uOttawa

L'Université canadienne
Canada's university

FACULTÉ DES ÉTUDES SUPÉRIEURES ET POSTOCTORALES

uOttawa

L'Université canadienne
Canada's university

FACULTY OF GRADUATE AND POSDOCTORAL STUDIES

Maher Arar
AUTEUR DE LA THÈSE / AUTHOR OF THESIS

Ph.D. (Electrical and Computer Engineering)
GRADE / DEGREE

School of Information Technology and Engineering
FACULTÉ, ÉCOLE, DÉPARTEMENT / FACULTY, SCHOOL, DEPARTMENT

A Parallel Low-Complexity Mimo Detection Algorithm Using QR Decomposition and Alamouti Space-Time Code
TITRE DE LA THÈSE / TITLE OF THESIS

Abbas Yongacoglu
DIRECTEUR (DIRECTRICE) DE LA THÈSE / THESIS SUPERVISOR

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

Claude D'Amours

Ali Miri

Reza Soleymani (Concordia University)

Halim Yanikomeroglu

Gary W. Slater
Le Doyen de la Faculté des études supérieures et postdoctorales / Dean of the Faculty of Graduate and Postdoctoral Studies

A PARALLEL LOW-COMPLEXITY MIMO DETECTION ALGORITHM USING QR DECOMPOSITION AND ALAMOUTI SPACE-TIME CODE

By

Maher Arar

A thesis submitted to the Faculty of Graduate and Postdoctoral Studies
in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY
in Electrical and Computer Engineering

Ottawa-Carleton Institute for Electrical and Computer Engineering
School of Information Technology and Engineering
UNIVERSITY OF OTTAWA
OTTAWA, CANADA
SEPTEMBER, 2009

Library and Archives Canada

Published Heritage Branch

395 Wellington Street
Ottawa ON K1A 0N4
Canada

Bibliothèque et Archives Canada

Direction du Patrimoine de l'édition

395, rue Wellington
Ottawa ON K1A 0N4
Canada

Your file Votre référence
ISBN: 978-0-494-61371-9
Our file Notre référence
ISBN: 978-0-494-61371-9

To Monia, Barâa and Houd who accompanied me patiently
and faithfully during my quest for justice

Abstract

In this thesis we present a parallel $N \times N$ MIMO detection algorithm that is suitable for implementation on multi-core low-power processors. This algorithm has multiple versions which, depending on the SNR and target FER, allows for the complexity, performance and power consumption to be traded off on a packet-by-packet basis.

The proposed hybrid algorithm essentially consists of coding the transmit symbols in pairs using the Alamouti space-time block code. At the receiver, one or multiple QR decompositions are performed in parallel after which alternate successive interference cancellation, Alamouti decoding and coherent combining are performed. Iterative detection can also be performed should better error performance be required.

Despite the use of the capacity-lossy Alamouti code we show that the outage capacity and FER performance of the proposed algorithm at practical SNR levels below 20dB are comparable to, or sometimes outperform, those offered by spatial multiplexing algorithms such as MMSE-VBLAST. Furthermore, we show that our algorithm is more robust than MMSE-VBLAST in the presence of channel imperfections such as spatial correlation and channel estimation errors, two commonly encountered imperfections in practice. We also show that the complexity of our algorithm is equivalent to that of a lower complexity version of MMSE-VBLAST. Through FER comparisons with other well-known hybrid algorithms we also show that our algorithm achieves similar or better performance at a much lower complexity.

The features offered by the proposed algorithm that is: parallel architecture, good FER/capacity performance, robustness against channel imperfections make it an ideal candidate for use in 4G and beyond-4G wireless standards.

Acknowledgements

I would like to thank my supervisor Dr. Abbas Yongacoglu for his unconditional support during the last four years. His availability, thoroughness and the trust he put in me allowed me to submit my dissertation as originally planned. I would equally like to thank Drs. Claude D'Amour and Halim Yanikomeroglu who provided me with very important and constructive feedback which allowed me to improve my understanding of certain subjects related to my research. Their feedback was also helpful in pushing me to explore various ways of improving the proposed algorithm.

Thanks also go to the University of Ottawa which granted me an admission scholarship for the entire four-year term. Without this scholarship I would not have been able to continue my studies on a full-time basis. I am also grateful to the Ontario Government for granting me the Ontario Graduate Scholarship in Science and Technology for the school year 2006-2007.

My parents deserve to be thanked as they took care of me throughout my childhood.

Finally, thanks go to my wife Monia. The emotional support I received from her has allowed me to overcome the obstacles I faced during my research. Her love and affection knew no bound. I owe her a lot.

Table of Contents

List of Figures

List of Tables

List of Abbreviations

3GPP	Third Generation Partnership Project
4G	Fourth Generation
AOA	Angle of Arrival
ASIC	Application-Specific Integrated Circuit
AWGN	Additive White Gaussian Noise
BER	BIT Error Rate
BPSK	Binary Phase Shift Keying
BS	Base Station
CCDF	Complementary Cumulative Distribution Function
CDMA	Code Division Multiple Access
CFLOPS	Complex Floating-Point Operations
CSI	Channel State Information
dB	Decibel
DBLAST	Diagonal Bell Labs Layered Space-Time
DFE	Decision Feedback Equalization

DL	Downlink
DSM	Distributed Scattering Model
EPP	Error Propagation Problem
FER	Frame Error Rate
FLOPS	Floating-Point Operations
FPGA	Field Programmable Gate Array
GIN	Group Interference Nulling
GLST	Group or Generalized Layered Space-Time
GSIC	Group Successive Interference Cancellation
IEEE	Institute of Electrical and Electronics Engineers
IC	Interference Cancellation
IES	Inter-element Spacing
LLR	Log Likelihood Ratio
LTE	Long Term Evolution
MIMO	Multiple Input Multiple Output
ML	Maximum Likelihood
MMSE	Minimum Mean Squared Error
MSE	Mean Squared Error
MSI	Multi-stream Interference
MU-MIMO	Multi-User MIMO

OFDM	Orthogonal Frequency Division Multiplexing
OFDMA	Orthogonal Frequency Division Multiplexing Access
OSTBC	Orthogonal Space-Time Block Code
QAM	Quadrature Amplitude Modulation
QoLST	Quasi-orthogonal Layered Space-Time
QPSK	Quadrature Phase Shift Keying
SIC	Successive Interference Cancellation
RB	Resource Block
RE	Resource Element
SD	Sphere Decoder
SDA	Sphere Decoder Algorithm
SISO	Single Input Single Output
SM	Spatial Multiplexing
SNR	Signal-to-Noise Ratio
SRAB	Square Root Algorithm for BLAST
STBC	Space-Time Block Code
STTC	Space-Time Trellis Code
SU-MIMO	Single-User MIMO
SVD	Singular Value Decomposition
TCM	Trellis Coded Modulation

TDD	Time Division Duplexing
UE	User Equipment
ULA	Uniform Linear Array
UL	Uplink
VBLAST	Vertical Bell Labs Layered Space-Time
WIMAX	Worldwide Interoperability for Microwave Access
WLAN	Wireless Local Area Network
ZF	Zero Forcing

List of Symbols

α_j	j^{th} i.i.d coefficient emanating from TX that is used to model keyhole channels
β_i	i^{th} i.i.d coefficient arriving at RX that is used to model keyhole channels
δ_i	i^{th} eigenvalue of Wishart matrix $\mathbf{\Xi}$
$\eta(\mathbf{\Psi}_t, \mathbf{\Psi}_r)$	additive contribution of TX and RX correlation on capacity
γ	rank of Wishart matrix $\mathbf{\Xi}$
λ	wavelength
$\mathbf{\Gamma}_q$	q^{th} 2×2 submatrix extracted from matrix $\mathbf{R}$
$\mathbf{\Psi}_r$	RX correlation matrix
$\mathbf{\Psi}_t$	TX correlation matrix
$\mathbf{\Xi}$	Wishart matrix
ψ_r	RX correlation coefficient
ψ_t	TX correlation coefficient
ρ	signal-to-noise ratio per symbol per receive antenna
θ	angle of arrival
Θ_r	receive angle spread
Θ_t	transmit angle spread
$(.)^+$	hermitian operator
$\bar{C}_d$	ergodic capacity due to diversity
$\bar{C}_m$	ergodic capacity due to multiplexing
det(.)	determinant operator

ϵ ..... parameter that represents channel estimation accuracy

$\mathbf{a}$ ..... set of modulation alphabet

$\mathbf{H}$ ..... MIMO channel matrix with N_r rows and N_t columns

$\mathbf{h}_j$ ..... j-th column vector of channel matrix $\mathbf{H}$

$\mathbf{H}_{r,iid}$ ..... i.i.d receive Gaussian matrix used in DSM

$\mathbf{H}_{t,iid}$ ..... i.i.d transmit Gaussian matrix used in DSM

$\mathbf{I}$ ..... identity matrix

$\mathbf{K}_r$ ..... receive correlation matrix used in DSM

$\mathbf{K}_s$ ..... virtual correlation matrix used ind DSM

$\mathbf{K}_t$ ..... transmit correlation matrix used in DSM

$\mathbf{n}$ ..... noise vector at input to receiver

$\mathbf{Q}$ ..... unitary matrix obtained from performing QR decomposition of channel matrix $\mathbf{H}$

$\mathbf{R}$ ..... upper triangular matrix obtained from performing QR decomposition of channel matrix $\mathbf{H}$

$\mathbf{r}_j$ ..... j-th column vector of matrix $\mathbf{R}$

$\mathbf{s}$ ..... transmit symbol vector

$\mathbf{v}_q$ ..... q^{th} modified noise vector obtained of size 2×1 obtained after QR decomposition and Alamouti decoding

$\mathbf{w}_k$ ..... zero forcing null vector used in VBLAST

$\mathbf{W}_{MMSE}$ ..... MMSE spatial filter

$\mathbf{W}_{ZF}$ ..... ZF spatial filter

$\mathbf{y}$ ..... receive vector

$\tilde{\mathbf{n}}$ ..... transformed receive noise vector which is equal to $\mathbf{Q}^{+}\mathbf{n}$

$\tilde{\mathbf{y}}$ ..... transformed receive vector obtained by multiplying $\mathbf{y}$ by $\mathbf{Q}^{+}$

C_E ..... ergodic MIMO capacity

C_H ..... instantaneous MIMO capacity

C_x ..... x-outage MIMO capacity

D ..... distance between transmit and receive arrays

D_r ..... half receive array aperture

d_r ..... inter-element spacing of receive array

D_t ..... half transmit array aperture

d_t ..... inter-element spacing of transmit array

$E(.)$ ..... expectation operator

G_d ..... diversity gain

G_d^{max} ..... maximum diversity gain

G_m ..... multiplexing gain

G_m^{max} ..... maximum multiplexing gain

h_{ij} ..... element of matrix $\mathbf{H}$ located at row i and column j

i ..... row index

j ..... column index

N_r ..... number of receive antennas

N_t ..... number of transmit antennas

N_{max} ..... $max(N_t, N_r)$

N_{min} ..... $min(N_t, N_r)$

N_{qr} ..... number of QR decompositions

P_b ..... probability of bit error

P_{out} ..... probability of outage

r_H ..... total transmission rate

$r_{i,j}$ ..... element of matrix $\mathbf{R}$ located at the i^{th} row and j^{th} column

s_j ..... data symbol transmitted from antenna j

T_c ..... channel coherence time

$tr(.)$ ..... trace operator

y_i ..... data received on antenna i

Chapter 1

Introduction

Given a bandlimited wireless channel and a regulatory maximum transmit power constraint, which are usually imposed by regulatory bodies, is it possible to reliably send and receive digital information at transfer rates far beyond those offered by traditional wireless communications systems? This question has preoccupied researchers especially for the past decade mainly because the demand over higher data rates that support high-speed wireless network access has been steadily increasing while licensed radio frequency spectrum has become scarce and more expensive.

It turns out that the answer to the above question is yes if certain channel conditions are met, multiple transmit (TX) and receive (RX) antennas are employed, and the signals at the transmitter and/or receiver are processed in a manner that exploits the temporal and spatial properties of the wireless channel [22][53]. The latter is usually achieved by introducing controlled redundancy across the time dimension as well as the space dimension (i.e. across antennas).

Systems that employ such space-time signal processing techniques to increase throughput or/and to improve the reliability of the wireless link are referred to in the literature as **M**ultiple-**I**nput-**M**ultiple-**O**utput, or simply MIMO. Figure 1.1 depicts

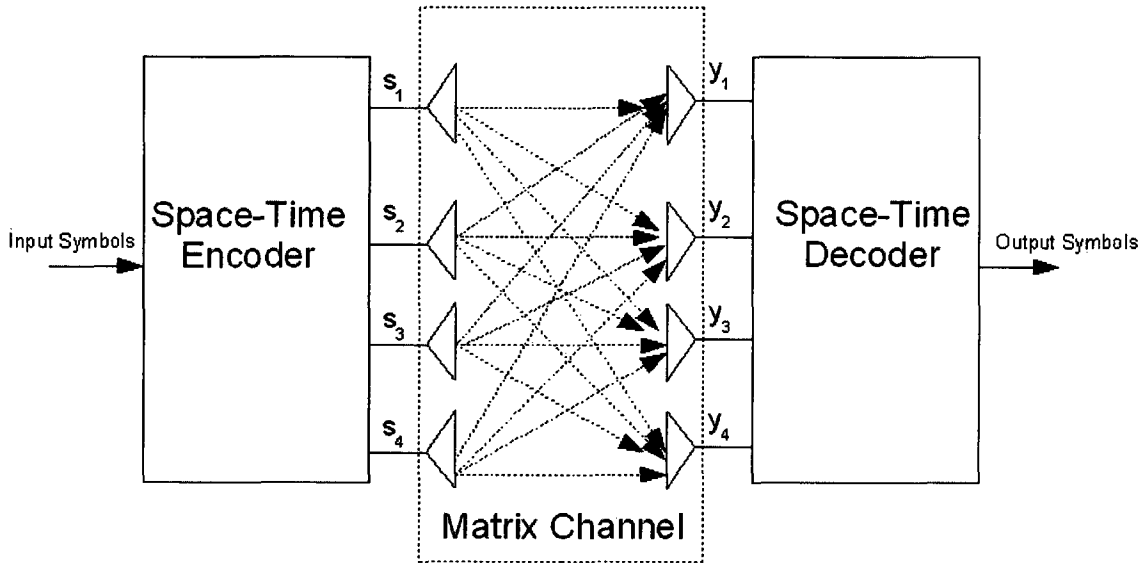

Figure 1.1: 4×4 MIMO system model

a 4×4 MIMO system model, where s_j represents the transmitted signal from TX antenna j, and y_i represents the received signal at RX antenna i.

Unlike in the Single-Input-Single-Output (SISO) wireless link, where the channel is simply modeled as a single impulse response, the MIMO baseband channel consists of a *matrix*, hereafter called $\mathbf{H}$, whose random entries h_{ij} represent the impulse response from the j-th TX antenna to the i-th RX antenna. When the channel is frequency-flat, the channel elements h_{ij} are simply complex-valued coefficients.

1.1 MIMO Benefits

The capacity (measured in bit/s/Hz) that could be achieved by a certain MIMO link depends primarily on the statistical properties of $\mathbf{H}$, with the greatest capacity offered when the various h_{ij} gains are uncorrelated to one another [22][53]. It seems

that the rich scattering that naturally exists in many environments, like in indoor and dense urban environments, greatly help synthesize such a channel. Put differently, MIMO exploits multipath to its advantage as opposed to traditional array systems which attempt to mitigate its effects. It will be shown in Chapter 2, that under the conditions described above, the MIMO channel capacity grows linearly with the minimum of the number of TX/RX antennas used. This latter advantage of MIMO systems is regularly referred to as the *multiplexing gain* in the literature. Other advantages include *diversity gain* (better reliability), *array gain* (enhanced signal-to-noise ratio) and interference suppression in multiuser applications[1]. Generally, those gains can not be obtained simultaneously as increasing one will generally lead to decreasing the other.

1.2 Existing MIMO Algorithms

Many algorithms have been proposed that attempted to achieve the dormant capacity and/or the diversity advantage offered by the MIMO channel. Early research efforts started at Stanford University by Paulraj and Kailath [40]. Some time later researchers at Bell Labs, led by Foschini, proposed Diagonal Bell Labs Layered Space-Time [23], also known as D-BLAST, and Vertical Bell Labs Layered Space-Time [19], also known as V-BLAST. Shortly after, this same team of researchers demonstrated a laboratory prototype of V-BLAST where they achieved an unprecedented capacity of 42 bits/s/Hz using 8 TX and 12 RX antennas [19]. One major drawback of V-BLAST is its dependence on matrix inversion which limits its performance due to the noise amplification that takes place [57].

[1] More precise definitions of some of these gains will be stated later.

In 1998, Alamouti proposed [1] a very simple space-time code for the two transmit and one receive antenna case. The Alamouti code possesses the distinctive feature of having a low-complexity maximum likelihood (ML) linear receiver. This makes it very attractive for use in battery-powered terminals. This explains the adoption of this particular space-time code by various 4G standards such as IEEE 802.11n (WLAN) and IEEE 802.16e (mobile WIMAX).

During the same year, Tarokh et al. [50] introduced the Space-Time Trellis Code (STTC) where they borrowed many concepts from Trellis Coded Modulation (TCM) and applied them to space-time coding. The main advantage of STTC is its ability to provide coding gain as well as the full diversity gain. Unfortunately, the complexity of STTC grows exponentially with the number of TX antennas and the modulation format used [56]. This disadvantage makes STTC unscalable and unattractive for practical implementation in modern systems where adaptive modulation is normally used to maximize system capacity. Realizing this deficiency in STTC, Tarokh et al. [48] introduced the theory behind Orthogonal Space-Time Block Codes (OSTBCs), where they generalized the Alamouti code to more than two transmit antennas. They also proved that, in order to preserve the orthogonality feature when more than two TX antennas were used, the transmission rate had to be sacrificed. This paper sparked further research interest in the field and some interesting full-rate and full-diversity 2×2 space-time codes were derived later [29][16][58][7][18]. These latter codes generally outperform OSTBCs at the cost of much higher receiver complexity due to the fact that ML decoding must be used to realize the full multiplexing-diversity gain [12].

Some of the algorithms mentioned above focus on maximizing the diversity gain

while others focus on maximizing the multiplexing gain. It seems that the best algorithms are those that can combine both of these gains [27]. Some of these algorithms, generally referred to as hybrid algorithms, are presented in [49][14]. The drawback of most of these proposed hybrid algorithms is that, similar to VBLAST, they still employ direct matrix inversion. As we mentioned previously, matrix inversion may enhance the noise level, and as a consequence limits the performance gain.

As most MIMO algorithms attempted to maximize either diversity gain or multiplexing gain, it was rather difficult to compare their performance in a fair way. Zheng was first to introduce [60] the concept of a fundamental trade-off where he proved the existence of an *optimal trade-off* between these two gains. One drawback of this latest benchmark is that it is only valid at infinite-SNR levels, which makes its applicability restricted to situations where extremely high SNR levels are available. In [37], a tradeoff curve is derived for finite-SNR levels where slightly different definitions of these gains were introduced.

1.3 Implementation Complexity

Since the introduction of space-time codes, the main challenge has been to come up with optimal or close-to-optimal (in capacity sense) detection algorithms that have a reasonable implementation complexity. At the higher end of the complexity spectrum is the Maximum Likelihood (ML) receiver. This algorithm provides the optimum Bit Error Rate (BER) performance, and is generally used as a reference for the best asymptotic error performance. Unfortunately, the ML receiver implementation complexity is often prohibitive as it increases exponentially with the number of TX antennas and the modulation order used [12].

ML-like algorithms, such as the sphere decoder algorithm (SDA), have been proposed. Unfortunately, the complexity of SDA varies with the Signal-to-Noise Ratio (SNR) and the modulation order used, which is not desirable from an implementation point-of-view. Attempts have been made to come up with SNR-independent and modulation-independent versions of SDA. A promising solution that meets these two conditions was proposed in [17] but its payload complexity is still high for high-speed real-time implementation. It is also important to mention that the complexity of SDA has been mostly analyzed in the absence of spatial correlation. In [5] it is shown that spatial correlation increases the expected complexity of SDA. Moreover, providing soft output[2] increases the complexity of SDA by about 50% [46]. All of the above deficiencies associated with SDA-type algorithms make them unattractive for use in 4G systems where sub-Gbps rates are required. For instance, the highly optimized implementation of a soft-output SDA that is described in [46] achieves only a maximum of 100Mbps throughput for a 4×4 MIMO system.

Due to their relatively lower complexity requirements, non-linear receivers, which use successive interference cancellation (SIC), are normally preferred for real-time implementation. Unfortunately, their error performance is much worse than the one provided by sphere decoders. This degradation is mainly caused by the self-generated error propagation problem (EPP) inherent in any SIC-type receiver. This normally can be slightly mitigated by improving the diversity of the first detected layer.

Finally, it is important to mention that due to the increasingly demanding throughput, energy and delay constraints as imposed by 4G standards' bodies, we believe that parallelizing MIMO algorithms will be the only way to meet these requirements.

[2]This is a must-have feature because efficient outer decoders expect soft inputs

1.4 MIMO in Wireless Standards and Commercial Products

Even though MIMO research started in 1994, it took almost six years to witness the first proprietary commercial products. The first systems which used MIMO technology to boost data rates were invented by Iospan Wireless Inc., a start-up that was later acquired by Intel Corporation. It took few more years until MIMO found its way into Third and Fourth Generation wireless standards such as 802.11n (wireless LAN), 802.16e (WIMAX) and 3GPP-Long Term Evolution (LTE), to name few.

In a nutshell, all of these standards allow for the optional use of various forms of MIMO namely: spatial coding (another name for space-time coding), spatial multiplexing and spatial beamforming. The latter is only possible to use in Time Division Duplexing (TDD) systems and when the channel changes slowly compared to the frame rate. Most of these standards have been ratified only recently (i.e. as of late 2008) and products based on these standards started to appear on the market as early as the first quarter of 2009. It is important to note that due to the complexity of the RF circuitry and baseband algorithms associated with MIMO most of these products currently use no more than two antennas. This is especially true for hand-held and battery-powered terminals with the physical dimensions of the device imposing an additional constraint on the use of multiple antennas. The use of more than two antennas still poses huge complexity and energy consumption challenges and it is expected that research efforts in these two areas will be the main focus for years to come. It is widely accepted that future MIMO algorithms must be inherently parallel so that lower clocking frequency can be used in order to conserve power and meet the required peak target rates.

Finally, it is also important to mention that, because of its linear complexity requirements, the Alamouti space-time code (or space-frequency code as in LTE) has already been adopted by these standards as the main choice for spatial coding.

1.5 Problem Definition

Research efforts over the past few years have focused on the 2×2 MIMO case. Optimum space-time codes such as tilted-QAM [58] have been proposed. For this simple 2×2 case, ML-type decoders with reasonable complexity have been proposed and implemented recently [42].

In order to achieve the high spectral efficiency specified for 4G and beyond-4G standards it is anticipated that 4G terminals will have the option to be equipped with up to 4 antennas (e.g. [51]). Beyond-4G products are expected to have the option to be equipped with as much as 8 antennas (e.g. [52]).

So it is evident from the above discussion that there is currently a need, from a practical point-of-view, to find a $N_t \times N_r$ ($N_r, N_t > 2$) MIMO algorithm that possesses the following desirable features:

- The architecture must be parallel in order to allow for high-speed low-power symbol processing.
- Should achieve ML-like capacity and error performance with reasonable complexity and without knowledge of noise variance, as is needed by some current MIMO algorithms. Obtaining accurate estimation of noise variance is not possible in many practical applications.

- Its complexity should not depend on the modulation order used or the operational SNR, as is the case in most SD algorithms.

- Should not suffer from the noise enhancement and/or the error propagation problems usually associated with SIC-based family of algorithms.

- Its improved performance should not be attained through the use of more RX antennas compared to the number of TX antennas , as is the case in V-BLAST. One can only think of a wireless downlink where the receiver, i.e. the user equipment (UE), is generally limited in the number of antenna elements that it can use due to the small form-factor of some of these devices.

- Should be robust against spatial correlation that naturally exists between antenna elements in many wireless environments. The majority of the proposed MIMO detection or space-time codes presented in the literature are derived with the simplistic assumption that the various channel coefficients are uncorrelated.

- Should be robust against channel estimation error, an imperfection that can not be avoided in practice due to the finite length of the training sequence and quantization effects caused by finite-precision arithmetic.

It is obviously challenging, if not impossible, to design an algorithm that achieves all of the above features. A tradeoff must be reached especially when it comes to performance and complexity. Because of the requirements stated previously (i.e. sub-Gbps rates, short delays and low power consumption), we believe that the first feature listed above is a must-have feature as it will allow for the efficient implementation of 4G and beyond-4G battery-powered terminals.

1.6 Thesis Contributions

The main contributions of this thesis are as follow:

1. In Chapter 4, we propose the use of a hybrid MIMO architecture where multiple instances of the Alamouti space-time code are used at the TX side. The RX decoding algorithm we propose possesses polynomial complexity in the number of TX/RX antennas regardless of SNR and the modulation format used. While capacity lossy when compared to MIMO capacity, we show that when the Alamouti code is compared to other algorithms such as V-BLAST, its capacity at practical SNR levels below 20dB is no worse than that of the best performing version of V-BLAST. The proposed algorithm is parallel in nature and as such should be able to support high data rates with low power consumption, a must-have feature for 4G wireless terminals. The latter advantage comes from the fact that lower clock frequencies can be used to implement this algorithm. Furthermore, when the operational SNR is high some processors can be turned off in order to achieve further power savings. This makes the proposed algorithm an ideal candidate for use in battery-powered 4G terminals.

2. We show in Chapter 2 that contrary to the popular statement made in the literature regarding the contribution of multiplexing gain to MIMO capacity, diversity gain contribution to capacity can surpass that offered by the multiplexing gain depending on the operational SNR and the number of TX/RX antennas used. In the same chapter, we also find a very good simple approximation for Laguerre polynomial. This in turn simplifies the MIMO ergodic capacity expression and allows us to derive a simple closed-form formula for the

finite-SNR multiplexing gain.

3. Through computer simulations in Chapters 3 and 5 we shed light on the multiplexing-diversity gain offered by MMSE-VBLAST. As such we reconcile many of the apparently conflicting conclusions reached in the literature regarding its performance. We basically show that its multiplexing-diversity gain changes with the transmission rate and the operational SNR.

1.7 Thesis Organization

The rest of this thesis is organized as follows. In Chapter 2, we briefly review the MIMO channel capacity and bounds on error performance. In Chapter 3, we present some of the most well known MIMO detection algorithms and space-time codes such as V-BLAST and the Alamouti space-time code. In Chapter 4, we present the proposed algorithm, derive its outage capacity and simulate its FER performance. In Chapter 5, we compare the complexity/performance of the proposed algorithm to other MIMO detection algorithms. In Chapter 6, we present a conclusion where we suggest further research work which can be undertaken to bring improvement to the proposed algorithm.

Chapter 2

MIMO Capacity and Theoretical Bounds

In order to understand why the use of antenna arrays on both ends of the wireless link has created so much research and commercial interest, we have to study the capacity offered by the MIMO channel. As this chapter will demonstrate, the MIMO capacity is enormous and to realize it, or at least realize a major fraction of it, attention must be paid to certain key parameters such as the inter-element spacing in the antenna arrays used.

To make our notation simpler, we will conduct our analysis of MIMO in discrete-time baseband, thereby dropping any dependency on the carrier frequency component from the expressions used. We will also assume that the channel matrix $\mathbf{H}$ is static over the burst duration, and as such, we will also ignore the dependency on time. We will also assume that the channel fading is frequency-flat.

2.1 MIMO System Model

Referring to Figure 1.1, we can express the received signal vector $\mathbf{y} = [y_1 \ y_2 ... y_{N_r}]^T$ as follows:

$$\begin{aligned} \mathbf{y} &= \sqrt{\frac{\rho}{N_t}}\mathbf{H}\mathbf{s} + \mathbf{n} \\ &= \sqrt{\frac{\rho}{N_t}}(\mathbf{h}_1 s_1 + \mathbf{h}_2 s_2 + ... + \mathbf{h}_{N_t} s_{N_t}) + \mathbf{n}, \end{aligned} \tag{2.1.1}$$

where $\mathbf{H}$ is the $N_r \times N_t$ random channel matrix whose elements h_{ij}, that represent the complex gains between TX antenna j and RX antenna i, have Gaussian distribution that are assumed to be independent and identically distributed (i.i.d.) with $\mathcal{N}(0, 0.5)$ in each dimension, $\mathbf{s} = [s_1 \ \ s_2....s_{N_t}]^T$ is the TX vector with each component representing a symbol, generally complex, belonging to a finite set of symbol alphabet $\mathbf{a} = [a_1 \ \ a_2....a_K]$ with $\frac{1}{K}\sum_{i=1}^{K} |a_i|^2 = 1$, and $\mathbf{n} = [n_1 \ \ n_2....n_{N_r}]^T$ represents the complex noise vector at the input to the receiver with elements n_i that are assumed to be uncorrelated with s_j and between themselves, i.e. $E(\mathbf{n}\mathbf{n}^+) = \mathbf{I}_{N_r}$, where $E(.)$ denotes the expectation operator. The normalization factor $\sqrt{\frac{\rho}{N_t}}$ ensures that the total transmit power is always the same regardless of the number of TX antennas and the average receive SNR per branch is always equal to ρ. The second form of the equation uses the columns of the channel matrix, i.e. $\mathbf{h}_j = [h_{1j} \ \ h_{2j} \ldots h_{N_r j}]^T$ $(j = 1, 2, \ldots N_t)$.

Unless otherwise specified we will assume that the receiver has full knowledge of $\mathbf{H}$ but the transmitter has no prior knowledge of it. In practice, one way $\mathbf{H}$ can be estimated is by transmitting a known sequence at the beginning of each frame and by employing ML or MMSE techniques at the receiver to compute an estimate of the channel. For our purposes we assume throughout this chapter that $\mathbf{H}$ has been perfectly estimated, i.e. the channel estimation error is zero.

2.2 MIMO Channel Capacity

Performance limits are important to establish as they allow us to measure how close to optimal a certain MIMO algorithm is. In MIMO applications, the capacity offered by a certain MIMO channel is the most important bound as it allows us to have *a priori* knowledge about the dormant throughput that we can potentially achieve.

Depending on whether we are operating in a fast or a slow fading channel, two different definitions of capacities arise: ergodic capacity and outage capacity. In what follows we will take a closer look at both of these definitions.

2.2.1 Ergodic Capacity

Ergodic capacity, sometimes called temporal mean capacity, gives an indication about the "average" capacity that can be achieved in fast fading scenarios where the random MIMO channel is changing frequently compared to the frame rate.

We present here the asymptotic MIMO channel ergodic capacity results that were first derived in [53][22]. It was shown that the ergodic capacity $\bar{C}$ is equal to $E\{C_H\}$ where the instantaneous capacity C_H is equal to:

$$C_H = \begin{cases} \log_2 \det\left(\mathbf{I}_{N_r} + \frac{\rho}{N_t}\mathbf{H}\mathbf{H}^+\right) & \text{for } N_r \leq N_t \\ \log_2 \det\left(\mathbf{I}_{N_t} + \frac{\rho}{N_t}\mathbf{H}^+\mathbf{H}\right) & \text{for } N_r > N_t. \end{cases} \tag{2.2.1}$$

In the above expression $\det(\cdot)$ represents the determinant operator and $\mathbf{H}^+$ represents the hermitian conjugate of $\mathbf{H}$.

