + \sum_{i=2}^{d} X_i(z)\frac{F_1(z)}{F_i(z)}}{F_1(z)}; \tag{5.20}$$

we notice the numerator cannot be a white regressor unless the polynomial $G_i(z) = \frac{F_1(z)}{F_i(z)}$ is of order zero in z or z^{-i-1}. To see this, denote the numerator as $X'(z)$. Since

$$EX'(z)X'^{*}(z) = 1 + \sum_{i=2}^{1} |G_i(z)|^2;$$

if and only if $G_i(z)$ is a zero-th order polynomial can $X'(z)$ be a white regressor, in other words, if all polynomials $fF_i(z)g$ have the same common roots. That is, all the individual AR processes have to be the same within a scalar. Under this condition, $(R_a)^{-i-1}$ is a band matrix of bandwidth L $(= L_i;\ i = 1;:::;d)$. If the noise is not zero, the discussion in case 2 can be applied and similar results can also be attained.

Case 4: Multiple digitally modulated channels. From above discussions we know that the necessary and sufficient condition for $(R_a)^{-i-1}$ being banded is that all the individual fading channels can be modeled by the same AR process (within a scalar). Thus in the following we will assume all channels are modeled by the same AR process (within a scalar) with a common order of L.

Since a sum of multiple modulated channels is not a AR process in general we expect $R^{-i-1}(S)$ is not banded. We now show explicitly that this is indeed the case. A necessary condition for $R^{-i-1}(S)$ being a band matrix of bandwidth $M < N$ is that there exist a vector p of order M such that

$$R_{M+1}(S)p = u_0 \tag{5.21}$$

for all S where $R_{M+1}(S)$ is a matrix consisting of the first $M+1$ columns of $R(S)$ and $u_0 = (1;0;0;:::;0)^T$. Let $fr^i(l)g$ denote the channel autocorrelation function of time lag l for user i. Since all fading processes are the same within a scalar, $r^i(l) = ®_i r(l)$. Consider a particular S as

$$S = \begin{pmatrix} ' & ' & ¢¢¢ & ' & ' & ¢¢¢ \\ \vdots & \vdots & \ddots & \vdots & \vdots & ¢¢¢ \\ ' & ' & ¢¢¢ & ' & ' & ¢¢¢ \\ ' & 1 & ¢¢¢ & 1 & ' & ¢¢¢ \end{pmatrix}_{d £ N} \tag{5.22}$$

where $1 6= '$, and is shown from the second to the $M+1$-th entries of last row. Substituting (5.22) into (5.21) gives

$$\begin{pmatrix} °r_0 & ¯r_1 & ¢¢¢ & ¯r_M \\ ¯r_1 & °r_0 & ¢¢¢ & °r_{M-1} \\ \vdots & \vdots & \ddots & \vdots \\ ¯r_M & °r_{M-1} & ¢¢¢ & °r_0 \\ °r_{M+1} & ¯r_M & ¢¢¢ & ¯r_1 \\ \vdots & \vdots & \vdots & \vdots \end{pmatrix} \begin{pmatrix} 1 \\ p_0 \\ p_1 \\ \vdots \\ p_M \end{pmatrix} = \begin{pmatrix} 1 \\ 0 \\ \vdots \\ \vdots \\ \vdots \end{pmatrix} \tag{5.23}$$

where $° = \sum_{i=1}^{d} ®_i$ and $¯ = \sum_{i=1}^{d-1} ®_i + ®_d 1¤'$: The first $M+1$ of above equations determines p uniquely as $p = c(a_0=¯; a_1=°; :::; a_L=°; 0; :::)^T$ where c is a constant. However, this p cannot satisfy the $M+2$-th equation, since it would require $° = ¯$. Thus a vector p of order M cannot be found for every S and consequently multiuser $R^{-1}(S)$ is a not band matrix in general.

Having shown that the multiuser $R^{-1}(S)$ generally is not banded we want to know if forcing $R^{-1}(S)$ to be a band matrix of bandwidth M by setting higher o®-diagonal entries to zero, or equivalently, truncating the channel length to M, will cause little performance degradation. To ease the analysis, we introduce the following three conditions.

Condition 5.3.1 The fade rate is not too high or M is not too large so that the (non-modulated) channel autocorrelation matrix of size $M £ M$ is near rank 1.

Condition 5.3.2 Data segment $S(n : M - 1 + n) = (s(n); s(n+1); :::; s(M - 1 + n))$; $n = 1; :::; N - M + 1$; has a full rank d.

Condition 5.3.3 SNR is high.

To proceed, we first use the conditional probability equality $p(A;B) = p(A|B)p(B)$ to expand (3.16) as

$$\begin{aligned} p(x_j(N)|S(N)) &= p(x_j(N)|x_j(N-1);S(N))p(x_j(N-1)|S(N-1)) \\ &= \prod_{n=1}^{N} p(x_j(n)|x_j(n-1);S(n)) \end{aligned} \tag{5.24}$$

with $p(x_j(1)|x_j(0);S(1)) = p(x_j(1)|S(1))$. Then partition the conditional data correlation matrix $R_n = R(S(n))$, (defined by (3.14)), according to

$$R_n = \begin{pmatrix} R_{n-1} & v \\ v^H & r_0 + \sigma_w^2 \end{pmatrix}. \tag{5.25}$$

It is straightforward to show that [95]

$$R_n^{-1} = \begin{pmatrix} R_{n-1}^{-1} + R_{n-1}^{-1} v v^H R_{n-1}^{-1}/\delta_n & -R_{n-1}^{-1} v/\delta_n \\ -v^H R_{n-1}^{-1}/\delta_n & 1/\delta_n \end{pmatrix} \tag{5.26}$$

and

$$|R_n| = |R_{n-1}|\delta_n \tag{5.27}$$

with

$$\delta_n = r_0 + \sigma_w^2 - v^H R_{n-1}^{-1} v. \tag{5.28}$$

Thus, each term $p(x_j(n)|x_j(n-1);S(n))$ in (5.24) can be expressed as

$$p(x_j(n)|x_j(n-1);S(n)) = \frac{1}{2\pi\delta_n} \exp\left[-\frac{|x_j(n) - \hat{x}_j(n)|^2}{2\delta_n}\right] \tag{5.29}$$

where $\hat{x}_j(n) = v^H R_{n-1}^{-1} x_j(n-1)$ is commonly known as the prediction of $x_j(n)$; $v^H R_{n-1}^{-1}$ the prediction filter of length $n-1$ and δ_n the prediction error variance [10]. It can be seen from (5.24) that if each $\hat{x}_j(n)$ in (5.29) depends on the last M data then maximizing $p(x_j(n)|S(n))$ can be approximated by maximizing each

$$p(x_j(n)|x_j(n-1);S(n)) = p(x_j(n)|x_j(n-M:n-1);S(n-M:n)),\ n = 1,\ldots,N$$

by dynamic programing techniques, such as VA. Under Condition 5.3.3, $\pm_k$ is mainly due to channel estimation error, so to maximize $p(x_j(n)|x_j(n-1); S(n))$ is equivalent to minimize $\pm_n = r_0 - v^H R_{n-1}^{-1} v$.

We now show that under Condition 5.3.1 – 5.3.3, $\pm_n = 0$. It is easy to show that under Condition 5.3.1 and 5.3.2, $\text{rank}(R_{n-1}(S(n-M:n-1))) = d$. Construct a $d \times d$ matrix from $R_{n-1}(S(n-M:n-1))$ by selecting only its linearly independent d columns and rows, starting from the lower right, and call it $\hat{R}_{n-1}$. Clearly $\hat{R}_{n-1}$ is full rank. Similarly, we construct $\hat{R}_n$ from (5.25), as

$$\hat{R}_n = \begin{pmatrix} \hat{R}_{n-1} & \hat{v} \\ \hat{v}^H & r_0 + \sigma_w^2 \end{pmatrix} \tag{5.30}$$

where $\hat{v}$ is constructed from the last column of R_n by keeping only the same d entries as the rows of $\hat{R}_{n-1}$. By using (5.27) on (5.30), we have $\hat{\pm}_n = r_0 - \hat{v}^H \hat{R}_{n-1}^{-1} \hat{v} = 0$ since $\text{rank}(\hat{R}_n) = d$ if Condition 5.3.1 – 5.3.3 are satisfied. Thus, if we replace the prediction filter $v^H R_{n-1}^{-1}$ by $\hat{v}^H \hat{R}_{n-1}^{-1}$, of length d, we can achieve $\hat{\pm}_n = 0$. Since $0 \le \pm_n \le \hat{\pm}_n$, that is, increasing the order of Wiener predictor will not increase the prediction variance (see, e.g., [10]), we finally have $\pm_n = 0$. Thus, under Condition 5.3.1 – 5.3.3, $b_j(n)$ depends only on the M preceding observations, $\{x_j(n-1), \ldots, x_j(n-M)\}$ with little performance degradation. We summarize these results into the following Theorem:

Theorem 5.3.1 Under Condition 5.3.1 – 5.3.3 , the predictor can be truncated to length M, or $R^{-1}(S)$ can be truncated as a band matrix of bandwidth M, with little performance degradation.

In other words, for any S, the associated predictor needs only to be long enough to span d linearly independent signal vectors $s(n)$ within S.

Conditions 5.3.1 and 5.3.3 are easily satisfied. For a typical Rayleigh fading channel of fade rate of 1%; $M = 20$ and SNR of 10dB, the first 5 normalized singular values of the (non-modulated) channel autocorrelation matrix is (1.0, 0.078, 0.0056, 0.0051, 0.0051). Thus, the channel matrix can be approximated as a rank one matrix. Note that as long as the noise is nonzero the channel autocorrelation matrix is always invertible.

Condition 5.3.2 is equivalent to the requirement that any $S(M)$ will have a full rank. As data $S(M)$ is a random variable, we want to know how large should M be such that with a high probability $S(M)$ will be of full rank or equivalently, with a very low probability

$p_{rank\ def}$ it will be rank deficient. Since an exact expression of the probability of S(M) being rank deficient is very difficult to obtain in general, its upper bound will be found instead. Notice that linearly independent signal vectors (column vectors in S) are all distinct, but the converse is not necessarily true. One way to find a bound is to express it as the sum of two terms: one with insufficient number of distinct signal vectors and the other with sufficient number of distinct but insufficient number of linearly independent signal vectors. Thus its upper bound (UB) for $p_{rank\ def}$ is given as

$$UB = \sum_{k=1}^{d-1} \binom{Q^{d-1}}{k} \left(\frac{k}{Q^{d-1}}\right)^M + \sum_{k=d}^{Q^{d-1}} C_k \left(\frac{k}{Q^{d-1}}\right)^M \tag{5.31}$$

where C_k is the number of different S(M)s' consisting of only k distinct but linearly dependent signals. The proof of (5.31) is given in Appendix F. As (5.31) shows, M depends not only on the channels like in one signal case but also on the signals – a distinct characteristic for BMVA. For small d and Q (5.31) can be calculated easily. Conversely, if UB is given, M can be estimated from (5.31). For example, for d = 3; Q = 2 and $UB = 10^{-5}$ we have $M \geq 20$.

In summary, $R^{-1}(S)$ is not exactly banded for multiple signals in general. However, if the fade rate is not too fast, data matrix segment (S(k : M − 1 + k)) has a full rank and SNR is reasonably high, with a very small probability of error, $R^{-1}(S)$ can be approximated as a banded matrix of bandwidth M; which is determined by the number of users, data constellation size and the channel fade rate. If these conditions are not satisfied, we can still use (5.31) to get a rough estimate of the channel length, which can be then verified through numerical simulations. Examples of these studies will be given in Section 5.4.

5.3.2 Memory reduced BMVA–Blind Multiuser PSP (BMPSP) Detector and its Memory Reduced Variant

From the last subsection we know that even if all the fading channels can be modeled exactly as AR the required effective channel memory M is still too large in some applications if standard VA is to be used since the size of the state set is $Q^{d(M-1)}$ which is about 10^{17} for 3 BPSK signals and M = 20. One state reduction technique is the per-survivor processing (PSP) [88]. The basic idea is to treat part of the M-symbol vector sequence needed for the branch metric calculation as the sequence associated to the survivor path terminating in the initial state. We can then change a standard VA into a PSP by dividing the memory

size M of the original VA into two parts: the memory reduced VA part of size M_{VA}; and the survivor part of size M_{surv} such that

$$M = M_{VA} + M_{surv}: \tag{5.32}$$

The memory reduced VA searches all the states (now de¯ned by the $M_{VA} ¡ 1$ most recent symbol vectors) while one survivor associated with each state is maintained for the branch metric calculation. The suboptimality of PSP is determined by the ratio of $M_{VA}=M_{surv}$: if it is increased PSP is closer to optimal but the computational load grows exponentially. For $M_{VA} = 1$ we have the blind multiuser decision-feedback detector (BMDF). For convenience we will use BMPSP for $M_{VA} > 1$ and refer it as BMDF for $M_{VA} = 1$. We will study the performance of BMPSP and BMDF in simulation.

We now derive the BMPSP algorithm. Assume in the sequel that $R(S)^{i1}$ is banded of bandwidth M. Then from (3.16) $p(Sjx_j(n))$ can be rewritten as

$$p(Sjx_j(n)) = \frac{1}{(2¼)^n jR_{n¡1}j±_n} \exp\left[¡\frac{jx_j(n) ¡ b_j(n)j^2}{2±_n} ¡ \frac{1}{2}x_j^H(n ¡ 1)R_{n¡1}^{i1}x_j(n ¡ 1)\right] \tag{5.33}$$

where $b_j(n)$ depends only on the last M input vectors,

$$b_j(n) = f^H(n)x_j(n ¡ M : n ¡ 1) \tag{5.34}$$

where $f^H(n)$ denotes the last M entries of $v^H R_{n¡1}^{i1}$. For the same reason $±_n$ is also a function of the last M input vectors. Taking the logarithm of (5.33) and summing it over j, we have

$$\begin{aligned} PM_n(S(n)jX(n)) &= PM_{n¡1}(S(n ¡ 1)jX(n ¡ 1)) + \sum_{j=1}^{K}\left[\frac{jx_j(n) ¡ b_j(n)j^2}{2±_n} + \ln(±_n)\right] \\ &= PM_{n¡1}(S(n ¡ 1)jX(n ¡ 1)) + BM_n(S(n ¡ M : n)): \end{aligned} \tag{5.35}$$

where $PM(¢)$ and $BM(¢)$ are referred to as the path and branch metric, respectively. The path metric (5.35) is the desired form when using BMVA with the state set de¯ned by the signal vectors $S(n ¡ M : n ¡ 1)$: Note the same expression can also be used for BMPSP and in this case the state is de¯ned as the M_{VA} most recent signal vectors and the remaining $M ¡ M_{VA}$ vectors are treated as a survivor for each state. The original problem (3.18) then becomes

$$\hat{S} = \arg \min_{S2fFg^N} fPM_N(S(N)jX(N))g \tag{5.36}$$

which is to be solved by BMPSP with branch metric de¯ned in (5.35) and number of states $Q^{d(M_{VA}-1)}$. For example, for 3 BPSK signals and $M_{VA} = 1;2$ the number of states equals 1 and 8, respectively. However, the predictor $f(n)$ needs to be evaluated for every state transition since it is a function of $S(n-M:n)$. It can be shown that $f(n)$ and $\pm_n$ can be determined from the following equations

$$f(n) = R^{-1}_{M-1}(n)v_{M-1}(n) \tag{5.37}$$

$$\pm_n = r_0 - v^H_{M-1}(n)R^{-1}_{M-1}(n)v_{M-1}(n) \tag{5.38}$$

where $R_{M-1}(n)$ denotes the last $(M-1)\times(M-1)$ sub-matrix of $R(n)$ along the diagonal and $v_{M-1}(n)$ the last $M-1$ entries of $v(n)$. A careful examination of (5.37) and (5.38) indicates that $R^{-1}_{M-1}(n)$ does not need to be recalculated for every state transition, as shown in Appendix D.

The BMPSP algorithm is summarized as follows:

1) Determine M based on the type and the number of signals so that the probability of signals being rank de¯cient $p_{rank\,def}$ is below a preset value.

2) Choose M_{VA} based on the performance/complexity consideration. Usually $M_{VA} \cdot 2$ for reasonable complexity.

3) Perform BMPSP detection with branch metric given by (5.35) and $f(n)$ and $\pm_n$ are updated for every state-transition using (5.37) and (5.38).

The complexity of BMPSP is dominated by the computation of $f(n)$, which is estimated as $Q^{d(M_{VA}-1)}(1=3(M-1)^3 + 2(M-1)^2Q^d)$ complex multiplications and additions (CMA) per step. More on the complexity of BMPSP is discussed in Appendix D. Clearly, the computational load is strongly M dependent. Thus, besides M_{VA}, M should also be chosen carefully so that the overall computational complexity is manageable. More e±cient ways of computing $f(n)$, e.g., the class of subspace tracking methods [97] { [99], are desirable to further reduce the computational load. However, we will not discuss them in this thesis. Instead, in the following, we present a less optimal but much less expensive way to compute $f(n)$.

The analysis in the last subsection suggests that as long as v is a linear combination of columns of R_{n-1} the prediction error variance can be very small. In other words, we need not compute $f(n)$ from a ¯xed $(M-1)\times(M-1)$ matrix R_{n-1}. If one can ¯nd a submatrix that has rank close to d then $f(n)$ can be computed from this dimension reduced matrix. By Condition 5.3.1, we need only ¯nd d linearly independent signal vectors from

$S(n - M : n - 1)$, starting from the end. A simple test routine can be developed to test and remove the linearly dependent components from $S(n - M : n - 1)$ until d vectors have been determined. Then $f(n)$ can be computed from a correlation matrix corresponding to these d signal vectors only.