It can be shown [27] that the maximum MIMO channel capacity is achieved when $E(\mathbf{H}\mathbf{H}^+) = I_{N_t}$ (or $E(\mathbf{H}^+\mathbf{H}) = I_{N_r}$), i.e. the different channel gains are uncorrelated and of equal magnitude. It should be noted that this definition implicitly assumes

the use of an infinite packet length and Gaussian input (with infinite modulation order) to achieve the ergodic capacity. It should also be noted that this capacity is the maximum that can be attained as all the degrees of freedom are used to transmit independent data. When redundancy is introduced, as in the case of space-time coding, the capacity of the resulting effective channel will be inevitably lower.

An alternative, but equivalent, expression for MIMO capacity in (2.2.1) can be obtained through the eigenvalue decomposition of the square matrix $\mathbf{\Xi} = \mathbf{H}\mathbf{H}^+$ when $N_r \leq N_t$ (or $\mathbf{\Xi} = \mathbf{H}^+\mathbf{H}$ when $N_r > N_t$). $\mathbf{\Xi}$ is generally referred to as the Wishart matrix in the literature. This matrix has many useful properties such as the one that states that all its eigenvalues are real-valued and positive. Using this information, the instantaneous capacity is obtained by [27]:

$$C_H = \sum_{j=1}^{\gamma} \log_2(1 + \frac{\rho}{N_t}\delta_j) \tag{2.2.2}$$

where δ_i are the real-valued positive eigenvalues of $\mathbf{\Xi}$ and γ is its rank. It should be noted that $\mathbf{\Xi}$ has full rank in general, i.e. $\gamma = \min(N_t, N_r)$. According to the above definition, capacity is reduced when $\gamma < \min(N_t, N_r)$, which is the case when there is heavy correlation or the MIMO channel is degenerate as we shall see later.

Figure 2.1 shows the ergodic capacity curves for various values of N_t and N_r at an average receive signal-to-noise ratio ρ of 15 dB. These curves were obtained by using (2.2.1) and running Monte Carlo simulations over 10,000 independent channel realizations for each simulation run. One conclusion we can draw from Figure 2.1 is that the theoretical capacity offered by the i.i.d. MIMO channel is limitless, and to realize it we must increase the number of antennas on both ends of the MIMO link. Increasing the number of antennas on one end offers only marginal increase in capacity

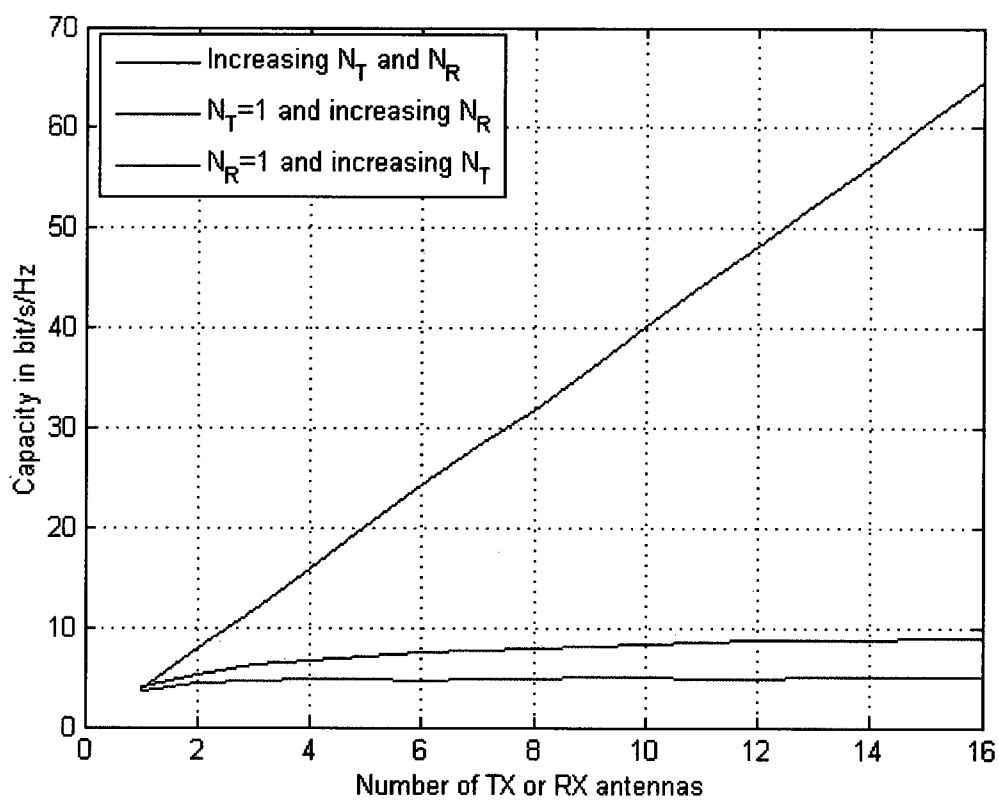

Figure 2.1: Ergodic capacity of an i.i.d. MIMO channel for $\rho = 15$dB.

(i.e. similar to the one offered by traditional array systems). Note that equation (2.2.1), similar to the famous Shannon equation for the SISO AWGN channel, does not tell us how to achieve this enormous capacity. The objective of any MIMO TX/RX processing algorithm is to attempt to achieve this capacity with a reasonable implementation complexity. In general, the capacity achieved by most existing MIMO algorithms is lower.

The huge capacity enhancement obtained with the use of multiple antennas is due to two important concepts: multiplexing and diversity. The former concept relates

to MIMO's ability to exploit the various degrees of freedom to send independent data across the TX/RX antennas. Intuitively, this ability is limited by the minimum number of TX or RX antennas used. The diversity concept relates to MIMO's ability to exploit the multiple independent paths created with the number of TX and RX antennas. The expressions in (2.2.1) and (2.2.2) do not provide us with insight about how these two concepts contribute to MIMO capacity. To see this we resort to the upper tight bound presented in [43]. It is shown that the ergodic capacity $\bar{C}$ is upper bounded as:

$$\bar{C} \leq \underbrace{N_{min}\log_2(\frac{\rho}{N_t}) + \log_2\left(L_{N_{min}}^{N_{max}-N_{min}}(\frac{-N_t}{\rho})\right)}_{\bar{C}_m} + \underbrace{\log_2\frac{N_{max}!}{(N_{max}-N_{min})!}}_{\bar{C}_d} \tag{2.2.3}$$

where $N_{min} = \min(N_t, N_r)$, $N_{max} = \max(N_t, N_r)$ and

$$L_{N_{min}}^{N_{max}-N_{min}}(\frac{-N_t}{\rho}) = \sum_{i=0}^{N_{min}} \frac{(-1)^i}{i!} \begin{pmatrix} N_{max} \\ N_{min}-i \end{pmatrix} (\frac{-N_t}{\rho})^i \tag{2.2.4}$$

is the Laguerre polynomial of order N_{min}. At high SNR, the contribution from the Laguerre polynomial is negligible and the expression in (2.2.3) reduces to:

$$\bar{C} \leq \underbrace{N_{min}\log_2(\frac{\rho}{N_t})}_{\bar{C}_m} + \underbrace{\log_2\frac{N_{max}!}{(N_{max}-N_{min})!}}_{\bar{C}_d} \tag{2.2.5}$$

We clearly see from the above expression that the MIMO ergodic capacity for the i.i.d. channel scales linearly with the number of antennas used. In other words, MIMO capacity at high SNR increases by N_{min} bit/s/Hz for every 3dB increase. It should be noted that this is only true for i.i.d. channels. A generalization to the

above that is valid for all channels is that $N_{min} = \min(N_t, N_r, \gamma)$, where γ is the rank of $\mathbf{H}$. $G_m^{max} = N_{min}$ is generally referred to as the multiplexing gain of the MIMO channel at infinite-SNR.

For our case of interest where $N_t = N_r = N$, the expressions in (2.2.3) and (2.2.5) reduce to:

$$\bar{C} \leq \underbrace{N \log_2(\frac{\rho}{N}) + \log_2\left(L_N^0(\frac{-N}{\rho})\right)}_{\bar{C}_m} + \underbrace{\log_2(N!)}_{\bar{C}_d} \tag{2.2.6}$$

and

$$\bar{C} \leq \underbrace{N \log_2(\frac{\rho}{N})}_{\bar{C}_m} + \underbrace{\log_2 N!}_{\bar{C}_d}, \tag{2.2.7}$$

respectively. Figure 2.2 compares the ergodic capacities obtained with (2.2.1), (2.2.6) and (2.2.7) for 4×4 and 8×8 MIMO systems. We clearly see that the upper bound in (2.2.6) is valid for all SNR while the high-SNR approximation as given by (2.2.7) is only valid for $\rho > 10$dB in the 4×4 case while the same is true for $\rho > 15$dB for the 8×8 case.

Multiplexing and Diversity Contribution to Capacity

Examining the expressions in (2.2.6) or (2.2.7) closely, one can safely conjecture that the quantities $\bar{C}_m$ and $\bar{C}_d$ represent the contributions to ergodic capacity due to multiplexing and diversity, respectively. We clearly see that the contribution of multiplexing is multiplicative and dependent on SNR while diversity contributes additively to capacity, and that is independently of SNR (i.e. vertical shift in the capacity curve).

It is then rather informative to compare the relative contributions of $\bar{C}_m$ and $\bar{C}_d$

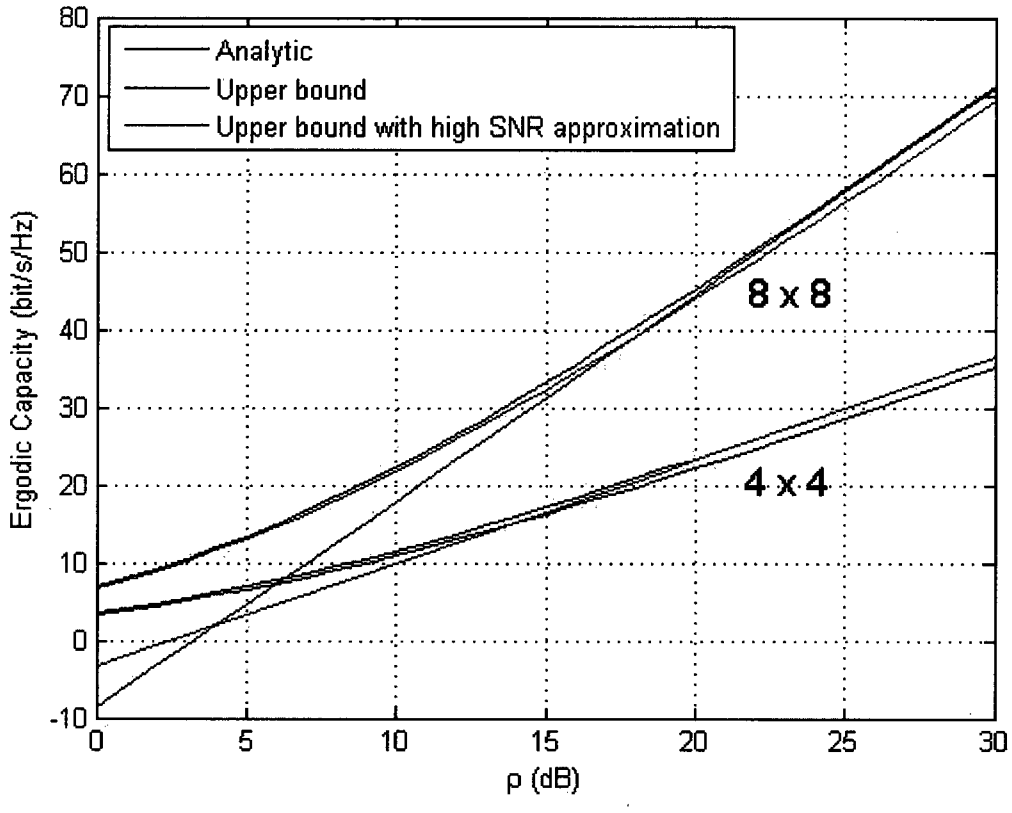

Figure 2.2: Comparison of ergodic capacity for 4×4 and 8×8 systems obtained by using (2.2.1) to that obtained with (2.2.6) and (2.2.7)

to ergodic capacity when $N_t = N_r = N$. To do this we resort to the high-SNR approximation in (2.2.7) in order to find the value of ρ where the ratio $\frac{\bar{C}_m}{\bar{C}_d}$ is equal to 1. That is:

$$\frac{\bar{C}_m}{\bar{C}_d} = \frac{N \log_2(\frac{\rho}{N})}{\log_2 N!} = 1. \tag{2.2.8}$$

After some simple mathematical manipulations we find that:

$$\rho = N(N!)^{\frac{1}{N}}. \tag{2.2.9}$$

Based on the above equation, Figure 2.3 shows the regions where $\bar{C}_m > \bar{C}_d$ and $\bar{C}_m < \bar{C}_d$. We clearly see that the diversity contribution to capacity can exceed that of multiplexing (i.e. when $\bar{C}_m < \bar{C}_d$) and this can happen at practical number of antennas and SNR levels. For example, when $N = 4$ and $\rho < 9.5$dB, $\bar{C}_d$ contributes more than $\bar{C}_m$ to the overall ergodic capacity. We have to point out that since the curve shown in Figure 2.3 was obtained with the high-SNR approximation of the upper bound the results are well approximated for $\rho > 10$dB. A more precise curve can be obtained by trial and error approach using the expression in (2.2.6). The more precise curve will be slightly shifted to the left at SNR below 10dB.

The important fact to highlight here is that MIMO algorithms that exploit diversity will be more effective at enhancing capacity at low-to-moderate SNR (and by default at low rates) while those that exploit multiplexing will be more effective at high SNR (and by default at high rates). Since a wide range of SNR levels are encountered in practice switching between both types should be employed. Another alternative is to use a hybrid algorithm.

Another interesting point to observe is that the threshold where $\bar{C}_m = \bar{C}_d$ increases as the number of TX/RX antennas increases. For example, when $N \geq 16$ and $\rho \leq 20$dB, we can see that the diversity contribution to capacity is always greater. This suggests that emphasis should be mainly placed on extracting the diversity contribution to capacity in this case.

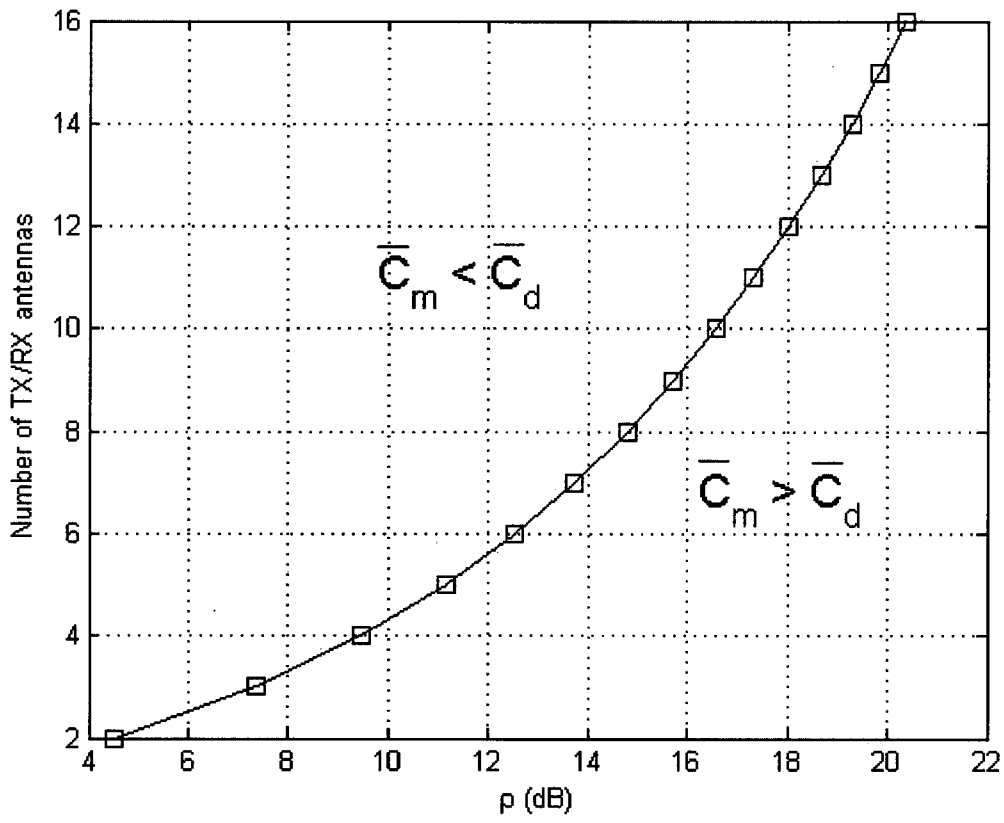

Figure 2.3: Regions where $\bar{C}_m > \bar{C}_d$ and $\bar{C}_m < \bar{C}_d$

Multiplexing Gain at Finite SNR

As mentioned previously, the $N \times N$ MIMO ergodic capacity grows by N bit/s/Hz for every 3dB increase in SNR. This is only valid at infinite SNR. Here we are interested in knowing the multiplexing gain at all SNR. To do this we resort to the tight upper bound formula in (2.2.6) and find an approximation for the the Laguerre polynomial. By using trial and error techniques, we found that the following simple function approximates the Laguerre polynomial:

$$L_N^0(\frac{-N}{\rho}) \approx 2^{(\frac{2N}{\sqrt{\rho}})}. \tag{2.2.10}$$

We now obtain a new approximate value for $\bar{C}_m$ in (2.2.6):

$$\begin{aligned} \bar{C}_m &\approx N\log_2(\frac{\rho}{N}) + \log_2(2^{(\frac{2N}{\sqrt{\rho}})}) && (2.2.11)\\ &= N\log_2(\frac{\rho}{N}) + \frac{2N}{\sqrt{\rho}}. && (2.2.12) \end{aligned}$$

To find the finite-SNR multiplexing gain G_m^ρ we basically find the capacity with a 3dB increase in SNR and then subtract it from the current one to obtain:

$$\begin{aligned} G_m^\rho &\approx N\log_2(\frac{2\rho}{N}) + (\frac{2N}{\sqrt{2\rho}}) - N\log_2(\frac{\rho}{N}) - (\frac{2N}{\sqrt{\rho}}) && (2.2.13)\\ &= N(\log_2 2 + \log_2(\frac{\rho}{N})) + (\frac{2N}{\sqrt{2\rho}}) - N\log_2(\frac{\rho}{N}) - (\frac{2N}{\sqrt{\rho}}) && (2.2.14)\\ &= N + (\frac{2N}{\sqrt{2\rho}}) - (\frac{2N}{\sqrt{\rho}}) && (2.2.15)\\ &= N\left(1 - \frac{\sqrt{2}-1}{\sqrt{\rho}}\right). && (2.2.16) \end{aligned}$$

Using the above expression we find that the multiplexing gain at $\rho = 10$ dB, for example, is only 81% of that reached at infinite SNR.

2.2.2 Outage Capacity

Outage capacity is a more useful metric to use in block fading scenarios (which is assumed in our study) where the channel stays constant during the entire frame. Because some of the channel realizations may not support a certain desired rate, it makes sense to speak of the percentage of the channel realizations that will support

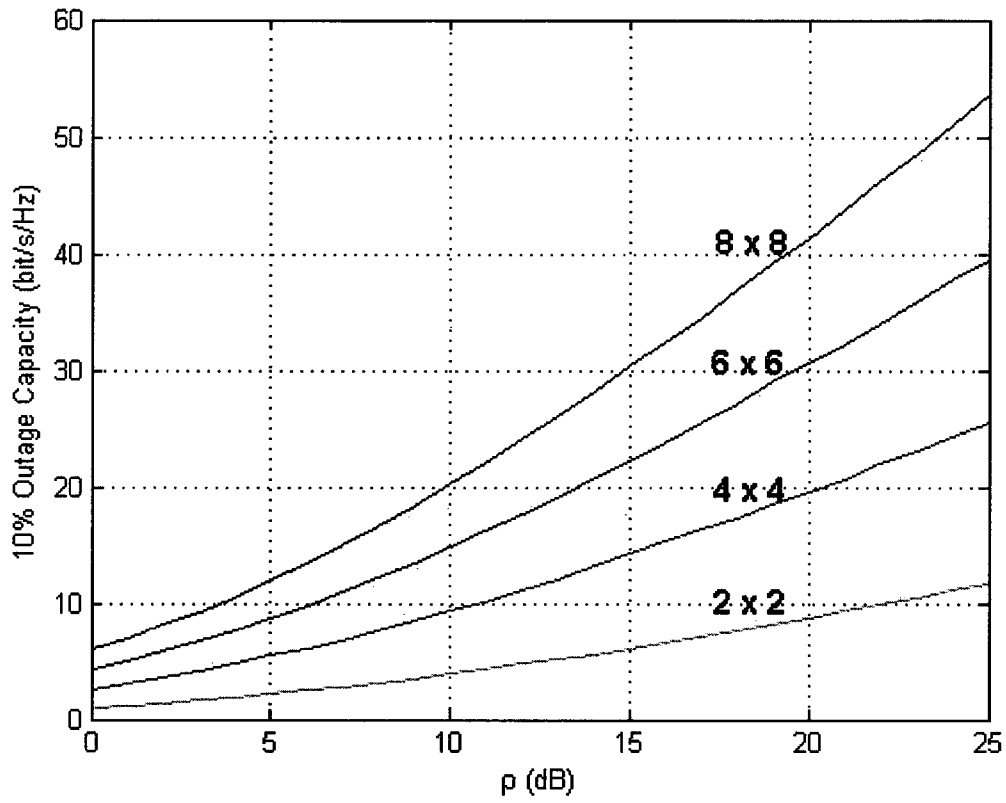

Figure 2.4: 10% outage capacity for various $N_r \times N_t$ i.i.d. MIMO channels as a function of SNR ρ.

this rate. The x outage capacity C_x is then defined as the capacity that will be supported $1 - x$ of the time. Mathematically, we have

$$P_{out}\left(C_H \leq C_x\right) = x \tag{2.2.17}$$

where P_{out} denotes the probability of outage. Figure 2.4 shows the 10% outage capacity of various $N \times N$ i.i.d. MIMO channels. It has to be pointed out that realistic

modulation orders can be employed to attempt to achieve these capacities. For instance, for a 4×4 MIMO channel and a typical SNR of 20dB, the use of 32-QAM modulation for each antenna (this translates into using 5 bits per antenna) will be sufficient.

In [39] an interesting interpretation of the x probability of outage is presented where it is argued that it is equal to the optimal x Frame Error Rate (FER). We will sometimes resort to this interpretation when comparing the performance of various MIMO algorithms. Figure 2.5 shows various probability of outage curves (i.e. optimal FER curves) for a given desired rate r_H for a 4 x 4 MIMO system.

Because of the randomness of the channel it is not possible to obtain a closed-form expression for the outage capacity. If this were possible it would have allowed us to conduct an analysis similar to the one presented in the previous section regarding the separate contributions of multiplexing and diversity to the overall capacity. But intuitively one would expect the same kind of trend to persist.

2.3 Effect of Correlation and Keyhole on MIMO Channel Capacity

In the previous sections we assumed that the MIMO channel is i.i.d. which, unfortunately, is rarely the case in real-world environments. Channel measurements conducted in [21] for a dense urban environment, show that the capacity offered by outdoor fixed-wireless MIMO channels is around 80% of the i.i.d. Gaussian one. Also a phenomenon, called keyhole (or pinhole), may physically occur in real-world wireless environments such as in hallways or in long-haul communications. As will be shown later, the keyhole phenomenon has a detrimental effect on MIMO capacity

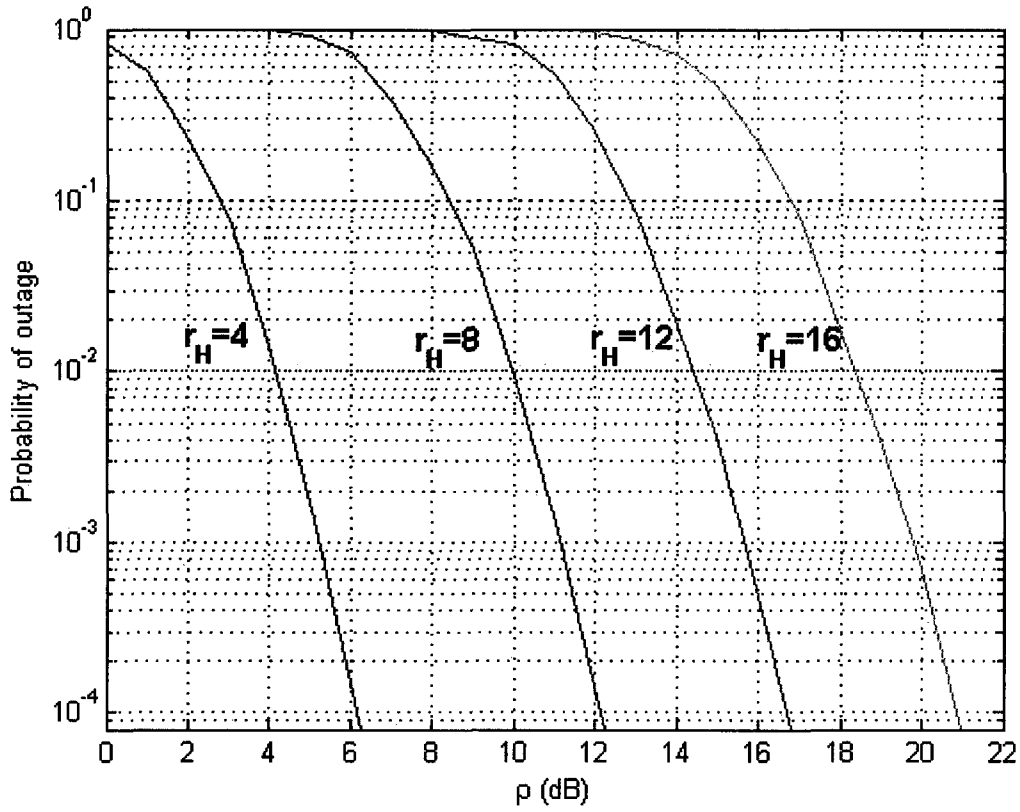

Figure 2.5: Probability of outage for a 4 x 4 i.i.d. MIMO channel for various rates r_H=4,8,12 and 16 bit/s/Hz

even in the absence of correlation. This possibly explains the considerable research interest in developing realistic MIMO channel models. A good review of various models is presented in [20]. Depending on many factors, such as the richness of scattering and mobility issues, these models have their own advantages and disadvantage when compared to one another.

Before diving into details it is rather important to gain insight on how correlation and keyhole phenomenon affect MIMO capacity. The following two sections treat

these two subjects in more detail using the $N \times N$ MIMO channel.

2.3.1 Effect of Correlation on MIMO Capacity

To incorporate the effect of correlation into MIMO channels we will use the Kronecker model described in [12]. This model assumes that the TX and RX correlations are independent and as such the resulting MIMO channel $\mathbf{H}$ can be obtained as:

$$\mathbf{H} = \mathbf{\Psi}_r^{1/2} \mathbf{H}_{iid} \mathbf{\Psi}_t^{1/2} \tag{2.3.1}$$

where $\mathbf{\Psi}_r$ and $\mathbf{\Psi}_t$ are the RX and TX correlation coefficient matrices, respectively, and $\mathbf{H}_{iid}$ is an $N \times N$ channel matrix with i.i.d. entries.

If we substitue this value of $\mathbf{H}$ in the instantaneous capacity C_H in (2.2.1) we obtain (assuming $N_t = N_r = N$):

$$\begin{aligned} C_H &= \log_2 \left[\det \left(I_N + \frac{\rho}{N} (\mathbf{\Psi}_r^{1/2} \mathbf{H}_{iid} \mathbf{\Psi}_t^{1/2})(\mathbf{\Psi}_r^{1/2} \mathbf{H}_{iid} \mathbf{\Psi}_t^{1/2})^+ \right) \right] \\ &\approx \log_2 \left[\det \left(\frac{\rho}{N} (\mathbf{\Psi}_r^{1/2} \mathbf{H}_{iid} \mathbf{\Psi}_t^{1/2})(\mathbf{\Psi}_r^{1/2} \mathbf{H}_{iid} \mathbf{\Psi}_t^{1/2})^+ \right) \right] \end{aligned} \tag{2.3.2}$$

where the approximation is valid at high SNR. Further simplifications of the above expression is possible by using various properties of the determinant and the logarithm operators such as $\det(AB) = \det(A)\det(B)$, $\det(A^+) = \det(A)$, and $\log_2(AB) = \log_2(A) + \log_2(B)$, where A and B are square matrices. The expression in (2.3.2) then reduces to:

$$C_H \approx \log_2 \left[\det \left(\frac{\rho}{N} \mathbf{H}_{iid} \mathbf{H}_{iid}^+ \right) \right] + \underbrace{\log_2 \det(\mathbf{\Psi}_t) + \log_2 \det(\mathbf{\Psi}_r)}_{=\eta(\mathbf{\Psi}_t, \mathbf{\Psi}_r) \leq 0}. \tag{2.3.3}$$

The quantity $\eta(\mathbf{\Psi}_t, \mathbf{\Psi}_r)$ above the brace in the right hand side of (2.3.3) represents the combined contribution of the TX and RX correlation. Since the sum of eigenvalues of $\mathbf{\Psi}_t$ is limited by N, their product which is equal to $\det(\mathbf{\Psi}_t)$ is ≤ 0. The same holds true for $\det(\mathbf{\Psi}_r)$. So the net effect of TX and/or RX correlation on MIMO capacity is to reduce it by a deterministic negative value that is equal to

$$\eta(\mathbf{\Psi}_t, \mathbf{\Psi}_r) = \log_2 \det(\mathbf{\Psi}_t) + \log_2 \det(\mathbf{\Psi}_r) \text{ bit/s/Hz}. \tag{2.3.4}$$

To gain further insight into the effect of correlation on MIMO capacity, we consider here the 2×2 MIMO channel scenario. In this case, and assuming that $\mathbf{\Psi}_t = \mathbf{\Psi}_r = \mathbf{\Psi}$, we have:

$$\mathbf{\Psi} = \begin{bmatrix} 1 & \psi \\ \psi & 1 \end{bmatrix}. \tag{2.3.5}$$

The correlation coefficient $\psi \in [0, 1)$ (0 for no correlation and 1 for full correlation). Substituting the determinant of $\mathbf{\Psi}$ in (2.3.4), we get

$$\eta(\psi) = 2\log_2(1 - \psi^2). \tag{2.3.6}$$

Since we are interested in knowing how the amount of correlation behaves with linear increase in ψ, we take the derivative of $\eta(\Psi)$ to obtain:

$$\frac{d}{d\psi}\eta(\psi) = -\frac{4}{\ln(2)}\left(\frac{\psi}{1 - \psi^2}\right). \tag{2.3.7}$$

Note that for values of ψ close to 0, the term ψ^2 becomes negligible and as a result the rate of decrease of $\eta(\psi)$ is linear with a slope equal to $-\frac{4}{\ln 2}$. When ψ is close to 1 the rate of decrease is inversely proportional to a quadratic function in ψ.

To confirm our analysis, we conduct here some Monte Carlo simulations for a 2×2 MIMO system. Figure 2.6 shows the results for various values of ψ using simulation and the analytical expression in (2.3.6). Various conclusions can be drawn from this figure. First, as expected, the effect of correlation is well approximated by (2.3.6) which gives an exact match at high SNR, i.e. $\psi \geq 40$dB. Second, the approximation is still valid for practical SNR levels ($\psi \leq 20$dB) and $\psi \leq 0.2$. Third, for $\psi \leq 0.7$ the degradation in capacity due to correlation is almost linear as opposed to when $\psi > 0.7$ after which a sharp decrease in capacity occurs, again validating our previous analysis. The good news is that the effect of correlation at practical SNR levels (i.e. $\rho < 20$dB) is less dramatic than that predicted by the high-SNR approximation.