If we carry this idea further, an even simpler approach can be derived. Notice that Condition 5.3.2 is only a sufficient condition to achieve $\pm_n = 0$. A necessary condition is that v is contained in the span of R_{n-1}. (Otherwise (5.28) will not be close to zero.) A simplest case would be that v is nearly equal to some columns of R_{n-1}. As before, we construct a matrix $\hat{R}_{n-1}$ from R_{n-1} by keeping only those columns that are close to v and then keeping only those rows with the same indices as that of columns. Following the similar analysis as above, one can show that $\hat{\pm}_n = r_0 - \hat{c}^H \hat{R}_{n-1}^{-1} \hat{c} = 0$. (The conditionality of $\hat{R}_{n-1}$ is determined by the underlying fading process and noise. If the fading process is AR of order L and noise is near zero, then to avoid ill-conditioning, the size of $\hat{R}_{n-1}$ should be reduced to L.) The algorithm is as follows. Suppose $S(n - 1)$ has been detected already by a hard decision and saved as $\hat{S}(n - 1)$. It first searches from $\hat{S}(n - 1)$ for each of Q^d distinct symbol vectors M_r times and record its time indices. Thus the predictor length is M_r. (The memory requirement for searching M_r times for each distinct signal vector is not a serious problem since we can fix the search length, say N_0 symbols past, and find each signal M_r^0 ($\cdot M_r$) times.) Then it computes the conditional correlation matrix for each of the Q^d distinct symbol vectors in accordance with their time indices. Following similar steps as those for BMDF, $f(n)$ and $\pm_n$ and the branch metric BM_n can then be calculated and the rest of the algorithm proceeds in a similar way as BMDF. We refer this memory reduction version of BMDF as MRDF since $M_{VA} = 1$.

Note that in the above (conditionality) discussion we have used a condition that channels have identical fade rate for simplicity, however, BMDF is not limited to this channel condition. The effectiveness of MRDF depends on the distribution of symbol vector. To increase the robustness of the algorithm, similar symbol vectors can also be included in calculating $\hat{R}_{n-1}$. Similar symbol vectors are those which are related to each other by a phase factor belonging to the signal constellation. If similar symbol vectors are distributed sparsely, the performance of MRDF will degrade, especially at higher fade rates. However, the gain in complexity reduction is significant, as will be discussed in Section 5.5.

5.3.3 Algorithm Initialization

Ideally BMPSP should start from a known initial state which can be inherited from the preceding block as in continuous mode transmission. The initial state for the burst mode transmission needs to be estimated blindly. In this work we discuss two possible ways to start BMPSP blindly.

The first state initialization method, referred to as the blind order update (BOU) initialization, runs a special BMPSP sequence to detect the first L signal vectors, S(L), which can then be used as the survivor state to start regular BMPSP or BMDF. More specifically, we first run one BMPSP with $M_{VA} = K$ and $M_{surv} = 0$ to detect the first K signal vectors S(K). (A larger K has a better performance but at a higher computational cost.) Then S(K) is used as the survivor state to start another BMPSP with a fixed M_{VA} and a variable M_{surv} from K to $L - K$ to detect next $L - K$ signal vectors successively. By the end of the second BMPSP S(L) is estimated.

The second method, is to start the regular BMPSP or BMDF from a random state and let it run for L_{init} steps. If L_{init} is chosen long enough the algorithm will converge, defined as when the branch metric is below some threshold. Then run the algorithm backward with the converged state as the initial state. We refer this procedure as random with backward detection (RBD) initialization.

The convergence behavior of RBD should be similar to that of an erroneous state converging to correct state in a standard single user VA detection [28, 89], as will be confirmed later in the simulation. In a single user VA the length of a diverged trellis is related to that of an error event, which has a probability that diminishes exponentially with its size. In [89] it was shown that the diverged trellis length is about 4-5 times of the VA memory size. From the similarity between BMVA and the single user VA, we would expect BMVA, starting from an erroneous state (corresponding to an incorrect autocorrelation matrix $R_{M-1}(n)$ in (5.37)), to converge to one of equivalent states (corresponding to the correct autocorrelation matrix $R_{M-1}(n)$ in (5.37)), very quickly. From Theorem 5.2.2 we know that a correct $R_{M-1}(n)$ means any corresponding solution segment $S(n - M : n - 1)$ belongs to the correct equivalent set. BMPSP, as a suboptimal variant of BMB, is expected to behave similarly. In the next section we will examine the convergence behavior of RBD further via simulation.

5.4 Simulation Results

In this section, we will present the simulation results of BMPSP with $M_{VA} = 2$, BMDF and MRDF for 3 BPSK cochannel signals of block size $N = 100$. For convenience we assume that all users have the same fade rate and all signals are transmitted synchronously and statistically independent. We will first study the detectors with a known initial (survivor) state. The effect of unknown initial state on the detectors will be studied later.

Test condition is very similar to that in Section 4.2.3. The time-varying channel and data are generated in the same way as that shown in Fig. 4.3.

The first test is to study the effect of the predictor length M on the detectors and results are depicted in Fig. 5.1. As shown, for both BMPSP and BMDF detectors and with equal or unequal power users some performance degradation is visible when $M < 20$; which suggests $M \geq 20$ should be used for $f_dT = 2\%$. This result agrees with our earlier estimate in Section 5.3.1, implying the analysis in Section 5.3.1 can be used as a guide in selecting M in practice. This figure also shows that both detectors work very well under this fast fading channel condition ($f_dT = 2\%$) and weaker signals can be detected effectively under strong interference signals, although performance degrades compared to the stronger one due mainly to a reduced SNR. (In this example, the weakest signal suffers from a +10dB interference and -5dB noise level relative to the strongest signal.) Moreover, this figure shows that BMPSP always outperforms BMDF due to an increased state set, as expected.

The second test is to evaluate BMPSP/BMDF/MRDF under different fading channels. The results in Fig. 5.2 show that both BMPSP and BMDF perform very well under these fading channels, and the performance improves for slower fading channels for a reduced prediction error variance (5.28) since Condition 5.3.1 becomes more accurate. Again the result shows the BMPSP is the better of the two. MRDF, on the other hand, is inferior to both BMPSP and BMDF. In this particular test, $M_r = 5$ was used so that the predictor length is 5. It shows that length reduction in prediction filter $f(n)$ results in some detection error, which is mainly caused by the sparse data distribution, as evident for faster fading channels. However, in some applications the gain in complexity reduction outweighs some performance loss.

The effect of the antennas on these algorithms are also evaluated, as shown in Fig. 5.3. The performance improves drastically with the number of antennas and the almost log-linear relationship between the number of antennas and BER is a clear indication of

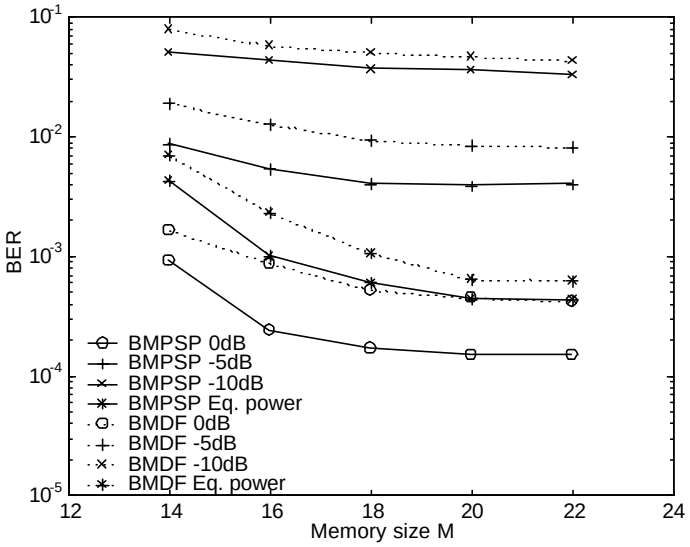

Figure 5.1: BMDF and BMPSP with M_{VA} = 2 for 3 BPSK users with m = 4; f_dT = 2%; SNR = 15dB (normalized to the 1st signal) and a variable memory size M. Unequal power levels (normalized to the 1st signal): 0, -5 and -10dB.

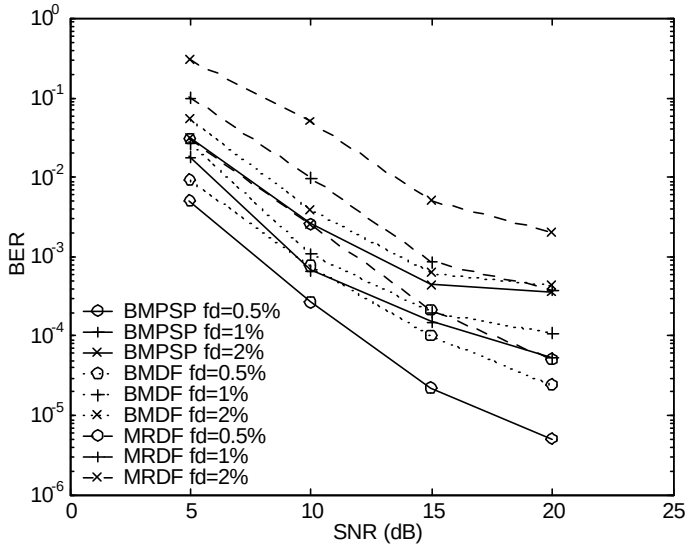

Figure 5.2: BMDF, MRDF and BMPSP with M_{VA} = 2 for 3 equipower users with m = 4 and various fade rates. In MRDF, filter length M_r = 5 and in BMPSP and BMDF M = 20.

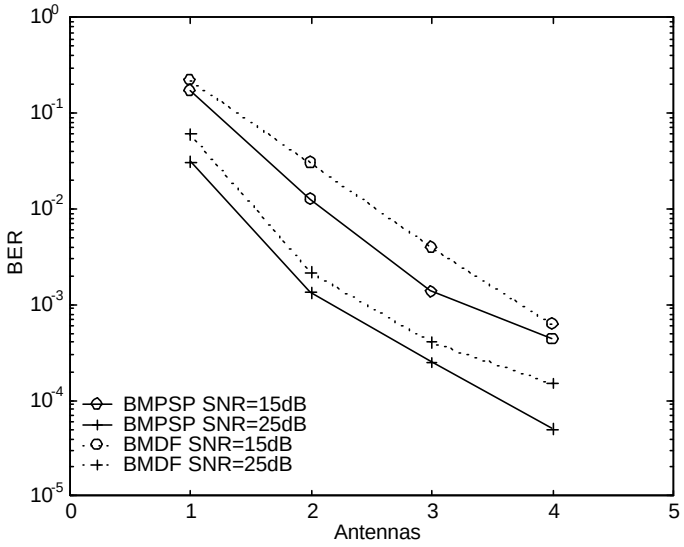

Figure 5.3: BMDF and BMPSP with $M_{VA} = 2$ and $M = 20$ for 3 equipower users with $f_dT = 2\%$ and various number of antennas.

diversity gain, a well-known fact for a single user MLSE detector[90]. Here, we observe a similar result for a multiuser MLSE. For a two diversity antennas system, it shows that with BMPSP 3 equipower BPSK signals can be detected with uncoded BER of 1% at $f_dT = 2\%$ and SNR=15dB. With slightly increased SNR or slower f_dT the same result can also be obtained for BMDF.

Finally we examine the two initialization routines presented in Section 5.3.3 through simulation. The result in Table 5.1 shows that indeed for RBD, BMPSP converges very quickly (after about $L_{init} \approx 100$ or $L_{init} \approx 5M$). This result is remarkable since it confirms our earlier analysis that the convergence behavior of BMVA or BMPSP should resemble that of a single user blind VA{algorithm, converging after about 4-5 times of the memory size of VA [89]. It also indirectly verifies that the memory size of BMPSP is about 20 if we use $L_{init} = 5M$, which has been shown in Fig. 5.1. Comparing RBD with a sufficiently long start up length L_{init}, BOU is less effective due to its insufficient initial memory size K. In addition, BOU requires more processing power than RBD and special BMPSPs. For these reasons, RBD is more suitable for a practical implementation. However, BOU has a short processing delay.

BOU	RBD			
K = 4	L_{init} = 60	L_{init} = 80	L_{init} = 100	L_{init} = 120
$3:3 \cdot 10^{-3}$	$4:4 \cdot 10^{-3}$	$9:2 \cdot 10^{-4}$	$4:8 \cdot 10^{-4}$	$4:1 \cdot 10^{-4}$

Table 5.1: BER performance of BMPSP with BOU or RBD initialization routines. f_dT = 2% and M_{VA} = 2

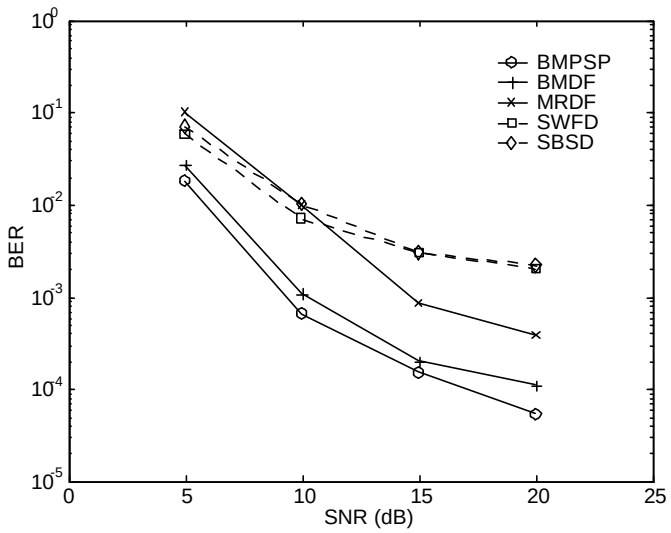

Figure 5.4: Performance comparison of BMB variants (solid lines) and BMML variants (dash lines) for d = 3 equipower users with m = 4 antennas. In BMML, p = 1. As before, only the averaged BERs are shown. f_dT = 0:01:

5.5 Comparison of BMML and BMB Detectors

We have developed blind multiuser maximum likelihood (BMML) detector in the last chapter and blind multiuser Bayes (BMB) detector in this chapter. It will be interesting to compare the performance/complexity of these two classes of detectors. It is obvious that BMB is better than BMML as more information is utilized in the former. However, since only the suboptimal forms of both BMB and BMML detectors are actually implemented, it is not clear as which of these suboptimal detectors has the best performance. For simplicity of the discussion, we $\bar{x}$ the channel fade rate at 1%. The results are plotted in Fig. 5.4.

As can be seen, BMPSP and BMDF are better than SBSD (the subblock-by-subblock detection) and SWFD (sliding window feedback detection) for most of the SNRs. MRDF

outperforms BMML variants at SNR > 10dB. However, at low SNRs, MRDF is slightly worse than BMML variants.

The computational complexity is another important aspect of these detection schemes. We now compare the complexity of these detectors in terms of the number of CMAs. We assume that conditional correlation matrix R(S) is the same for all antennas and the fade rate is the same for all users. Since the convergence of BMML variants depends on the fade rates and number of antennas, the complexity for these detectors is the product of number of iterations and the number of CMAs in each iteration. First, we estimate CMAs for each detector in detecting a block of data, including or excluding the initialization procedures, then examine a numerical example. In Appendix D we give expressions for these estimates excluding initialization. In burst mode communications, initialization procedure may cost a signi¯cant part of overall complexity for blind detection schemes.

The complexity for these detectors including initialization are estimated in this section. For BMPSP, BMDF and MRDF, we use RBD running them backward ¯rst and using the ¯nal state to run them forward again. For SWFD and SBSD, we divide the block into several subblocks, each N_{sub} symbols long, and run the blind initialization routine (detailed in Section 4.2.1) for the ¯rst subblock, which costs about mdQN CMAs. We denote the number of iterations required for the ¯rst subblock and subsequent ones as I and I_{sub}, respectively. In the following we list the estimated complexity for these detectors with initialization routine.

BMPSP: $2C_{BMPSP}$;

BMDF: $2C_{BMDF}$;

MRDF: $2C_{MRDF}$;

SWFD: $C_{BMML}(\frac{N_{sub}}{N} + \frac{N \, j \, N_{sub}}{I_{1st}})$;

SBSD: $C_{BMML}(\frac{N_{sub}}{N} + \frac{I_{sub}}{I_{1st}}\frac{N \, j \, N_{sub}}{N})$.

In the above expressions a factor of 2 is used for the ¯rst three detectors because we run those detectors twice per block (one backward for initialization and one forward for detection).