One can ask the legitimate question of how correlation affects the capacity of MIMO channels with greater number of antennas. To do this we propose to use the exponential correlation matrix. In this case, the elements of the correlation matrix are found as:

$$[\mathbf{\Psi}]_{i,j} = \psi^{|i-j|} \quad i,j \in \{1,2,....,N\} \text{ and } \psi \in [0,1). \tag{2.3.8}$$

It is clear from the above expression that the farther apart the antennas are from each other the smaller is the correlation coefficient. This agrees well with both intuition and real-world measurements. Another nice property of the matrix in (2.3.8) is the fact that its determinant can be shown to have a simple closed-form expression:

$$\det(\Psi) = (1-\psi^2)^{N-1}. \tag{2.3.9}$$

From the expressions in (2.3.8) and (2.3.9) it is easy to see that this matrix is a generalization to the 2×2 one in (2.3.5). Assuming again that $\Psi_t = \Psi_r = \Psi$ the

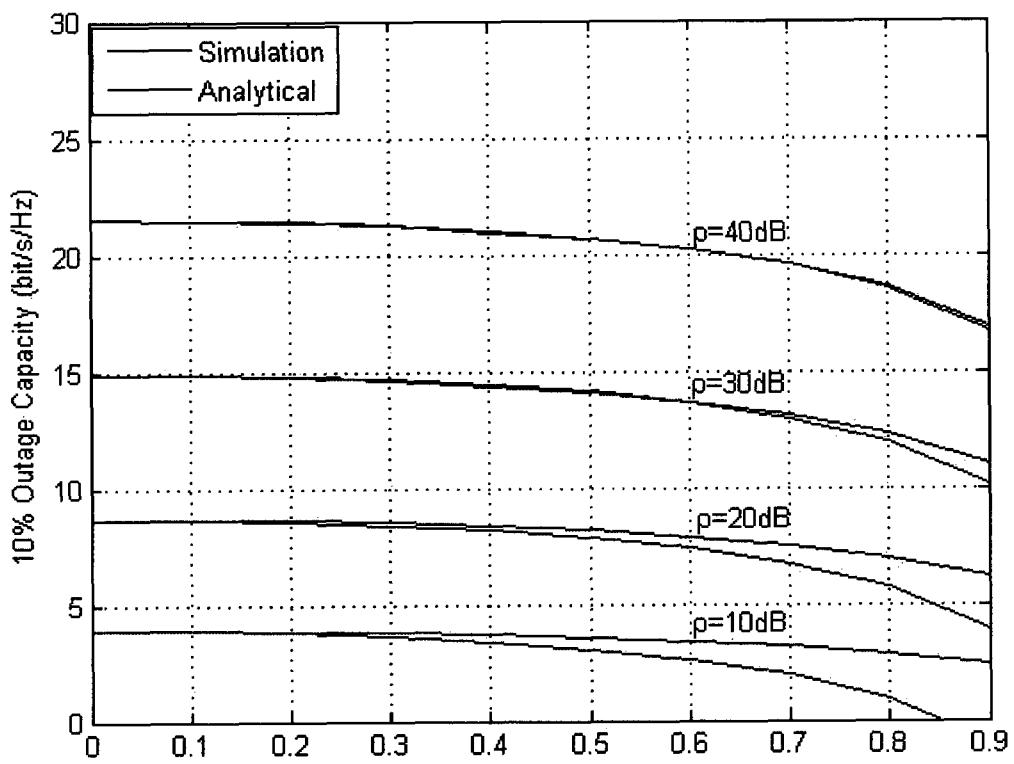

Figure 2.6: Effect of correlation on a 2×2 MIMO channel capacity for $\rho = 10, 20, 30, 40$ dB

total effect of correlation is:

$$\eta(\psi) = 2(N-1)\log_2(1-\psi^2). \tag{2.3.10}$$

To see the effect of a linear increase of ψ on the general $N \times N$ MIMO capacity, we take the derivative of $\eta(\psi)$ with respect to ψ to obtain the following slope:

$$\frac{\partial}{\partial\psi}\eta(\psi) = -\frac{4}{\ln(2)}(N-1)\left(\frac{\psi}{1-\psi^2}\right). \tag{2.3.11}$$

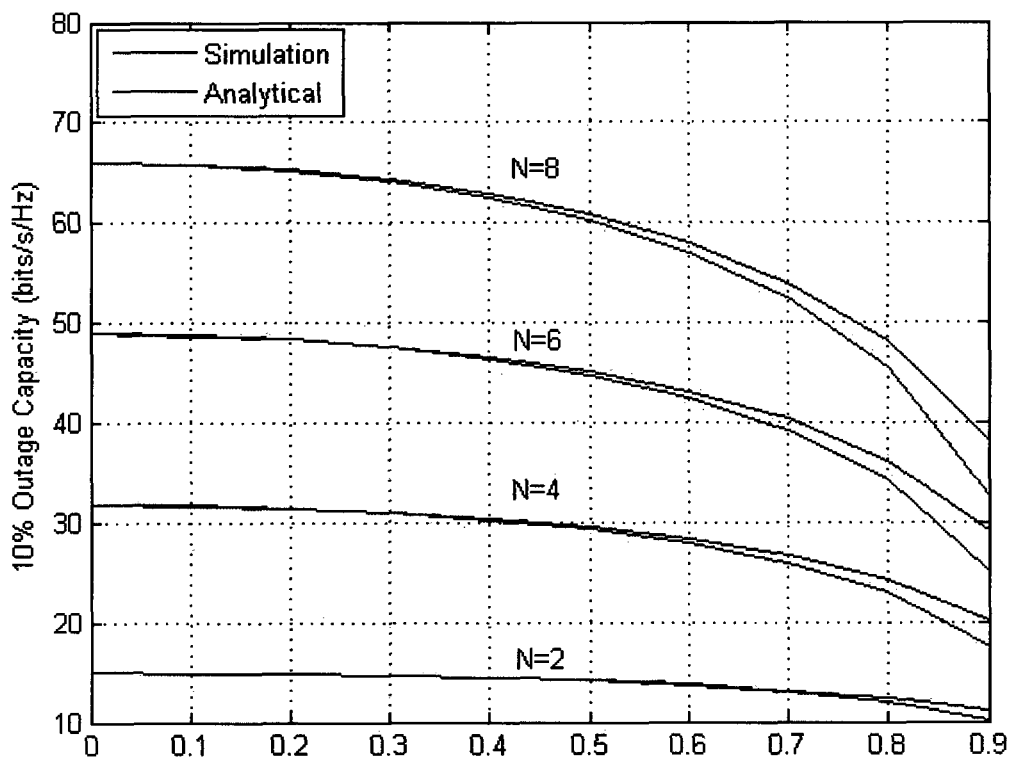

Figure 2.7: Effect of correlation on a $N \times N$ MIMO capacity for $\rho = 30$ dB

For values of ψ close to 0, the term ψ^2 can be simply ignored. In this case, the slope becomes $-\frac{4}{\ln(2)}(N-1)$. As in the case of the 2×2 channel, the slope for small values of ψ decreases linearly but this linear decrease becomes worse with increasing N. When ψ is close to 1 the slope decreases quadratically (because ψ^2 can not be ignored any more) and the sharpness of this exponential decent becomes worse as N increases. Figure 2.7 confirms these last points by showing the capacity for different values of N.

To summarize, correlation generally reduces the capacity offered by the MIMO

channel. Because the reduction in capacity due to correlation is additive (versus multiplicative), we conclude that correlation reduces the diversity contribution to capacity and as such does not affect the multiplexing gain. The effect of correlation becomes worse with increasing ψ and N. It is also to be noted that the effect of correlation is negligible for $\psi \leq 0.2$. That explains why MIMO channels with $\psi \leq 0.2$ can be considered uncorrelated for all practical purposes. Indoor-to-indoor MIMO channels are representatives of such channels.

2.3.2 Keyhole MIMO capacity

The phenomenon of keyhole (or pinhole) can occur in many natural environments even in the absence of TX and/or RX correlation [12]. As we will show shortly, keyhole has a far worse negative effect on MIMO capacity than correlation does.

An illustration of what happens in a typical keyhole scenario is depicted in Figure 2.8. Basically, signals transmitted from all antennas pass through a virtual hole. On the RX side, the antennas see the TX signals only through this hole. In practice, this situation can occur when there is a single dominant scatterer or the distance between the transmitter and receiver is large and the radius of the effective scatterers is extremely small compared to that distance. As can be easily derived from Figure 2.8, the overall channel matrix has the following structure:

$$\mathbf{H}_p = \begin{bmatrix} \alpha_1\beta_1 & \alpha_2\beta_1 & \dots & \alpha_N\beta_1 \\ \alpha_1\beta_2 & \alpha_2\beta_2 & \dots & \alpha_N\beta_2 \\ \vdots & \vdots & \ddots & \vdots \\ \alpha_1\beta_N & \alpha_2\beta_N & \dots & \alpha_N\beta_N \end{bmatrix} \tag{2.3.12}$$

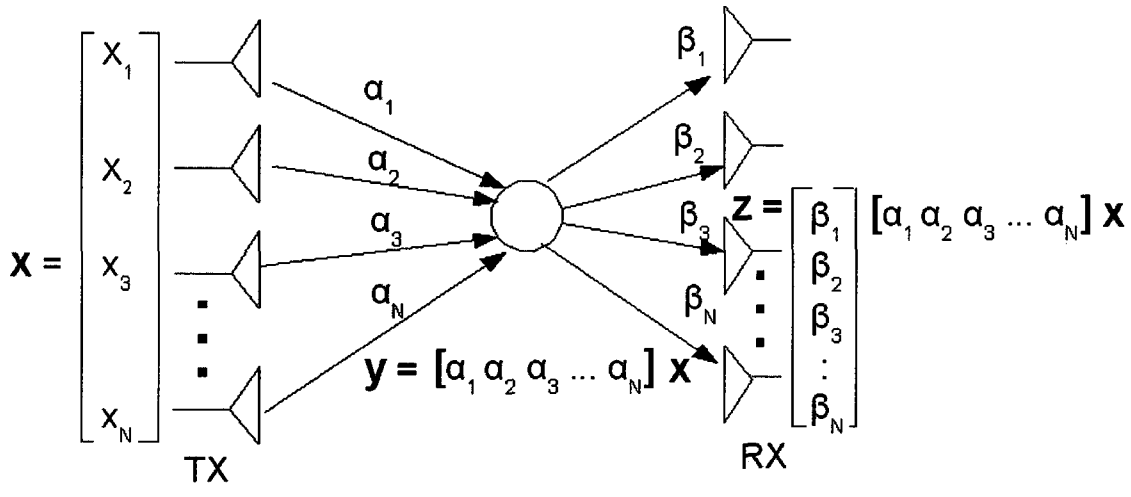

Figure 2.8: Illustration of keyhole effect on $N \times N$ MIMO channel

where the complex channel gains $\{\alpha_j, \beta_i\}$ are assumed to be i.i.d. with $\mathcal{N}(0, 0.5)$ in each dimension.

Since every column of the above channel matrix is linearly dependent on every other column, we conclude that the rank of the pinhole channel is equal to one. As a result there is a single non-zero eigenvalue δ equal to $tr(\mathbf{H}_p) = \sum_{i=1}^{N} \alpha_i \beta_i$, where $tr(.)$ is the trace operator. The latter result follows from the property of any square matrix which states that the trace of such matrix equals the sum of its eigenvalues.

In order to compute the keyhole capacity of the MIMO channel we need to compute the eigenvalues of the Wishart matrix $\mathbf{\Xi} = \mathbf{H}_p \mathbf{H}_p^+$. This is easily done by realizing that $\mathbf{\Xi}$ has also rank one and as a result only possesses a single non-zero positive eigenvalue $\delta^{keyhole}$ equal to $tr(\mathbf{\Xi}) = \sum_j^N |\alpha_j|^2 \sum_i^N |\beta_i|^2 = \|\alpha\|^2 \|\beta\|^2$. With this information we can easily find the instantaneous capacity of a keyhole channel by substituting the value of $\delta^{keyhole}$ in (2.2.2) to obtain:

$$C_{H_p} = \log_2(1 + \frac{\rho}{N}\|\alpha\|^2\|\beta\|^2). \tag{2.3.13}$$

In [43], it shown that a tight upper bound on the ergodic keyhole capacity $\bar{C}_{H_p}$ is:

$$\bar{C}_{H_p} \leq \log_2(1 + N\rho). \tag{2.3.14}$$

At high SNR the above expression can be rewritten as:

$$\begin{aligned} \bar{C}_{H_p} &\leq \log_2(N\rho) \\ &\leq \log_2(\frac{N^2\rho}{N}) \\ &\leq \log_2(\frac{\rho}{N}) + \log_2 N^2. \end{aligned} \tag{2.3.15}$$

Comparing (2.3.15) to (2.2.7) shows that, in keyhole scenarios, MIMO looses its multiplexing capability (i.e. multiplexing gain is equal to 1) but still contributes to diversity. It should be mentioned that this contribution to diversity becomes marginal with increasing N since $\frac{d(\log_2 N^2)}{dN} = 0$ as $N \to \infty$.

Figure 2.9 shows the ergodic capacity of an $N \times N$ keyhole channel. The simulated curves were obtained by resorting to (2.2.1) and by running Monte Carlo simulations using 10,000 independent channel realizations. Each element of the channel matrix is a product of two random variables as in (2.3.12). Each of these variables has Rayleigh distribution with zero mean and variance of 0.5 in each dimension. The analytical curves were directly obtained from (2.3.15). As expected, the multiplexing gain (i.e. the slope of the capacity curve) remains constant and is equal to one regardless of the number of antennas used. Also, it is clear from the same figure that the only advantage we get from increasing the number of antennas in this special case is to

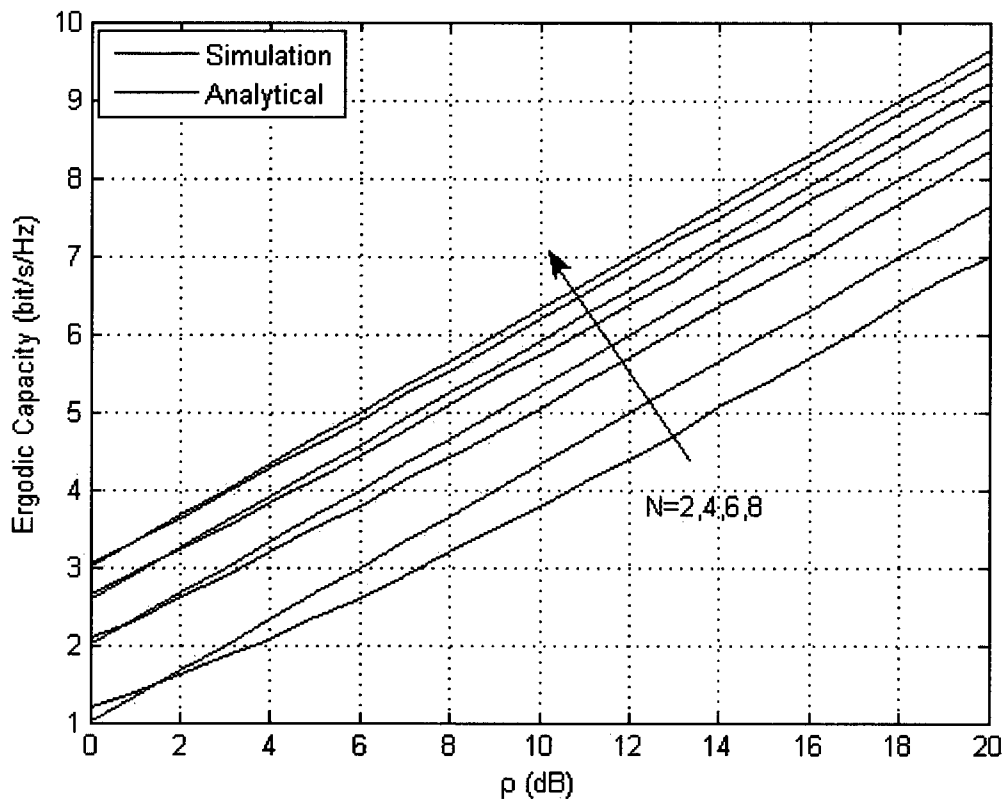

Figure 2.9: Ergodic keyhole capacity of $N \times N$ MIMO channel

obtain an exponentially decreasing SNR-independent shift in capacity due to the diversity contribution.

To summarize, while channel correlation only reduces the diversity contribution to capacity, the keyhole phenomenon affects both the multiplexing and diversity contributions to capacity, albeit in different proportions: the multiplexing contribution is reduced by a factor of N while the diversity contribution is roughly reduced by a factor of $\frac{N}{3}$. This latter result can be easily verified using the following approximation: $\log_2 N! \approx \frac{N}{3} \log_2 N^2$.

2.4 Distributed Scattering Model

The MIMO channel models presented in the previous sections do not allow us to directly tie the physical array parameters to the effect of correlation. Many models have been proposed to address this issue. The Distributed Channel Model, hereafter referred to as DSM, was first introduced in [26] for outdoor fixed-wireless MIMO channels. It is a generalization to the Kronecker model presented in a previous section. The main advantage of DSM is that it captures the effects of correlation and keyhole simultaneously. Also, the correlation matrix is obtained through the specification of physical parameters such as the inter-element spacing (IES) and the distance separating the TX and RX ends of the wireless link, among others.

DSM assumes the existence of effective scatterers at both ends of the wireless link. Figure 2.10 shows an illustration of this model. The S TX scatterers collect the reflections of all signals transmitted from N_t TX antennas and radiate them to an equal number of RX virtual scatterers which in turn send these signals to N_r RX antennas. The distance separating the TX and RX scatterers is denoted by D. The TX and RX angle spreads Θ_t and Θ_r will heavily influence the amount of correlation between the TX and RX signals, respectively. As we will show later the same can be said of the TX and RX inter-element spacing d_t and d_r.

In DSM the channel matrix $\mathbf{H}$ is obtained as:

$$\mathbf{H} = \frac{1}{\sqrt{S}} \mathbf{K}_r^{1/2} \mathbf{H}_{r,iid} \mathbf{K}_s^{1/2} \mathbf{H}_{t,iid} \mathbf{K}_t^{1/2} \tag{2.4.1}$$

where $\mathbf{\Psi}_t = (\mathbf{K}_t^{1/2})(\mathbf{K}_t^{1/2})^+$, $\mathbf{\Psi}_r = (\mathbf{K}_r^{1/2})(\mathbf{K}_r^{1/2})^+$ and $\mathbf{\Psi}_s = (\mathbf{K}_s^{1/2})(\mathbf{K}_s^{1/2})^+$ are the TX, RX and scatterers correlation coefficient matrices, respectively, $\mathbf{H}_{t,iid}$ and $\mathbf{H}_{r,iid}$ are i.i.d. Gaussian matrices. Note that if any of the matrices just mentioned has a

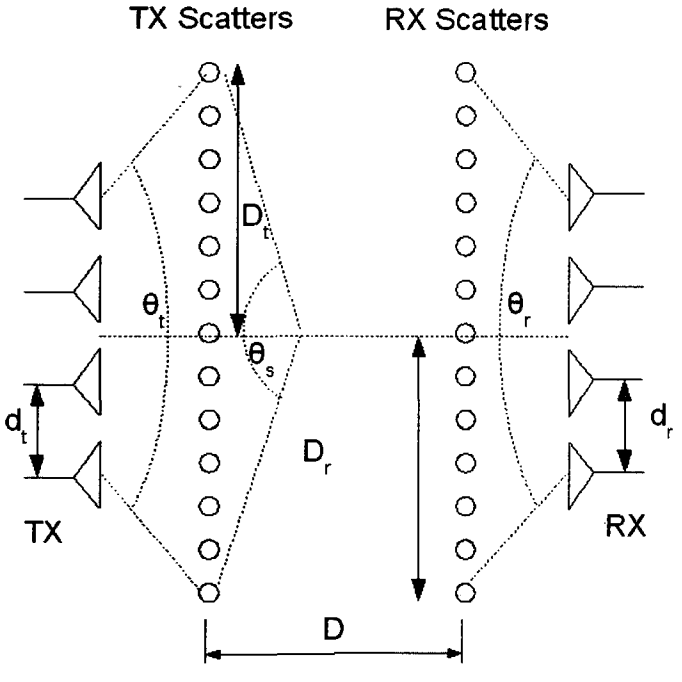

Figure 2.10: The distributed scattering model.

rank deficiency the same can be said of $\mathbf{H}$. Rank deficiency occurs in $\mathbf{K}_s$ when D is much larger than the scatterers aperture, i.e. $D >> 2D_t$ or $D >> 2D_r$. The same can be said of matrix $\mathbf{K}_t$ (or $\mathbf{K}_r$) when d_t (or d_r) is not large enough.

For a uniform linear array (ULA) of inter-element spacing d and a uniform distribution for TX and RX scatterers, element K_{lm} of $\mathbf{K}$ can be computed as [26]:

$$K_{lm} = \frac{1}{S} \sum_{i=-\frac{S-1}{2}}^{i=\frac{S-1}{2}} e^{-2\pi j}(m-l)\frac{d}{\lambda}\cos(\frac{\pi}{2}+\theta_i) \tag{2.4.2}$$

where λ is the wavelength and θ_i is the angle of arrival (AOA) with respect to the horizontal axis for scatterer i.

While DSM was derived for the outdoor MIMO channel, nothing prevents its use in

Parameter	**Symbol**	**Typical Indoor**	**Typical Urban**	**Typical Rural**
Number of TX antennas	N_t	4	4	4
Number of RX antennas	N_r	4	4	4
Number of TX scatterers	S_t	21	21	21
Number of RX scatterers	S_r	21	21	21
IES for TX array (in multiple of λ)	d_t	0.5	0.5	0.5
IES for RX array (in multiple of λ)	d_r	0.5	5	5
Half TX array aperture (m)	D_t	10	35	50
Half RX array aperture (m)	D_r	10	3.1	8
TX angle spread (degrees)	Θ_t	126.8°	120.5°	28°
RX angle spread (degrees)	Θ_r	126.8°	7°	1.83°
Distance between TX and RX arrays (m)	D	50	1000	10000

Table 2.1: DSM parameters used to produce Figure 2.11

simulating indoor MIMO channels as long as no direct line-of-sight is assumed. Also, the original model assumes that the angles of arrivals θ_i have a uniform distribution. Measurements in [2] indicate the angles of arrivals in outdoor channels have rather a Gaussian distribution with typical angle spreads of 5-7 degrees at the Base station (BS). Using computer simulations we found that DSM can also be used when Gaussian distribution and small values (below 10 degrees) of θ_i are assumed, which, as we mentioned above, is typically the case at the BS site in outdoor environments.

2.5 Real-World MIMO Capacities

The DSM was used to simulate the uplink 10% outage capacity of indoor and outdoor 4×4 MIMO channels. The DSM parameters used are listed in Table 2.1. The TX and RX correlation matrices associated with these parameters were computed and

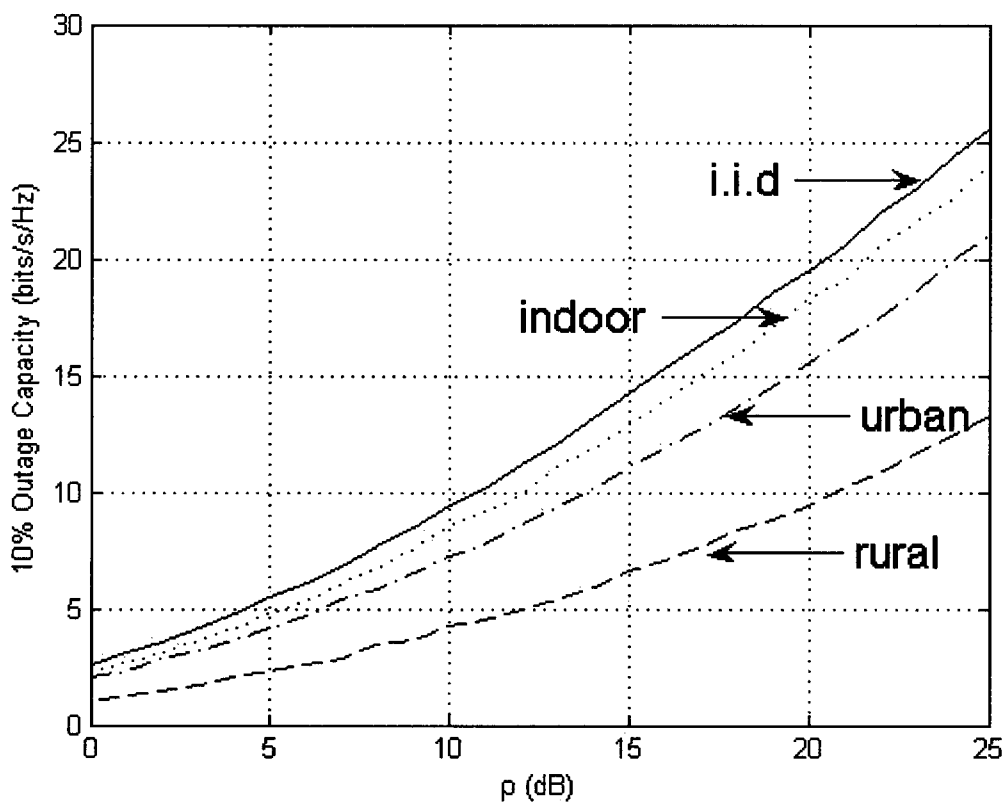

Figure 2.11: 10% outage capacity of a 4 x 4 MIMO channel in different environments obtained by simulation using DSM.

included in Appendix A. Figure 2.11 shows the simulation results. The curves shown in Figure 2.11 confirm the following about DSM:

- Matching real-world measurements such as the one in [21], MIMO capacities are in general lower than the i.i.d. Gaussian case with the indoor channels representing the highest (close to 95% of the i.i.d. case) dormant MIMO capacity.
- Simulated MIMO capacity obtained for a typical urban environment is around 80% of the i.i.d. This matches field measurements that were reported in [21].

- The low MIMO capacity obtained for a typical rural environment, where poor scattering exists and where the distance between BS and user equipment (UE) is large, confirms the "keyhole" effect observed in real-world measurements. In [33], capacity measurements were conducted in a rural setting. It was shown that unless the array spacing is very large ($d_t, d_r \gg \lambda$) on both ends of the wireless link, the capacity gain from MIMO processing is nowhere near the expected one. In fact, as shown in Figure 2.12, the outage capacity as predicted by DSM decreases exponentially with distance until it reaches the keyhole capacity at extremely large distances (not shown in the figure). This finding and the fact that the SNR generally decreases with distance tells us that emphasis should be placed on minimizing error probability rather than multiplexing gain when the UE is relatively far away from the BS.

- As mentioned in Section 2.3.1, correlation does not affect the multiplexing gain of the MIMO channel. This is evidenced by the fact that at high SNR the slope of the curves for the various outage capacities still increases at a rate of roughly N bits for 3dB increase in SNR.

From the above we conclude that DSM accurately models the indoor and outdoor frequency-flat MIMO channels.

2.6 Effect of Inter-Element Spacing on MIMO Capacity

Before we conclude this chapter it is rather important to examine the effect of inter-element spacing on the correlation as it is the only parameter that is under direct control of the system designer. For doing so we will resort to (2.4.2) for computing

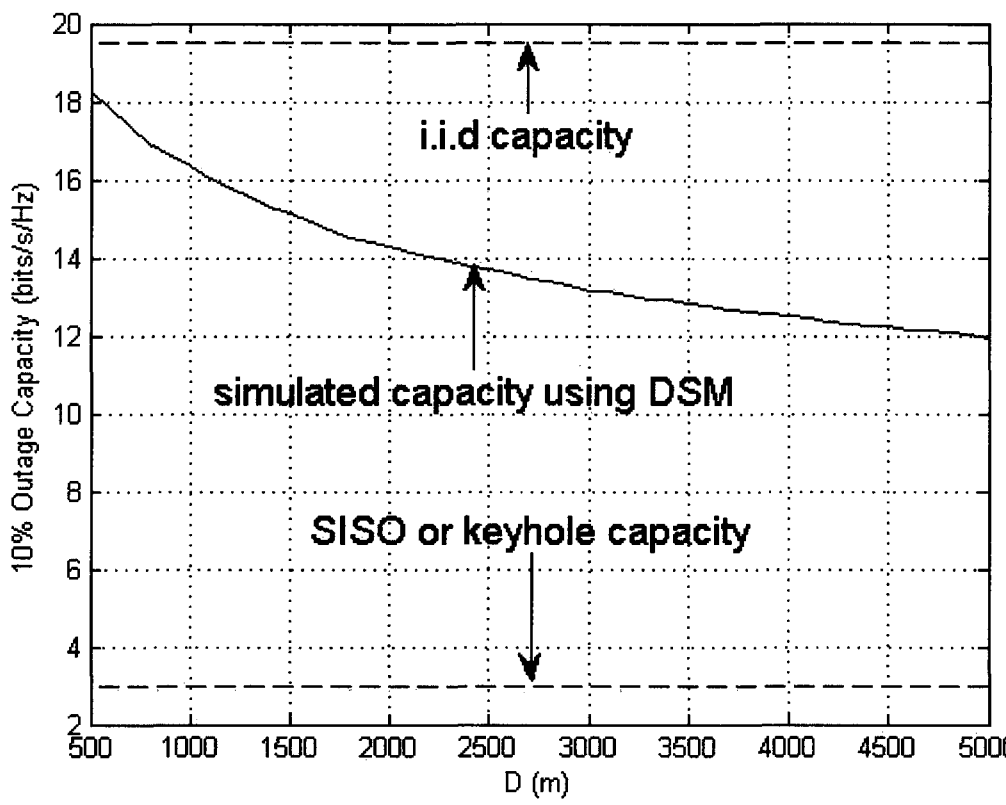

Figure 2.12: Effect of distance between BS and UE on a 4×4 MIMO capacity at $\rho = 20$dB

the correlation coefficient between a pair of antennas. Figure 2.13 shows the results for a case when poor scattering exists (such as at the BS in an outdoor environment) and the results for a case where rich scattering exists (such as in indoor environments or around the UE in outdoor environments). We clearly see the dependence of correlation on the random distribution assumed: the curve that belongs to uniform distribution decays faster while it suffers from residual correlation even at large IES. When Gaussian distribution is assumed, the rate of decay is slower but now the tail of

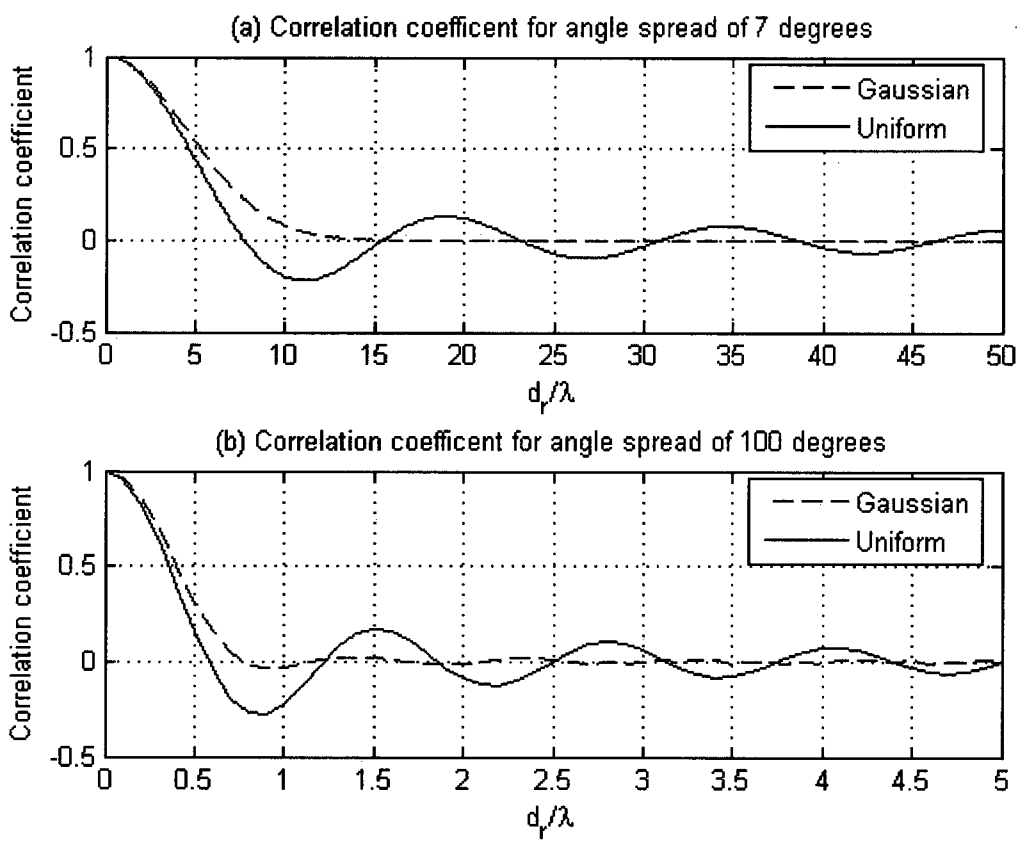

Figure 2.13: Correlation coefficient for a typical (a) poor scattering environment and (b) rich scattering environment

the curve vanishes rather quickly. In the worst case scenario, we can see that in order to achieve sufficient decorrelation (i.e. $K_{lm} \leq 0.3$) in poor scattering environments an inter-element spacing of at least 8λ is required while an inter-element spacing of 0.5λ is sufficient in rich scattering environments. Luckily, and for a typical uniform linear array (ULA) of 4 elements, the physical space (such as on rooftops) available

for the BS ULA allows for a minimum separation of 8λ between antennas[1]. Depending on the carrier frequency and the physical dimensions of the UE, a ULA of 2 to 4 elements with minimum separation of 0.5λ between elements is practically feasible at sub-10GHz frequencies. Other array geometries may allow for using a higher number of array elements but that depends on the physical shape of the UE. One can, for instance, think of a laptop, where a planar array can be used instead.