As a numerical example, let m = 6; d = 3, Q = 2, M_r = 5, M = 20, M_{VA} = 2, p = 1, f_dT = 1%, N = 100, N_{sub} = 25, and I = 8 (according to Table 4.1) and I_{sub} = 4 (based on simulation). The results for detecting a block of data (N = 100) are listed in Table 5.2. As we can see from the table, BMPSP is the most expensive one of all and its performance compared to BMDF is only slightly better, suggesting BMDF is more suitable

Complexity in CMAs	BMPSP	BMDF	MRDF	SBSD	SWFD
Without Init.	$9.7£10^6$	$1.2£10^6$	$9.4£10^4$	$1.8£10^5$	$1.1£10^6$
With Init.	$1.9£10^7$	$2.4£10^6$	$1.9£10^5$	$2.2£10^5$	$3.4£10^6$

Table 5.2: Complexity comparison of suboptimal BMML and BMB detectors, with and without initialization procedure. N = 100:

for implementation. MRDF has more than 10 times lower computational load than BMDF but its performance is also inferior to BMDF. For BMML variants{SBSD and SWFD, the performance is similar but the former is 10 times cheaper in computation. Finally, both MRDF and SBSD have similar complexity and similar performance for SNR < 10dB, but at moderate to high SNRs (> 10dB) MRDF has a much better performance. However, BMML variants have an advantage over BMB ones in that they do not require the statistical knowledge of the fading channel. More optimization is of course possible, but a detailed study will not be given in this thesis.

5.6 Summary

We have presented a blind multiuser Bayes detector and its suboptimal but implementable form { BMPSP (including BMDF) and a less optimal but memory reduced form, MRDF. We have shown that all solutions of BMB are equivalent with probability one under a set of easily met conditions. Although the exact memory size of BMB is shown to be the same as the block size, it can be truncated to a much shorter value without much performance degradation. An expression for this truncated memory size, which depends on the channel order, the type and the number of signals, is also provided. Further memory reduction is provided by incorporating PSP with BMB resulting in an implementable form. For applications which have a limited computational power and can tolerate some degraded performance, an even simpler form of BMDF { MRDF can be used. Two blind initialization methods are also presented. The performance of the detectors and initialization routines are studied via simulation and results show that BMPSP and the blind initialization routines are highly e®ective in a fast fading environment. Finally, we compare two classes of detectors, BMML and BMB, and show that BMB variants generally have a better performance. Among BMB detectors, BMPSP is the best in terms of performance but only slightly better than BMDF and requires as much as 10 times processing power as BMDF. BMDF is in turn 10

times as expensive as MRDF, but the performance is much improved. MRDF, the least e®ective one in the BMB class, is still better than the BMML class at SNR > 10dB and also has the least computational requirement of all, suggesting it may be the choice of practical systems if the computational requirement is critical.

Chapter 6

Multiuser Detection in Multipath Channels

6.1 Introduction

In the preceding chapters, we investigated two classes of blind detection schemes in time-varying fading channels. To simplify the discussion, we have made two assumptions: multiple users are symbol synchronous and multipaths have a delay spread that are negligible with respect to symbol period. Unfortunately, these are not realistic assumptions in the fading uplink channels of a personal communication system. Thus a more complete solution to this more difficult problem is needed. In this chapter, we solve the blind detection problem of multiple asynchronous cochannel users in a fast time-varying multipath channel with long delays.

This chapter is organized as follows. First, we develop the Bayes and maximum likelihood approaches based on generalization of the BMML and BMB detectors to the delay spread channel model. Although optimal, these approaches involve an enumerative search over high-dimensional spaces, which is computationally expensive. As an alternative, we present a sub-optimal approach next. This approach is a cascade of the subspace-based blind equalization approach developed in [31], [50] and BMML developed in Chapter 4. We then present simulation results and a summary.

6.2 Optimal Detectors in Multipath Channels

In this section, we consider the extension of BMML and BMB detectors to the multipath channels with delay spread. For simplicity of discussion, we assume all cochannel sources transmit from a linearly modulated signal set of the same constellation size and symbol rate. We also assume that the multipath channel has a finite impulse response and the transmission filter has a pulse duration of L_{pulse} symbols such that $g(t) \neq 0$ for $0 \cdot t \cdot L_{pulse}$. For notational convenience, we set the symbol period to one.

6.2.1 BMML Approach

As in Chapter 4, the time-varying channel is assumed to be perfectly characterized by a polynomial of order p.

In Chapter 4 BMML is shown to be equivalent to a LS problem (4.5). The goal of this subsection is then to rearrange the system equations into a matrix form so that we can obtain a similar LS problem. To proceed, we first sample the array output at the symbol rate and express (2.14) in a vector form,

$$x(n) = \sum_{i=1}^{d} \sum_{n'=-1}^{1} h_i(n; n') s_i(n - n') + w(n) \tag{6.1}$$

where

$$h_i(n; n') = \sum_{l=1}^{L_i} a_{il}(n) g(n' - \tau_{il}) \tag{6.2}$$

where L_i is the number of paths for the ith user. A scalar form of (6.1) is shown in Fig. 6.1. Each entry $h_{ji}(t; t')$ in $h_i(t; t')$ represents the channel impulse response from the ith source to the jth antenna at time t from all L_i paths of delays τ_{il}. We assume without loss of generality that all m channel responses, $h_{ji}(t; t')$, $j = 1:::m$; associated with source i, have a finite impulse response of length at most T_i symbols:

$$h_{ji}(t; t') = 0; \quad t' \in [0; T_i) \quad j = 1; :::; m; \; i = 1; :::; d: \tag{6.3}$$

Define the maximal channel length among all sources,

$$T = \max_i T_i$$

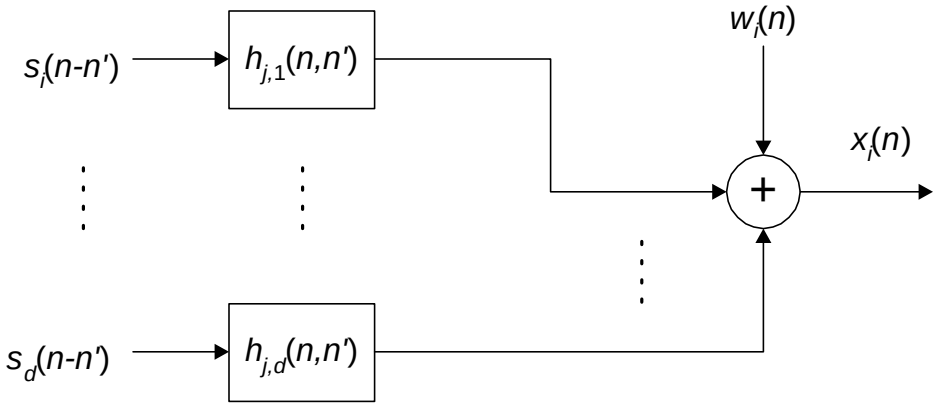

Figure 6.1: A illustration of the time-varying multipath channel model for antenna j.

and the maximal number of multiple paths

$$L = \max_i L_i:$$

Under the polynomial basis expansion,

$$a_{jil}(n) = a^0_{jil} + (n)a^1_{jil} + ¢¢¢ + (n)^p a^p_{jil}$$

and

$$\begin{aligned} h_i(n;n^0) &= \sum_{l=1}^{L_i} a^0_{¢il} + na^1_{¢il} + ¢¢¢ + n^p a^p_{¢il}\ ¢\ g(n^0 - ¿_{il}) \\ &= h^0_i(n^0) + nh^1_i(n^0) + ¢¢¢ + n^p h^p_i(n^0) \end{aligned} \tag{6.4}$$

where for notational convenience, we have de¯ned

$$h^k_i(n^0) = \sum_{l=1}^{L_i} a^k_{¢il} g(n^0 - ¿_{il}):$$

From (6.3), (6.4) and (6.1), we have

$$\begin{aligned} x(n) &= \sum_{n'=1}^{T} \sum_{i=1}^{K} [h_i^0(n') + n h_i^1(n') + \cdots + n^p h_i^p(n')] s_i(n - n') + w(n) \\ &= \sum_{n'=1}^{T} [H^0(n') + n H^1(n') + \cdots + n^p H^p(n')] s(n - n') + w(n) \\ &= \sum_{n'=1}^{T} [H^0(n') \; H^1(n') \; \cdots \; H^p(n')] \begin{pmatrix} s(n-n') \\ n s(n-n') \\ \vdots \\ n^p s(n-n') \end{pmatrix} + w(n) \\ &= \sum_{n'=1}^{T} H_{aug}(n') s_{aug}(n - n') + w(n) \end{aligned} \tag{6.5}$$

where

$$H^k(n') = [h_1^k(n'), h_2^k(n'), \ldots, h_d^k(n')]'.$$

In matrix form, (6.5) can be expressed as

$$X = H_{aug} S_{aug} + W \tag{6.6}$$

where

$$X = [x(1) \quad \cdots \quad x(N)]';$$

$$H_{aug} = [H_{aug}(1) \quad H_{aug}(2) \quad \cdots \quad H_{aug}(T)]; \tag{6.7}$$

and

$$S_{aug} = \begin{bmatrix} s_{aug}(0) & s_{aug}(1) & \ddots & s_{aug}(N-1) \\ s_{aug}(-1) & s_{aug}(0) & & \\ \ddots & \ddots & \ddots & \vdots \\ s_{aug}(1-T) & s_{aug}(2-T) & \cdots & s_{aug}(N-T) \end{bmatrix}. \tag{6.8}$$

Note that if the channels do not all have the same length, certain columns of H_{aug} are zero. Since entries w_{ji} are independent Gaussian variables, it is straightforward to show that BMML is equivalent to the following LS problem:

$$\min_{H_{aug}, S_{aug}} \|X - H_{aug} S_{aug}\|_F^2 \tag{6.9}$$

with S_{aug} satisfies the constraint that it is block-Toeplitz structure (6.8) with s_{aug} defined in (6.5).

At this point, we could use the BMML algorithms developed in Chapter 4 to detect S_{aug}. The channel length T_i can also be estimated theoretically with this detector by observing the structure of $\hat{H}_{aug}$. Specifically, if after some channel length index l, some columns of $\hat{H}_{aug}$ are much smaller comparing with previous ones, we may regard the channel length for that particular one is l.

Note that temporal oversampling is not required to obtain unique solutions (within the equivalent set) since it can shown similarly that with probability one the channel matrix H_{aug} satisfies the unique mapping condition (condition (4.2.1) of Chapter 4). Since each signal (column) vector in $S_{aug}(N)$ is no longer independent of others, it cannot be estimated individually. This suggests that the signal vectors have to be detected jointly for optimal detection, such as by a Viterbi algorithm, as was done in the case of CPM signal detection in Section 4.3. However, the size of state set here is much larger than that in Section 4.3. Thus, although feasible, this approach may not be attractive due to its high computational complexity.

6.2.2 BMB Approach

In this subsection we outline a BMB approach to combat the multipath time-varying fading channels. The key to this approach is to obtain the conditional data autocorrelation matrix, conditioned on data S(N). For easy reference, we restate the system equation (2.14) and channel autocorrelation function (2.18), as

$$x_j(n+\epsilon) = \sum_{i=1}^{K} \sum_{l=1}^{L} \sum_{n'=-\infty}^{\infty} a_{j,i,l}(n+\epsilon)\, g(n+\epsilon - \epsilon_{i,l} - n' T)\, s_i(n') + w_j(n+\epsilon) \tag{6.10}$$

and

$$\tfrac{1}{2} E[a_{j,i,l}(t)\, a_{j',i',l'}(t')] = \delta_{j,j'}\delta_{i,i'}\delta_{l,l'}\tfrac{1}{2}\sigma_{j,i,l}^2 J_0\left(2\pi f_d (t-t')\right) \tag{6.11}$$

under the WSSUS channel model, in which information on $\sigma_{j,i,l}^2$ is assumed to be known a priori. If the prior pdf of $\epsilon_{i,l}$ is also assumed to be known, e.g., constant over some duration, then we can treat $\epsilon_{i,l}$ as a nuisance parameter and integrate it out of the following conditional

data autocorrelation function:

$$r_{xx;j}(t - t') = \frac{1}{2}E[x_j(t)x_j^*(t')]$$
$$= \sum_{i=1}^{P}\sum_{l=1}^{L_i}\sum_{m;m'=-1}^{N} \sigma_{jil}^2 J_0\left(2\pi f_d (t - t')\right) \int g(t - m - \tau_{il})g(t' - m' - \tau_{il})p(\tau_{il})d\tau_{il} + \sigma_w^2 \pm_{tt'}$$

where $p(\tau_{il})$ is the prior pdf of τ_{il}. (For a typical urban and rural cell, the delay spread is about 10^{-1}s which is about 0.5 and 2.5 symbols duration for IS136 and GSM, respectively.) When sampled at symbol rate, we obtain the conditional data autocorrelation matrix as

$$R_j(S) = \sum_{i=1}^{P}\sum_{l=1}^{L_i}\sum_{m;m'=-1}^{N} D(s_i(m : N + m))R_{jilmm'}D(s_i^*(m' : N + m')) + \sigma_w^2 I \tag{6.13}$$

where

$$\left(R_{jilmm'}\right)_{n;n'} = \sigma_{jil}^2 J_0\left(2\pi f_d (n - n')\right) \int g(n - m - \tau_{il})g(n' - m' - \tau_{il})p(\tau_{il})d\tau_{il} + \sigma_w^2 \pm_{nn'}. \tag{6.14}$$

Then the likelihood function for the multipath channel can be written as

$$p(x_j|S) = \frac{1}{(2\pi)^N |R_j(S)|} \exp\left(-\frac{x_j^* R_j(S)^{-1} x_j^T}{2}\right)$$

with $R_j(S)$ given in (6.13). A similar procedure as that described in Chapter 5 can then be developed to detect S under the Bayes criterion. The uniqueness of the detection can also be analyzed similarly as that in Chapter 5 and the detail is beyond the scope of this thesis. From (6.13), we can show that the number of the state set in this approach is $Q^{d(T + L_{pulse})}$ times larger than that of the memoryless channel, because of the multipath. Thus, it may not be suitable for applications with limited computational power.

The results of both the BMML and BMB approaches in this section show that FA is a very powerful signal property by which multiple signals can be theoretically detected in a frequency selective and time-varying channel by using a antenna array of arbitrary array size and geometry. Moreover, samples need not be taken more than once per symbol. However, the computational complexity problem may hinder them from being practical. In the next section, we use a different approach that has a lower complexity.

6.3 A Sub-Optimal Detector - SBJD

In the last section, two optimal detectors have been proposed for detecting asynchronous multiple users in a time-varying frequency selective channel. However, the advantages of these optimal detectors might be negated by a higher computational cost. To overcome the complexity problem, in this section we use an alternative approach by exploiting another important property of the signals{the cyclostationarity (CS){by temporal oversampling. Oversampling does not merely improve the performance because of its insensitiveness to timing errors in the time recovery circuit, as used in a fractionally spaced equalizer. More importantly, it provides a new way to solve the problem by exploiting the cyclostationarity of the signals. CS was ¯rst introduced in terms of "statistical" correlation and other moments and later generalized to the "deterministic" correlation (sampled correlation) by Tong el at. [56], [57], which generalized the idea of "statistical" CS to temporal oversampling to increase the dimensionality of the measurements. More speci¯cally, oversampling allows the channel to be identi¯ed uniquely and blindly using the "deterministic" autocorrelation of the oversampled antenna output [56], [57], or the "deterministic" cross correlation between the oversampled antenna output and the channel [49], or the "deterministic" cross correlation between the oversampled antenna output and data [31]. In the latter two approaches the second order channel identi¯cation problem is converted to a set of ¯rst order (linear) equations which can be handled with ease. In the last approach, multipath channel is also blindly equalized. After the blind channel identi¯cation or equalization, the problem can be solved using either BMB or BMML algorithms, which are discussed in the last two chapters. This is the approach we will adopt in this chapter. It makes use of a diversity array and is subspace-based, hence, the name subspace blind joint detection (SBJD). By utilizing FA and CS properties together, SBJD is less expensive. It appears that we are the ¯rst to use FA-CS based method to solve the multiuser detection problem in time-varying multipath channels.

6.3.1 Problem Formulation

For clarity of the presentation, we defer the inclusion of the additive white Gaussian noise term until later. Let oversample rate be P, such that $¿ = 0; 1=P; :::; (P ¡ 1)=P$. Note that not all users have the same e®ective channel length. That is, the set of samples in one symbol period, $x_j(n); :::; x_j(n + (P ¡ 1)=P)$; may not involve the same number of adjacent

symbols for each user. Denote d_z; $z = 2, ..., Z$, as the number of users having the same effective channel length of z symbols. Clearly $\sum d_z = d$. Note that intersymbol interference (ISI) exists even for memoryless channel if sample timing is not at symbol boundaries, that is, $z > 1$. In general, however, ISI is caused by multipath interference and incorrect timing. We then rearrange (2.14), according to z, as

$$x_j(n+\tau) = \sum_{z=2}^{Z} \sum_{i \in \{d_z\}} \sum_{l=1}^{L_z} \sum_{n'=1} a_{jil}(n+\tau) g(n+\tau+n'-\tau_{il}) s_i(n-n'). \tag{6.15}$$

where $\{d_z\}$ denotes the set in which users have the same z, and L_z the number of paths that have the same z. Under the linear channel assumption,

$$a_{jil}(n+\tau) = a^0_{jil} + (n+\tau)a^1_{jil}; \quad \tau = 1/P;\ 2/P;\ :::;\ (P-1)/P. \tag{6.16}$$

To simplify the derivation, we further assume (6.16), for $k = 1; :::; K$, can be approximated as

$$a_{jil}(n+\tau+k) \approx a^0_{jil} + (n+k-1)a^1_{jil}; \tag{6.17}$$

which indicates that the channel does not depart significantly from our linear channel model over K symbol periods, where the integer K is called the smoothing factor and its meaning is explained later in this section. This assumption is reasonable if the channel fade rate is not too high compared to the data rate and K is small.

As in [31], we formulate $x_j(t)$ as a matrix $X(K)$ of dimension $KmP \times (N-K+1)$,

$$X(K) = \begin{bmatrix} x(1) & x(2) & \cdots & x(N-K+1) \\ x(2) & x(3) & \cdots & x(N-K+2) \\ \vdots & \vdots & \ddots & \vdots \\ x(K) & x(K+1) & \cdots & x(N) \end{bmatrix} \tag{6.18}$$

where $x(n) = (x_1(n); :::; x_m(n); x_1(n+1/P); ::; x_m(n+1/P); :::; x_m(n+(P-1)/P))^T$. With (6.16) and (6.18), (6.15) can be rewritten compactly as

$$\begin{aligned} X(K) &= \left[H^2_{aug}(K);\ H^3_{aug}(K); :::; H^Z_{aug}(K) \right] \begin{bmatrix} S^2_{aug}(K+1) \\ S^3_{aug}(K+2) \\ \vdots \\ S^Z_{aug}(K+Z-1) \end{bmatrix} \\ &= H_{aug}(K) S_{aug}(K) \end{aligned} \tag{6.19}$$

where components are defined below. First, $H^z_{aug}(K) = \left[H^z_{aug;1}(K); :::; H^z_{aug;d_z}(K)\right]$ with $H^z_{aug;i}(K) = (H^{z;0}_{aug;i}(K); H^{z;1}_{aug;i}(K))$ where $H^{z;\alpha}_{aug;i}(K)$ is a $KmP \pounds (K+z)$ matrix, defined as

$$H^{z;\alpha}_{aug;i}(K) = \begin{pmatrix} h^{z;\alpha}_i(z) & \cdots & h^{z;\alpha}_i(1) & 0 \\ \ddots & & \ddots & \\ 0 & h^{z;\alpha}_i(z) & \cdots & h^{z;\alpha}_i(1) \end{pmatrix};$$

which is constructed by first stacking the channel vector

$$h^{z;\alpha}_i(k) = \left[h^{z;\alpha}_{1i}(k); :::; h^{z;\alpha}_{mi}(k); h^{z;\alpha}_{1i}(k+1=P); :::; h^{z;\alpha}_{mi}(k+1=P); :::; h^{z;\alpha}_{mi}(k+\frac{P-1}{P})\right]^T$$

K times and then shifting each of them one place to the left over its lower one followed by setting zeros to the rest of cells, where

$$h^{z;\alpha}_{ji}(k+\tau) = \sum_{l=1}^{L} a^{\alpha}_{jl} g(\tau + k - \tau_{il}):$$

Next,

$$S^z_{aug}(K+z-1) = \left[(S^z_{aug;1}(K+z-1))^T; :::; (S^z_{aug;d_z}(K+z-1))^T\right]^T$$

with

$$\left[S^z_{aug;i}(K+z-1)\right]^T = \left[(S^z_i(K+z-1))^T; (S^z_i(K+z-1)D(K))^T\right]$$

where the smoothed data matrix $S^z_i(K+z-1)$, of dimension $(K+z-1) \pounds (N-K+1)$, is defined as

$$S^z_i(K+z-1) = \begin{pmatrix} s_i(2-z) & s_i(3-z) & \cdots & s_i(N-K+2-z) \\ \vdots & & \ddots & \vdots \\ s_i(K) & s_i(K+1) & \cdots & s_i(N) \end{pmatrix}$$

with $D(K)=\mathrm{diag}(1; :::; N-K+1)$.

6.3.2 SBJD Algorithm

Our aim is to estimate $\{s_i(n)\}$ from (6.18) without using training sequences. This is achieved by the proposed SBJD consisting of two parts: the first part synchronizes the user sequences while the second part detects users jointly and blindly. The data synchronization is based on the multiuser channel deconvolution (MCD) [31] which performs a subspace projection followed by a null subspace smoothing. The time-variation of the channels is incorporated

into MCD by the augmented channel matrix $H_{aug}(K)$ and data matrix $S_{aug}(K)$. We refer to this time-varying MCD as TVMCD. For joint data detection we use BMML developed in Chapter 4 for memoryless channels and refer to it as the zero memory BMML or ZBMML to distinguish it from the BMML approach developed in Section 6.2.1.

We first detail TVMCD. If for some K, the channel matrix $H_{aug}(K)$ has full column rank, then the rank property of $S_{aug}(K)$ is preserved through mapping to $X(K)$. The condition for $H_{aug}(K)$ to be smoothed to full column rank is given in [57]. Denoting the null subspace of $X(K)$ as $V_0(K)$ gives

$$V_0(K)S_{aug}(K)^H = 0: \tag{6.20}$$

By (6.17) and the structure of $S_{aug}(K)$; (6.20) is approximately equivalent to

$$V(K+k)S^z_{aug}(K+z-k)^H = 0; \quad k = 1;:::;K+z-2 \tag{6.21}$$

where $V(K+k)$ is constructed by first stacking the matrix $V_0(K)$ k times and then shifting each $V_0(K)$ one place to the left over its lower one followed by padding zeros to the remaining cells, as follows,

$$V(K+k) = \begin{bmatrix} V_0(K) & 0 & \cdots & 0 \\ 0 & V_0(K) & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \cdots & 0 & V_0(K) \end{bmatrix}: \tag{6.22}$$

$V(K+k)$ can be interpreted as the smoothed $V_0(K)$ and $S^z_{aug}(K+z-k)$ as the 'de-smoothed' signal matrices $S^z_{aug}(K+z)$, for $z = 2;:::;Z$. Then (6.21) can be interpreted as that the row subspace of the de-smoothed data matrix $S^z_{aug}(K+z-k)$ is orthogonal to the smoothed signal null subspace $V(K+k)$. If the effective channel lengths are the same for all users, say $d_z \neq 0$, then by setting $k = K+z-1$ in (6.21) we have $V(2K+z-1)S^z_{aug}(1)^H = 0$: If $S_{aug}(K)$ has full row rank, then it can be shown [31] that the rows of $S^z_{aug}(1)$ are the only vectors in the null space of $V(2K+z-1)$. Thus we must have

$$V^?(2K+z-1) = TS^z_{aug}(1) \tag{6.23}$$

where T is a full rank but unknown square matrix and $V^?(2K+z-1)$ is the null subspace of $V(2K+z-1)$. By the definition of $S^z_{aug}(1)$, we see that nth column of $S^z_{aug}(1)$ contains only

user symbols at time n. This fact can be interpreted as the users have been synchronized or the channels have been equalized. Suppose now the e®ective channel lengths are di®erent and $d_z = 0$ only for $z > 3$. We ¯rst set $z = 3$, $k = K + z - 1$ in (6.21) and obtain (6.23) which holds only for users with $z = 3$. Then by setting $k = K + 1$ and $z = 2$ and 3 in (6.21), we obtain

$$V^?(2K+1) = T^0 \begin{pmatrix} S^2_{aug}(1) \\ S^3_{aug}(2) \end{pmatrix} \tag{6.24}$$

where T^0 is another full rank but unknown square matrix. However, from Chapter 4 we know that ZBMML can be used to detect $S^3_{aug}(1)$ from (6.23). Once $S^3_{aug}(1)$ is detected and its smoothed form $\hat{S}^3_{aug}(2)$ is substituted in (6.24) $S^2_{aug}(1)$ is readily detected using ZBMML again since the only unknown data in (6.24) now is $S^2_{aug}(1)$ which is in the synchronized form.

The above discussion can be easily extended to more general cases of multipath of more than two delays. Readers may refer to the original paper [31]. The discussion so far has assumed the system is noise free. When noise is present we calculate numeric subspaces and replace (6.23), (6.24) with their least squares solutions. We summarize the SBJD algorithm as follows:

1. Set $K = 1; 2; :::; Z - 1$; and construct $X(K)$ according to (6.18). Compute $V_0(K)$ the null space of $X(K)$.

2. For any $z = Z; Z - 1; :::; 2$; if $d_z \neq 0$; construct $V(2K + z - 1)$ from (6.22) and calculate its null subspace $V^?(2K + z - 1)$.

3. Solve (6.23) and/or (6.24) using ZBMML to solve $\{s_i(n)\}$.

In step 1, setting $K = 1; 2; :::; Z - 1$ allows us to obtain $Z - 1$ equations from the singular value distribution of $X(K)$ so that d_z can be resolved and used in step 2. If for some z $d_z = 0$, step 2 can be saved for this z.

De¯ne the data frame as a record of data with a single I.D. sequence for each user. If the data frame size is greater than the sub-block of size N_{sub} an additional step is taken to ensure correct detection over the entire frame. Due to the detection order ambiguity inherent in blind detection schemes [29, 32], each sub-block needs to be joined correctly with its predecessor. One way to remedy this problem is to start the SBJD with the initial

gain estimate projected from the previous sub-block. Refer to Section 4.2.1 for further detail. Simulation results there showed that this scheme works very well for ZBMML.

6.3.3 Implementation

This section concerns the implementation of SBJD. First we need to know the size of data window N in SBJD. Figure 6.2 illustrates a snapshot of the singular value distribution of a typical single X(1). It can be seen from the ¯gure that even at low fade rates there is a signi¯cant energy associated with higher modes. A mode is de¯ned here as one of the dominant eigen modes of the channel autocorrelation matrix, shown roughly as a stair in the ¯gure. As the fade rate rises more modes (indicated as the number of stairs in the ¯gure) are excited. It is interesting to note that when the fade rate is not too high each mode is associated with one monomial approximation of the channel starting from the zero-th order. See Section 2.2.3 for more detail. In other words, we can roughly associate each mode with one monomial of the channel, corresponding to static, linear and quadratic channel models, and so on. Note in the ¯gure $N = 40$ was used because it allows us to see more modes (a longer data record with a higher fade rate excites more modes). This ¯gure also suggests that for $N = 40$ and $f_dT \cdot 0{:}5\%$ the linear channel approximation is adequate since only one higher mode is excited. Accordingly, if we reduce the data window size by half and increase the fade rate by two, we would expect about the same channel model error or performance. However, N can not be too small, since $S_{aug}(K)$ needs to be of full rank to ensure the unique subspace projection, which is less probable for smaller N and larger K.

Although the smoothing factor K can take a range of values, the simulation shows that the gap between signal and noise singular values is narrower for larger K, especially at higher fade rate. For this reason it should be selected as small as possible; e.g., in our application, $K \cdot L ¡ 1$.

In step 3 of the algorithm we use ZBMML rather than the iterative least squares with projection (ILSP) [29] or its variants since it requires $V^?(K + k)$ to have the same rank as that of $((S^2_{aug}(K + z ¡ k))^T; (S^3_{aug}(K + z ¡ k))^T; :::; (S^Z_{aug}(K + z ¡ k))^T)^T$. Depending on the data, sometimes $V^?(K + k)$ su®ers from the numeric rank de¯ciency problem. That is, the numeric rank of $V^?(K + k)$ may be less than full and thus ILSP does not work. ZBMML, on the other hands, works very well since it does not impose the rank requirement on $V^?(K + k)$. In fact, $\dim(V^?(K + k))$ is adjusted dynamically in the algorithm. (The dimensionality of $V^?(K + k)$ is determined by the number of singular values of $V(K + k)$

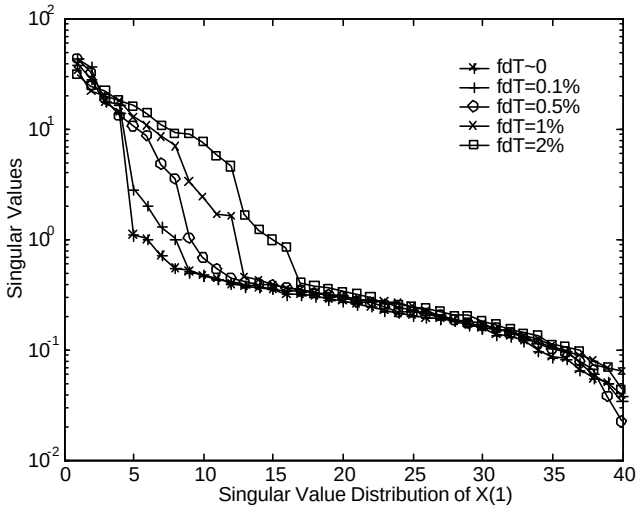

Figure 6.2: Singular values distribution of $X(1)$ at different fade rates with $d = 2$, SNR=25dB and $N = 40$.

whose values are below a threshold which is determined experimentally.) Thus, whenever $V^{?}(K + k)$ suffers from the rank deficiency problem a lower-dimensioned null subspace of $V(K + k)$ is used in (6.23) or (6.24). The simulation also indicates that with this change SBJD actually converges faster and performs better in some cases when a lower-dimensioned $V^{?}(K+k)$ is used. More study is needed to quantify when a lower-dimensioned null subspace $V^{?}(K + k)$ should be used.

6.3.4 Numerical Simulation

In this section we evaluate the performance of the SBJD algorithm in a time-varying multipath channel by simulation. In particular, we study two equipower cochannel BPSK signals under independent fading channels of identical fade rate. 12The pulse shape is a raised cosine with roll-off factor $\beta = 0.35$. In the first set of experiments, we study the asynchronous transmission only. In other words, there is no multipath and ISI is caused by incorrect timing only. The delays of these two users are 0.6; 0.85, respectively. The delays are chosen so that the two users have the same effective channel length. The channel is modeled as a Rayleigh process with Jakes' power spectrum [73]. The sample rate $P = 2$ and number of antennas $m = 4$. SNR is defined as the ratio of user #1 power over the noise variance. For

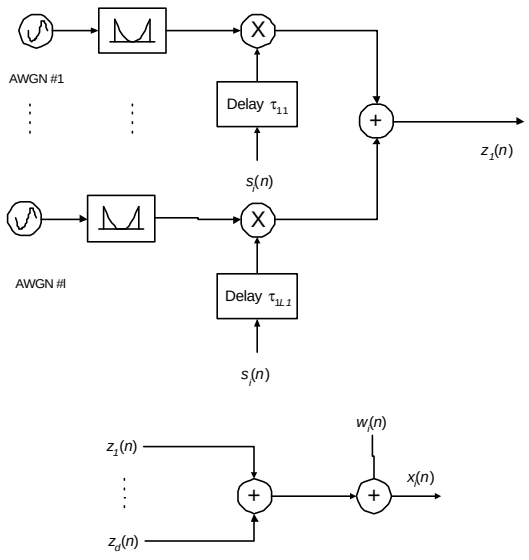

Figure 6.3: Signal $x_j(n)$ generation in a time-varying multipath channel. Each signal $s_i(n)$ has L_i paths with delay ¿$_{il}$, $l = 1; :::; L_i$.

display clarity, only the averaged BER for the two equal powered signals are shown.

Figure 6.4 shows the result. Also shown as comparison is another SBJD that treats channels as static over the entire data window, which is referred to as SBJD/S to distinguish it from SBJD. As we can see from the ¯gure, SBJD performs well for various fade rates and SNRs. In comparison, SBJD/S su®ers from fast fading; e.g., at $f_dT = 1\%$; SBJD/S is clearly ine®ective. However at higher SNR and very low fade rate, such as $f_dT = 0{:}1\%$; SBJD/S is better due to an improved global convergence, which is explained in Chapter 4. Similarly, the performance °oors in both detectors is also mainly due to the channel model error and global convergence of ZBMML. However, delay spreading does not contribute the performance °oors, as can be seen in the next experiment.

In the second experiment, we study the e®ects of multipath as well as asynchronous transmission. There are two paths for each user and they fade independently with the same average power. The path delays for user #1 and #2 are set to as [0.5 0.95] and [0.8 0.6], respectively. The results are shown in Figure 6.5. It shows similar results as that for the asynchronous user case. It is interesting to compare these two results. For example, the result clearly shows that multipath propagation actually improves the performance and

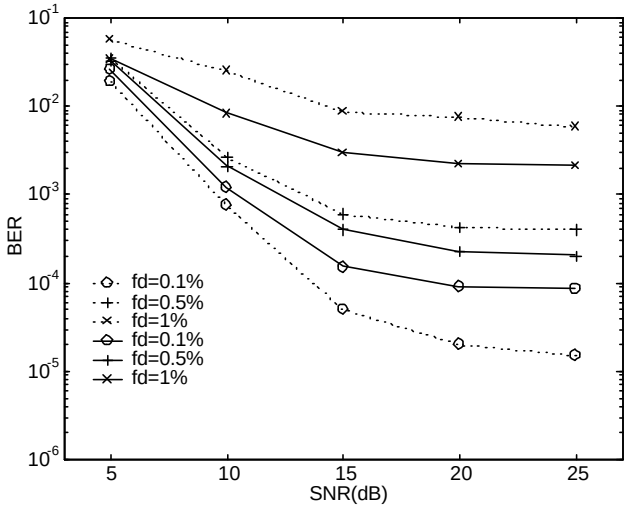

Figure 6.4: The performance of SBJD (solid lines) and SBJD/S (dashed lines) for two asynchronous equipower users at di®erent fade rates and SNRs. N = 20 and m = 4.

reduce the performance °oor. This is a well known result in single user detection case due to the time-diversity e®ect. Here we observe a similar diversity gain in a multiuser detection scheme. It also shows that at a low fade rate, performance di®erence between SBJD and SBJD/S is smaller in the multipath channel than the memoryless channel. At a high fade rate (1%), the improvement of SBJD over SBJD/S is also larger in a multipath channel, which indicates that the diversity gain is more evident if the channel model is more accurate.