[1]For a 2.4GHz carrier frequency this corresponds to an overall ULA size of 3m. The size of the ULA shrinks even further at higher frequencies.

Chapter 3

MIMO Detection Algorithms

The previous chapter discussed the tremendous dormant capacity offered by the MIMO channel. The main challenge is then how to achieve, or at least approach, this capacity using algorithms that have a reasonable implementation complexity. To address this issue, researchers have proposed many detection algorithms mainly derived from space-time coding theory. Before we dive into details, it is important to restate few important assumptions.

1. Rayleigh flat fading is assumed.

2. It will also be assumed that $\mathbf{H}$ is quasi-static, meaning that it remains invariant during the entire frame period and only changes from one frame to the next (i.e. it undergoes block fading only). This is a very reasonable assumption since even in the most hostile mobile environments, where vehicular speeds can reach up to 300 Km/h (one can think of high-speed trains), the worst Doppler frequency does not exceed 500 Hz at a typical 1.8 GHz carrier frequency. The channel coherence time T_c, in this worst case scenario, is equal to $0.5/500 = 1$ ms [44]. At a typical transmission rates of 20M symbols/s, the symbol duration, in this case is equal to $T_s = 50ns$, which is several orders of magnitude less than 1 ms.

3. Unless otherwise stated, it will be assumed that the channel matrix $\mathbf{H}$ has been perfectly estimated, possibly through the use of a training sequence transmitted at the beginning of each frame. One possible channel estimation method is described in [36].

4. The focus in this chapter and next will be on the case where $N_r = N_t = N$. The main reason behind this choice is twofold. First, while some MIMO algorithms perform better when $N_r > N_t$, a plausible assumption on the uplink, this is not the case in the downlink where generally higher data rates are required and the small form-factor of the UE makes satisfying this condition rather impossible. Second, most 4G standards define MIMO profiles where the same number of TX and RX antennas are used.

When designing MIMO detection algorithms, there are usually two competing objectives to meet: capacity maximization, generally achieved by maximizing the use of the various degrees of freedom (i.e. multiplexing capability), and reliability maximization, generally achieved by maximizing the use of the various independent channel paths to transmit the same information (i.e. diversity capability). One can ask the legitimate question of whether there is a tradeoff between these two extremes. This problem has been addressed in [60]. The following section discusses this topic in more details.

3.1 Diversity versus Multiplexing Tradeoff

The optimal tradeoff at infinite SNR levels was derived in [60], where a formal mathematical SNR-dependent definitions of these quantities, namely *multiplexing gain* and

diversity gain, were introduced. The multiplexing gain G_m and the diversity gain G_d for a given rate r_H (in bit/s/Hz) are respectively defined as [60]:

$$G_m = \lim_{\rho \to \infty} \frac{r_H(\rho)}{\log_2(\rho)} \tag{3.1.1}$$

$$G_d = -\lim_{\rho \to \infty} \frac{\log_2(P_{out}(r_H))}{\log_2(\rho)}. \tag{3.1.2}$$

Note that in this definition the rate r_H is allowed to change with SNR. In other words, the multiplexing gain is the rate of increase of the throughput with SNR at a given probability of outage while the diversity gain is equal to the negative of the slope of the probability of outage curve at infinite SNR on a log-log scale.

In [60], it was shown that the optimal diversity-multiplexing tradeoff (DMT) curve is piecewise-linear connecting the points $(m, (N_t - m)(N_r - m))$, where $m = 0, 1, \ldots, min(N_t, N_r)$. This curve is plotted in Figure 3.1. For comparison purposes, we also include on the same figure representative finite-SNR DMT curves that we will discuss shortly.

Before we proceed any further we would like to emphasize the difference between commonly used terms in the literature such as "diversity order" versus "diversity gain", and "degrees of freedom" versus "multiplexing gain". Somehow related but not exactly the same, these terms have been used interchangeably.

The diversity order is a constant number and it refers to the total number of independent paths offered by the MIMO channel. For instance, in a 4×4 i.i.d. flat MIMO channel, there is a total number of 16 independent scalar channels. The diversity gain as defined in (3.1.2) is a variable number that refers to the rate of decrease of the probability of outage P_{out} with SNR and r_H. For i.i.d. MIMO channels this gain is equal to the diversity order at infinite SNR (while keeping r_H fixed), i.e.

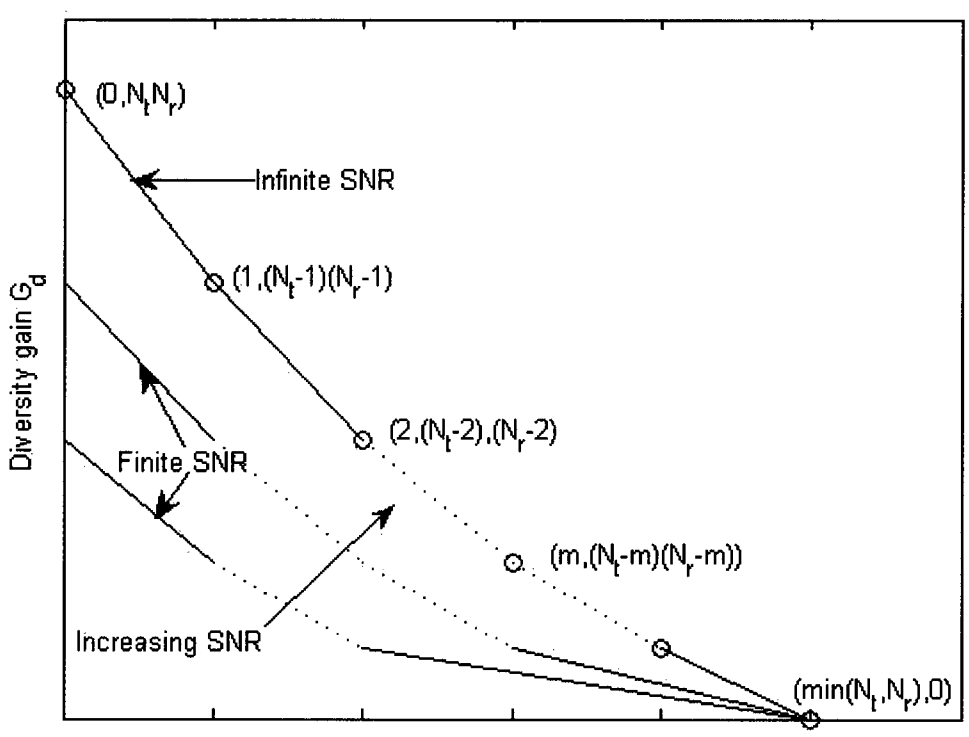

Figure 3.1: Exact infinite-SNR optimal tradeoff curve for multiplexing gain G_m and diversity gain G_d. Optimal tradeoff curves at finite SNR are also illustrated.

the point $(0, N_t N_r)$ on the DMT curve. In other words, we can only take advantage of the maximum diversity gain (i.e. the diversity order) if we keep the rate r_H constant as SNR increases to infinity. In this case we clearly see that $G_m \to 0$. Because of this asymptotic relationship between diversity gain and diversity order many authors use the former term to refer to the latter one and vice versa.

The degrees of freedom is a constant number and refers to the number of $N_{min} = \min(N_t, N_r)$ in i.i.d. MIMO channels. The multiplexing gain G_m, as defined in (3.1.1),

is a variable number that refers to the ability of the i.i.d. MIMO channel to support a certain incremental rate with increasing SNR, which at infinite SNR and fixed P_{out} reaches a maximum equal to the degrees of freedom offered by the MIMO channel, i.e. the point $(min(N_t, N_r), 0)$ on the DMT curve. In this extreme case the probability of outage is kept constant while the rate G_m increases by $min(N_t, N_r)$ for every 3dB increase in SNR at infinite SNR. Because of this asymptotic relationship between multiplexing gain and degrees of freedom many authors use the former term to refer to the latter one and vice versa.

The other points on the DMT curve refer to how much diversity gain can be optimally traded off for multiplexing gain. In other words, we say that a certain space-time code is optimal in DMT sense if its behavior at infinite SNR levels follows that of the i.i.d. MIMO channel as described by the DMT curve. The 2×2 space-time codes presented in [58][18] are examples of space-time codes that achieve the infinite-SNR DMT curve. For illustration purposes, Figure 3.2 shows the various optimal probability of outage curves for a 2×2 i.i.d. MIMO channel. These curves were obtained by Monte Carlo simulations using (2.2.1) and (2.2.17), and by allowing the rate $r_H(\rho)$ to change according to (3.1.1). We clearly see from the figure that as the multiplexing gain G_m increases, the rate of decay of P_{out} decreases and vice versa.

Finite-SNR DMT

The infinite-SNR DMT curve is a useful benchmark for comparing the diversity-multiplexing performance of various space-time codes. Unfortunately, it is only valid at extremely high SNR levels (and correspondingly at extremely low probability of outage). Practical SNR and probability of outage levels are typically 3-20dB and 0.1-0.01, respectively [37]. Also, the fact that a certain space-time code is optimal

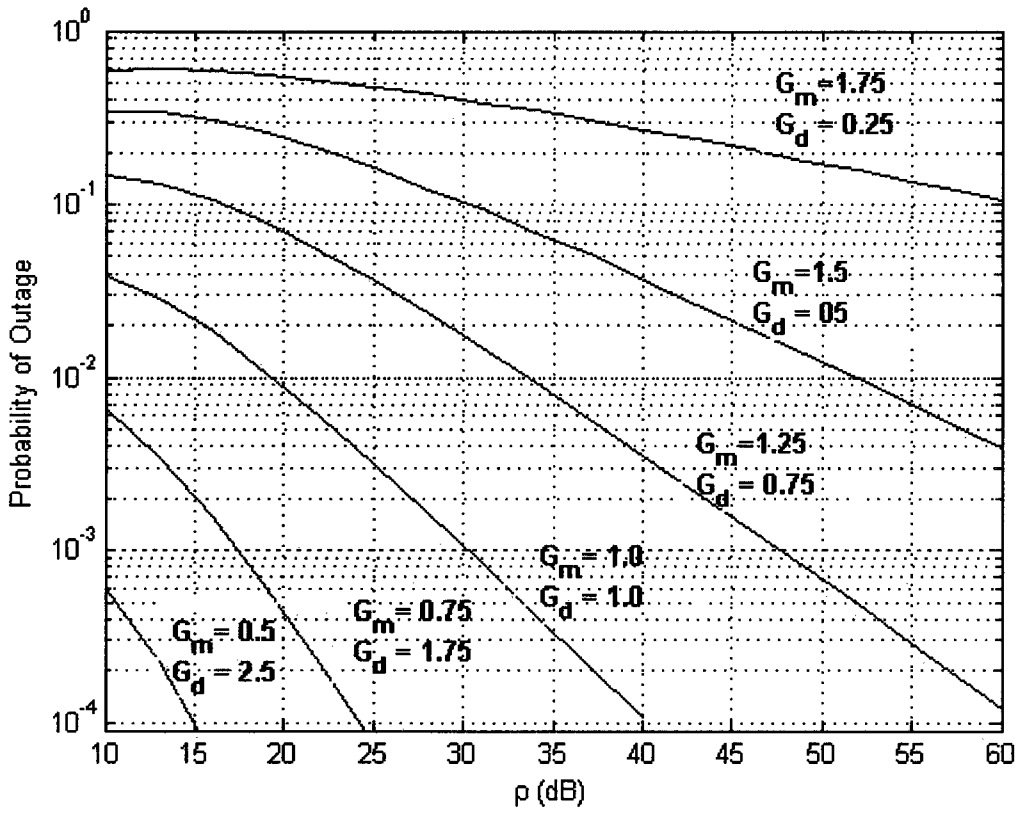

Figure 3.2: Optimal probability of outage curves for various (G_m, G_d) points of the infinite-SNR DMT curve for a 2×2 i.i.d. MIMO channel.

in the infinite-SNR DMT sense does not guarantee that it is also optimal at finite-SNR. For instance, the finite-SNR multiplexing and diversity gains of two space-time codes, that are optimal according to the infinite-SNR DMT, may exhibit a completely different convergence behavior.

We have two tools that allow us to examine the behavior of the MIMO channel (or any space-time code) at finite-SNR. The first one uses the probability of outage curves at different fixed rates and the second uses the finite-SNR DMT curve derived

in [37][35]. As we will see next the first tool is a practical one especially in modern communication systems where adaptive modulation is employed while the second one is more useful from an analysis point of view.

The first tool looks at how the various probability of outage curves behave for different fixed rates. Figure 3.3 shows the probability of outage curves for a 2×2 i.i.d. MIMO channel for $r_H = 2, 4, 6, 8$ bit/s/Hz. It is clear that all curves have the same slope, i.e. $G_d^{max} = 4$ at sufficiently high SNR. In other words, for every 10dB increase in SNR, P_{out} decreases by $10^{G_d^{max}} = 10^4$ (or alternatively requires 2.5dB per decay). At low SNR we definitely see that we need more than 2.5dB per decay. G_m^{max} can be visualized by looking at the incremental SNR needed to increase the rate by $min(N_t, N_r) = 2$ bit/s/Hz. We notice that this incremental SNR difference very much depends on the the probability of outage levels: at high probability of outage levels like $P_{out} = 10^{-1}$ this optimum value is approached rather quickly as only 3.2dB is needed to go from $r_H = 6$ to $r_H = 8$ while this value is 4.5dB at $P_{out} = 10^{-3}$. This behavior which can not be explained by the DMT curve is explained further in [4]. Using the above analysis we can then say that a space-time code is optimal in the finite-SNR DMT sense if its probability of outage curves at different rates and finite SNR behave the same as that of the MIMO channel.

The second tool resorts to approximations of the finite-SNR DMT curves that were first derived in [37][35], where slightly different definitions of the multiplexing and diversity gains were introduced. These new definitions converge to the SNR-asymptotic ones at infinite SNR levels. According to [35], and using size-asymptotic derivations and approximations, an approximate expression for the finite-SNR DMT when $N_t = N_r = N$ is:

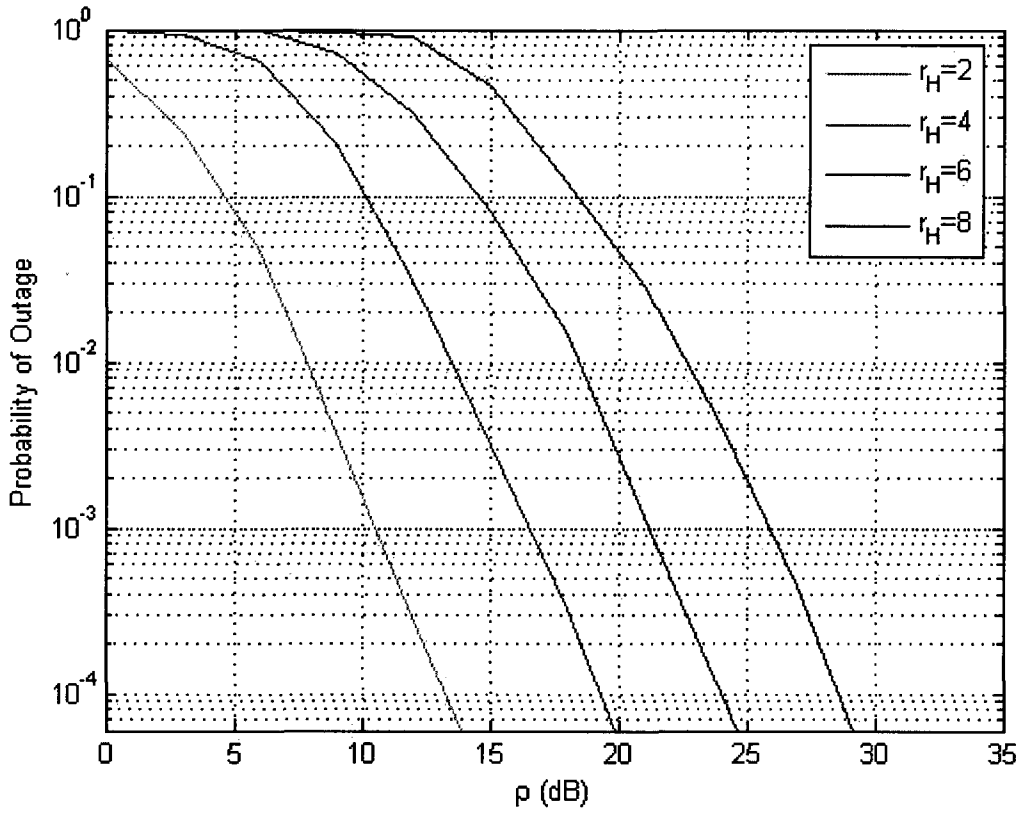

Figure 3.3: Probability of outage curves of a 2×2 i.i.d. MIMO channel for different fixed rates $r_H = 2, 4, 6, 8$ bit/s/Hz

$$G_d(\rho) \approx (N - G_m)^2 \left[1 - \frac{1}{2\sqrt{\rho}}\right], \quad G_m \in [0, N]. \tag{3.1.3}$$

Even though the above expression slightly overestimates the actual finite-SNR diversity gain for $G_m < 1$ [35], it does give a useful insight of the effect of finite SNR on the diversity gain: the term $\frac{1}{2\sqrt{\rho}}$, which represents the contribution of finite SNR, decays rather quickly with increasing SNR to reach arbitrary small values for $\rho \geq 15$dB. Figure 3.4 shows various finite-SNR DMT curves for a 2×2 MIMO channel.

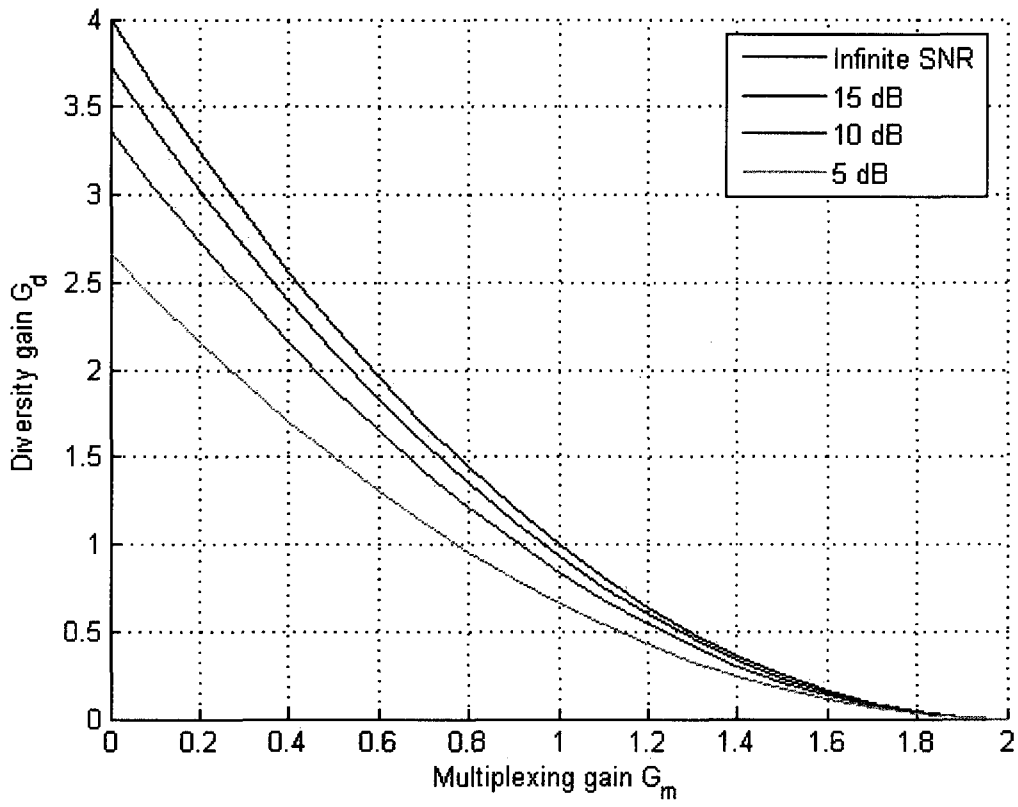

Figure 3.4: Various finite-SNR DMT curves for a 2×2 i.i.d. MIMO channel. The infinite-SNR DMT curve is also shown.

To be consistent with the first tool presented above, and since the DMT curve does not factor in the rate, we conclude that the above finite-SNR DMT results are only valid for high probability of outage levels.

3.2 MIMO Detection Algorithms

The following sections will present various well known MIMO and space-time coding algorithms. We will first look at those algorithms that focus on maximizing capacity

(i.e. maximizing multiplexing gain) followed by those that focus on minimizing the probability of error (i.e. maximizing diversity gain).

3.2.1 Algorithms for Maximizing Capacity

Spatial Multiplexing (SM) algorithms are usually designed to maximize multiplexing gain. That is, incoming symbols are multiplexed and independently transmitted across antennas. The most known algorithms in this category are the Maximum Likelihood (ML) detector, the Zeros Forcing (ZF) and Minimum-Mean-Square-Error (MMSE) linear receivers, and Vertical Bell Labs Layered Space-Time (VBLAST) based on ZF nulling (ZF-VBLAST) or MMSE nulling (MMSE-VBLAST) [39]. The following sections describe these algorithms in more details.

The Spatial Maximum Likelihood Detector

The spatial ML detector is the most optimal algorithm in the BER/FER sense [39]. Its BER/FER curve is normally used as a lower bound on the performance that could potentially be achieved by these algorithms.

In the ML detector, an estimate $\hat{\mathbf{s}}$ of the transmitted vector $\mathbf{s}$ is found through the following optimization problem:

$$\hat{\mathbf{s}} = \arg \min_{s \in \Re^{N_t}} \left\| \mathbf{y} - \sqrt{\frac{\rho}{N_t}} \mathbf{H}\mathbf{s} \right\|^2 . \tag{3.2.1}$$

The ML detector basically compares all possible transmitted vector symbols, multiplied by a scaled channel matrix, to the received vector $\mathbf{y}$ and chooses the one that minimizes the l_2-norm of the difference.

It is not hard to see that when L-QAM modulation is used the detection complexity of the ML detector is $O(L^{N_t})$ per symbol vector. While this complexity could be

justified for low values of L and N_t, it will be practically impossible to employ such a detector when high-order modulations (such as 16-QAM or higher) and/or more than two transmit antennas are used. Consider, for example, when $N_t = 4$ and 64-QAM modulation is used. In this case we have to compute $64^4 = 16777216!$ norms per symbol vector and then chooses the smallest one among all these possibilities.

Reduced complexity ML-like Sphere Decoding Algorithms (SDA) have been proposed [15]. The main idea that is exploited to reduce the ML complexity is to limit the search to potential vector candidates that lie within a sphere of certain radius in N_t space. The size of the radius depends on the actual SNR, which makes the complexity of this type of algorithms variable. In addition, the depth of the search makes its complexity dependant on the modulation order. These shortcomings make SDAs unsuitable for practical implementation [6]. In [6], an SDA with SNR-independent complexity was proposed but its complexity still depends on the modulation order. A promising SDA that is independent of the modulation order was proposed in [17] but its payload complexity is $O(N_t^6)$ which is still extremely high especially for high data rate communication. Another disadvantage that the SDA algorithm suffers from is that its complexity increases in the presence of spatial correlation [5].

The infinite-SNR DMT curve for the spatial ML detector is [12]:

$$G_d = N_r(1 - \frac{G_m}{N_t}), \quad G_m \in [0, N_t]. \tag{3.2.2}$$

Spatial Filtering Techniques

There are two standard spatial filtering techniques, one that is based on ZF optimization and another one that is based on MMSE optimization.

In the ZF method, we basically find a set of nulling row vectors $\mathbf{w}_k$ $(k = 1, 2, ...N_t)$ such that:

$$\mathbf{w}_k \mathbf{h}_j = \begin{cases} 1 & \text{if } k = j \\ 0 & \text{otherwise} \end{cases} \tag{3.2.3}$$

Basically, $\mathbf{w}_k$ keeps the contribution from symbol s_k while completely nulling contributions from all other symbols s_k $(k \neq i)$. One then derives the following spatial filtering matrix, commonly referred to as the pseudo-inverse of $\mathbf{H}$ [39]:

$$\mathbf{W}_{ZF} = \sqrt{\frac{N_t}{\rho}} \left(\mathbf{H}^+ \mathbf{H}\right)^{-1} \mathbf{H}^+, \tag{3.2.4}$$

whose k-th row vector $\mathbf{w}_k$ is used to null out the contribution of all i-th symbols (i $\neq$ k) from $\mathbf{y}$ before slicing takes place. Using the above ZF spatial filter one can then obtain the the pre-slicing symbol vector estimate as:

$$\hat{\mathbf{s}} = \mathbf{s} + \underbrace{\mathbf{W}_{ZF}\mathbf{n}}_{\textit{noise enhancement}} = \mathbf{s} + \tilde{\mathbf{n}}. \tag{3.2.5}$$

It is not hard to see that, while $\mathbf{W}_{ZF}$ succeeds in completely eliminating multi-stream interference (MSI), it does enhance noise in the process. This is generally referred to as the noise enhancement problem in the literature.

The MMSE method attempts to reach a balance between canceling MSI and reducing the effect of noise enhancement. We basically find a set of weight vectors $\mathbf{w}_k$ $(k = 1, 2, ...N_t)$ such that each Mean Square Error (MSE) term

$$\varepsilon_k^2 = E\left(|s_k - \mathbf{w}_k \mathbf{y}|^2\right) \tag{3.2.6}$$

is minimized. One then obtains the following spatial filtering matrix [39]:

$$\mathbf{W}_{MMSE} = \sqrt{\frac{N_t}{\rho}} \left(\underbrace{\mathbf{H}^{+}\mathbf{H}}_{MSI\ cancellation} + \underbrace{\frac{N_t}{\rho}\mathbf{I}_{N_t}}_{noise\ suppression} \right)^{-1} \mathbf{H}^{+}. \tag{3.2.7}$$

The first term in the bracket takes care of canceling the effect of interference originating from undesired symbols and the second one reduces the effect of noise enhancement. One added advantage of the MMSE method, compared to the ZF method, is that the $\frac{N_t}{\rho}\mathbf{I}_{N_t}$ term guarantees that the expression inside the bracket is always invertible even when channel conditioning is poor such as when strong spatial correlation exists. Note that at sufficiently high SNR ($\rho \gg 1$), we have $\mathbf{W}_{MMSE} \approx \mathbf{W}_{ZF}$ and consequently both methods offer similar diversity performance.

The preprocessing complexity of these linear receivers (ZF or MMSE), which consists of finding the spatial filtering matrix, is mainly dominated by finding the pseudo-inverse of a matrix of size $N_r \times N_t$. Efficient matrix inversion algorithms allow us to achieve this with $O(N_t^3)$ complexity. The payload complexity which mainly consists of multiplying the RX symbol vector by the filtering matrix requires $O(N_r^2)$ complexity.

The infinite-SNR DMT curve for the ZF and the MMSE linear recievers is given by the same following equation [12]:

$$G_d = (N_r - N_t + 1)(1 - \frac{G_m}{N_t}), \quad G_m \in [0, N_t]. \tag{3.2.8}$$

The VBLAST Algorithm

A better way to decode the received vector $\mathbf{y}$ is to use a feedback loop, similar to the one used in time-domain Decision Feedback Equalization (DFE), by which

the contribution of the already detected symbol is subtracted out from the received vector $\mathbf{y}$ before the next nulling and detection steps are performed. Assuming the correct detection decision has been made, the about-to-be detected symbol enjoys an additional degree of diversity compared to the previous one. One might ask the question: which symbol should we attempt to detect first? It turns out [19] that optimal BER/FER performance is achieved when we choose the symbol that has the highest post-detection SNR. This non-linear decoding technique, along with optimal ordering, are the main ideas that are exploited in most VBLAST-type of algorithms. As we shall see later, these techniques offer improved diversity performance (compared to linear receivers) at low-to-moderate SNR.

The VBLAST algorithm consists of the following three steps:

1. Optimal ordering (OO)
2. Interference nulling (IN)
3. Interference cancellation (IC)

VBLAST starts by keeping the contribution of the symbol that enjoys the best post-detection SNR while eliminating the contributions from other symbols (IN step using ZF or MMSE). We then slice the decision statistics to obtain an estimate of that transmitted symbol. Assuming we made the right decision, we then remove the contribution of that particular symbol from the transmitted vector $\mathbf{y}$ (IC step), and then we repeat the same process again, choosing to decode the symbol that enjoys the next best post-detection SNR. We do this until all symbols have been decoded. Table 3.1 summarizes the ZF version of VBLAST, hereafter referred to as ZF-VBLAST.

The MMSE version, i.e. MMSE-VBLAST, differs mainly in the computation of the spatial matrix.

Since VBLAST relies on the assumption that previously decoded symbols were detected correctly, it runs the risk of carrying (or propagating) additional "noise" terms (in addition to $\mathbf{n}$) when decision errors have been made. This deficiency in the VBLAST decoding process is referred to in the literature as the Error Propagation Problem (EPP) and it has been recognized as the main performance bottleneck in any MIMO algorithm that depends on successive interference cancellation (SIC). Note that even if symbols detected at later stages enjoy a better diversity, the overall error performance is still dominated by the first detected symbol. Many solutions have been proposed to improve the performance of the first detected symbol. The easiest one consists of using higher number of RX antennas to increase the diversity of that stage[1].

The preprocessing complexity of the original ZF-VBLAST (or MMSE-VBLAST), hereafter described in Table 3.1, is $O(N_t^4)$, which is more complex than that of the ZF or MMSE linear receivers. The payload complexity is $O(N_r^2)$ which is similar to the one provided by linear receivers but a lot less complex than the payload processing required by the ML receiver.