6.4 Summary

In this chapter we ¯rst extend both the BMB and BMML results of previous two chapters to both frequency and time selective fading channels. Although optimal, these detectors appear to be computationally expensive due to the fact that only FA property is exploited in them. However, it shows that with FA, multiple users can be detected in a both time- and frequency-selective channel than just frequency-selective channels [29] or time-selective channels (Chapter 4 and 5). To overcome the computational complexity problem, we presented a subspace based joint blind detection (SBJD) scheme by exploiting both FA and CS properties of the signals. Simulations demonstrate that SBJD can be applied e®ectively in

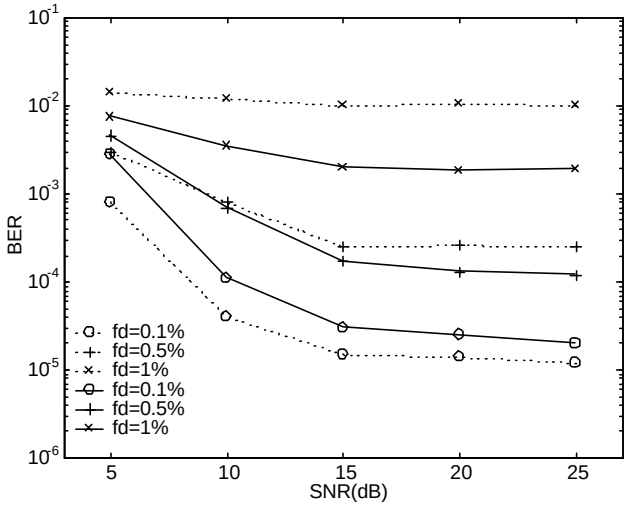

Figure 6.5: The performance of SBJD (solid lines) and SBJD/S (dashed lines) for two equipower users in a time-varying multipath channel at di®erent fade rates and SNRs. N = 20 and m = 4.

detecting cochannel asynchronous PSK signals in time-varying multipath fading channels. It also shows that SBJD preforms better in a frequency and time-selective fading channel than in the time-selective one due to an implicit time diversity gain, a well known result in a single user detection.

Chapter 7

Conclusions

This thesis has focused on the development and analysis of advanced multiple signal detection techniques for mobile communication systems. It has been motivated by the continuous demand for more and faster wireless services under relative limited bandwidth resources. The key is then to provide such services while maintaining the spectrum bandwidth specified by the regulatory organizations. The proposed techniques in this work present solutions to this problem in digital cellular networks by using antenna arrays and novel signal processing techniques. The spectral efficiency is achieved by developing techniques that use reduced or no training sequences (and hence are blind), as described in Chapter 1.

In a typical wireless network, the transmitted signal is impeded by three major sources of impairments: 1) intersymbol interference caused by multiple delayed paths in the transmission, 2) cochannel interferences caused by inter- and intra-cell users that share the same frequency and time slot as the desired user and 3) Doppler fading caused by user's mobility. These channel impairments together with the lack of sufficient training sequences present a most challenging signal detection problem in wireless communication systems. In this thesis, we develop techniques that are able to detect blindly multiple cochannel users in a time-varying multipath channel by using an antenna array of arbitrary geometry and number of sensing elements and exploiting the finite alphabet property of the digitally modulated signals.

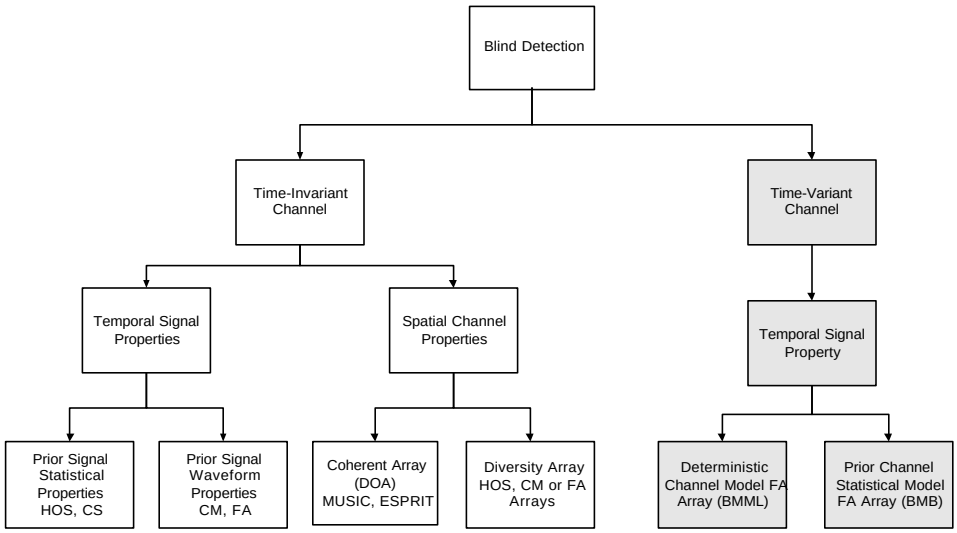

Figure 7.1: Blind Multiuser Detectors Review. Gray colored boxes represent the contribution of this thesis and white colored boxes are the results by other researchers.

7.1 Thesis summary

To better summarize the thesis work, Figure 7.1 shows the perspective of blind multiuser detection methods, with gray colored ones representing the contribution of the thesis:

- To the best of our knowledge, we are the first and only one to solve the blind multiuser detection problem in a time-varying multipath channel.
- To the best of our knowledge, we are also the first to apply the diversity reception technique to the above mentioned detection problem and the channel condition.
- Also to the best of our knowledge, we are the first to apply the Bayes and maximum likelihood criteria to solve the problem, which results in the statistical and deterministic based detection algorithms.

In Chapter 3, we developed two general blind multiuser detection techniques based on the channel models described in Chapter 2. We focused on the development of blind techniques in the time-varying but frequency non-selective channel model in Chapter 4 and 5 and generalized them to the time- and frequency-selective channel model in Chapter 6

in which we also developed an alternative detection algorithm with reduced computational complexity.

We developed the blind multiuser maximum likelihood (BMML) and blind multiuser Bayes (BMB) detection algorithms in Chapter 4 and 5, respectively. In BMML, the time-variation of the channel was modeled as a low order polynomial, whose order was determined by the fade rate of the channel and the observation duration (data window size). The detection is performed by alternating minimizations for jointly estimating the channel and data matrix. Both linear and nonlinear digital modulations are considered. PSK and CPM modulations, as examples of linear and nonlinear modulations, were studied in detail. We showed that cochannel signals can be identified up to a scalar without the training sequences under some easily satisfied conditions for the channel and data matrix. While symbol rate processing is sufficient for linear modulations, oversampling is required for nonlinear modulations.

In the blind multiuser Bayes (BMB) detection scheme, a priori information on the channel statistics is assumed. BMB is a VA type detector, whose state set is determined by the channel memory as well as the number of users and the signal constellation, contrary to the single user detections, where the channel memory solely determines the state set. The study also shows that for optimal detection the state set is as long as the block size, which limits its usefulness in practice. However, for slightly sub-optimal detection, the state set of VA can be reduced drastically. To reduce its complexity further, several sub-optimal versions of BMB have also been proposed. The analysis shows that some variant has a complexity which is lower than that of BMML. Unlike the BMML, BMB has a fixed complexity load since it is noniterative. Signal detectability is also studied fully in BMB. It shows that, with probability one, all solutions are equivalent if channel statistics satisfy some simple conditions and data possess some simple structures.

The performance of BMML and BMB detectors are analyzed numerically in this thesis. It shows that both classes of detectors work very effectively in time-varying fading channels. BMB has a better performance than BMML although some variants of BMB have a higher computational load. For both, the number of antennas does not impose a limit on the number of signals that can be detected, however, performance improves when more antennas are used due to diversity gain.

In Chapter 6, BMB and BMML are extended to the time-varying multipath channels. It shows that FA is a very powerful signal property with which multiple cochannel users can

be identi¯ed in a both time- and frequency-selective communication channel, which has an important theoretical value. However, high computational complexities of these detectors may o®set their practical value in some applications. A subspace multiuser detector, based on CS for equalization and FA for detection, was developed to overcome the computational complexity. Channel lengths and the number of signals can be identi¯ed e®ectively. The proposed approach was evaluated numerically and found that the detector works better in a time- and frequency-selective channel than in a time-selective-only channel due to an implicit time diversity gain in the former.

7.2 Future Directions

This thesis has set forth a foundation for blind multiuser detection in mobile communication channels by exploiting FA and CS properties. However, there are several issues that remain to be explored. In the following we discuss some important areas which require further study.

7.2.1 Performance analysis

One important task is to analyze the performances of BMML and BMB detectors. We have evaluated the performance of proposed algorithms via numerical simulations. However, it is also important to obtain analytic performance results. If the channel is time-invariant, some analytic results [30] have obtained by assuming the channel is known, as in the non-blind detection case. However, treating the channel as known to obtain an upper bound on the error performance is not too useful in BMML since the errors are predominated by the lack of convergence and less by the noise, which suggests some new approach is required. For BMB, a direct application of standard transfer function techniques [90] requires a large amount of computation, which makes the approach unattractive. This also suggests some new approach is needed to compute the error performance e±ciently.

7.2.2 Macro diversity reception

One important direction for future research is to improve the performance for users near the cell edge by applying blind multiuser detection. (For example, in a circular cell of uniform user density, half of the users are distributed outside 0.7 radius of the cell.) We can achieve

this by adopting the detection schemes developed in the thesis in conjunction with macro diversity reception. Macro diversity also provides a natural way of hando®, that is, the central processing unit monitors the signal strength (power of the channel gain estimate) and decides when to switch a user to another cell. E®ective hando® will not only improve system's performance but also reduce the system's load since when a signal is very weak in a cell it can be regarded as absent from that cell and therefore the processing power can be saved by not including it in detection. A major di®erence between the micro-diversity (antennas are located at center of one cell) and macro-diversity (antennas are distributed in several cells) is that not all users are seen by all the antennas in the latter. This results in a new problem and perhaps requires a new solution, to e®ectively determine users seen by each antenna and jointly detect users seen by more than one antenna. In addition, signals received from di®erent base stations will have to be aligned and sent to a central location for signal processing which requires the cooperation of all base stations and perhaps additional hardware cost.

7.2.3 More e±cient variant of algorithms

As mentioned in Chapter 1, one of the major disadvantages of joint detection is the computational complexity. Unfortunately, for most of the existing joint detection algorithms, the complexity grows exponentially with the number of users. Although in this thesis attempts have been made to reduce the complexity while maintaining performance at a reasonable level, there is still room for improvement. For example, more e®ective blind initialization procedures are yet to be found for BMML detections and more e®ective state reduction techniques, such M-algorithm, are to be explored for BMB. This may be an important future research direction.

7.2.4 Application to existing systems

Perhaps one of the most important future directions is to apply blind detection techniques to all the existing wireless communication systems, since they are consistent with most of them. However, since access standards are di®erent, there will be a di®erent solution to each access standard, depending on the access method, modulation, symbol rate, Doppler fade rate and delay spread. Among the existing cellular telephone systems, IS-136, GSM and IS-95 [102], are the most popular ones.

IS-136 is a TDMA system with ¼=4-shift di®erentially encoded phase shift keying (DQPSK) modulation, a special case of PSK modulation, and a square-root raised cosine pulse shape of roll-o® factor ® = 0:35. The symbol period T = 41:6 ¹secs and the slot time 6.67 ms, or block size 324 symbols. Due to the long duration of a slot, the channel may change signi¯cantly over a slot, depending on the mobile speed. Therefore the time-variation of the channel can not be ignored. On the other hand, the long symbol period makes the channel equalization a relative simple task, since the delay spread is in the order of 10 ¹s, and therefore a equalizer of up to 1 symbol period is su±cient. Thus blind methods developed in Chapter 6, or even that in Chapter 4, can be directly applied to identify multiple cochannel users. In fact, the numerical simulation study presented in Chapter 6 is very similar to this multipath channel environment.

GSM is another TDMA system with Gaussian-¯ltered minimum-shift keying, a special case of CPM modulation. (Although GSM contains a mid amble, for the cochannel GSM signals, this information is hard to extract due to asynchronous and multipath transmission and therefore blind detection can be used here to overcome this problem.) The symbol period in this case is T = 3:7¹secs and slot time 577 ¹s, or block size of about 156 bits. For this channel environment, the time-variation of the channel over each time slot is not signi¯cant so that linear channel approximation or even static channel model is su±cient. However, a shorter symbol period may corresponds a longer delay spread, in terms of symbols. For example, a delay spread of 10 ¹s corresponds to more than 2 symbols in delay. Hence it is necessary to apply the detection technique of Chapter 6 for channel equalization and signal detection.

Exploration of blind techniques in a CDMA system is also interesting. In a CDMA system, multiple users share the same frequency band at the same time, but with di®erent orthogonal spreading codes. In practice, however, these codes are only quasi-orthogonal due to multipath and other channel impairments. To improve performance, some spreading codes are utilized as training sequences in the systems. Blind techniques might be used here to reduce or eliminate the need for these training codes so that extra codes can be saved to transmit data, resulting in a higher data rate.

More recently, wireless LAN (WLAN) has becoming increasingly popular and one of the key performance index is the data transmission rate since one application is the wireless digital video transmission. Undoubtedly, blind techniques can be used in this area as well because they are standard independent and bandwidth e±cient.

Hence, blind techniques proposed in this thesis are directly applicable to the existing cellular based wireless communication systems as well as future systems to support multiple cochannel users and hence reduce cochannel interference with reduced or no training sequences. At present, these blind techniques have been tailored to each particular system. However, to fully take advantage of antenna arrays processing, new air-interface standards are needed to allow more e±cient space-time processing with increased spectral e±ciency at a moderate system (computational) complexity.

Appendix A

Alternative proof of Equation 3.16

Given

$$p(x_j|a_j;S) = \frac{1}{(2\pi\sigma_w^2)^N}\exp\left(-\frac{|x_j - \sum_{i=1}^{d} a_{ji}D(s_i)|^2}{2\sigma_w^2}\right)$$

and

$$p(a_j) = \prod_{i=1}^{\infty} p(a_{ji})$$

with

$$p(a_{ji}) = \frac{1}{(2\pi)^N\sqrt{R_{a_{ji}}}}\exp\left(-\frac{\sum_{i=1}^{d} a_{ji}^{*}R_{a_{ji}}^{-1}a_{ji}^{T}}{2}\right):$$

We want to calculate

$$I = \int_{a_j \in C^{dN}} p(x_j|a_j;S)p(a_j)da_j \qquad (A.1)$$

by direct integration, where $da_j = da_{j1}\cdots da_{jd}$ with $da_{ji} = da_{ji}(1)\cdots da_{ji}(N)$.

Let $a_{ji}D(s_i) = r_{ji}$, we have $da_{ji} = \frac{1}{|D(s_i)|}dr_{ji}$ and $R_{r_{ji}} = \frac{1}{2}E[r_{ji}^T r_{ji}^{*}] = D^T(s_i)R_{a_{ji}}D(s_i^{*})$. We then have $\sqrt{R_{a_{ji}}}^{-1/2}|D(s_i)| = \sqrt{R_{r_{ji}}}^{-1/2}$ and $a_{ji}^{*}D(s_i^{*})D^{-1}(s_i^{*})R_{a_{ji}}^{-1}D^{-T}(s_i)D^T(s_i)a_{ji}^T = r_{ji}^{*}R_{r_{ji}}^{-1}r_{ji}^T$. Thus

$$p(a_{ji})da_{ji} = \frac{1}{(2\pi)^N\sqrt{R_{r_{ji}}}}\exp\left(-\frac{\sum_{i=1}^{d} r_{ji}^{*}R_{r_{ji}}^{-1}r_{ji}^T}{2}\right)dr_{ji} = p(r_{ji})dr_{ji}$$

and

$$p(x_j|a_j;S) = \frac{1}{(2\pi\sigma_w^2)^N}\exp\left(-\frac{|x_j - \sum_{i=1}^{d} r_{ji}|^2}{2\sigma_w^2}\right) = p(x_j|r_j;S):$$

Then Equation (A.1) becomes

$$
\begin{aligned}
I &= \int_{r_{ji}\in C^N} \frac{1}{(2\pi\sigma_w^2)^N} \exp\left(-\frac{\|x_j - \sum_{i=1}^{d} r_{ji}\|^2}{2\sigma_w^2}\right) \\
&\quad \prod_{i=1}^{d} \frac{1}{(2\pi)^N \overline{|R_{r_{ji}}|}} \exp\left(-\frac{r_{ji}^* R_{r_{ji}}^{-1} r_{ji}^T}{2}\right) dr_{ji} \\
&= \frac{1}{(2\pi\sigma_w^2)^N} I_2
\end{aligned} \tag{A.2}
$$

with

$$
I_2 = \frac{1}{(2\pi)^{Nd}\prod_{i=1}^{d}\overline{|R_{r_{ji}}|}} \int_{r_{ji}\in C^N} \exp\left(-\frac{\|x_j - \sum_{i=1}^{d} r_{ji}\|^2}{2\sigma_w^2} - \frac{\sum_{i=1}^{d} r_{ji}^* R_{r_{ji}}^{-1} r_{ji}^T}{2}\right) \prod_{i=1}^{d} dr_{ji}. \tag{A.3}
$$