On the DMT Performance of VBLAST

There does not seem to be an agreement in the literature regarding the DMT performance of VBLAST. We attempt next to reconcile many of the apparently conflicting conclusions regarding VBLAST's DMT performance. Our analysis will be validated by simulations presented in this chapter and the next. We will strictly focus on

[1]We will see later that VBLAST's maximum diversity gain is equal to $N_r - N_t + 1$

initialization :

$$\imath = 1 \tag{3.2.9a}$$
$$\mathbf{G}_\imath = \mathbf{H}^{-1} \tag{3.2.9b}$$
$$k_1 = \arg\min_\jmath \|(\mathbf{G}_1)_\jmath\|^2 \tag{3.2.9c}$$

recursion :

$$\mathbf{w}_{k_\imath} = (\mathbf{G})_{k_\imath} \tag{3.2.9d}$$
$$y_{k_\imath} = \mathbf{w}_{k_\imath}^T \mathbf{r}_\imath \tag{3.2.9e}$$
$$\hat{s}_{k_\imath} = Q(y_{k_\imath}) \tag{3.2.9f}$$
$$\mathbf{r}_{\imath+1} = r_\imath - \hat{s}_{k_\imath}(\mathbf{H})_{k_\imath} \tag{3.2.9g}$$
$$\mathbf{G}_{\imath+1} = \mathbf{H}_{\bar{k}_\imath}^{-1} \tag{3.2.9h}$$
$$k_{\imath+1} = \arg \min_{\jmath \notin \{k_1 \ldots k_\imath\}} \|(\mathbf{G}_{\imath+1})_\jmath\|^2 \tag{3.2.9i}$$
$$\imath = \imath + 1 \tag{3.2.9j}$$

Table 3.1: The ZF VBLAST algorithm. Note that here $\mathbf{H}^{-1}$ is the pseudo-inverse of $\mathbf{H}$, i.e. $\mathbf{H}^{-1} = \left(\mathbf{H}\mathbf{H}^+\right)^{-1}\mathbf{H}^+$

MMSE-VBLAST as it is the algorithm of interest in our study. Also we will assume $N_t = N_r = N$.

In [60], it is shown that an upper bound on the DMT curve for unordered MMSE-VBLAST is:

$$G_d \leq (N-1)(1 - \frac{G_m}{N}), \quad G_m \in [0, N]. \tag{3.2.10}$$

It is not clear how tight this bound is. It is to be noted that in deriving the above expressions the authors did not take into account the rate being used.

In [9], it is shown that for the 2×2 case the diversity of ordered MMSE-VBLAST actually depends on the rate-per-layer; if this rate does not exceed 1 (such as when using BPSK) then ordered MMSE-VBLAST enjoys a diversity of 2.

Contrary to the conclusion reached in [60], it is shown in [30] that at high-SNR ordering does not improve the diversity order, i.e. the DMT curve for unordered or ordered MMSE-VBLAST is the same.

While the above presentation seems to offer conflicting conclusions on the DMT performance of MMSE-VBLAST, we offer the following explanation that will help clarify the issue. First of all, we have to distinguish between the low-to-moderate SNR region, where the SNR suppression capability of MMSE-VBLAST helps improve its finite-SNR diversity, and the high SNR region where it looses this advantage. Second, it is to be noted that when the modulation order increases the effect of noise on performance becomes worse because the Euclidian distance between the various QAM symbols become smaller. These two facts make the diversity-multiplexing performance of MMSE-VBLAST very much dependant on the rate used and on the operational SNR. This explanation will be validated in the next chapter through computer simulations.

On the Outage Capacity of MMSE-VBLAST

In the literature (see [12] for example) it is frequently mentioned that MMSE-VBLAST attains the full MIMO capacity. We have to emphasize here that this is only true in fast fading scenarios. In slow fading scenarios, as is assumed in our study, MMSE-VBLAST does not attain the MIMO outage capacity [54].

The outage capacity of MMSE-VBLAST can be calculated by considering the capacity achieved by the worst detected layer as [11]:

$$C_{MMSE-VBLAST} = N_t \times \min_{q \in \{1,\ldots,N_t\}} \{\log_2(1+\rho_q)\} \tag{3.2.11}$$

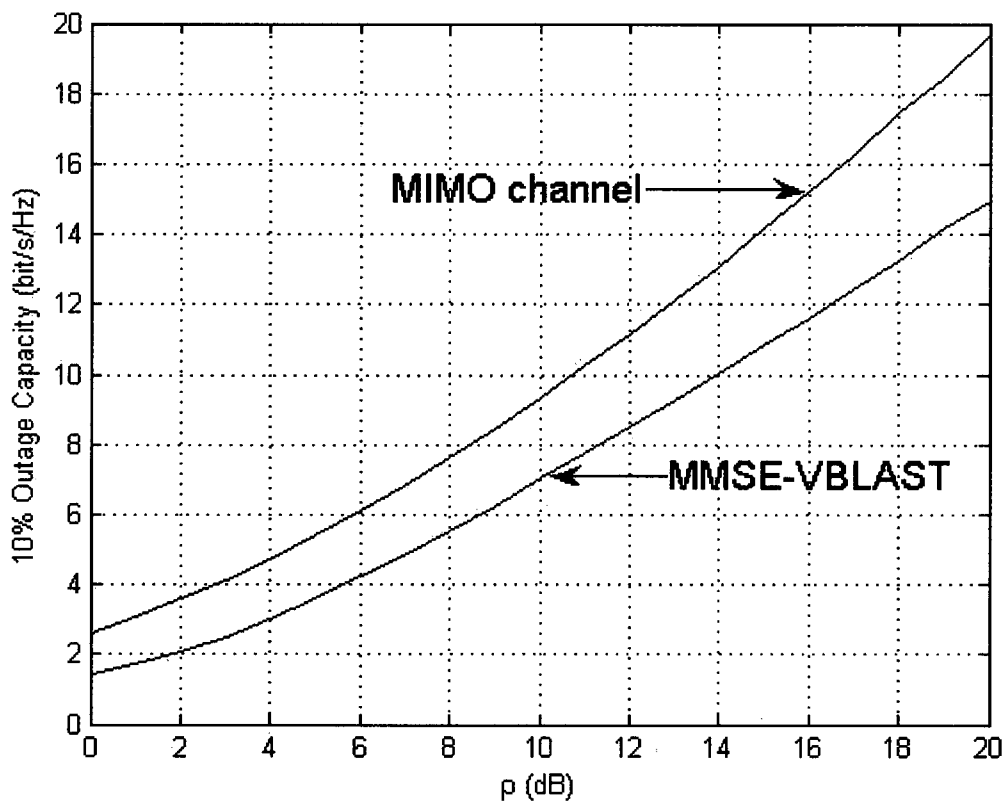

Figure 3.5: 10% Outage capacity comparison between MMSE-VBLAST and the i.i.d. MIMO channel for $N_t = N_r = 4$

where

$$\rho_j = \frac{1}{\left[(\frac{\rho}{N}\mathbf{H}_{\bar{q}}^{+}\mathbf{H}_{\bar{q}} + \mathbf{I})^{-1}\right]_{q,q}} - 1 \tag{3.2.12}$$

is the post-detection SNR for q^{th} TX detected stage [32]. Note that in the above expression $\mathbf{H}_{\bar{q}}$ represents a reduced dimension channel where the q^{th} column has been deleted. Also note that q refers to the detection order rather than the actual TX antenna number. Because the first detected stage in SIC-type of receivers has the worst outage performance [34] we can simply replace q with one in the above

Algorithm	**Infinite-SNR DMT Performance $G_d =$**	**Preprocessing Complexity**	**Payload Complexity**
ZF	$1 - \frac{G_m}{N}$	$O(N^3)$	$O(N^2)$
MMSE	$1 - \frac{G_m}{N}$	$O(N^3)$	$O(N^2)$
ZF-VBLAST	$1 - \frac{G_m}{N}$	$O(N^4)$	$O(N^2)$
MMSE-VBLAST	$1 - \frac{G_m}{N}$	$O(N^4)$	$O(N^2)$
ML	$N(1 - \frac{G_m}{N})$	0	$O(L^N)$

Table 3.2: Comparison between various $N \times N$ SM MIMO detection algorithms

expression.

Figure 3.5 shows the 10% outage capacity attained by MMSE-VBLAST. It is evident that MMSE-VBLAST is capacity lossy as it only achieves about 70% of the MIMO channel capacity.

Comparison Between the Various Spatial Multiplexing Algorithms

Table 3.2 summarizes the infinite-SNR DMT performance and the complexity of the various SM detection algorithms we have presented in this section.

To see how different SM detection algorithms compare in terms of BER performance, we ran some MATLAB simulations for a 4×4 i.i.d. MIMO channel with an overall bandwidth efficiency of 8 bit/s/Hz (i.e. QPSK modulation is used in each spatial dimension). Figure 3.6 shows the results. As expected we clearly see that the ZF spatial receiver has the worst performance while the ML receiver has the best one. We also see that MMSE-VBLAST performs considerably better than ZF-VBLAST, attaining a BER of 10^{-3} with a considerable 8dB energy advantage. Also we can clearly

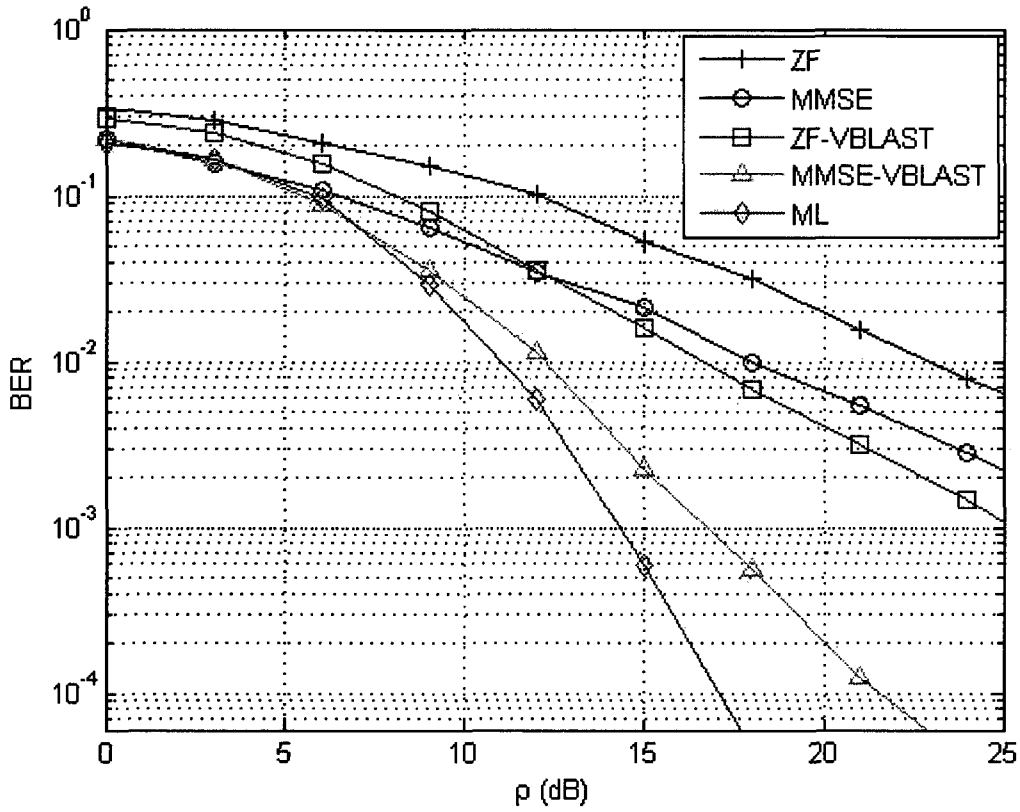

Figure 3.6: BER comparison between various detection algorithms for a 4×4 MIMO channel with bandwidth efficiency of 8 bits/s/Hz

see that while MMSE-VBLAST diversity at low SNR is high, as demonstrated by the sharpness of the curve, the diversity decreases with increasing SNR as the curve starts to loose its sharpness. It is remarkable that MMSE-VBLAST's performance at BER $< 5\%$ almost matches that of the ML decoder. This demonstrates that the BER performance at low SNR (where BER is high) is mainly by dominated the noise which MMSE-VBLAST is able to cancel effectively. The gap between the two algorithms

widens with increasing SNR as MMSE-VBLAST starts to loose its diversity advantage while the ML decoder maintains its full diversity advantage of 4 (as the curve only needs 2.5dB per decay at high SNR).

We conclude from the above complexity and error performance analysis that MMSE-VBALST reaches a tradeoff between complexity and error performance. This explains why we choose it in our study as a good representative SM algorithm.

3.2.2 Algorithms for Minimizing Probability of Error

In contrast to standard VBLAST-type detection where spatial multiplexing is maximized, space-time coding techniques focus on minimizing the probability of error in fading environments by using redundancy[2]. That is, the same symbol (or a transformed form of it) that was transmitted at time t, is retransmitted from different antennas at times $t+i$ $(i = 1, 2, ...T-1)$, hence reducing the probability that a given symbol is completely lost[3].

There are two ways to introduce redundancy. The first one consists of using a trellis where memory is introduced. We refer to this type of codes as space-time trellis codes (STTC). The more trellis states we employ the better the performance we obtain. Unfortunately, the decoding complexity grows exponentially with the modulation order and the depth of the trellis. For more in-depth description of STTC the reader is referred to [56]. The second category of space-time codes introduces redundancy by simply retransmitting a transformed form of the same symbol vector at subsequent time instants. We refer to this category as space-time block codes (STBC).

[2]One can view SM algorithms as a special case of space-time codes where redundancy is not used.
[3]This very much depends on how independent the fading each subchannel undergoes.

In general, both STBC and STTC provide full diversity gain. While it is true that the amount of transmitted symbols in a given frame is reduced when introducing redundancy, the diversity gain achieved usually allows us to use higher order modulation to compensate for the rate loss.

The decoding of STBC is a lot simpler than the decoding of STTCs while the performance of STTC generally outperforms that offered by STBCs because STTCs provide coding gain in addition to diversity gain. A special category of STBC is orthogonal STBC (OSTBC). These codes possess an additional important property of being able to decouple the MIMO channel into simple SISO channels using simple linear processing.

In what follows we will discuss in more details the most simple orthogonal STBC: the Alamouti code proposed in [1].

Alamouti Space-Time Block Code

The Alamouti OSTBC, first proposed in [1], is a rate-1 space-time code. Not only it achieves ML detection with full diversity [27], it also allows the use of linear processing techniques at the receiver, hence greatly reducing the baseband processing requirements. This feature makes it particularly attractive for battery-powered terminals, and that probably explains its adoption by various 3G and 4G wireless standards such as 3GPP-LTE and IEEE-802.16e (mobile WIMAX).

To help us understand how the Alamouti space-time code achieves the full MIMO channel diversity with linear processing at the receiver, we reproduce here the fundamental equations used by the algorithm. The Alamouti TX processing consists of transmitting the following simple space-time codeword:

$$\mathbf{X}_{ALAMOUTI} = \begin{bmatrix} s_1 & s_2 \\ -s_2^* & s_1^* \end{bmatrix} . \tag{3.2.13}$$

The columns of this code represent space dimension while the rows represent time dimension. That is, at time t odd-indexed symbols are transmitted from the first antenna while even-indexed ones are transmitted from the second antenna. At time $t+1$, the complex conjugate of the even-indexed symbols with opposite polarity are transmitted from the first antenna while the complex conjugate of the odd-indexed symbols are transmitted from the second antenna. One can then express the received vector at RX antenna i as follows:

$$\begin{bmatrix} y_{1,i} \\ y_{2,i} \end{bmatrix} = \sqrt{\frac{\rho}{2}} \begin{bmatrix} s_1 & s_2 \\ -s_2^* & s_1^* \end{bmatrix} \begin{bmatrix} h_{i,1} \\ h_{i,2} \end{bmatrix} + \begin{bmatrix} n_{1,i} \\ n_{2,i} \end{bmatrix}, i = 1, 2, ..., N_r \tag{3.2.14}$$

or written differently

$$\underbrace{\begin{bmatrix} y_{1,i} \\ y_{2,i}^* \end{bmatrix}}_{\mathbf{y}_i} = \sqrt{\frac{\rho}{2}} \underbrace{\begin{bmatrix} h_{i,1} & h_{i,2} \\ h_{i,2}^* & -h_{i,1}^* \end{bmatrix}}_{\tilde{\mathbf{H}}_i} \underbrace{\begin{bmatrix} s_1 \\ s_2 \end{bmatrix}}_{\mathbf{s}} + \underbrace{\begin{bmatrix} \tilde{n}_{1,i} \\ \tilde{n}_{2,i} \end{bmatrix}}_{\tilde{\mathbf{n}}_i}, i = 1, 2, ..., N_r \tag{3.2.15}$$

where the subscripts 1 and 2 have been used to refer to time instant 1 and time instant 2, respectively. We can easily verify the orthogonality of the effective channel matrix $\tilde{\mathbf{H}}_i$, i.e. $\tilde{\mathbf{H}}_i^+ \tilde{\mathbf{H}}_i = (|h_{i,1}|^2 + |h_{i,2}|^2)\,\mathbf{I}_2$. It is indeed this property that allows the use of linear processing at the receiver. To see this, multiply the received vector in (3.2.15) by $\tilde{\mathbf{H}}_i^+$ to get the following pre-detection estimate:

$$\hat{\mathbf{s}} = \sqrt{\frac{\rho}{2}} \sum_{i=1}^{N_r} (|h_{i,1}|^2 + |h_{i,2}|^2)\mathbf{s} + \sum_{i=1}^{N_r} \tilde{\mathbf{n}}_i. \tag{3.2.16}$$

Equation (3.2.16) clearly demonstrates that the Alamouti TX and RX processing automatically eliminates multi-stream interference and adds the symbol energy coherently while adding the noise components incoherently. Note that this is done without resorting to matrix inversion, hence eliminating the noise enhancement problem associated with the SIC-type of algorithms. Also note that since we have a total of $2N_r$ independent channel coefficients contributing to boosting the symbol energy the diversity order achieved by the Alamouti space-time code is equal to $2N_r$

To gain further insight on the BER performance achieved by the Alamouti scheme, we have conducted some simulations and compared its performance to that of the SISO case. Figure 3.7 shows the simulation results. We clearly see from this figure that the $2 \times N_r$ Alamouti scheme offers huge BER improvement over traditional SISO systems. For example, a 2×1 Alamouti space-time code offers a significant 10dB coding advantage at a BER of 10^{-3}. We also note that as N_r grows beyond 3, we start to get marginal improvement.

The biggest advantage of the Alamouti code is that it has zero preprocessing complexity while its payload complexity is only $O(N_r)$ which makes it attractive for use in hand-held battery-powered wireless terminals.

Capacity of the Alamouti STBC

The capacity of the $2 \times N_r$ Alamouti STBC can be calculated without resorting to basic principles of information theory if we can rearrange the TX and RX Alamouti equation (3.2.15) in a form similar to (2.1.1). This is indeed possible by stacking the

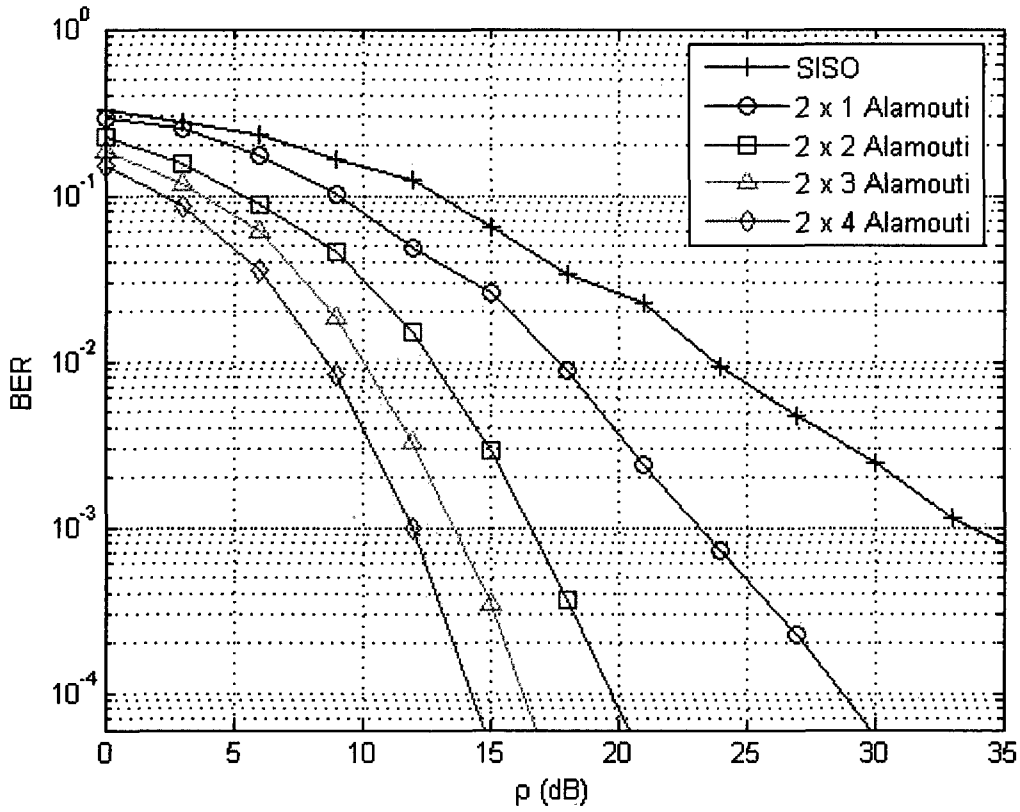

Figure 3.7: BER performance of $2 \times N_r$ Alamouti STBC with 16QAM modulation giving a total bandwidth efficiency of 4 bit/s/Hz

various channel matrices, receive and noise vectors one on top of the other to obtain:

$$\underbrace{\begin{bmatrix} \mathbf{y}_1 \\ \mathbf{y}_2 \\ \vdots \\ \mathbf{y}_{N_r} \end{bmatrix}}_{\mathbf{y}} = \sqrt{\frac{\rho}{2}} \underbrace{\begin{bmatrix} \mathbf{H}_1 \\ \mathbf{H}_2 \\ \vdots \\ \mathbf{H}_{N_r} \end{bmatrix}}_{\tilde{\mathbf{H}}} \mathbf{s} + \underbrace{\begin{bmatrix} \tilde{\mathbf{n}}_1 \\ \tilde{\mathbf{n}}_2 \\ \vdots \\ \tilde{\mathbf{n}}_{N_r} \end{bmatrix}}_{\tilde{\mathbf{n}}} \tag{3.2.17}$$

where the $2N_r \times 2$ matrix $\tilde{\mathbf{H}}$ is normally referred to as the effective channel matrix.

We can now calculate the instantaneous capacity of the Alamouti scheme as:

$$\begin{aligned} C_{Alamouti} &= \frac{1}{2}\log_2 \ \det\left(\mathbf{I}_2 + \frac{\rho}{2}\tilde{\mathbf{H}}^{+}\tilde{\mathbf{H}}\right) \\ &= \frac{1}{2}\log_2 \ \det\left(\mathbf{I}_2 + \frac{\rho}{2}||\mathbf{H}||^2\mathbf{I}_2\right) \\ &= \frac{1}{2}\log_2 \ \left(1 + \frac{\rho}{2}||\mathbf{H}||^2\right)^2 \\ &= \log_2 \ \left(1 + \frac{\rho}{2}||\mathbf{H}||^2\right) \end{aligned} \tag{3.2.18}$$

where $||\mathbf{H}|| = \sqrt{\sum_{i=1}^{N_r}\sum_{i=1}^{N_t}|h_{ij}|^2}$ is the Frobenius norm of $\mathbf{H}$ and the factor $\frac{1}{2}$ has been included to compensate for the fact that new symbols are only delivered every second symbol interval.

We can easily show that $C_H \geq C_{Alamouti}$ by rewriting the value of C_H in (2.2.1) in a different form as follows [12]:

$$\begin{aligned} C_H &= \log_2 \ \det\left(\mathbf{I}_2 + \frac{\rho}{2}\mathbf{H}\mathbf{H}^{+}\right) \\ &= \log_2 \ \left(1 + \frac{\rho}{2}||\mathbf{H}||_F^2 + \left(\frac{\rho}{2}\right)^2 |\det(\mathbf{H})|^2\right) \\ &\geq \log_2 \ \left(1 + \frac{\rho}{2}||\mathbf{H}||^2\right) \end{aligned} \tag{3.2.19}$$

where the inequality in the last expression is derived from the fact that the $\log_2$ function is monotonically increasing and $|\det(\mathbf{H})|^2$ is always a positive number. Equality holds when $\det(\mathbf{H}) = 0$ which happens when $rank(\mathbf{H}) = 1$, i.e. when the channel is fully correlated or in keyhole scenarios. In summary, the Alamouti code is capacity lossy when compared to the MIMO channel capacity but is capacity optimal in fully-correlated or keyhole channels.

3.3 Performance Comparison Between the Alamouti STBC and MMSE-VBLAST

It is rather informative to compare the performance of MMSE-VBLAST to that of the Alamouti STBC. For fair comparison, the same bandwidth efficiency should be used in both cases. We choose here a medium rate of 4 bit/s/Hz. MMSE-VBLAST can achieve this rate by using QPSK in each subchannel while the Alamouti code needs to use 16QAM in each subchannel. We will also test the robustness of each algorithm against spatial correlation and channel estimation errors, two imperfections commonly encountered in practice. Channel estimation errors are generally caused by the following independent factors:

1. Insufficient length of the training sequence (or number of pilot symbols).
2. Insufficient word length when implementing the algorithm in finite precision
3. Channel deviation during the payload processing phase

While the effect of the first two factors mentioned above can be minimized by proper system design the third factor can not be controlled due to the dynamic nature of wireless environments channels.

For evaluation purposes, we will use the following commonly-used channel estimation error model [59]:

$$\tilde{\mathbf{H}} = \sqrt{1-\epsilon^2}\mathbf{H} + \epsilon\mathbf{H}_e \tag{3.3.1}$$

where $\tilde{\mathbf{H}}$ is the estimated channel matrix, $\mathbf{H}_e$ is a channel estimation error matrix that is independent of $\mathbf{H}$ and with elements that are i.i.d. with $\mathcal{CN}(0,1)$. The parameter

$\epsilon \in [0, 1]$ represents the accuracy of channel estimation. Note that the scaling factors used in the above equation ensure that the variance of each entry of the estimated channel matrix $\tilde{\mathbf{H}}$ is still equal to 1.

Correlation is introduced by using the exponential channel model discussed in Section 2.3.1. We next conduct BER comparison followed by capacity comparison.

BER comparison

Figure 3.8 compares the BER performance of the 2×2 Alamouti to that of the 2×2 MMSE-VBLAST with and without channel estimation errors. We see that for $\rho < 8$dB, both algorithms have almost the same performance. As SNR increases the Alamouti OSTBC manifests its diversity advantage through its sharper curve. We also see that the Alamouti OSTBC is more robust than MMSE-VBLAST against channel estimation errors. In fact, as channel estimation error increases to 5%, we start to witness an error floor around a BER of 10^{-3} for MMSE-VBLAST. In contrast, the Alamouti curve continues its sharp descent with only slight degradation. For instance, comparing the zero estimation error scenario to the 5% estimation error scenario, one can easily see that the Alamouti OSTBC looses only a fraction of a dB compared to SNR loss of approximately 2dB for MMSE-VBLAST at a BER of 10^{-3}. The above results can be explained as follows: since MMSE-VBLAST focuses on canceling noise its performance at low SNR is almost unaffected. At high SNR where noise is negligible, its performance mainly depends on the conditioning of the channel matrix which becomes poor due to the additional noise term created by the channel estimation error. The Alamouti OSTBC does not depend on matrix inversion and as such does not suffer from these deficiencies.

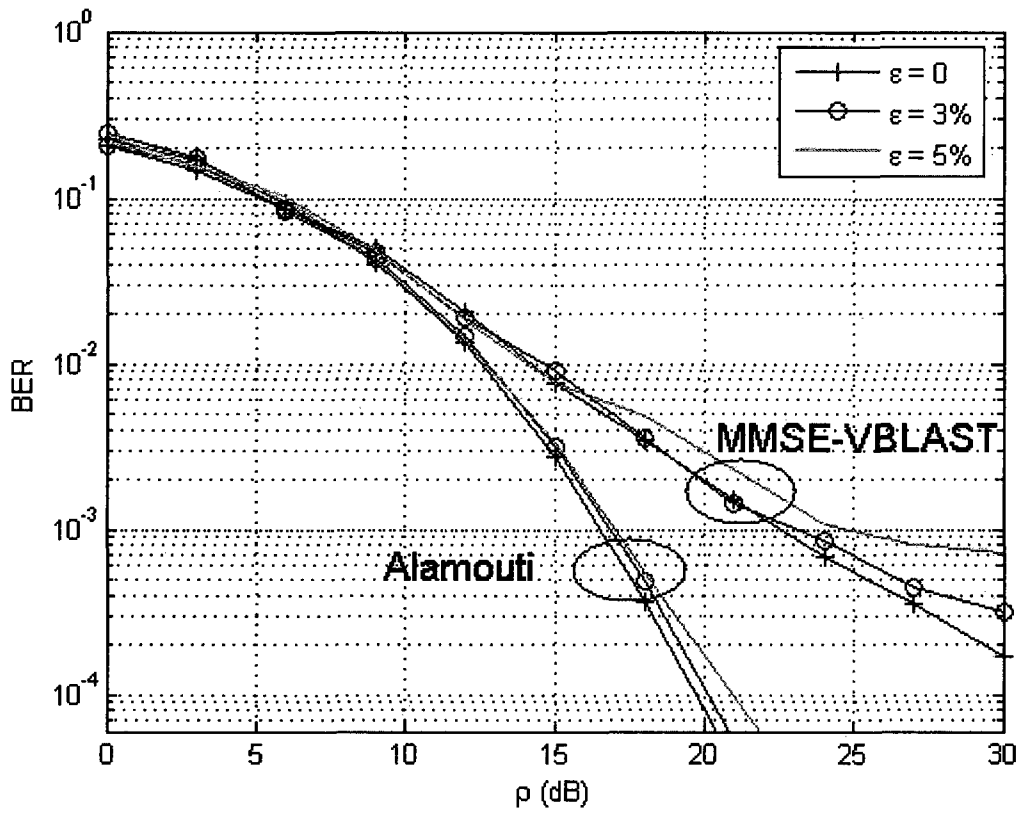

Figure 3.8: BER performance comparison with and without channel estimation error between 2×2 Alamouti OSTBC with 16QAM and 2×2 MMSE-VBLAST with QPSK for a total same bandwidth efficiency of 4 bits/s/Hz

Figure 3.9 compares the BER performance of both algorithms when channel correlation is present. Note that the correlation coefficients chosen represent typical values found on the downlink. In this case the link is mostly limited by the correlation at the BS side since the correlation at the UE side is generally very low due to rich scattering that exists around the UE.