The exponent of (A.3)

$$
\begin{aligned}
&-\frac{\|x_j - \sum_{i=1}^{d} r_{ji}\|^2}{2\sigma_w^2} - \frac{\sum_{i=1}^{d} r_{ji}^* R_{r_{ji}}^{-1} r_{ji}^T}{2} \\
&= -\frac{x_j^* x_j^T}{2\sigma_w^2} + \frac{y_j^* r_j^T}{2\sigma_w^2} + \frac{r_j^* y_j^T}{2\sigma_w^2} - \frac{r_j^* \left(1_I + R_{r_j}^{-1}\sigma_w^2\right) r_j^T}{2\sigma_w^2} \\
&= -\frac{x_j^* x_j^T}{2\sigma_w^2} + \frac{y_j^* r_j^T}{2\sigma_w^2} + \frac{r_j^* y_j^T}{2\sigma_w^2} - \frac{r_j^* U^H U r_j^T}{2\sigma_w^2} \\
&= -\frac{x_j^* x_j^T}{2\sigma_w^2} + \frac{y_j^* U^{-1} v_j^T}{2\sigma_w^2} + \frac{v_j^* U^{-H} y_j^T}{2\sigma_w^2} - \frac{v_j^* v_j^T}{2\sigma_w^2} \\
&= -\frac{x_j^* x_j^T}{2\sigma_w^2} - \frac{v_j^* - y_j^* U^{-1}}{2\sigma_w^2}\frac{v_j^T - U^{-H} y_j^T}{} + \frac{y_j^* U^{-1} U^{-H} y_j^T}{2\sigma_w^2}.
\end{aligned} \tag{A.4}
$$

where y_j is a row vector by repeating x_j d times, i.e., $y_j = (x_j; x_j; \ldots; x_j)$, $r_j = (r_{j1}; \ldots; r_{jd})$; 1_I is a $dN \times dN$ matrix with identical $d \times d$ blocks, each of $N \times N$ identity matrix. R_{r_j} is de-ned as a block diagonal matrix with $R_{r_{ji}}$ as its ith diagonal block, i.e., $R_{r_j} = \mathrm{diag}(R_{r_{j1}}; \ldots; R_{r_{j2}})$: Matrix $1_I + R_{r_j}^{-1}\sigma_w^2$ is a symmetrical and invertible and therefore can be written as $U^H U$: After changing of variables, $U r_j^T = v_j^T$, and $dv_j = \frac{1}{|U|} dv_j$, and integrating (A.3), we have

$$
I_2 = \frac{(\pi\sigma_w^2)^{Nd}}{\prod_{i=1}^{d}\overline{|R_{r_{ji}}|}\,|U|} \exp\left(-\frac{x_j^* x_j^T}{2\sigma_w^2} + \frac{y_j^* \left(U^H U\right)^{-1} y_j^T}{2\sigma_w^2}\right) \tag{A.5}
$$

By matrix inversion lemma, we have

$$\begin{aligned}\left(U^H U\right)^{-1} &= \left(1_I + R_{r_j} \sigma_w^{-2}\right)^{-1} \\ &= \frac{R_{r_j}}{\sigma_w^2} - \frac{R_{r_j}}{\sigma_w^2}\left(I + \sum_{i=1}^{N} \frac{R_{r_{ji}}}{\sigma_w^2}\right)^{-1} \frac{R_{r_j}}{\sigma_w^2}\end{aligned}$$

The exponent of (A.5) is therefore

$$\begin{aligned}&-\frac{1}{2\sigma_w^2} x_j^* \left(I - \sum_{i=1}^{N} \frac{R_{r_j}}{\sigma_w^2} + \frac{R_{r_j}}{\sigma_w^2}\left(I + \sum_{i=1}^{N} \frac{R_{r_{ji}}}{\sigma_w^2}\right)^{-1} \frac{R_{r_j}}{\sigma_w^2}\right) x_j^T \\ &= -\frac{1}{2\sigma_w^2} x_j^* \left(I - \left(I + \sum_{i=1}^{N} \frac{R_{r_{ji}}}{\sigma_w^2}\right)^{-1}\right) x_j^T \\ &= -\frac{1}{2\sigma_w^2} x_j^* \left(I + \sum_{i=1}^{N} \frac{R_{r_{ji}}}{\sigma_w^2}\right)^{-1} x_j^T = -\frac{1}{2} x_j^* \left(I\sigma_w^2 + \sum_{i=1}^{N} R_{r_{ji}}\right)^{-1} x_j^T \qquad \text{(A.6)}\end{aligned}$$

by applying the matrix inversion lemma twice. From (A.2) and (A.5), we see the linear term N is

$$\begin{aligned}N &= \frac{1}{(2\pi\sigma_w^2)^N} \frac{\sigma_w^{2N_d}}{\prod_{i=1}^{d} \bar{R}_{r_{ji}} |U|} \\ &= \frac{\sigma_w^{2N_d}}{(2\pi\sigma_w^2)^N} \frac{1}{\bar{R}_{r_j} |U|}.\end{aligned}$$

Since

$$\begin{aligned}|U| &= \left|\bar{U}^H \bar{U}^{-1/2}\right| \\ &= \left|\bar{1}_I + R_{r_j} \sigma_w^{-2}\right| \left|\bar{U}^{-1/2}\right|;\end{aligned}$$

we have

$$\begin{aligned}N &= \frac{1}{(2\pi\sigma_w^2)^N} \frac{\left|\bar{U}^{-1/2}\right|}{\left|\bar{R}_{r_j}\sigma_w^{-2}\right| \left|\bar{1}_I + R_{r_j}\sigma_w^{-2}\right| \left|\bar{U}^{-1/2}\right|} \\ &= \frac{1}{(2\pi\sigma_w^2)^N} \frac{1}{\left|\bar{I} + 1_I \frac{R_{r_j}}{\sigma_w^2}\right|}. \qquad \text{(A.7)}\end{aligned}$$

Expanding $\mathbf{I} + \mathbf{1}_l \frac{\mathbf{R}_{\mathbf{r}_{ji}}}{\sigma_w^2}$ as

$$\begin{bmatrix} \frac{\mathbf{R}_{\mathbf{r}_{j1}}}{\sigma_w^2} + \mathbf{I} & \frac{\mathbf{R}_{\mathbf{r}_{j2}}}{\sigma_w^2} & \cdots & \frac{\mathbf{R}_{\mathbf{r}_{jd}}}{\sigma_w^2} \\ \frac{\mathbf{R}_{\mathbf{r}_{j1}}}{\sigma_w^2} & \frac{\mathbf{R}_{\mathbf{r}_{j2}}}{\sigma_w^2} + \mathbf{I} & \cdots & \frac{\mathbf{R}_{\mathbf{r}_{jd}}}{\sigma_w^2} \\ \vdots & \vdots & \ddots & \\ \frac{\mathbf{R}_{\mathbf{r}_{j1}}}{\sigma_w^2} & \frac{\mathbf{R}_{\mathbf{r}_{j2}}}{\sigma_w^2} & & \frac{\mathbf{R}_{\mathbf{r}_{jd}}}{\sigma_w^2} + \mathbf{I} \end{bmatrix};$$

and subtract the first row blocks from all other ones, we obtain

$$\begin{bmatrix} \frac{\mathbf{R}_{\mathbf{r}_{j1}}}{\sigma_w^2} + \mathbf{I} & \frac{\mathbf{R}_{\mathbf{r}_{j2}}}{\sigma_w^2} & \cdots & \frac{\mathbf{R}_{\mathbf{r}_{jd}}}{\sigma_w^2} \\ -\mathbf{I} & \mathbf{I} & \cdots & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots \\ -\mathbf{I} & \mathbf{0} & & \mathbf{I} \end{bmatrix}$$

whose value can be easily obtained by using the block determinant formula [82], which equals $\mathbf{I} + \sum_{i=1}^{d} \frac{\mathbf{R}_{\mathbf{r}_{ji}}}{\sigma_w^2}$. Thus

$$\begin{aligned} N &= \frac{1}{(2\pi\sigma_w^2)^N} \frac{1}{\mathbf{I} + \sum_{i=1}^{d} \frac{\mathbf{R}_{\mathbf{r}_{ji}}}{\sigma_w^2}} \\ &= \frac{1}{(2\pi)^N} \frac{1}{\mathbf{I}\sigma_w^2 + \sum_{i=1}^{d} \mathbf{R}_{\mathbf{r}_{ji}}}; \end{aligned} \tag{A.8}$$

Combining with (A.5), we have

$$I = \frac{1}{(2\pi)^N} \frac{1}{\mathbf{I}\sigma_w^2 + \sum_{i=1}^{d} \mathbf{R}_{\mathbf{r}_{ji}}} \exp\left\{ -\frac{1}{2} \mathbf{x}_j^T \left(\mathbf{I}\sigma_w^2 + \sum_{i=1}^{d} \mathbf{R}_{\mathbf{r}_{ji}} \right)^{-1} \mathbf{x}_j \right\},$$

which is identical to (3.16).

Appendix B

Proof of lemmas

B.1

Lemma 4.2.4: The rows of S_k, (k = 0:::p) are linearly independent under the DS condition.

Proof. Suppose there exist non-zero row vectors $\bar{}_k$, (k = 0:::p) that satisfy

$$\sum_{k=0} \bar{}_k S_k = 0: \tag{B.1}$$

Suppose also that some signal vector s in S appears (p + 1) times, at $n = n_0; n_1; :::; n_p$. We have, from (B.1),

$$\sum_{k=0} n^k \bar{}_k s = 0;\ n = n_0; n_1; :::; n_p$$

which can be also written as

$$\begin{pmatrix} 1 & n_0 & \cdots & n_0^p \\ 1 & n_1 & \cdots & n_1^p \\ \vdots & \vdots & \ddots & \vdots \\ 1 & n_p & \cdots & n_p^p \end{pmatrix} \begin{pmatrix} \bar{}_0 s \\ \bar{}_1 s \\ \vdots \\ \bar{}_p s \end{pmatrix} = 0: \tag{B.2}$$

The matrix in (B.2) is a Vandermonde matrix and therefore is invertible, resulting in

$$\bar{}_k s = 0; \qquad k = 0:::p: \tag{B.3}$$

Since (B.3) holds for any s; if we can ¯nd d or more linearly independent s vectors from the columns of S then we would have $\bar{}_k = 0$; k = 0:::p: However, this would contradict the assumption and therefore prove lemma 4.2.4.

We now verify that such a collection of d or more linearly independent column vectors, each appearing (p + 1) times in S, indeed exists. Under the DS condition, S_{sp_2} is such a collection and has a full row rank. To see this, we need only to subtract the $(j-1)$-th from the jth row of S_{sp_2}; starting from the last one, and observe that the resulting matrix can be converted to a lower triangular form with nonzero diagonal elements. ■

B.2

Lemma 4.2.5: If $\hat{S} = T\hat{S}$ and $\hat{S}$ satisfies the DS condition, then T is an equivalence transformation.

Proof. By the DS condition, $\hat{S}$ contains S_{sp}. Re-order the columns of $\hat{S}$ by post-multiplying $\hat{S} = T\hat{S}$ by a transformation P_1 such that $\hat{S}P_1 = (S_{sp}; S_r)$ where S_r is the remaining part of $\hat{S}$ after the re-ordering. It is then sufficient to show that T is an equivalence transformation in

$$\hat{S}_1 = TS_{sp} \tag{B.4}$$

where $\hat{S}_1$ is the corresponding sub-matrix of $\hat{S}$ after the re-ordering. Expanding (B.4) for the first row of $\hat{S}_1$ results in a set of linear equations, with the first one and a particular $(Q-1)$ equations listed below:

$$\begin{cases} q_1t_1 + \cdots + q_1t_j + \cdots + q_1t_d = q_0 \\ q_1t_1 + \cdots + q_2t_j + \cdots + q_1t_d = q_1 \\ \qquad\qquad \cdots \\ q_1t_1 + \cdots + q_Qt_j + \cdots + q_1t_d = q_{Q-1} \end{cases} ; \; j = 1; \ldots; d \tag{B.5}$$

where $q_i \in Q$ and $(t_1; t_2; \ldots; t_d)$ is the first row of T. If some of q_i in (B.5) are the same, we have

$$t_j = 0 \text{ and } q_1 \sum_{n \neq j} t_n = q_0; \text{ if } q_i = q_k \text{ for } i \neq k; \; j = 1; \ldots; d. \tag{B.6}$$

If all q_i are different we must have $\sum_{i=0}^{Q-1} q_i = 0$ since for PSK modulation the sum of all Q different data symbols is zero. Thus we have

$$t_j = q_0/q_1 \text{ and } \sum_{k \neq j} t_k = 0; \text{ if } \sum_{i=0}^{Q-1} q_i = 0; \; j = 1; \ldots; d. \tag{B.7}$$

It is easy to see from (B.6) and (B.7) that $t_j = 0$ or $e^{j\theta_0}/e^{j\theta_1}$. We now show that only one $t_j = e^{j\theta_0}/e^{j\theta_1}$ and the rest are zeros. Suppose, without losing generality, that $t_j = e^{j\theta_0}/e^{j\theta_1}$ for $j = 1;:::;l$ and $t_j = 0$ for $j = l+1;:::;d$. Then from the first equation of (B.5) we have $le^{j\theta_0} = e^{j\theta_0}) l = 1$ and therefore $t_j = e^{j\theta_0}/e^{j\theta_1}$ for some j and $t_i = 0$ for $i \neq j$.

Identical proof is given for other rows of T and since T is invertible, we have T is an equivalence transformation matrix, i.e., $T = T_E$.

Although the proof given is for PSK, it can be generalized to other linear modulation schemes satisfying the similar DS condition. ■

B.3

Lemma 4.3.2: If two solutions Φ and Ψ are related by an invertible matrix T and either one satisfies the DS condition then T is an equivalence transformation.

Proof. First rearrange the columns of Φ and Ψ in (4.42) by moving P columns at a time along the symbol time boundaries (or symbol columns, defined in Section 4.3.1) in such a way that $\Phi \to (Y_{DS_1}; Y_r)$. It is then sufficient to show that T is an equivalence transformation in

$$\Psi_1 = T Y_{DS_1} \tag{B.8}$$

where Ψ_1 is the corresponding sub-matrix of Ψ after the column re-ordering.

Next we show that T in (B.8) is an equivalence transformation by induction on the number of signals d. When d = 2, (B.8) becomes

$$\begin{pmatrix} \xi_{1;n_1+s_1(n_1)}(i) & \xi_{1;n_2+s_1(n_2)}(i) \\ \xi_{2;n_1+s_2(n_1)}(i) & \xi_{2;n_2+s_2(n_2)}(i) \end{pmatrix} = \begin{pmatrix} t_{1;1} & t_{1;2} \\ t_{2;1} & t_{2;2} \end{pmatrix} \begin{pmatrix} \eta_{1;n_1+q_1}(i) & \eta_{1;n_2+q_2}(i) \\ \eta_{2;n_1+q_1}(i) & \eta_{2;n_2+q_2}(i) \end{pmatrix}:$$

Expanding the top left component of the above equations with $i = 0;1$ and 2 and letting $s_1(n_1) = q_1$;

$$\xi_{1;n_1} = t_{1;1}\eta_{1;n_1} + t_{1;2}\eta_{2;n_1} \tag{B.9a}$$

$$\xi_{1;n_1+q_1}(1) = t_{1;1}\eta_{1;n_1+q_1}(1) + t_{1;2}\eta_{2;n_1+q_1}(1) \tag{B.9b}$$

$$\xi_{1;n_1+q_1}(2) = t_{1;1}\eta_{1;n_1+q_1}(2) + t_{1;2}\eta_{2;n_1+q_1}(2): \tag{B.9c}$$

$(t_{1;1}; t_{1;2})$ can be uniquely determined by (B.9a) and (B.9b). We want to show that $(t_{1;1}; t_{1;2})$ obtained is inconsistent with the (B.9c) if $q_1 \neq (q_1; q_1)$. Substituting $(t_{1;1}; t_{1;2})$ obtained

from (B.9a) and (B.9b) into (B.9c), the right-hand side becomes

$$\sum_{1;n_1} \frac{[\pm_{q_1}(1) \; ¡ \; \pm_{\phi_1}(1)]\pm_{\phi_1}(2) + [\pm_{\phi_1}(1) \; ¡ \; \pm_{\phi_1}(1)]\pm_{q_1}(2)}{\pm_{q_1}(1) \; ¡ \; \pm_{\phi_1}(1)} = \sum_{1;n_1} \frac{b\pm_{\phi_1}(2) + c\pm_{q_1}(2)}{a} \tag{B.10}$$

which is required to have a unit norm since the left-hand side of (B.9c) does. It is clear that $(b + c)=a$ has a unit norm since vectors a, b and c form a triangle. It is easily seen that (B.10) will have a unit norm if and only if the inner products are the same,

$$(b; c) = (b\pm_{\phi_1}(2); c\pm_{q_1}(2)) \tag{B.11}$$

which is true only if

$$\pm_{\phi_1}(2) = \pm_{q_1}(2) \tag{B.12}$$

if $(b; c) \neq 0$. If $(b; c) = 0$, we can show geometrically that it will also result in (B.12). Thus (B.11) holds if and only if (B.12) does. Similarly we can show $\pm_{\phi_1}(¿) = \pm_{q_1}(¿)$ for $¿ = 1; :::; P \; ¡ \; 1$, or equivalently,

$$h(\phi_1 \; ¡ \; q_1)Á(¿) = 2l; \tag{B.13}$$

where l is an integer. For most of the practical cases, however, (B.13) is not satis¯ed for every $¿ \in (1; :::; P \; ¡ \; 1)$ and therefore (B.11) does not hold and (B.9a) and (B.9c) are inconsistent if $\phi_1 \in (q_1; \phi_1)$. If $\phi_1 = q_1$ we have $t_{1;1} = 0$, $t_{1;2} = \sum_{1;n_1}=\sum_{2;n_1}$. Likewise $t_{1;1} = \sum_{1;n_1}=\sum_{1;n_1}$ and $t_{1;2} = 0$ if $\phi_1 = \phi_1$. Similarly we can calculate $t_{2;1}$ and $t_{2;2}$. The results are summarized as follows:

$$T = \begin{cases} \begin{pmatrix} 0 & \sum_{1;n_1}=\sum_{2;n_1} \\ \sum_{2;n_1}=\sum_{1;n_1} & 0 \end{pmatrix}; & \text{if } \mathbf{e}(n_1) = \begin{pmatrix} q_1 \\ \phi_1 \end{pmatrix} \\ \begin{pmatrix} \sum_{1;n_1}=\sum_{1;n_1} & 0 \\ 0 & \sum_{2;n_1}=\sum_{2;n_1} \end{pmatrix}; & \text{if } \mathbf{e}(n_1) = \begin{pmatrix} \phi_1 \\ q_1 \end{pmatrix} \end{cases} \tag{B.14}$$