For moderate correlation, i.e. when $\psi_t = 0.5$, the Alamouti OSTBC suffers a fraction of a dB degradation at a BER of 10^{-3} while MMSE-VBLAST suffers a significant

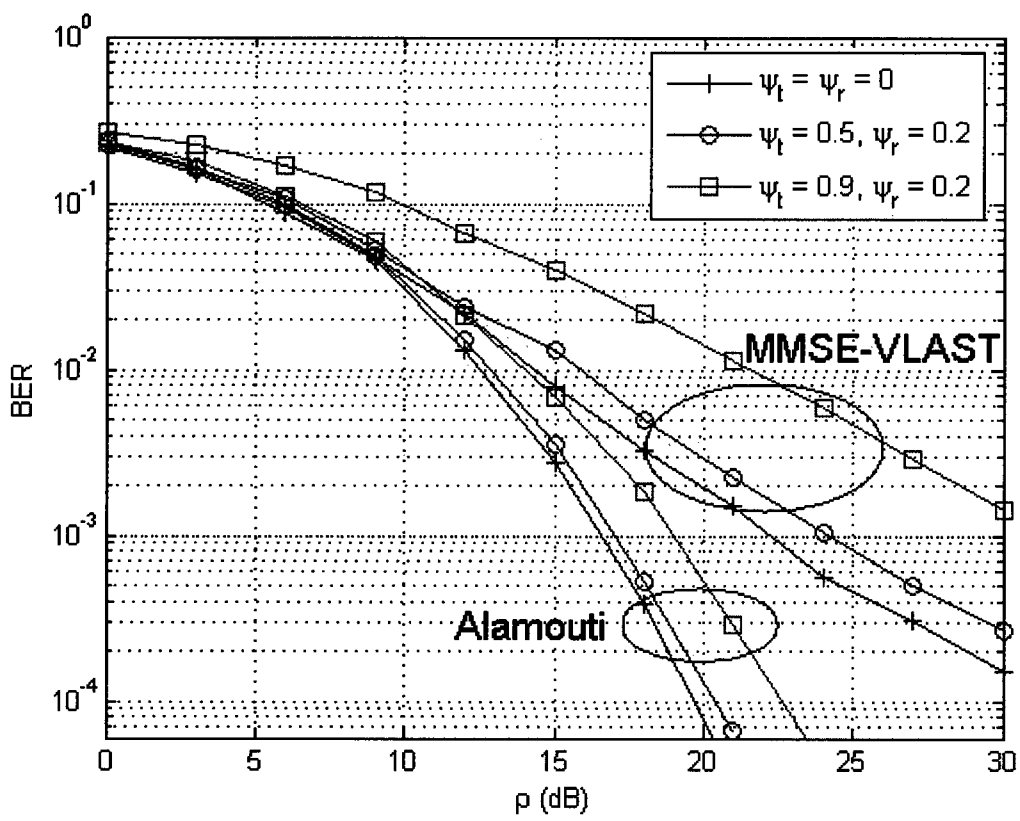

Figure 3.9: BER performance comparison with and without channel correlation between 2×2 Alamouti OSTBC with 16QAM and 2×2 MMSE-VBLAST with QPSK for a total same bandwidth efficiency of 4 bits/s/Hz

SNR degradation of about 2dB at the same BER. When correlation becomes severe, i.e. when $\psi_t = 0.9$, the Alamouti OSTBC suffers a 2.5dB SNR degradation while MMSE-VBLAST suffers a loss of over 7dB at the same BER. The above results are expected since MMSE-VBLAST's dependance on channel inversion makes it susceptible to channel conditioning: when channel correlation increases the channel rank becomes closer to 1. The Alamouti OSTBC is more resistant to channel correlation because it does not depend on matrix inversion. In fact, when the channel is fully

correlated the Alamouti code is capacity optimal (as we demonstrated earlier) while MMSE-VBLAST completely breaks down.

The simulation results presented in this section make it clear that the Alamouti STBC is a lot more robust against correlation and channel estimation errors than MMSE-VBLAST. A more complete comparison would include simulation results at different rates since the Alamouti STBC and MMSE-VBLAST have different multiplexing and diversity gains. Also, the inclusion of outer channel codes may give us different results. Both of these can be studied by comparing the capacity curves. We do this in the following section.

Capacity Comparison

Figure 3.10 shows the 1% outage capacity for the 2×2 MIMO channel as well as the corresponding Alamouti and the MMSE-VBLAST curves. It is clear that the Alamouti scheme does not achieve the full MIMO channel capacity, as predicted by (3.2.19), but the same applies to MMSE-VBLAST. It is also clear that the outage capacity offered by the Alamouti STBC is greater than that offered by MMSE-VBLAST in both the i.i.d. and the correlated channel scenarios. Note that in the correlated case, the slope of the MMSE-VBLAST's curve loses its sharpness considerably, i.e. its multiplexing gain at high SNR becomes poor. In contrast, Alamouti's multiplexing gain is maintained at 1 for high SNR. These findings are consistent with the uncoded BER results presented in the previous section: adding channel coding does not help MMSE-VBLAST bridge the performance gap with the Alamouti STBC. We have to mention that channel estimation errors will introduce further degradation in capacity performance and from the BER results we can deduce that this effect will be more severe on MMSE-VBLAST's capacity.

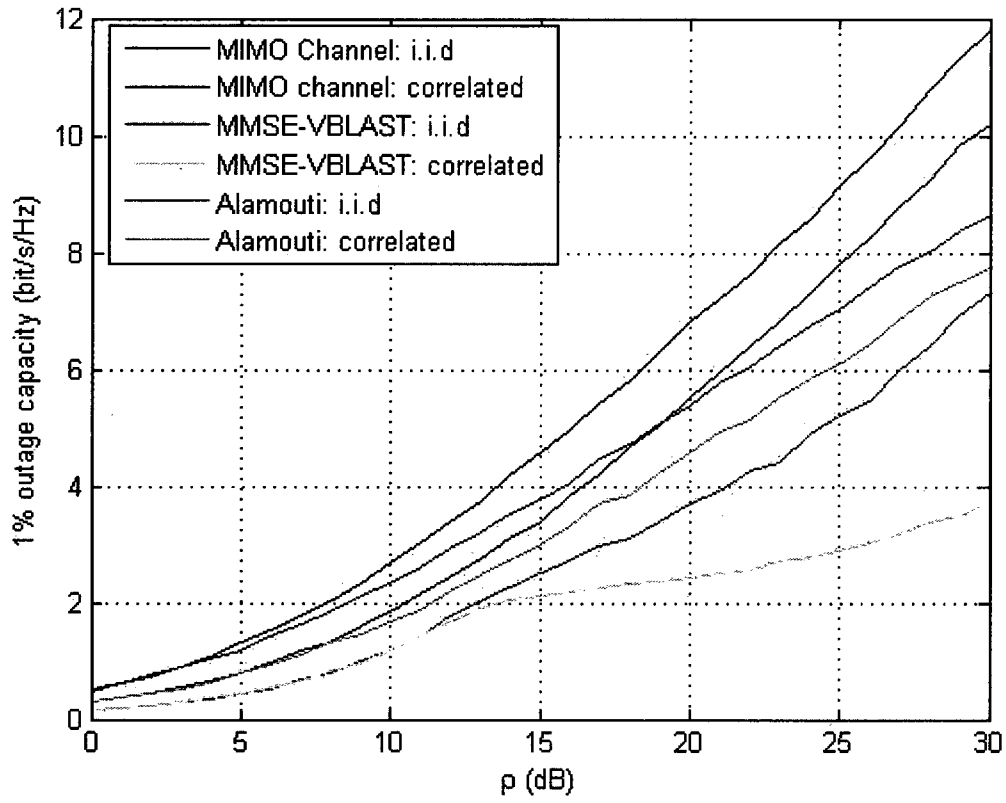

Figure 3.10: 10% outage capacity for Alamouti and MMSE-VBLAST for a 2×2 MIMO channel. For the correlated scenario: the exponential model was used with $\psi_t = 0.9$ and $\psi_r = 0.2$

To summarize, the Alamouti STBC possesses the following desirable features:

1. It achieves excellent error and capacity performance using a low complexity linear encoding and decoding

2. It is robust in the presence of channel estimation errors.

3. It is robust in the presence of spatial correlation.

4. While it does not achieve the full MIMO capacity it outperforms MMSE-VBLAST in terms of both BER and capacity performance.

The question is then: *is it possible to extend all, or at least most, of the benefits offered by the* 2×2 *Alamouti STBC to the general* $N \times N$ *MIMO case?* This is one of the questions we will attempt to answer in the next chapter.

Chapter 4

The Proposed Architecture and Associated Algorithms

It is evident from the material that was presented in previous chapters that a robust $N \times N$ MIMO architecture is needed where sufficient multiplexing and diversity gains are achieved simultaneously. The associated algorithms with this architecture should be robust against spatial correlation and channel estimation errors. Yet these algorithms should have reasonable implementation complexity and should avoid the use of direct matrix inversion in order to avoid the noise amplification problem. It is also important to emphasize that the algorithm should be inherently parallel in order to enable the use of low clock rates for implementation purposes. No feedback channel from the receiver to transmitter shall be assumed. The following sections present an architecture and algorithms that will allow us to achieve these objectives and most of the ones outlined in the introductory chapter.

Before we present the new technique, it is important to answer the following question: *how much diversity is sufficient in MIMO applications?* In the next section we present a discussion which is a step in the direction of answering this question.

4.1 Probability of Error with Transmit Diversity

Even though we restrict our analysis here to transmit diversity and BPSK modulation, the conclusion we arrive at applies as well to receive diversity and higher order modulation schemes.

The probability of bit error P_b for BPSK with N_t TX diversity is [56]

$$P_b = \left[\frac{1}{2}\left(1-\sqrt{\frac{\xi}{1+\xi}}\right)\right]^{N_t} \sum_{k=0}^{N_t-1} \binom{N_t-1+k}{k} \left[\frac{1}{2}\left(1+\sqrt{\frac{\xi}{1+\xi}}\right)\right]^{k} \quad (4.1.1)$$

where $\xi = \frac{\rho}{N_T}$ is the normalized SNR per bit. The probability of error P_b is plotted in Figure 4.1 for $N_t = 1, 2, 3, 4, 5, 6, 7$ and 8. The probability of error for AWGN channel (when $N_t = \infty$) is also plotted on the same figure. We clearly see that as N_t increases P_b is reduced dramatically. We also notice that when N_t increases beyond four we start to get marginal improvement and the sharpness of decent of the various curves almost becomes the same[1]. For instance, increasing N_t from one to two, two to three and three to four, gives us a differential SNR gain of 19dB, 6dB and 2.5dB at a BER of 10^{-5}, respectively. Increasing N_t from four to five gives only 1dB additional SNR gain. This leads us to conclude that a diversity gain of 4 is reasonably sufficient for providing relatively low error probabilities.

For a general $N \times N$, this means that as long as the individual symbols emanating from each TX antenna enjoy at least a diversity order of 4 we should be able to obtain relatively acceptable error performance at finite SNR.

[1]This is only true at finite SNR

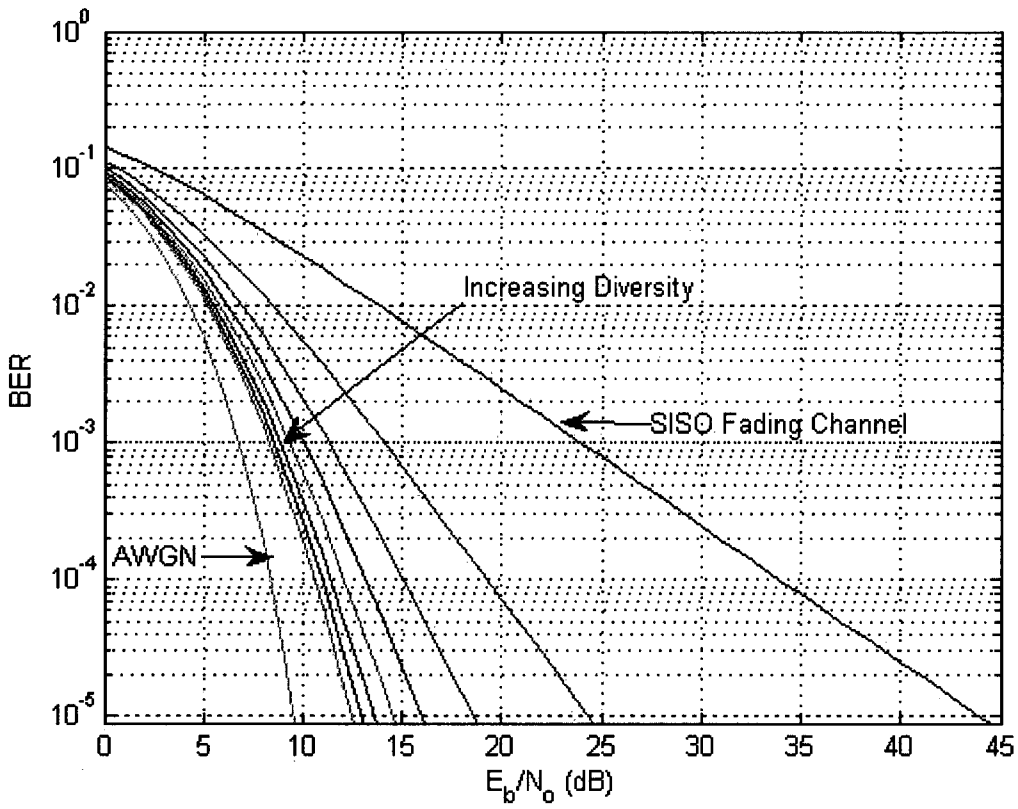

Figure 4.1: Effect of increasing TX diversity on the probability of error P_b of an $N_t \times 1$ MIMO system.

4.2 Proposed Architecture

Similar to [49] [14] we propose to combine STC with SM techniques. On the receiver side, in contrast to [49] [14], and in order to avoid the noise amplification problem, we propose to use the QR decomposition instead of resorting to matrix inversion. Figure 4.2 shows an illustration of the proposed architecture where two groups of Alamouti codes were used in a $N \times N$ MIMO system, where N is assumed to be an even number. Basically, the TX symbols are grouped in pairs and space-time coded

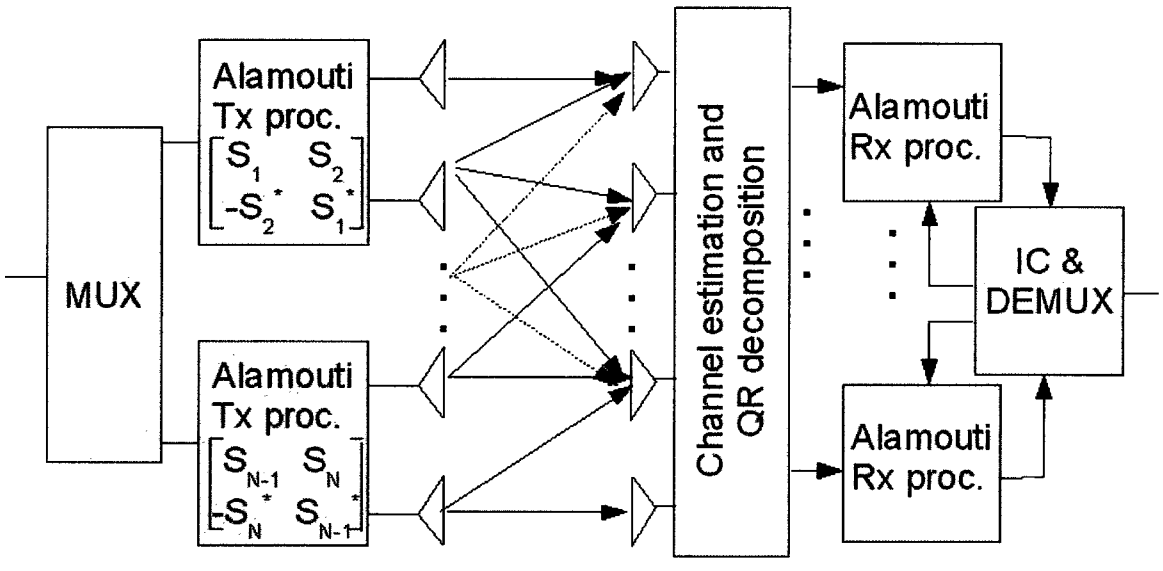

Figure 4.2: Model of the proposed architecture

using the Alamouti STBC. On the RX side channel estimation is performed first[2], one or more QR decompositions are computed and finally successive interference cancellation (IC) followed by Alamouti decoding is employed.

In order to understand the benefits offered by the proposed architecture we first present a brief overview of the QR decomposition which will allow us to reformulate the MIMO detection problem.

4.2.1 QR Matrix Decomposition

When performing MIMO detection, the QR matrix decomposition allows us to eliminate some multi-stream interference without resorting to matrix inversion. To see this, let us review how this decomposition works. We will restrict our analysis to our case of interest where $N_t = N_r = N$.

[2]Channel estimation is assumed perfect in our study unless otherwise stated

Any $N \times N$ matrix $\mathbf{H}$ can be decomposed as $\mathbf{H} = \mathbf{QR}$, where $\mathbf{Q}$ is unitary, i.e. $\mathbf{Q}^{+}\mathbf{Q} = \mathbf{I}$, and

$$\mathbf{R} = \begin{bmatrix} r_{11} & r_{12} & \dots & r_{1N} \\ 0 & r_{22} & \dots & r_{2N} \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & r_{NN} \end{bmatrix} \tag{4.2.1}$$

is an upper triangular matrix. The $\mathbf{Q}$ and $\mathbf{R}$ matrices can be obtained through various techniques the most common of which is the Householder transformation. A brief overview of this method is presented in Appendix B. Using QR decomposition, equation (2.1.1) can then be rewritten as

$$\mathbf{y} = \sqrt{\frac{\rho}{N}}\mathbf{QRs} + \mathbf{n}. \tag{4.2.2}$$

By multiplying the RX vector $\mathbf{y}$ from the left by $\mathbf{Q}^{+}$ we get the following transformed RX vector

$$\begin{aligned} \tilde{\mathbf{y}} \;= \mathbf{Q}^{+}\mathbf{y} &= \sqrt{\frac{\rho}{N}}\mathbf{Rs} + \tilde{\mathbf{n}} \\ &= \sqrt{\frac{\rho}{N}}(\mathbf{r}_1 s_1 + \mathbf{r}_2 s_2 + ... + \mathbf{r}_N s_N) + \tilde{\mathbf{n}}, \end{aligned} \tag{4.2.3}$$

where we have used the property that $\mathbf{Q}^{+}\mathbf{Q} = \mathbf{I}$. Note that all $\tilde{n}_i$ still have zero mean and unity variance, i.e. no noise amplification takes place. It is also easy to verify that $\mathbf{R}$ has the same eigenvalue spread as $\mathbf{H}$, and finding $\mathbf{Q}$ and $\mathbf{R}$ is always possible even when $\mathbf{H}$ is ill-conditioned. i.e. when its eigenvalue spread is extremely

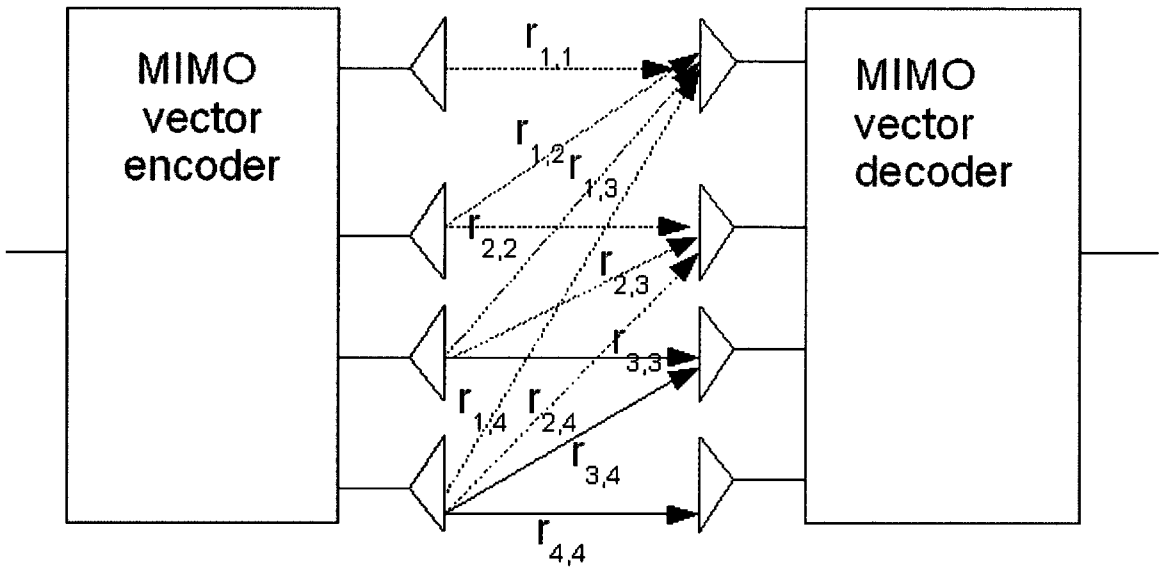

Figure 4.3: Transformed 4×4 MIMO system with QR decomposition

large. This is in contrast to matrix inversion where the inverse may not exist in this particular situation [3]. So one added advantage of using QR decomposition is its numerical stability.

One rather useful interpretation of the QR factorization is that it allows us to reformulate the MIMO detection problem using a "virtual" channel matrix $\mathbf{R}$ rather than the actual channel matrix $\mathbf{H}$. This transformation has indeed some useful properties, and to see this we have redrawn a 4×4 MIMO system model (see Figure 4.3) using $\mathbf{R}$ as the channel matrix. Looking at Figure 4.3 we clearly see that the QR transformation has eliminated interference from certain TX antennas. Of particular interest are the last two RX antennas. These antennas now "see" contributions only from the last two TX antennas. In other words, in the context of MIMO detection,

[3]This problem may be compounded in fixed-point processors.

the QR decomposition allows us to achieve partial interference cancellation without amplifying noise. We have to mention that even though we are discussing the 4×4 MIMO case this is also true for the general $N \times N$ case.

Another important property of the QR decomposition is that we can obtain distinct $\mathbf{Q}$ and $\mathbf{R}$ matrices by permuting the columns of $\mathbf{H}$ for a total number of N! QR decompositions. When the TX signals are coded and decoded in pairs, using space-time coding and decoding, the total number of permutations reduce to (N/2)!. For example, for a 8×8 MIMO channel matrix we can have 24 different QR decompositions when the TX signals are space-time coded in groups of two symbols. Unless otherwise specified, $\mathbf{Q}$ and $\mathbf{R}$ will refer to the base matrices obtained with the default permutation, i.e. $\mathbf{Q}(1,2,..., j, N-1, N)$ and $\mathbf{R}(1,2,..., j, N-1, N)$ for an $N \times N$ MIMO system. Throughout this chapter we will use left superscript index to refer to a specific permutation. When omitted, it means we are refereing to the base permutation.

In order to gain further insight into the proposed algorithm we have to take a closer look at the random distribution of the non-zero elements of the upper triangular matrix $\mathbf{R}$. Assuming that all $h_{i,j}$ are i.i.d. with $\mathcal{CN}(0,1)$, the various non-zero elements $r_{i,j}$ $(i \neq j)$ of $\mathbf{R}$ are also i.i.d. with $\mathcal{CN}(0,1)$ and the magnitude square $|r_{j,j}|^2$ of each diagonal entry $r_{j,j}$ is Chi-squared distributed with $2(N-j+1)$ degrees of freedom[4] [55]. In other words, most of the energy is concentrated in the diagonal elements of $\mathbf{R}$. To see this, we next present further analysis on the relationship between matrices $\mathbf{H}$ and $\mathbf{R}$.

Using the orthogonality property of matrix $\mathbf{Q}$, we can easily verify that $\mathbf{H}^{+}\mathbf{H} =$

[4]Each imaginary or real dimension counts as one degree of freedom

$\mathbf{R}^{+}\mathbf{R}$. In other words, because $\mathbf{R}$ has an upper triangular structure, we can easily see that:

$$\sum_{i=1}^{N} |r_{ij}|^2 = \sum_{i=1}^{N} |h_{ij}|^2, \quad j = 1, 2, ..., N. \tag{4.2.4}$$

Since all the elements of the first column of $\mathbf{R}$ are zero except r_{11}, we have:

$$|r_{11}|^2 = \sum_{i=1}^{N} |h_{i1}|^2. \tag{4.2.5}$$

The previous equation clearly shows that $|r_{11}|^2$ has Chi-square distribution with $2N$ degrees of freedom. In practical terms, this means that $|r_{11}|^2$ will only fade when *all* the $|h_{i1}|^2$ $(i = 1, 2, \ldots, N)$ fade simultaneously. In mathematical terms, we know that $P(|h_{i,j}|^2 < \varepsilon) \approx \varepsilon$ [12], where ε is a small positive constant representing the energy threshold under which reliable communication is not possible. Because the channel coefficients are i.i.d., we have $P(|r_{1,1}|^2 < \varepsilon) \approx \epsilon^N$. In general, we have:

$$\begin{aligned} P(|r_{i,j}|^2 < \varepsilon) &\approx \varepsilon^{N-j+1} \quad i = j \\ P(|r_{i,j}|^2 < \varepsilon) &\approx \varepsilon \qquad\quad i \neq j. \end{aligned} \tag{4.2.6}$$

The above equation establishes that the magnitude square of each diagonal element, i.e. $|r_{jj}|^2$ $(j = 1, 2, \ldots, N)$, has a diversity order of $N - j + 1$ while the magnitude square of each off-diagonal element has a diversity order of 1.

4.3 Algorithm Description

As illustrated in Figure 4.2, we essentially apply the Alamouti space-time code to the transmitted symbols in groups of two for a total of $\frac{N}{2}$ codes. On the RX side, the

description of the proposed detection algorithm differs slightly depending on whether we use a single or multiple QR decompositions.

As the reader will realize soon, our description will be general but there will be a stronger emphasize on the 4×4 and 8×8 cases. The main motivation behind this choice is that these two cases represent the strongest possible candidates for future implementations of 4G standards. For instance, the recently ratified 4G Long Term Evolution (LTE) standard, which is part of the 3rd Generation Partnership Project (3GPP), now supports the 4×4 case as an option [8] while the advanced version, called LTE-advanced (not ratified yet), is moving towards incorporating the 8×8 case in order to meet a target peak bandwidth efficiency of 30 bit/s/Hz [52].

In what follows we describe the algorithm with a single QR decomposition followed by a description of the one where multiple QR decompositions are employed. A less detailed description of the proposed algorithm can also be found in [3].

4.3.1 Algorithm with a Single QR Decomposition

In this base algorithm, we start by computing the $\mathbf{Q}$ and $\mathbf{R}$ matrices which we use to compute a transformed RX vector $\tilde{\mathbf{y}}$. We then decode the interference-free symbols corresponding to the last two rows and columns of $\mathbf{R}$. After decoding these symbols, we subtract their contribution from $\tilde{\mathbf{y}}$. We then successively repeat these two steps until no more symbols are left for decoding. The flow chart shown in Figure 4.4 helps in understanding the various steps followed by the algorithm. Note that the calculation of the $\mathbf{Q}$ and $\mathbf{R}$ matrices are only required once per frame (preprocessing) while all other operations are repeated for every symbol vector (payload processing).

To gain further insight into the proposed algorithm let us write the expression for

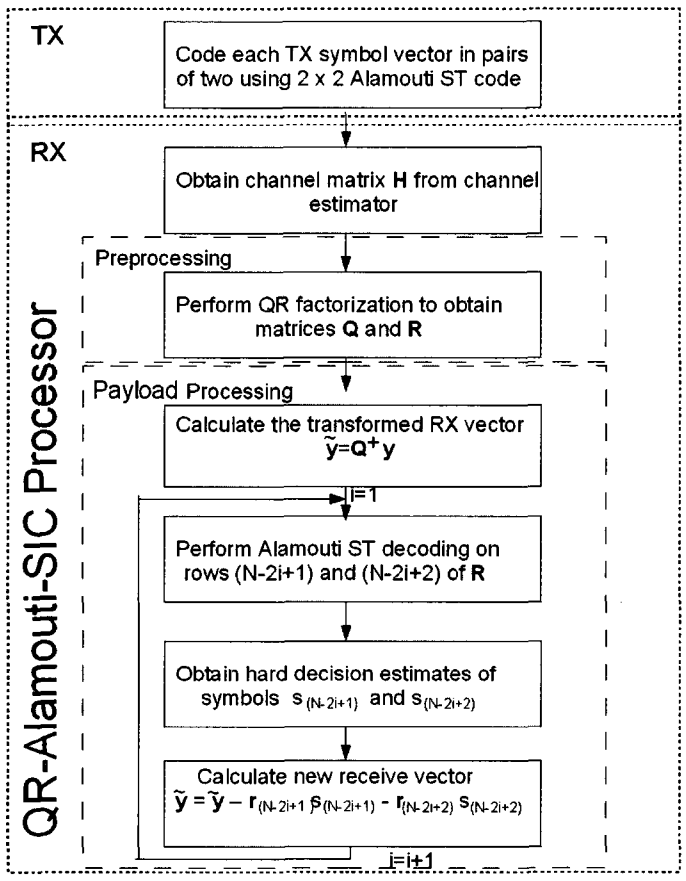

Figure 4.4: Description of the proposed algorithm with a single QR decomposition

the resultant post-Alamouti processing subvector for the q^{th} stage[5]. Ignoring error propagation we have:

$$\underbrace{\begin{bmatrix} \hat{s}_{N-2q+1} \\ \hat{s}_{N-2q+2} \end{bmatrix}}_{\hat{\mathbf{s}}_q} = \sqrt{\frac{\rho}{N}}||\mathbf{\Gamma}_q||^2 \underbrace{\begin{bmatrix} s_{N-2q+1} \\ s_{N-2q+2} \end{bmatrix}}_{\mathbf{s}_q} + \underbrace{\begin{bmatrix} v_{N-2q+1} \\ v_{N-2q+2} \end{bmatrix}}_{\mathbf{v}_q}, q = 1, 2, ..., \frac{N}{2} \qquad (4.3.1)$$

where $||\mathbf{\Gamma}_q||^2$ is the square of the Frobenius norm of submatrix $\mathbf{\Gamma}_q$ defined as:

[5]Note that 'stage' here refers to the decoding of a pair of layers as opposed to a single layer

$$\mathbf{\Gamma}_q = \begin{bmatrix} r_{(N-2q+1)(N-2q+1)} & r_{(N-2q+1)(N-2q+2)} \\ 0 & r_{(N-2q+2)(N-2q+2)} \end{bmatrix} \tag{4.3.2}$$

and the noise subvector $\mathbf{v}_q$ is:

$$\begin{aligned} \mathbf{v}_q &= \begin{bmatrix} r^*_{(N-2q+1)(N-2q+1)} & r_{(N-2q+1)(N-2q+2)} \\ r^*_{(N-2q+1)(N-2q+2)} & -r_{(N-2q+1)(N-2q+1)} \end{bmatrix} \begin{bmatrix} \tilde{n}_{1,(N-2q+1)} \\ \tilde{n}^*_{2,(N-2q+1)} \end{bmatrix} \\ &+ \begin{bmatrix} 0 & r_{(N-2q+2)(N-2q+2)} \\ r^*_{(N-2q+2)(N-2q+2)} & 0 \end{bmatrix} \begin{bmatrix} \tilde{n}_{1,(N-2q+2)} \\ \tilde{n}^*_{2,(N-2q+2)} \end{bmatrix}. \end{aligned} \tag{4.3.3}$$

Note that each component of the noise subvector $\mathbf{v}_q$ contains two independent noise components from two different RX antennas at time instant 1 and time instant 2.

Since the outage performance mainly depends on the symbols decoded first, i.e. s_N and s_{N-1}, it is important to establish their diversity performance. This can be done in a very straightforward manner by observing that these two symbols go through three independent channel paths namely r_{NN}, $r_{(N-1)(N-1)}$ and $r_{(N-1)N}$ and by resorting to (4.2.6). It is not hard to see that the actual diversity enjoyed by these symbols is equal to 4 as $|r_{NN}|^2$ enjoys a diversity of 2 while $|r_{(N-1)(N-1)}|^2$ and $|r_{(N-1)N}|^2$ each enjoys a diversity of 1.

Ignoring the effect of EPP and resorting to (4.2.6), one can easily find that each symbol in symbol-pair (s_{N-2q+1}, s_{N-2q+2}), where $q = 1, 2, ..., N/2$, enjoys a diversity order of $4q$. For example, using this algorithm in an 8×8 MIMO system allows symbols s_1 and s_2 (decoded last) to enjoy a diversity order of 16. Unfortunately,

when a standard SIC receiver is used, as is the case here, the overall diversity order is limited by the diversity order of the first stage [13] (i.e. to 4 in our case). Nevertheless, this increasing diversity order allows the symbols decoded at later stages to enjoy a better SNR so that the probability of outage P_q ($q = 1, 2, \ldots, \frac{N}{2}$), and consequently the FER, improve as q increases [34]. Mathematically we have:

$$P_{\frac{N}{2}} \leq \ldots \leq P_q \leq \ldots \leq P_1. \tag{4.3.4}$$

Since outage performance is dominated by the first stage, we can say that outage approximately occurs when the instantaneous capacity of the first stage falls below a certain desired capacity C_x. That is:

$$P_{out} \approx P_1 = P\left(\log_2(1 + \frac{\rho}{N}||\mathbf{\Gamma}_1||^2) < C_x\right). \tag{4.3.5}$$

The algorithm described above does not optimize the order in which decoding is performed. Similar to any SIC algorithm, one can optimize the outage performance by selecting a proper decoding order. There exist various techniques that can be implemented at either the transmitter or at the receiver that allow us to achieve this objective. A common RX technique is to order the decoding process based on the SNR level starting with the group that possesses the best SNR followed by the one that possesses the next best SNR and so on. We discuss this subject further in the following two subsections.

Optimum Ordering

The optimum decoding order is to decode the various groups based on their instantaneous SNR level. That is we start the decoding with the stage that possesses the

best SNR followed by the one that possesses the next best SNR and so on until we finish the decoding with the one that possesses the worst SNR.

Because there are $\frac{N}{2}!$ possible pair-wise column permutations, one would normally have to perform $N_{qr} = \frac{N}{2}!$ QR factorizations in order to find the optimum decoding sequence. This may not pose a problem for small N but the complexity of optimum ordering becomes prohibitive for large N. Fortunately, using various properties of the QR decomposition, this number can be reduced substantially. We show in Appendix B that a sufficient number of QR decompositions N_{qr} for optimum ordering is:

$$N_{qr} = \begin{cases} 2 & \text{for } N = 4; \\ 4 & \text{for } N = 6; \\ \frac{1}{16}N^2 + \frac{1}{4}N + 1 & \text{for } N \geq 8 \text{ and } \frac{N}{2} \text{ is even}; \\ \frac{1}{16}N^2 + \frac{1}{4}N + \frac{9}{4} & \text{for } N \geq 8 \text{ and } \frac{N}{2} \text{ is odd}. \end{cases} \tag{4.3.6}$$

For instance when $N = 14$ and $N = 16$ we have to perform $N_{qr} = 18$ and $N_{qr} = 21$ QR factorizations, respectively. Note that, since we are only interested in calculating the instantaneous SNR for the various stages, the explicit computation of matrix $\mathbf{Q}$ is not required. The explicit computation of this matrix is only needed for the chosen QR candidate. In the sequel we refer to this ordering method as OPT.

Suboptimum Ordering

The optimum ordering method described above requires the computations of $O(\frac{N^2}{16})$ QR decompositions. A suboptimum method can be obtained by realizing that the FER performance is mostly dominated by the stage that is decoded first. The number of required QR decompositions in this case is $N_{qr} = \frac{N}{2}$ (see Appendix B). Note that computation of the matrix $\mathbf{Q}$ is only necessary for the selected factorization candidate.

In the sequel we refer to this method as SUBOPT1.

The simplest way to perform ordering, which does not require any additional QR factorization, is to base our decision on the Frobenius norm of the various column-vector pairs of the channel matrix $\mathbf{H}$. This method is motivated by relationship (4.2.4) and by the fact that a significant amount of energy in the corresponding column-vectors in matrix $\mathbf{R}$ are in the last two non-zero rows that correspond to our desired SNR. Obviously this will become less and less true as N increases since the energy above the last two rows becomes greater compared to our desired SNR. We ran Monte Carlo simulations to test the validity of this method. We found out, that based on 10,000 independent channel matrices, only 17% and 28% of total orderings are incorrect (compared to the optimum one) for $N = 4$ and $N = 8$, respectively. In the sequel we refer to this method as SUBOPT2.

4.3.2 Algorithm with Multiple QR Decompositions

The ordering methods described above improve the performance of the proposed algorithm by selecting to start the decoding with the group that has the best SNR. While this improves the SNR performance, it does fail to give additional diversity gain as the overall performance is still dominated by the diversity order of the first detected stage.

Another method for improving the performance of the proposed algorithm consists of combining the energy contained in various branches. In a way, this is similar to coherent combining techniques. As we shall see later, this has the effect of improving the overall diversity.

To understand this method better, let us take a closer look at the effect of permuting the various pair-wise columns of matrix $\mathbf{H}$. We take the 4×4 as an example.

The only possible permutations in this case are $(1,2,3,4)$ and $(3,4,1,2)$. Note that column permutation does not change the row ordering whose indices correspond to the RX antenna numbers. So when we perform QR decomposition based on permutation $(1,2,3,4)$, the noise components seen by symbols (s_1, s_2) (s_3, s_4) belong to RX branches $(1,2)$ and $(3,4)$, respectively. Now performing QR decomposition based on permutation $(3,4,1,2)$ will make (s_1, s_2) (s_3, s_4) see noise components belonging to branches $(3,4)$ and $(1,2)$, respectively. Ignoring EPP, the post-Alamouti soft symbol estimates are:

$$\begin{bmatrix} \hat{s}_3 \\ \hat{s}_4 \end{bmatrix} = \sqrt{\frac{\rho}{N}} \left(||^1\mathbf{\Gamma}_1||^2 + ||^2\mathbf{\Gamma}_2||^2 \right) \begin{bmatrix} s_3 \\ s_4 \end{bmatrix} + \begin{bmatrix} {}^1v_3 + {}^2v_1 \\ {}^1v_4 + {}^2v_2 \end{bmatrix} \tag{4.3.7}$$

$$\begin{bmatrix} \hat{s}_1 \\ \hat{s}_2 \end{bmatrix} = \sqrt{\frac{\rho}{N}} \left(||^1\mathbf{\Gamma}_2||^2 + ||^2\mathbf{\Gamma}_1||^2 \right) \begin{bmatrix} s_1 \\ s_2 \end{bmatrix} + \begin{bmatrix} {}^1v_1 + {}^2v_3 \\ {}^1v_2 + {}^2v_4 \end{bmatrix} \tag{4.3.8}$$

where a left superscript index was used to denote the permutation number. Note that $||^1\mathbf{\Gamma}_2||^2$ (or $||^2\mathbf{\Gamma}_2||^2$) has a diversity of 8. The diversity provided by $||^2\mathbf{\Gamma}_1||^2$ (or $||^1\mathbf{\Gamma}_1||^2$) is redundant as they are extracted from the same column pairs. But in the absence of error propagation these redundant Frobenius norms still provide SNR gain.

The above analysis is very optimistic as it does not account for EPP. One way to account for EPP is to resort to relationship (4.3.5). In this case a lower bound on outage rate (or FER) can be obtained if we assume that ${}^1P_1 = {}^1P_2$ and ${}^2P_1 = {}^2P_2$. Using this relationship and because ${}^1P_1$ is independent of ${}^2P_1$ (as they are computed using distinct column pairs) one can for the purpose of analysis write:

$$\begin{bmatrix} \hat{s}_3 \\ \hat{s}_4 \end{bmatrix} = \sqrt{\frac{\rho}{N}} \left(||^1\mathbf{\Gamma}_1||^2 + ||^2\mathbf{\Gamma}_1||^2 \right) \begin{bmatrix} s_3 \\ s_4 \end{bmatrix} + \begin{bmatrix} {}^1v_3 + {}^2v_1 \\ {}^1v_4 + {}^2v_2 \end{bmatrix} \quad (4.3.9)$$

$$\begin{bmatrix} \hat{s}_1 \\ \hat{s}_2 \end{bmatrix} = \sqrt{\frac{\rho}{N}} \left(||^1\mathbf{\Gamma}_1||^2 + ||^2\mathbf{\Gamma}_1||^2 \right) \begin{bmatrix} s_1 \\ s_2 \end{bmatrix} + \begin{bmatrix} {}^1v_1 + {}^2v_3 \\ {}^1v_2 + {}^2v_4 \end{bmatrix}. \quad (4.3.10)$$

The above equation demonstrates that all symbols have a diversity of 8 even after we account for EPP. It is important to emphasize that the noise components after post-Alamouti processing and column permutation stay mutually independent.

Before we can generalize the above to the case where $N > 4$ it is important to realize that, in order to have uncorrelated noise components at the input of the combiner, the maximum number of permutations we can have is $N_{qr}^{max} = \frac{N}{2}$. This is because we can only cycle through a maximum of $\frac{N}{2}$ row-wise pairs (which represent RX antenna numbers) while maintaining mutual independence between the noise components of post-processed Alamouti subvectors.

When choosing the additional permutations, we must do this so as to maximize the total average Frobenius norm seen by each symbol pair. This will make sure that SNR gain is extracted in addition to diversity gain.

To illustrate this point, let us take the 8×8 as an example. Let us assume that our base permutation is $(1, 2, 3, 4, 5, 6, 7, 8)$. Note that in this base permutation symbol pairs $(s_7, s_8), (s_5, s_6), (s_3, s_4)$ and (s_1, s_2) are associated with an average square Frobenius norms of 4,8,12 and 16, respectively.

When we only need to perform two QR decompositions, i.e. N_{qr}=2, the second permutation that should be chosen is $(7, 8, 5, 6, 3, 4, 1, 2)$. This permutation consists of switching the outer column-pairs and the inner column-pairs, respectively. This

choice is unique and ensures that the worst total Frobenius norm is maximized while also ensuring that each symbol pair sees an independent noise vector . In this case, all the average square Frobenius norms are equal to 20. When $N_{qr} = 3$, and since one of the symbol pair has to be permuted to the last two columns in order to ensure noise independence, the maximum worst square Frobenius norm is equal to 24. Note that for $N = 8$ there are two choices for the the third permutation when $N_{qr} = 3$. An arbitrary choice can be made or the choice can be made based on optimum or suboptimum ordering at the expense of additional complexity. Through simulations we found out that a fraction of a dB SNR gain was obtained which in our opinion does not justify the additional complexity. Finally, when $N_{qr} = 4$, and because all symbol pairs cycle through all columns, one unique choice remains and it is easily verified that all average square Frobenius norms are equal to 40.

We should mention that when $\frac{N}{2}$ is odd and $N_{qr} = 2$ it is not possible to obtain an optimum permutation as the column-pair in the middle does not have a corresponding column-pair to switch with. In this case an arbitrary choice between two permutations can be made for the second permutation or one that is based on optimum or suboptimum ordering. Table 4.1 lists the optimum permutations for $N = 4, 6, 8$ and $N_{qr} = 1, 2, ...\frac{N}{2}$. The way to read this table is that for a given N and $N_{qr} \leq \frac{N}{2}$ the permutations that we should use are the ones listed in the N column starting from row one down until the N_{qr} row.

While we restricted our analysis here to $N \leq 8$, as it is the more practical case, similar analysis can be carried out for $N > 8$.

Finally we have to mention that the choice of the base permutation can be arbitrary as ordering would not have an effect on performance because of cycling and

$\mathbf{N_{qr}/N}$	**4**	**6**	**8**
1	$(1,2,3,4)$	$(1,2,3,4,5,6)$	$(1,2,3,4,5,6,7,8)$
2	$(3,4,1,2)$	$(5,6,1,2,3,4)$	$(7,8,5,6,3,4,1,2)$
3	N/A	$(3,4,5,6,1,2)$	$(5,6,7,8,1,2,3,4)$
4	N/A	N/A	$(3,4,1,2,7,8,5,6)$

Table 4.1: Optimum permutations

combining, and because of the properties of the QR decomposition stated in Appendix B. Monte Carlo simulations (not shown here) have been conducted to confirm what we have just stated.

Iterative Detection

We can obtain further improvement in performance by reiterating the SIC procedure in a zigzag fashion in order to help reduce the effect of EPP[6] and consequently help increase the SNR gain (which in turn helps faster convergence to the actual diversity) with each additional iteration. In this case each improved estimate obtained after combining can be reused. Alamouti decoding is bypassed for the first stage and these estimates are directly used in canceling MSI contributed by symbol pairs $({}^{p}s_N, {}^{p}s_{N-1}), p = 1, 2, \ldots, N_{qr}$. Then alternate SIC and Alamouti can be used until no symbols are left for decoding. The multiple-QR version of the algorithm is illustrated in Figure 4.5.

[6]Reducing the effect of EPP in the multiple-QR version has a huge impact on reducing FER as we are not any more limited by the performance of the first decoded stage as in the single-QR version

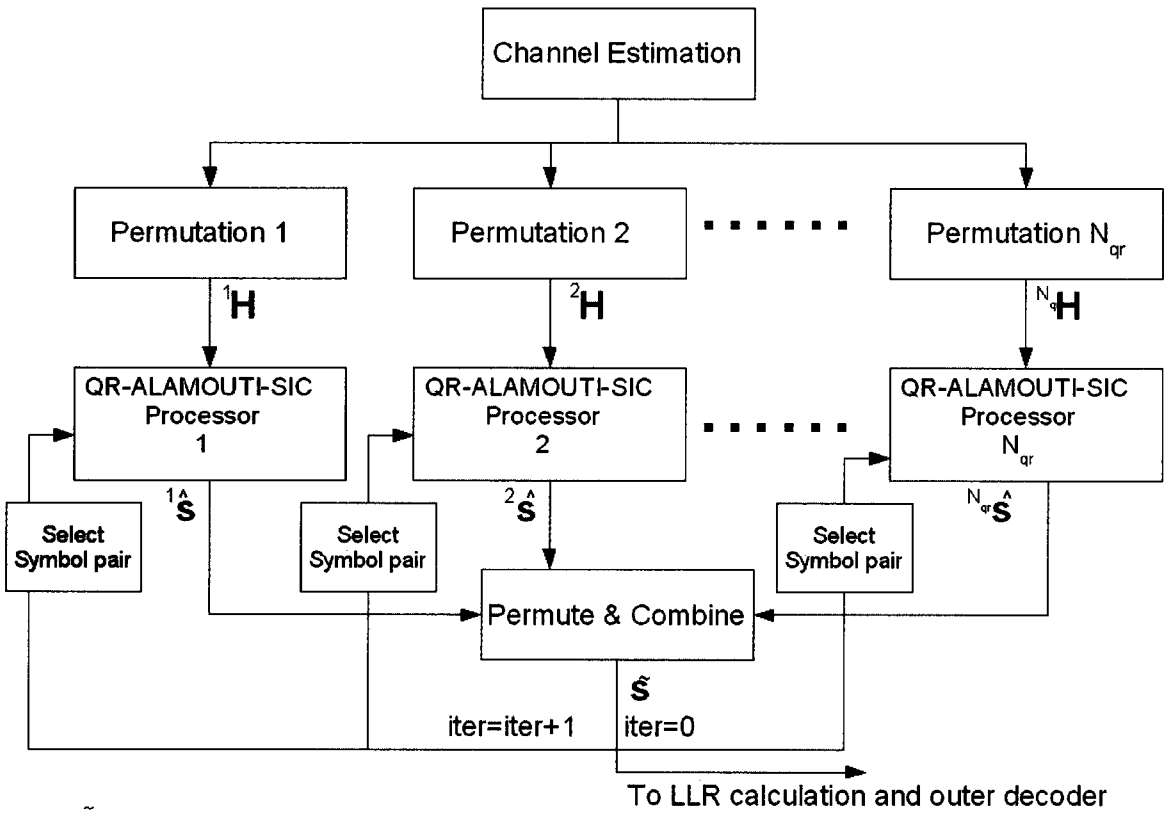

Figure 4.5: Description of the proposed algorithm with multiple QR decompositions

4.4 Information Theoretic Aspects of the Proposed Architecture and Associated Algorithms

In the sequel we will derive the capacity expression of the effective MIMO channel and the capacity expressions of the various versions of the proposed algorithm.

4.4.1 Capacity of the Effective MIMO Channel

It is rather important to derive the capacity of the effective channel that results from the proposed architecture. This will allow us to find out how optimal, in a capacity

sense, the various versions of the proposed algorithm perform. Again, as we did in Section 3.2.2, there is no need to resort to basic principles of information theory if we can find the resulting effective channel matrix $\tilde{\mathbf{H}}$.

Realizing that the proposed architecture is nothing but the stacking of $\frac{N}{2}$ Alamouti codes one on top of the other, we can rewrite the MIMO equation as follows:

$$\underbrace{\begin{bmatrix} \mathbf{y}_1 \\ \mathbf{y}_2 \\ \vdots \\ \mathbf{y}_N \end{bmatrix}}_{\mathbf{y}} = \sqrt{\frac{\rho}{N}} \underbrace{\begin{bmatrix} \mathbf{H}_{1,1} & \mathbf{H}_{3,1} & \cdots & \mathbf{H}_{N-1,1} \\ \mathbf{H}_{1,2} & \mathbf{H}_{3,2} & \cdots & \mathbf{H}_{N-1,2} \\ \vdots & & & \\ \mathbf{H}_{1,N} & \mathbf{H}_{3,3} & \cdots & \mathbf{H}_{N-1,N} \end{bmatrix}}_{\tilde{\mathbf{H}}} \mathbf{s} + \underbrace{\begin{bmatrix} \tilde{\mathbf{n}}_1 \\ \tilde{\mathbf{n}}_2 \\ \vdots \\ \tilde{\mathbf{n}}_N \end{bmatrix}}_{\tilde{\mathbf{n}}} \tag{4.4.1}$$

where

$$\mathbf{H}_{i,j} = \begin{bmatrix} h_{j,i} & h_{j,i+1} \\ h^*_{j,i+1} & -h^*_{j,i} \end{bmatrix}. \tag{4.4.2}$$

Using the expression of the instantaneous capacity in (2.2.1), the capacity of the proposed effective channel can now be obtained as:

$$C_{\tilde{\mathbf{H}}} = \frac{1}{2} \log_2 \ \det\left(\mathbf{I}_N + \frac{\rho}{N} \tilde{\mathbf{H}}^{+} \tilde{\mathbf{H}}\right) \tag{4.4.3}$$

where the a scaling factor of $\frac{1}{2}$ was used because we only send half the amount of symbols when compared to SM algorithms. It is very important to emphasize that the capacity just derived is that of the effective channel $\tilde{\mathbf{H}}$ and not that of the proposed algorithm. In other words, it represents the upper capacity bound that can be achieved by any algorithm that is used to decode the pair-wise Alamouti coded signals.

4.4.2 Capacity of the Algorithm with Single QR Decomposition

Because the proposed algorithm employs successive interference cancellation, we expect that the symbols detected at later stages enjoy a better diversity order (in the absence of EPP) as long as the the first detected stage has a sufficiently high-order diversity. In other words, the worst signal-to-noise ratio SNR_w experienced by these symbols is that of the last two symbols s_N and s_{N-1} (detected first). In this case, and inspired by the analysis presented in [11], the capacity is calculated as:

$$C_1 = \frac{N}{2}\log(1 + SNR_w). \tag{4.4.4}$$

Because of the structure of the proposed architecture and the use of the QR decomposition, it is easily found that:

$$\begin{aligned} SNR_w &= \frac{\rho}{N}||\mathbf{\Gamma}_1||^2 \\ &= \frac{\rho}{N}\left(|r_{NN}|^2 + |r_{(N-1)(N-1)}|^2 + |r_{(N-1)(N)}|^2\right). \end{aligned} \tag{4.4.5}$$

4.4.3 Capacity of the Algorithm with Multiple QR Decompositions

Deriving the exact expression for the instantaneous capacity in this case is not possible due to the non-linear nature of the SIC operation and the existence of EPP. We can however derive a lower bound by resorting to (4.3.5). Using this relationship, one can easily find that the instantaneous capacity $C_{N_{qr}}$ is lower bounded as:

$$C_{N_{qr}} \geq \frac{N}{2}\log(1 + SNR_{N_{qr}}) \tag{4.4.6}$$

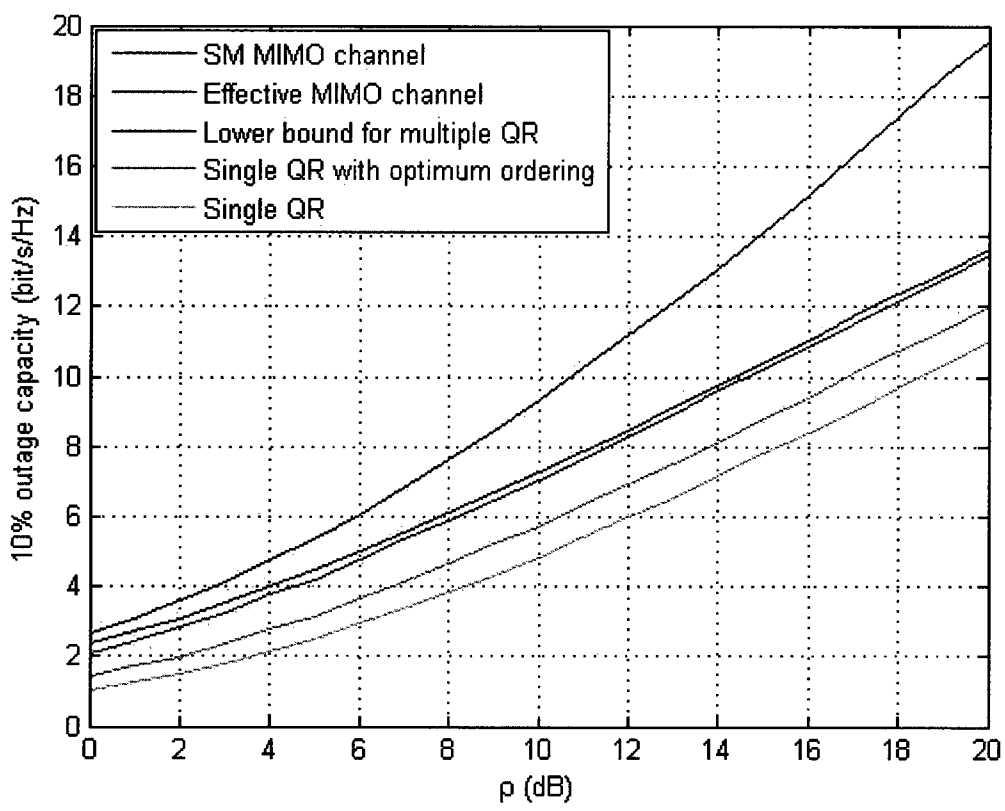

Figure 4.6: 10% Outage capacity of various MIMO channels and proposed algorithms for 4×4 i.i.d. MIMO channel

where

$$SNR_{N_{qr}} = \frac{\rho}{N} \sum_{i=1}^{N_{qr}} ||^{i}\mathbf{\Gamma}_1||^2. \tag{4.4.7}$$

Note that to maximize capacity ${}^{i}\Gamma_1$ must be chosen according to Table 4.1.

4.4.4 Capacity Comparison

In this section we compare the capacities offered by the proposed algorithm to the MIMO channel and to the effective MIMO channel assuming the channel coefficients are i.i.d. The various capacities are obtained by Monte Carlo simulations where 10,000 independent channel realizations are used for each SNR.

Figure 4.6 shows the 10% outage capacity of the various channels and various versions of the proposed algorithm for the 4×4 i.i.d. case. As expected, the capacity offered by the single-QR version offers the poorest capacity performance. Optimum ordering enhances capacity by 1 bit/s/Hz while the multiple-QR version (with $N_{qr} = N_{qr}^{max} = 2$) offers the best capacity with 2.2 bit/s/Hz improvement. It is interesting to note that the lower bound on capacity achieved by the multiple-QR version almost matches the capacity that is offered by the effective channel. This demonstrates that the lower bound in (4.4.6) is very tight.

Figure 4.7 shows the 10% outage capacity of the various channels and various versions of the proposed algorithm for the 8×8 case. In this case, we can clearly see the incremental capacity improvement with each additional QR decomposition. Again we see that the version with a single-QR decomposition offers the worst capacity. Optimum ordering offers substantial improvement (about 4 bit/s/Hz) in capacity in this case compared to the 4×4 case. This improvement is due to the fact that we have more independent choices regarding the stage with which to start our decoding. We also clearly see that when we perform additional QR decompositions we get diminishing returns. For instance we get capacity enhancement of 5,8,10 bit/s/Hz when $N_{qr} = 2, 3, 4$, respectively. This is indeed expected; according to (4.4.6) the increase in diversity obtained from performing multiple-QR decompositions contributes

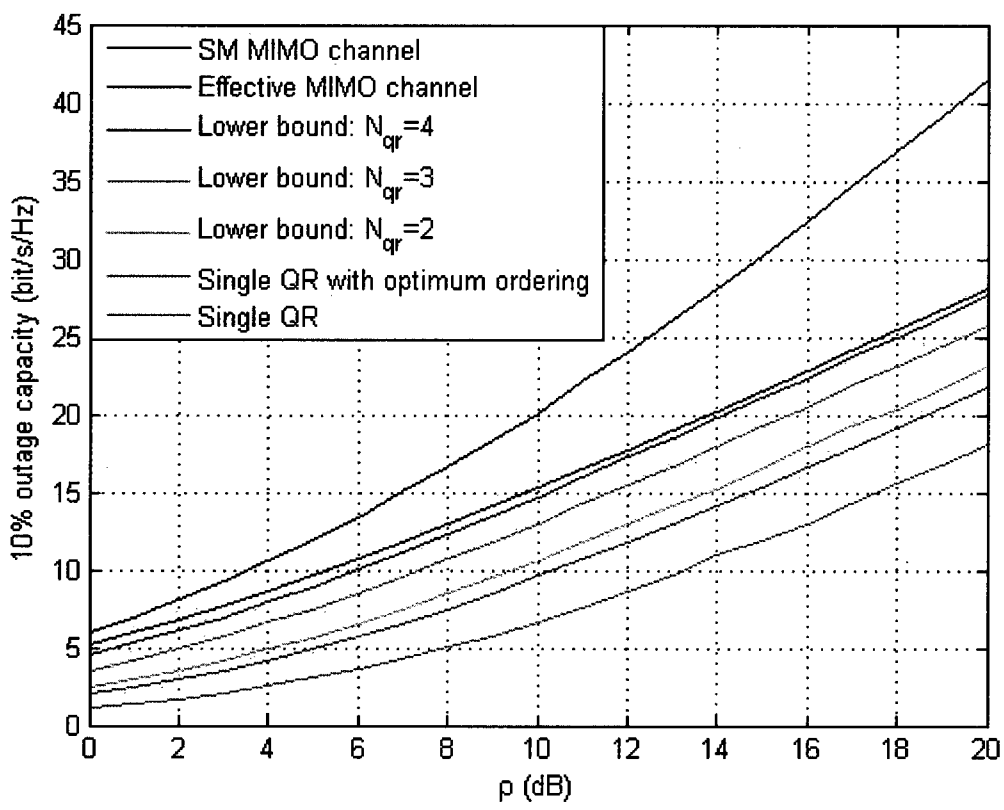

Figure 4.7: 10% Outage capacity of various MIMO channels and proposed algorithms for 8×8 i.i.d. MIMO channel

only logarithmically to capacity enhancement. Again we see that the lower bound on capacity almost matches that of the effective channel leading us to conclude that the multiple-QR version with N_{qr}^{max} achieves the full capacity of the effective MIMO channel.

4.5 FER Performance for the i.i.d. Channel

We present here Frame Error Rate (FER) simulation results for 4×4 and 8×8 MIMO systems with a bandwidth efficiency of 8 bit/s/Hz and 16 bit/s/Hz, respectively. In order to achieve this bandwidth efficiency, the proposed algorithm uses 16QAM modulation in each subchannel. We use a transmission duration of 130 payload symbols per subchannel. For each SNR we run the simulation until we have 100 frame errors. Note that frame here is defined as the entire block of symbols transmitted from all antennas, i.e. $N_t \times 130$. An error is declared when at least one RX symbol is decoded incorrectly. This definition of FER allows us to find out how far we are operating from capacity by comparing the required SNR, at a given FER, to the outage capacity curve. It should be mentioned that this same definition is used in the literature but sometimes it is understandably referred to as BLER (Block Error Rate) instead.

The choice of a 130-symbol time slot is mainly motivated by our desire to have a fair comparison with the most important proposed algorithms which used this frame size. This frame size is the length of a payload time slot[7] in IS-136, a TDMA-based North American cellular standard. Historically, this standard was still dominant when space-time coding research intensified (between mid-to-late 1990's).

4.5.1 FER of the Single-QR Version with or without Ordering

Here we are interested in examining the effect of ordering on the FER performance of the single-QR version. Figure 4.8 shows the results for the 4×4 system while

[7]In fact the size of a time slot in IS-136 is 162 symbols, 32 symbols of which are used for synchronization and channel estimation purposes

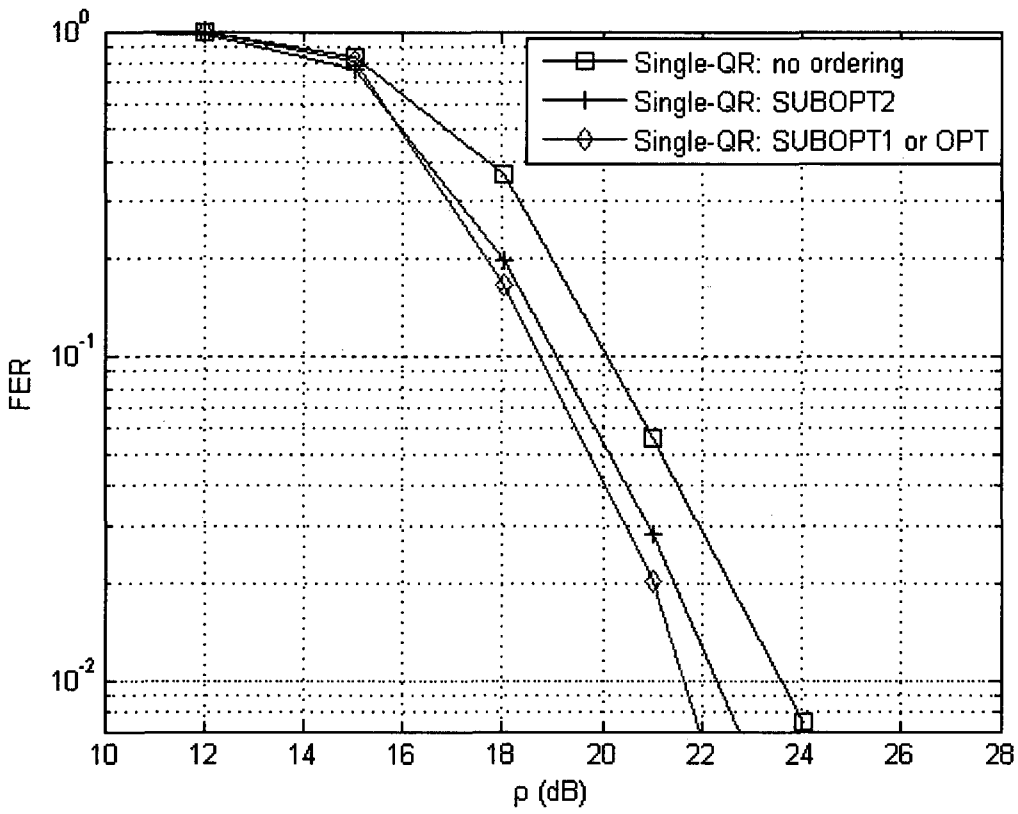

Figure 4.8: FER of the single-QR version with or without ordering for a 4×4 i.i.d. MIMO channel and a bandwidth efficiency of 8 bits/s/Hz

Figure 4.9 shows the results for the 8×8 system.