Other $\mathbf{e}(n_1)$ values are not allowed since they are not consistent with the full rank property of T. Thus $T = E_p$ for $d = 2$. Assume for $d = k$ that $T = E_p$ and thus $s_i(n_j) \in (q_j; \phi_j)$. For $d = k + 1$, the ¯rst row of (B.8) is

$$\begin{aligned} & \left[\sum_{1;n_1}\pm_{s_1(n_1)}(¿); \sum_{1;n_2}\pm_{s_1(n_2)}(¿); ¢¢¢ ; \sum_{1;n_{k+1}}\pm_{s_1(n_{k+1})}(¿)\right] \\ = \; & (t_{1;1}; ¢¢¢ ; t_{1;k+1}) \begin{pmatrix} \sum_{1;n_1}\pm_{\phi_1}(¿) & \sum_{1;n_2}\pm_{q_2}(¿) & ¢¢¢ & \sum_{1;n_{k+1}}\pm_{q_{k+1}}(¿) \\ \sum_{2;n_1}\pm_{q_1}(¿) & \sum_{2;n_2}\pm_{\phi_2}(¿) & ¢¢¢ & \sum_{2;n_{k+1}}\pm_{q_{k+1}}(¿) \\ \vdots & \vdots & \ddots & \vdots \\ \sum_{k+1;n_1}\pm_{q_1}(¿) & \sum_{k+1;n_2}\pm_{q_2}(¿) & ¢¢¢ & \sum_{k+1;n_{k+1}}\pm_{\phi_{k+1}}(¿) \end{pmatrix} \end{aligned} \tag{B.15}$$

Examine the first one. As in the $d = 2$ case we can show $s_1(n_1) \geq (q_1, q_1)$. If $s_1(n_1) = q_1$; $t_{1;1} = 0$. In general, if any of $s_1(n_i) = q_i$ for $i = 1; :::; k+1$,) $t_{1;i} = 0$ and (B.15) reduces to the case for $d = k$ which by assumption is true. If $s_1(n_i) = q_i$ for $i = 1; :::; k+1$, we have

$$t_{1;i} = \alpha_{1;n_i} / \lambda_{i;n_i} \quad i = 1; :::; k+1 \tag{B.16}$$

and

$$\begin{pmatrix} 0 & \lambda_{2;n_1} & \cdots & \lambda_{k+1;n_1} \\ \lambda_{1;n_2} & 0 & \cdots & \lambda_{k+1;n_2} \\ \vdots & \ddots & \ddots & \vdots \\ \lambda_{1;n_{k+1}} & \lambda_{2;n_{k+1}} & \ddots & 0 \end{pmatrix} \begin{pmatrix} t_{1;1} \\ t_{1;2} \\ \vdots \\ t_{1;k+1} \end{pmatrix} = 0$$

$$\text{or } \Delta t = 0: \tag{B.17}$$

By the DS condition, $\det \Delta \neq 0$ and we have $t = 0$ but this contradicts (B.16). Thus $(t_{1;1}; \cdots; t_{1;k+1})$ obtained from (B.16) are inconsistent with (B.17) and therefore we must have $t_{1;i} = 0$ for at least one $i \in (1; :::; k+1)$ which reduces to the case for $d = k$. The identical discussion is given for other rows of T and the induction completes. Thus for any d, $T = E_p$.

■

Appendix C

Eigenvalues of $R(\hat{\theta})^{-1}R(S)$ are positive real numbers

From the properties of the covariance matrix, both $R(\hat{\theta})$ and $R(S)$ are positive definite Hermitian matrix (with non-vanishing noise variance). Hence, there exists an unitary transformation U such that $U^H R(\hat{\theta})U = D^2$, where D is a diagonal matrix with positive real diagonal elements. Let $Q = UD^{-1}$, we have

$$Q^H R(\hat{\theta})Q = I. \tag{C.1}$$

Since $Q^H R(\hat{\theta})Q$ is a positive definite Hermitian, there exist a unitary transformation T such that

$$\begin{aligned} & T^H Q^H R(S) Q T \\ = \; & P^H R(S) P \\ = \; & \Lambda \end{aligned}$$

where $P = QT$ and Λ is diagonal matrix with positive real elements. From (C.1), we have

$$P^H R(\hat{\theta})P = I \;\Rightarrow\; R(\hat{\theta})P = (P^{-1})^H. \tag{C.2}$$

Using (C.2) in (C.2), we have $R(\hat{\theta})^{-1}R(\hat{\theta})P = P\Lambda$. Hence $R(\hat{\theta})^{-1}R(S)$ has positive real eigenvalues as the diagonal elements of Λ and the corresponding eigenvectors as columns of P.

Appendix D

Complexity estimate for BMML and BMB detectors

In the following, we consider linearly modulated signals only. We assume that conditional correlation matrix R(S) is the same for all antennas and the fade rate is the same for all users. To simplify the discussion, we also assume the transmission is continuous, that is, the computation for initialization is ignored.

- For BMML

The BBML detection algorithm iteratively estimates channel matrix $A_{aug}(S(N))$ and the signal matrix $S_{aug}(N)$ from (4.8) and (4.7), respectively. $\hat{A}_{aug}(S(N))$ can be obtained from solving the following set of linear equations

$$\left[\hat{S}_{aug}^H \hat{A}_{aug}^H\right]_k = X^H \qquad \text{(D.1)}$$

for given $\hat{A}_{aug}$. To solve (D.1), we can first compute the Householder or Givens QR decomposition, each requiring $2z^2(N - z/3)$ CMAs ([82], p. 223), then solve an equivalent triangular system of linear equations which needs $mz^2 + Nmz$ CMAs ([82], p. 88), where $z = d(p + 1)$. $\hat{S}_{aug}(N)$ is obtained from (4.7) with a fixed $\hat{A}_{aug}$ by enumeration. For each of the Q^d distinct signal vectors, we need to perform $m(z + 1)$ multiplications to get $\hat{A}_{aug}\hat{S}_{aug}(n)$. Thus, we need a total of $NmQ^d(z + 1)$ CMAs to get $\hat{A}_{aug}\hat{S}_{aug}(N)$. Assume BMML takes I iterations to converge. Then BMML requires

$$C_{BMML} = \left[2z^2(N - z/3) + mz^2 + Nmz + NmQ^d(z + 1)\right] I \qquad \text{(D.2)}$$

CMAs to detect a block. From this we can easily obtain the complexities for SBSD and SWFD detectors, which are described in Chapter 4. They are given, respectively, as

$$C_{SBSD} = C_{BMML} \tag{D.3}$$

and

$$C_{SWFD} = C_{BMML} N_{sub}=I \tag{D.4}$$

where N_{sub} is the subblock length and it takes one iteration to converge.

- For BMB

For BMB detectors, BMPSP, BMDF and MRDF, the most computational intensive parts are the computation of prediction ¯lters and the branch metric BM_n in (5.35). From (5.37), the prediction ¯lter f can be computed in each state transition from

$$R_{M\,i\,1} f = v_{M\,i\,1}: \tag{D.5}$$

First, it needs about $2d(M\ i\ 1)^2 Q^{d(M_{VA}\,i\,1)}$ and $2d(M\ i\ 1)Q^{dM_{VA}}$ CMAs to construct $R_{M\,i\,1}$ and $v_{M\,i\,1}$ from (3.14), respectively, at each step of the VA. Since $R_{M\,i\,1}$ is a positive de¯nite Hermitian, we can solve (D.5) by ¯rst performing Choleskey decomposition on $R_{M\,i\,1}$ ($(M\ i\ 1)^3=3$ CMAs) then solving two subsequent triangular system problems, each requires only $(M\ i\ 1)^2$ CMAs [82]. Metric computation for BM_n in (5.35) is straightforward once f is obtained. At each step of the VA, it requires about $m(M + 1)Q^{dM_{VA}}$ CMAs. Put these results together, we have, for a block of data,

$$C_{BMPSP} = \left[\frac{(M\ i\ 1)^3}{3} + 2(M\ i\ 1)^2(d + Q^d) + 2(M\ i\ 1)dQ^d + m(M + 1)Q^d\right] Q^{d(M_{VA}\,i\,1)} N \tag{D.6}$$

and

$$C_{BMDF} = \left[\frac{(M\ i\ 1)^3}{3} + 2(M\ i\ 1)^2(d + Q^d) + 2(M\ i\ 1)dQ^d + m(M + 1)Q^d\right] N: \tag{D.7}$$

Similarly, we obtain a CMA estimate for MRDF. As described in Chapter 5, MRDF ¯rst searches each of the $Q^{d\,i\,1}$ distinct signal vectors M_r times from $\mathbf{b}(n\ i\ 1)$; which requires on average a data block of length $Q^{d\,i\,1}M_r$ symbols by performing an inner product of desired

signal vector and a column of $\mathbf{S}(n-1)$; which needs d CMAs. Thus this step takes about $(Q^{d-1})^2 M_r d$ CMAs. The second step is to construct $R_{M_r,i-1}$ and $v_{M_r,i-1}$, (see (5.25)), which costs about $O(M_r)$ that can be ignored. The third step is to compute the prediction filter $\bar{f}$ which needs $(M_r^3/3 + 2M_r^2)Q^{d-1}$ CMAs. The fourth step is the metric computation, which takes $m(M_r + 2)Q^d$ CMAs. Together, we have, again for a block of data,

$$C_{MRDF} = Q^{d-1}\left[M_r d + 2M_r^2 + \frac{M_r^3}{3} + m(M_r + 2)Q\right] Q^{d-1} N. \qquad \text{(D.8)}$$

Appendix E

An upper bound on probability of satisfying DS property

The probability that S satisfies the data structure condition can be calculated as follows.

First, define A as the event that all distinct signals s_i in S_{sp} are selected, each pQ + 1 times, in N independent trials. The sought after probability is then P (A).

For ease of computation, we first calculate $P(\bar{A})$. Note $\bar{A} = \sum_{i=1}^{K_A} \bar{A}_i$, where $\bar{A}_i$ is the sub event that signal s_i is not selected pQ + 1 times and K_A is the number of distinct signals in S_{sp}, i.e., $K_A = (Q - 1)d + 1$. $P(\bar{A})$ can be found [100] as

$$P(\bar{A}) = P(\bigcup_{i=1}^{K_A} \bar{A}_i) = S_1 - S_2 + S_3 + \cdots \pm S_{K_A} \tag{E.1}$$

where $S_k = \sum_{i_1 < i_2 < \cdots < i_k} P(\bar{A}_{i_1} \cap \bar{A}_{i_2} \cap \cdots \cap \bar{A}_{i_k})$. As $\bar{A}_i$ is independent and equally likely, it follows that

$$S_k = \binom{K_A}{k} P(\bigcap_{i=1}^{k} \bar{A}_i) = \binom{K_A}{k} \sum_{i_1=0}^{pQ} \cdots \sum_{i_k=0}^{pQ} \frac{(K-k)^{N-i_1-\cdots-i_k}}{K^N} \frac{N!}{i_1! \cdots i_k! (N-i_1-\cdots-i_k)!} \tag{E.2}$$

where K is the total number of distinct signal vectors, $K = Q^{d-1}$. If K is not too large (E.1) is approximated by the first one or two terms. Since $S_k > S_{k'}$ for $k < k'$ we have $S_1 - S_2 + S_3 > P(\bar{A}) > S_1 - S_2$. Thus, an upper bound UB on the probability of not satisfying the DS condition is given by

$$1 - S_1 + S_2 - S_3. \tag{E.3}$$

The numerical result of (E.3) is plotted in Figure 4.2.

Appendix F

Proof of Equation 5.31

We want to show an upper bound on the probability that a random signal matrix S(M) of size d £ M is rank de¯cient. As stated in the main text, this bound can be expressed as the sum of two terms, one of which contains less than d distinct (within a phase factor) signal vectors s(n) and the other contains more than d ¡ 1 distinct vectors but they are linearly dependent. There are Q^{d-1} distinct vectors, so that the probability of forming S(M) by a particular set of k (< d) distinct signal vector is $< \left(\frac{k}{Q^{d-1}}\right)^M$. There are $\binom{Q^{d-1}}{k}$ sets of distinct vectors, so the probability of selecting any k distinct signal vectors in M trials is $P_1(k) < \binom{Q^{d-1}}{k}\left(\frac{k}{Q^{d-1}}\right)^M$. The probability of selecting a particular k (, d) distinct signal vectors are linearly dependent, in M trails, is $< \left(\frac{k}{Q^{d-1}}\right)$. Let C_k be the number of di®erent ways of selecting k distinct but linearly dependent signals in M trials, we have the probability of selecting any k distinct but linearly dependent signal vectors in M trials is $P_2(k) < C_k \left(\frac{k}{Q^{d-1}}\right)^M$. Thus the desired probability is bounded by

$$\sum_{k=1}^{d-1} P_1(k) + \sum_{k=d}^{Q^{d-1}} P_2(k) < \sum_{k=1}^{d-1} \left(Q^{d-1}\right)\left(\frac{k}{Q^{d-1}}\right)^M + \sum_{k=d}^{Q^{d-1}} C_k \left(\frac{k}{Q^{d-1}}\right)^M$$

which is (5.31).

Bibliography

[1] Strategies Group, \World Cellular and PCS Markets Study," Washington DC, 1997.

[2] Electronic Industries Association, EIA Interim Standard, IS-136, Rev 0, \Dual-Mode Mobile Station - Base Station Compatibility Standard," Washington, DC, 1995.

[3] European Telecommunications Standardization Institute. 1988. Group Special Mobile or Global System of Mobile Communication (GSM) Recommendation, ETSI Secretariat, Sophia Antipolis Cedex, France.

[4] C.E. Shannon, \A Mathematical Theory of Communication," Bell Syst. Tech. J., vol. 27, pp. 623-656, 1948.

[5] W.C.Y. Lee, \Smaller Cells for Greater Performance," IEEE Comm. Mag., pp. 19-30, Nov. 1991.

[6] S.J. Grant and J.K. Cavers, \Performance Enhancement Through Joint Detection of Cochannel Signals Using Diversity Arrays," IEEE Trans. Comm., vol. 46, no. 8, pp. 1038-1049, Aug. 1998.

[7] J.H. Winters, \Optimum Combining in Digital Mobile Radio with Cochannel Interference," IEEE J. Select. Area in Comm., vol. SAC-2, no. 4, pp. 528-539, July 1984.

[8] H. Murata and S. Yoshida, \Maximum-Likelihood Sequence Estimation for Coded Modulation in the Presence of Co-Channel Interference and Intersymbol Interference," in Proc. IEEE Vehicular Tech. Conf. pp. 701-705, 1996.

[9] Motorola Inc., \ReFLEX Protocol Speci¯cation Document," Motorola Internal Document Rev. 2.7, Dec. 17, 1999.

[10] Simon Haykin, Adaptive Filter Theory, 2nd Ed, Prentice-Hall, New York, 1986.

[11] Y. Sato, \A Method of Self-Recovering Equalization for Multilevel Amplitude-Modulation Systems," IEEE Trans. Comm. vol. COM-23, pp. 679-682, June 1975.

[12] D.M. Godard, \Self-Recovering Equalization and Carrier Tracking in Two-Dimensional Data Communications Systems," IEEE Trans. Comm. vol. COM-28, pp. 1867-1875, Nov. 1980

[13] A. Benveniste and M. Goursat, \Blind Equalizers," IEEE Trans. Comm. COM-32, pp. 871-883, 1994.

[14] G. Picchi and G. Prati, \Blind Equalization and Carrier Recovery Using a `Stop-and-Go' Decision-Direct Algorithm," IEEE Trans. Comm., COM-35, pp. 877-887, 1987.

[15] J. R. Treichler and B.G. Agee, \A New Approach to Multipath Correction of Constant Modulus Signals," IEEE Trans. ASSP, vol. ASSP-31, pp. 459-471, Apr. 1983

[16] B. G. Agee, \Convergent Behavior of Modulus-Restoring Adaptive Arrays in Gaussian Interference Environments," in Proc. 21st Asilomar Conf. on Signals, Systems and Computer, pp. 589-593, 1987.

[17] Z. Ding, R.A. Kennedy, B.D.O. Anderson, and C.R. Johnson, Jr., \Ill-Convergence of Godard Blind Equalizers in Data Communications," IEEE Trans. Comm. vol. 39 pp. 1313-1328, 1991.

[18] B.G. Agee, \Blind Separation and Capture of Communication Signals Using a Multitarget Constant Modulus Beamformer," IEEE Military Com. Conf. pp. 340-346, 1989.

[19] A. van der Veen and A. Paulraj, \An Analytical Constant Modulus Algorithm," IEEE Trans. Signal Processing, vol. 44, no. 5, pp. 1136-1155, May 1996.

[20] A.J. van der Veen, S. Talwar and A. Paulraj, \A Subspace Approach to Blind Space-Time Signal Processing for Wireless Communication Systems," IEEE Trans. Signal Processing, vol. 45, no. 1, pp. 173-190, Jan. 1997.

[21] A.J. van der Veen, S. Talwar and A. Paulraj, \Blind Estimation of Multiple Digital Signals Transmitted over FIR Channels," IEEE Signal Proc. Lett., vol. 2, no.5, pp. 99-102, May 1995.

[22] D. Yellin and B. Porat, \Blind Identi¯cation of FIR Systems Excited by Discrete-Alphabet Inputs," IEEE Trans. Signal Proc. vol. 41, no. 3, pp. 1331-1339, March 1993.

[23] N. Seshadri, \Joint Data and Channel Estimation Using Fast Blind Trellis Search Techniques," in Proc. Globecomm, pp. 1659-1663, 1990.

[24] N. Seshadri, \Joint Data and Channel Estimation Using Fast Blind Trellis Search Techniques," IEEE Trans. Comm. vol. 42, no. 2/3/4, pp. 1000-1011, Feb./Mar./Apr. 1994.

[25] J.H. Lodge and M.L. Moher, \Maximum Likelihood Sequence Estimation of CPM Signals Transmitted Over Rayleigh Flat-Fading Channels," IEEE Trans. Comm. vol. 38, no. 6, pp. 787-794, 1990.

[26] X. Yu and S. Pasupathy, \Innovations-Based MLSE for Rayleigh Fading Channels," IEEE Trans. Comm. vol. 43, no. 2/3/4, pp. 1534-1544, Feb./Mar./Apr. 1995.

[27] G.D. Forney, Jr. \Maximum-Likelihood Sequence Estimation of Digital Sequences in the Presence of Intersymbol Interference," IEEE Trans. Info. Theory, vol. 18, no. 3, pp. 363-378, May 1972.

[28] G.D. Forney, Jr. \The Viterbi Algorithm," IEEE Proc., vol. 61, pp. 268-278, March 1972.

[29] S. Talwar, M. Viberg, and A Paulraj, \Blind Separation of Synchronous Co-Channel Digital Signals Using an Antenna Array-Part I: Algorithms," IEEE Trans. Signal Processing, vol. 44, no. 5, pp. 1184-1197, May 1996.

[30] S. Talwar and A Paulraj, \Blind Separation of Synchronous Co-Channel Digital Signals Using an Antenna Array-Part II: Performance Analysis," IEEE Trans. Signal Processing, vol. 45, no. 3, pp. 706-717, March 1997.

[31] H. Liu and G. Xu, \Smart Antennas in Wireless Systems: Uplink Multiuser Blind Channel And Sequence Detection," IEEE Trans. Comm. vol. 45, no. 2 pp. 187-199, Feb. 1997.

[32] J. Wang and J.K. Cavers, \Block Based Blind Joint Detection of Co-Channel Signals over Rayleigh Flat-Fading Channels with an Antenna Array," in Proc. IEEE Veh. Technol. Conf., pp. 857-861, May 1999.

[33] J. Wang and J.K. Cavers, \Block Based Blind Joint Detection of Cochannel Signals Over Time-Varying Channels," submitted for publishing, July 1999.

[34] J. Wang and J.K. Cavers, \Blind Joint Detection of Cochannel CPFSK Signals Over Rayleigh Flat-Fading Channels with A Diversity Antenna Array," in Proc. IEEE Veh. Technol. Conf. -Spring, pp. 177-181, Tokyo, May 2000.

[35] J. Wang and J.K. Cavers, \Blind Joint Detection of Cochannel CPFSK Signals Over Rayleigh Flat-Fading Channels with a Diversity Antenna Array," in Proc. IEEE Veh. Technol. Conf. -Fall, pp. 768-773, Boston, September 2000.

[36] J. Wang and J.K. Cavers, \Joint Blind MLSE for Co-Channel PSK Signals over Rayleigh Flat-Fading Channels with an Antenna Array," in Proc. IEEE Veh. Technol. Conf., pp. 470-474, Ottawa, May 1998.

[37] J. Wang and J.K. Cavers, \TDMA Signals Over Fast Rayleigh Flat-Fading Channels With Diversity Antenna Arrays," in preparation.

[38] J. Wang and J.K. Cavers, \Joint Blind MLSE for Co-Channel CPM Signals over Rayleigh Flat-Fading Channels with An Antenna Array," submitted in September, 2000.

[39] J-F. Cardoso, \Source Separation Using Higher Order Moments," in Proc. IEEE ICASSP, pp. 2109-2112, 1989.

[40] L. Tong, \Blind Sequence Estimation," IEEE Trans. Comm. vol 43, no 12, pp. 2986-2994, Dec. 1995

[41] D. Hatzinakos and C. Nikias, \Estimation of multipath channel response in frequency selective channels," IEEE J. Select. Areas Comm. vol. 7, pp. 12-19, Jan. 1989.

[42] B. Porat and B. Friedlander, \Blind Equalization of Digital Communication Channels Using High-Order Moments," IEEE Trans. Signal Processing, vol. 39, no. 2, pp. 2016-2024, 1991.

[43] O. Shalvi and E. Weinstein, \New Criteria for Blind Deconvolution of Non-Minimum Phase Systems (Channels)," IEEE Trans. Info. Theory, IT-36, pp. 312-321, 1990.

[44] O. Shalvi and E. Weinstein, \Super-exponential Methods for Blind Deconvolution," IEEE Trans. Info. Theory, IT-39, no. 2, pp. 504-519, Mar. 1993.

[45] W.A. Gardner, Introduction to Random Processes with Applications to Signals and Systems, Macmillan Publishing Company, New York, 1986.

[46] W.A. Gardner, Statistical Spectral Analysis: A Nonprobabilistic Theory, Prentice-Hall, Englewood Cli®s, NJ, 1987.

[47] W.A. Gardner and C.M. Spooner, \The Cumulant Theory of Cyclostationarity Time-Series, Part I: Foundation.," IEEE Trans. Signal Processing, SP-42, no. 12, pp. 3387-3408, Dec. 1994.

[48] C.M. Spooner and W.A. Gardner, \The Cumulant Theory of Cyclostationarity Time-Series, Part II: Development and Application.," IEEE Trans. Signal Processing, SP-42, no. 12, pp. 3409-3429, Dec. 1994.

[49] G. Xu, H. Liu, L. Tong and T. Kailath, \A Least-Squares Approach to Blind Channel Identi¯cation," IEEE Trans. Signal Processing, vol. 43, no. 12, pp. 2982-2993, Dec. 1995.

[50] H. Liu and G. Xu, \Closed-Form Blind Symbol Estimation in Digital Communications," IEEE Trans. Signal Processing, vol. 43, no. 11, pp. 2714-2723, Nov. 1995.

[51] E. Moulines, J.F. Cardoso and S. Mayrargue, \Subspace Methods for Blind Identi¯cation of Multichannel FIR Filters," IEEE Trans. Signal Processing, vol. 43, pp. 516-525, 1995.

[52] L. Tong, V.C. Soon, Y.F. Huang and R. Liu, \Indeterminacy and Identi¯ability of Blind Identi¯cation," IEEE Trans. Circuits Syst., pp. 499-509, May 1991.

[53] L. Tong, Y. Inouye and R. Liu, \Waveform-Preserving Blind Estimation of Multiple Independent Sources," IEEE Trans. Signal Processing vol. 41, no. 7, pp. 2461-2470, July 1993.

[54] L. Tong, Y. Inouye and R. Liu, \A Finite-Step Global Convergence Algorithm for the Parameter Estimation of Multichannel MA Processes," IEEE Trans. Signal Processing, vol. 40, no. 10, pp. 2547-2558, Oct. 1992.

[55] L. Tong, G. Xu and T. Kailath, \Fast Blind Equalization via Antenna Arrays," in Proc. IEEE ICASSP, pp. 751-755, 1993.

[56] L. Tong, G. Xu and T. Kailath, \A New Approach to Blind Identi¯cation and Equalization of Multipath Channel," in Proc. 25th Asilomar Conf. Signals, Systems and Computers, pp. 856-860, 1991.

[57] L. Tong, G. Xu and T. Kailath, \Blind Identi¯cation and Equalization Based on Second-Order Statistics: A Time Domain Approach," IEEE Trans. Info. Theory, IT-40, no.2, pp. 340-349, 1994.

[58] L. Tong, G. Xu, B. Hassibi and T. Kailath, \Blind Channel Identi¯cation Based on Second-Order Statistics: A Frequency Domain Approach," IEEE Trans. Info. Theory, vol. 41, no. 1, pp. 329-334, Jan. 1995.

[59] E. Serpedin and G.B. Giannakis, \Blind Channel Identi¯cation and Equalization with Modulation-Induced Cyclostationarity," IEEE Trans. Signal Processing, SP-46, pp. 1930-1944, 1998.

[60] M.K. Tsatsanidis and G.B. Giannakis, \Transmitter Induced Cyclostationarity for Blind Channel Equalization," IEEE Trans. Signal Processing, SP-45, pp. 1785-1794, 1997.

[61] B. Agee, S.V. Schell and W.A. Gardner, \Self-Coherence Restoral: A New Approach to Blind Adaptation of Antenna Arrays," in Proc. 21st Asilomar Conf. on Signals, Systems, and Computers, pp. 274-278, 1988.

[62] B. Agee, S.V. Schell and W.A. Gardner, \Spectral Self-Coherence Restoral: A New Approach to Blind Adaptive Signal Extraction Using Antenna Arrays," in Proc. IEEE, vol. 78, no. 4, pp. 753-767, April, 1990.

[63] B. Widrow, et al., \Adaptive Antenna Systems," in Proc. IEEE, vol. 55, pp. 2143-2159, Dec. 1967.

[64] IEEE Trans. Antennas Propagat., Special Issue on Adaptive Antennas, vol. AP-24, Sept. 1976.

[65] R.A. Monzingo and T.W. Miller, Introduction to Adaptive Arrays, New York: Wiley, 1980.

[66] V.F. Pisarenko, \The Retrieval of Harmonics From a Covariance Function," Geophys. J. Roy. Astron. Soc., vol. 33, pp. 347-366, 1973.

[67] R.O Schmidt, \Multiple Emitter Location and Signal Parameter Estimation," in Proc. RADC Spectrum Estimation Workshop, Gri±ss Air Force Base, NY, 1979; repr. IEEE Trans. Antennas Propagat., vol. AP-34, pp. 276-280, Mar. 1986.

[68] R. Roy, A. Paulraj, and T. Kailath, \Estimation of Signal Parameters via Rotational Invariance Techniques{ESPRIT," in Proc. 19th Asilomar Conf. Circ. syst. and Computers, Asilomar, pp. 83-87, CA, 1985.

[69] T.J. Shan, M. Wax, and T. Kailath, \On Spatial Smoothing for Direction-of-Arrival Estimation of Coherent Signals," IEEE Trans. Acoust., Speech, Signal Process., ASSP-33, pp. 806-811, 1985.

[70] B. Friedlander and A.J. Weiss, \E®ects of Model Errors on Waveform Estimation Using the MUSIC Algorithm," IEEE Trans. Signal Processing, vol. 42, no. 1, pp. 147-155, Jan. 1994.

[71] A. L. Swindlehurst, S. Daas and J. Yang, \Analysis of a Decision Directed Beamformer," IEEE Trans. Signal Processing, vol. 43, no. 12, pp. 2920-2027, Dec. 1995.

[72] A. Swindlehurst and J. Yang, \Using Least Squares to Improve Blind Signal Copy Performance," IEEE Signal Processing Lett. vol. 1, no. 5, pp. 80-82, May, 1994.

[73] W.C. Jakes, Microwave Mobile Communications, Wiley, New York, 1974.

[74] R.P. Gooch and J.D. Lundell, \The CM array: An adaptive beamformer for constant modulus signals," in Proc. IEEE Int. Conf. Acoust., Speech, Signal Processing, pp. 2523-2526, Tokyo, Japan, Apr. 1986.

[75] J.J. Shynk, and R.P. Gooch, \The Constant Modulus Array for Cochannel Signal Copy and Direction Finding," IEEE Trans. Signal Processing, vol. 44, no. 3, pp. 652-660, March 1996.

[76] Q. Wu, K.M. Wong and R. Ho, \Fast Algorithm for Adaptive Beamforming of Cyclic Signals," IEE Proc.-Radar, Sonar Navig., vol. 141, no. 6, pp. 312-320, Dec. 1994.

[77] J.H. Winters, J. Salz and R. D. Gillin, \The Impact of Antenna Diversity on the Capacity of Wireless Communication Systems," IEEE Trans. Comm. vol. 42, no. 2/3/4, Feb./Mar./Apr. pp. 1740-1751, 1994.

[78] J. Salz and J.H. Winters, \E®ect of Fading Correlation on Adaptive Arrays in Digital Mobile Radio," IEEE Trans. Veh. Tech., vol. 43, no. 4, pp. 10491057, Nov. 1994.

[79] S.J. Orfanidis, Optimal Signal Processing, 2nd Ed, New York: McGraw Hill, 1988.

[80] G. B. Giannakis and C. Tepedelenlioglu, \Basis Expansion Models and Diversity Techniques for Blind Identi¯cation and Equalization of Time-Varying Channels," Proc. IEEE, vol. 86, no. 10, pp. 1969-1986, Oct. 1998.

[81] M. Martone, \Wavelet-Based Separating Kernels for Sequence Estimation with Unknown Rapidly Time-Varying Channels," IEEE Comm. Lett. vol. 3, no. 3, pp. 78-80, March 1999.

[82] G.H. Golub and C.F. Van Loan, Matrix Computations, 3rd Ed, Johns Hopkins University Press, 1996.

[83] R. Steele, Mobile Radio Communications, Pentech and IEEE Press, New York, 1994.

[84] J.K. Cavers, Mobile Channel Characteristics, Kluwer Academic Publishers Inc., New York, 2000.

[85] A.J. van der Veen, \Resolution Limits of Blind Multi-User Multi-Channel Identi¯cation Schemes { The Bandlimited Case," in Proc. IEEE ICASSP, pp. 2722-2725, Atlanta, GA, May 1996.

[86] W. P. Osborne and M. B. Luntz, \Coherent and Noncoherent Detection of CPFSK," IEEE Trans. Comm., vol. 22, no. 8, pp. 1023-1036, Aug. 1974.

[87] S. Verdu, \Minimum Probability of Error for Asynchronous Gaussian Multiple Access Channels," IEEE Trans. Info. Theory, vol. 32, pp. 2461-2470, Jan. 1986.

[88] R. Raheli, A.Polydoros and C. K. Tzou, \Per-Survivor Processing: A General Approach to MLSE in Uncertain Environments," IEEE Trans. Comm. vol. 43, no. 2/3/4, Feb/Mar/Apr, pp. 354-364, 1995.

[89] A. J. Viterbi and J. K. Omura, Principles of Digital Communication and Coding, McGraw-Hill, New York, 1979.

[90] J.G. Proakis, Digital Communications, 2nd Ed. New York: McGraw Hill, 1989.

[91] D. Middleton, An Introduction to Statistical Communication Theory, McGraw-Hill Book Co., New York, N.Y. 1960.

[92] W. H. Press B. P. Flannery, S. A. Teukplsky and W. T. Vetterling, Numerical Recipes in C, Cambridge University Press, New York, 1988.

[93] J. M. Mendel, Maximum-likelihood Deconvolution : A Journey Into Model-Based Signal Processing, Springer-Verlag, New York, 1990.

[94] T. Bayes, \An Essay Toward Solving a Problem in the Doctrine of Chances," Phil. Trans. Roy. Soc., vol. 53, pp. 370-418, 1763. Reprinted in Biometrika, vol. 45, pp. 293-315 (1958).

[95] R. Horn and C.R. Johnson, Matrix Analysis, Cambridge University Press, New York, 1985.

[96] A. Paulraj, \Smart Antennas in Wireless Communication," The 8th International Conference on Wireless Communications, Calgary, AB. Canada. July 1996.

[97] B. Yang, \Projection Approximation Subspace Tracking," IEEE Trans. Signal Processing, vol. 43, no. 1, pp. 95-107, Jan. 1995.

[98] P. Comon and G.H. Golub, \Tracking a Few Extreme Singular Values and Vectors in Signal Processing," Proc. IEEE, pp. 1327-1343, Aug. 1991.

[99] G. Xu and T. Kailath, \Fast Subspace Decomposition," IEEE Trans. Signal Processing, pp. 539-551, Mar., 1994.

[100] G. A. Korn and T.M. Korn, Mathematical Handbook for Scientists and Engineers, McGraw-Hill, New York, 1968.

[101] A.J. Van der Veen and A. Paulraj, \Singular Value Analysis of Space-Time Equalization in the GSM Mobile System," in Proc. IEEE ICASSP, 1996.

[102] Electronic Industries Association, EIA Interim Standard, IS-95, Rev 0, \Dual-mode Mobile Station - Base Station Compatibility Standard," Washington, DC, 1994.