For the 4×4 case, the effect of simple ordering provided by SUBOPT2 is a gain of 1.2dB SNR at 1% FER. Optimum ordering (OPT), which is also equivalent to suboptimum ordering SUBOPT1 in this special case[8], provides a gain of 1.9dB SNR at the same FER. Because the sharpness of the curves with ordering increases with increasing SNR, we may erroneously conclude that ordering provides diversity

[8]This is only true when N=4

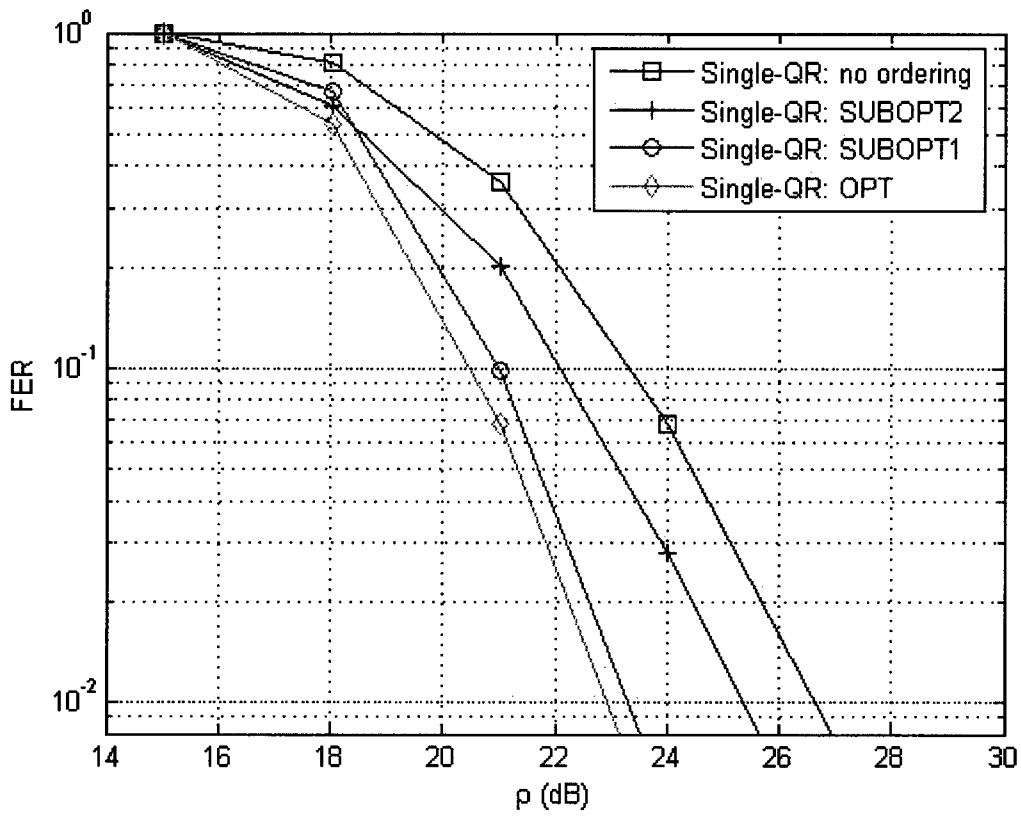

Figure 4.9: FER of the single-QR version with or without ordering for a 8×8 i.i.d. MIMO channel and a bandwidth efficiency of 16 bits/s/Hz

gain. Notice that the curve of the single-QR without ordering does not decrease proportionally to ρ^{-4} at finite SNR (the diversity is approximately 2.9 as 3.5dB is needed per decay). It only achieves this rate of decay at infinite-SNR levels. So in that respect we can say that ordering allows the single-QR version to converge to its asymptotic decay value faster. Comparing this figure to Figure 4.6 we conclude that the single-QR version without ordering and the single-QR version with OPT (or SUBOPT1) are 4.5dB, 5.25dB away from the 10% outage capacity, respectively.

For the 8×8 case the effect of the simple ordering provided by SUBOPT2 is 1.2dB at an FER of 1%. This SNR gain is the same as that provided in the 4×4 case. This is despite the fact that a greater percentage of false choices of orderings are made. This can be explained as follows: while there are more wrong ordering choices there are also more choices compared to the 4×4 case. This increased number of independent choices seems to compensate for the greater number of wrong ordering choices. This explanation is supported by looking at the curves obtained with SUBOPT1 and OPT. With these ordering methods, where we have the correct choice for the first stage (with SUBOPT1) and correct choices for all stages (with OPT), we obtain SNR gains of 3.3dB and 3.8dB at FER of 1% (compared to 1.4dB for the 4×4 case), respectively.

The incremental SNR gain of 0.5dB obtained with OPT compared to SUBOPT1 in the 8×8 case does not, in our opinion, justify the added complexity of computing three additional $\mathbf{R}$ matrices. We conclude that ordering method SUBOPT1 offers a good tradeoff between complexity and performance. Note that the curves with OPT and SUBOPT2 attain the asymptotic diversity order of 4 as only 2.5dB is needed per decay. Comparing this figure to Figure 4.7 we conclude that the single-QR version without ordering and single-QR version with OPT are 4.75dB, 5.25dB away from the 10% outage capacity, respectively.

4.5.2 FER of the Different Stages for the Single-QR Version

We mentioned in Section 4.2.1 that the diversity increases after each stage. This diversity theoretically reaches an order of $2N$ for the last stage[9]. We also mentioned that this would only be true if the effect of EPP is completely eliminated. For illustration purposes, Figure 4.10 shows the FER of the individual stages of an 8×8

[9]This is a direct consequence of the relationship in (4.2.6) and the use of Alamouti decoding

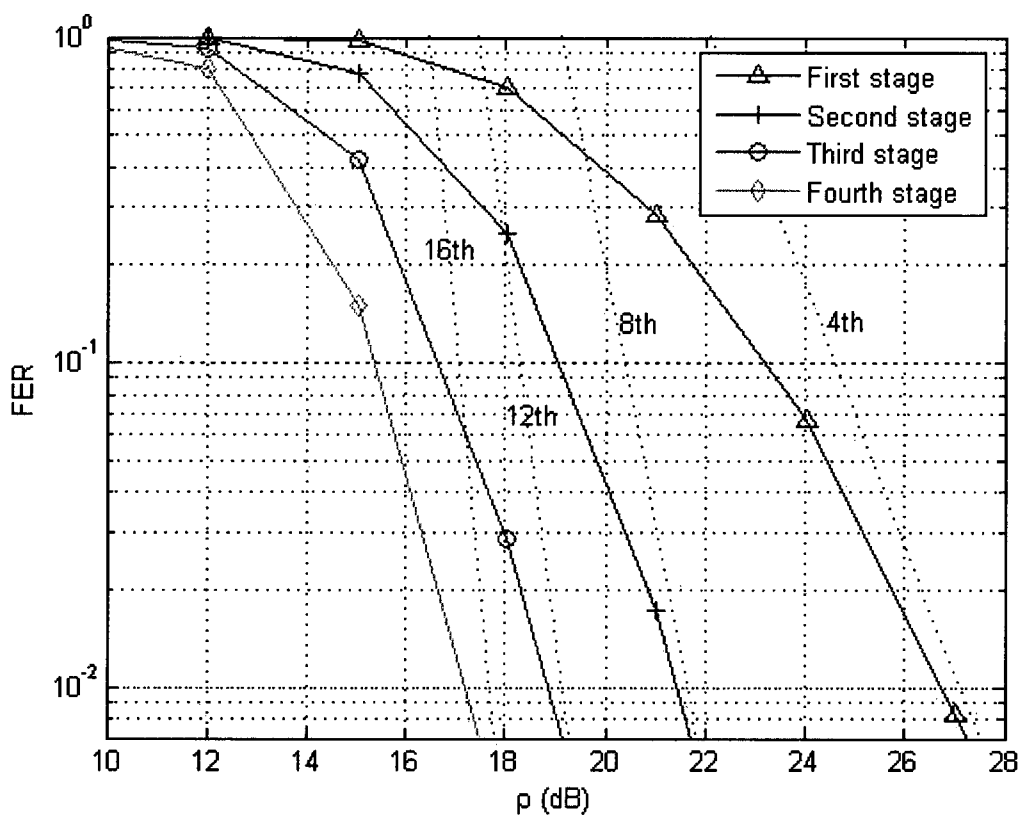

Figure 4.10: EPP-free FER of different stages for 8×8 i.i.d MIMO using the single-QR version of the algorithm. The dotted lines represent the theoretical asymptotic diversity curves.

system using the "genie" concept. That is, we use the actual symbols instead of the detected ones for the interference cancellation phase. This basically eliminates the effect of EPP. We clearly see in the figure that the diversity increases according to the relationship in (4.2.6).

When EPP is accounted for, as is the case in practice, the diversity enjoyed by each stage is the smallest of all stages, i.e. the diversity of the first detected stage. Figure 4.11 shows the FER of each individual stage obtained in practice. Clearly,

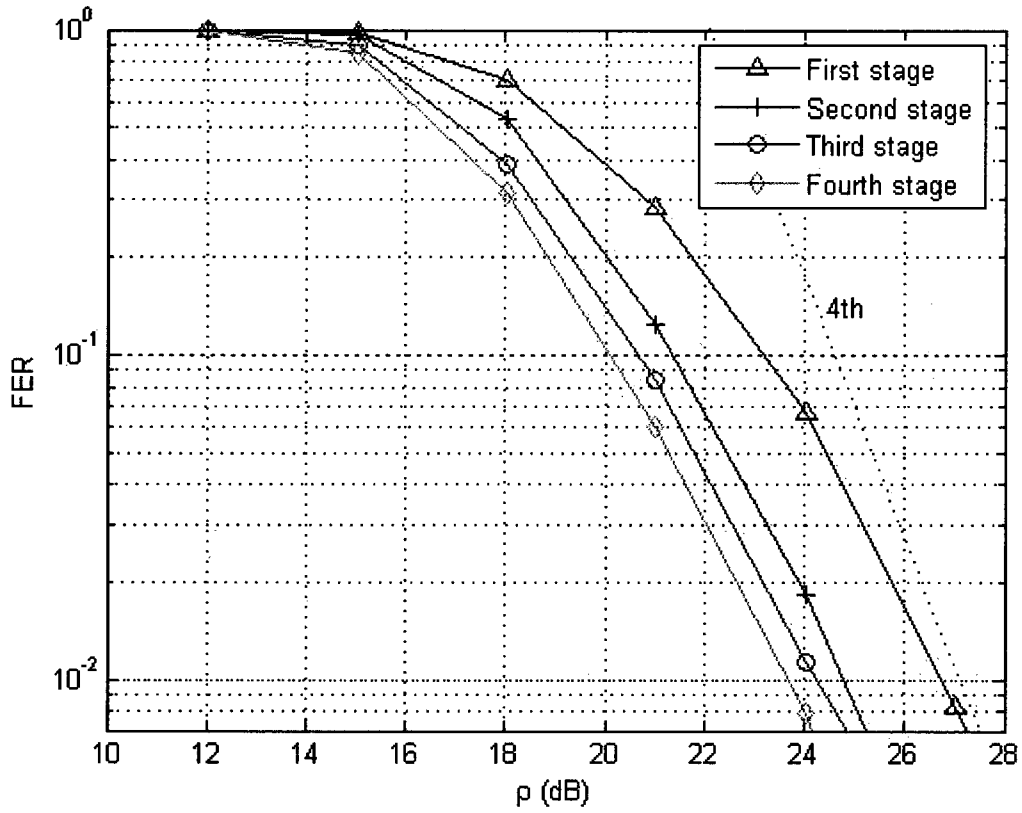

Figure 4.11: FER of different stages for 8×8 i.i.d MIMO using the single-QR version of the algorithm. The dotted line represents the theoretical asymptotic diversity curves.

all stages enjoy the same diversity order of 4. It is important to note that while each stage enjoys only a diversity of 4, the SNR gain increases after each stage. This "feature" can be exploited in modern wireless modems where the type of transmitted information (i.e. voice, video or data) requires different FER and delay requirements. For example, voice data, which does not require stringent FER but rather requires the smallest delay possible, can be transmitted using the first stage while data, which has opposite requirements to voice, can be transmitted using the last stage. Video,

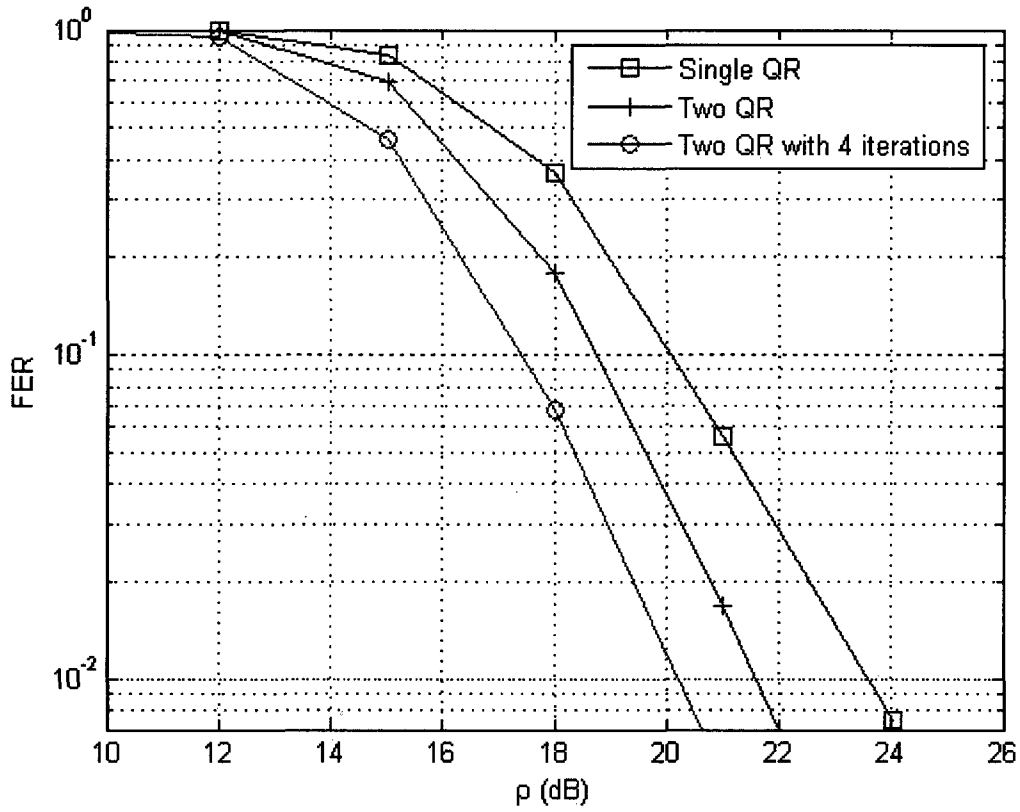

Figure 4.12: FER of the various versions of the proposed algorithm for a 4×4 i.i.d. MIMO channel and a bandwidth efficiency of 8 bits/s/Hz

which has medium requirements between voice and data, can be transmitted using the intermediate stages.

It is important to note that, because of how FER is defined, the overall FER for the EPP-free version and the actual one will be close as the error rate is dominated by the first stage which has the worst outage performance.

4.5.3 FER of the Multiple-QR Version

In this section we are interested in examining the effect of performing additional QR decompositions and iterative detection. Figure 4.12 shows the results for the 4×4 case. First, we clearly see that the single-QR version has the worst FER. The one with two QR decompositions has a better performance. For instance, the latter outperforms the former by 2dB at 1% FER. The positive effect of turbo detection is also evident in the figure; an additional 1.4dB SNR gain is obtained at the same FER. Note that the curve with two QR decompositions, with or without turbo iterations, has a diversity slightly higher than 4. It is expected that the sharpness of both curves will increase with SNR to reach its final value of 8 at extremely high SNR levels. Comparing this figure to Figure 4.6 we conclude that the single-QR and the two-QR (with 4 iterations) versions are 4.5dB and 5.25dB away from the 10% outage capacity, respectively.

The results for the 8×8 are shown in Figure 4.13. In general, we see the same trends we saw in the 4×4 case. But in this case, we have the benefit of examining the incremental SNR advantage each QR decomposition brings: at 1% FER we get SNR gains of 3dB, 3.8dB and 4.7dB from performing one additional QR decomposition, two additional QR decompositions and three additional QR decompositions, respectively. The effect of turbo iterations is significant as we get an additional SNR gain of 3dB at the same FER. It is also evident that in this case turbo iterations help the four-QR curve converge to its final diversity of 16 quicker as demonstrated by the sharpness of the curve. By comparing the EPP-free FER performance of the first two symbols (which possess a diversity of 16) in Figure 4.11 to that of the four-QR version with turbo iterations in Figure 4.13 we can conclude that the full diversity of 16 will

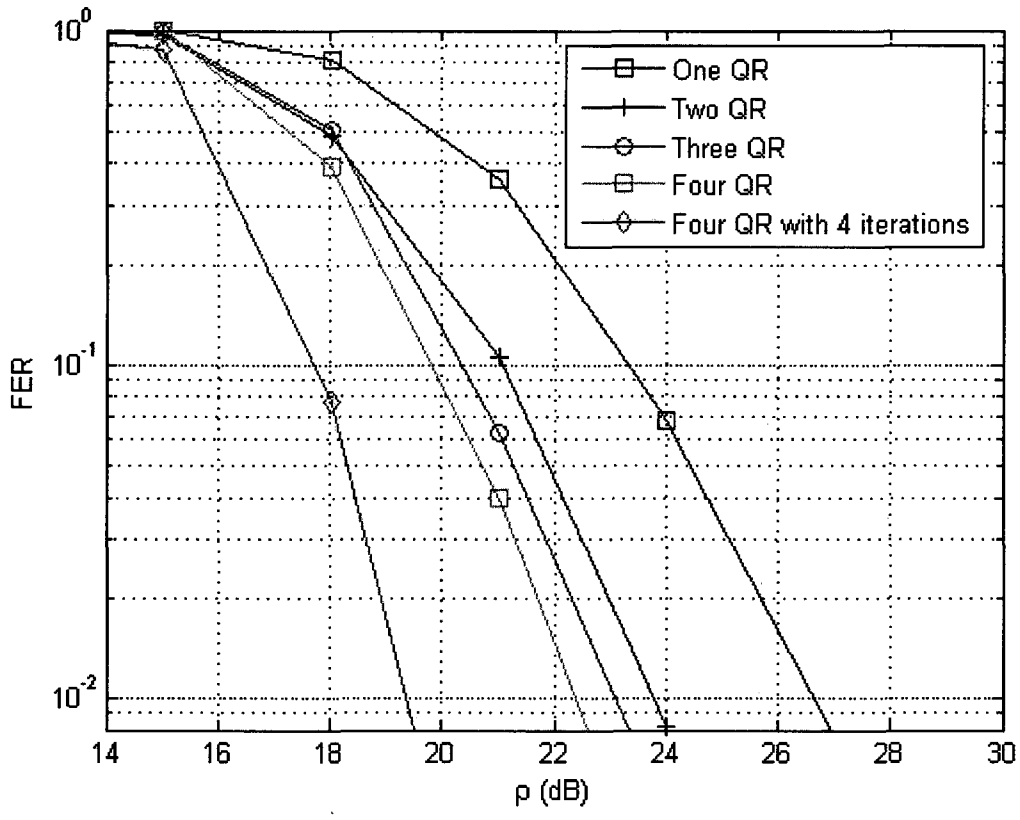

Figure 4.13: FER of the various versions of the proposed algorithm for a 8×8 i.i.d. MIMO channel and a bandwidth efficiency of 16 bits/s/Hz

be achieved at high SNR as the curves possess the same slope at finite SNR. It is noteworthy that only a uniform SNR gain of 2.25dB is needed to achieve the EPP-free performance of the first two symbols. This is indeed quite remarkable. Only the 16×8 ML algorithm[10] can achieve this EPP-free performance but its complexity is extremely prohibitive.

Another point to highlight in the 8×8 case is the significant improvement obtained

[10]This is because the effective MIMO channel is of size 16×8 for the 8×8 case

with the use of turbo iterations (3dB at 1%FER) compared to the 4×4 case (only 1.4dB at 1% FER). This can be explained by the fact that EPP becomes worse with higher number of antennas and this proves that turbo detection is an effective way to reduce its effects on the overall performance.

Finally, it is important to highlight that which version to use during the processing of any given frame will depend on the available SNR and FER required. For instance, given the 8×8 as with an available SNR of 24dB and a required FER of 10%, the use of the single-QR version is sufficient. If the SNR drops to 20dB then the use of the multiple-QR with $N_{qr} = 4$ becomes essential to maintain the same FER. This feature is essential to have in modern wireless systems where optimizing power consumption is of prime importance. Note also that parallelism is inherent in the multiple-QR version which makes it suitable for high speed implementations of battery-powered terminals.

How Much Turbo Iterations are Needed?

Here we are interested in finding out the incremental SNR gain obtained with turbo iterations. Even though turbo iterations can be used when $N_{qr} < \frac{N}{2}$, we limit our analysis to the case where the maximum number of QR decompositions are used, i.e. $N_{qr} = \frac{N}{2}$. Simulations shown in Figure 4.14 and Figure 4.15 show that for the 4×4 and the 8×8 we do not get any more improvement beyond 4 iterations. It is worth noting that in both cases only one iteration is needed to achieve at least 70% of the SNR gain obtained with 4 iterations. We conclude that the multiple-QR version with one iteration offers a good compromise between performance and complexity.

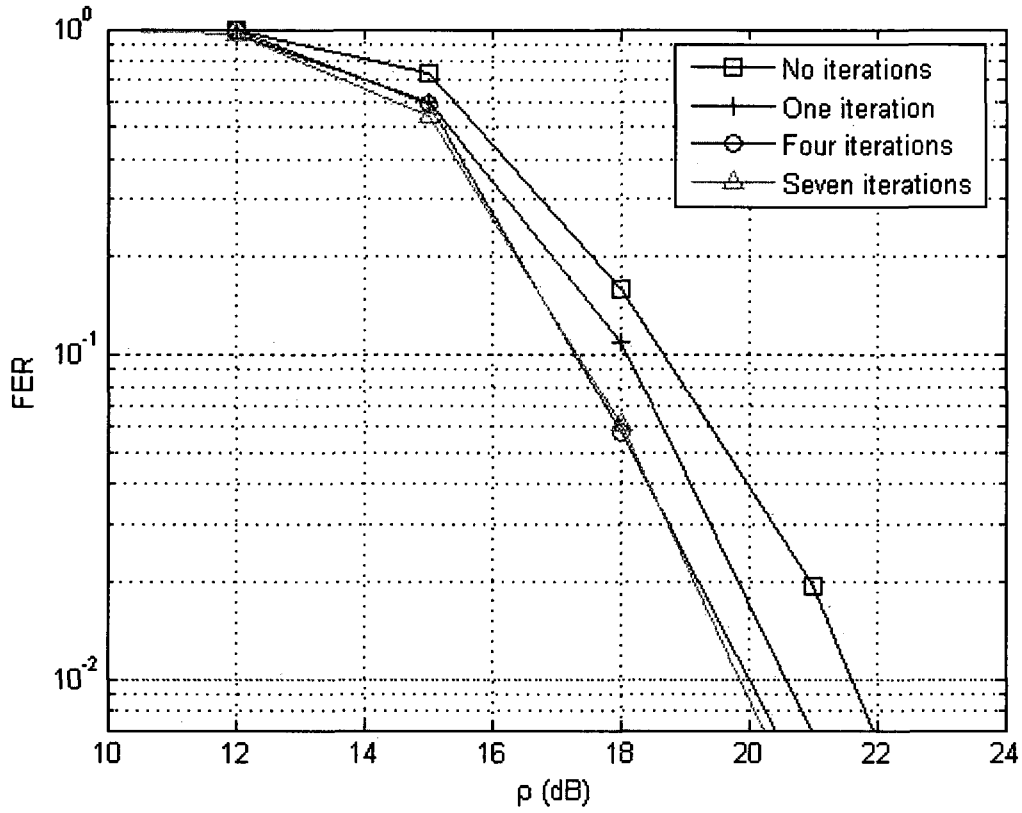

Figure 4.14: Effect of the number iterations in the 4×4 MIMO case

4.6 Complexity Analysis

It is important to establish the complexity of the proposed algorithm so that a fair comparison can be made with other algorithms. A straight forward way to do this is to compute the amount of complex floating-point operations (CFLOPS) that are required.

We will strictly consider multiplications and additions as they are the dominant

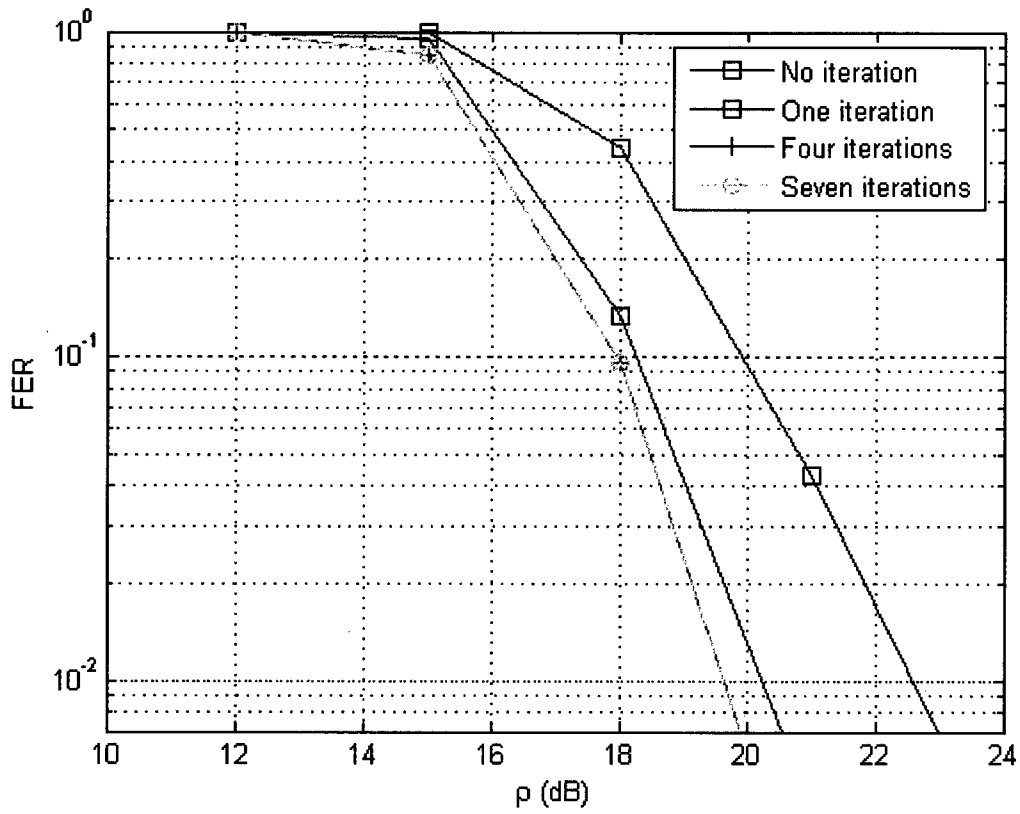

Figure 4.15: Effect of the number iterations in the 8×8 MIMO case

operations. When comparing the obtained complexity to the complexity of other algorithms it is essential that we convert the CFLOPS to real floating-point operations (RFLOPS or FLOPS for short). This is easily done by realizing that each complex multiplication is composed of six FLOPS (four real multiplications and two real additions) and each complex addition is composed of two FLOPS (two real additions). Since we are assuming block fading where the channel only changes on a frame-by-frame basis, it is important to separate the payload processing complexity which is required for every symbol vector from that of preprocessing which is only required

once per frame.

4.6.1 Complexity of the Single-QR Version

In the sequel we will derive the complexity of the single-QR version of the proposed algorithm. Deriving the complexity of the multiple-QR version then becomes straightforward as we shall see later. The complexity of TX Alamouti processing is ignored as it is extremely simple and does not involve any multiplication/addition operations.

The single-QR version of the proposed algorithm is composed of the following signal processing stages: computation of matrices $\mathbf{Q}$ and $\mathbf{R}$, RX vector projection, SIC, and Alamouti RX processing. The complexity of each is discussed next:

1. **Computation of Q and R**: There are three main techniques used to obtain the matrix $\mathbf{R}$, namely: Gram-Schmidt, Givens Rotations, and Householder Transformation [24]. The latter method seems to be the most desirable as it is the most stable and offers a reasonable complexity [31]. Assuming the use of Householder Transformations, the cost of obtaining the matrix $\mathbf{R}$ is $\frac{4}{3}N^3 + O(N^2)$ multiplications and $\frac{4}{3}N^3+O(N^2)$ additions [24][25]. To obtain the matrix $\mathbf{Q}$ costs the same number of operations for a total number of $\frac{8}{3}N^3 + O(N^2)$ multiplications and $\frac{8}{3}N^3 + O(N^2)$ additions. In the sequel we will drop the dependance of the above complexity expressions on $O(N^2)$ as its contribution to the overall complexity is almost negligible for $N \geq 4$. It is important to note that the computation of these matrices is required only once every frame as the channel is considered to be quasi-static.

2. **RX vector projection**: To obtain the new transformed RX vector requires

multiplying the original RX vector by $\mathbf{Q}^+$. This transformation costs N^2 multiplications and $N^2 - N$ additions. This operation is required for every new symbol vector.

3. **SIC**: Because of the upper triangular structure of $\mathbf{R}$, and the fact that SIC is not performed for the last stage, SIC requires only $\frac{N^2+N}{2} - 3$ multiplications and $\frac{N^2+N}{2} - 3$ additions. This operation is required for every new symbol vector.

4. **Alamouti RX processing**: Each individual Alamouti processing requires six multiplications and four additions. Because we have $\frac{N}{2}$ Alamouti RX processing to perform every two symbol periods, the total adds up to $6 \times \frac{N}{2} \times \frac{1}{2} = \frac{3}{2}N$ multiplications and $4 \times \frac{N}{2} \times \frac{1}{2} = N$ additions. This operation is required for every new symbol vector.

It should be emphasized again that operation 1 is performed once per frame (preprocessing) while operations 2 to 4 are performed once every symbol vector (payload processing). Table 4.2 summarizes the complexity of the single-QR version of proposed algorithm.

Complexity of Ordering

We here find the complexity of the optional ordering methods described in Section 4.3.1. We restrict our analysis to SUBOPT1 and SUBOPT2 as the OPT method offers only minimal improvement for a substantial increase in complexity (see Figure 4.9).

1. SUBOPT1: Ignoring the cost of comparators, we require the computation of $\frac{N}{2} - 1$ additional $\mathbf{R}$ matrices and the computation of $\frac{N}{2}$ square Frobenius norms.

Stage	**Multiplications**	**Additions**	**Flops**
Comp. of **Q** and **R**	$\frac{8}{3}N^3$	$\frac{8}{3}N^3$	$\frac{64}{3}N^3$
RX vector proj.	N^2	$N^2 - N$	$8N^2 - 2N$
SIC	$\frac{N^2+N-6}{2}$	$\frac{N^2+N-6}{2}$	$4N^2 + 4N - 24$
Alamouti RX proc.	$\frac{3}{2}N$	N	$11N$
Total: Preprocessing	$\frac{8}{3}N^3$	$\frac{8}{3}N^3$	$\frac{64}{3}N^3$
Total: Payload processing	$\frac{3}{2}N^2 + 2N - 3$	$\frac{3}{2}N^2 + \frac{1}{2}N - 3$	$12N^2 + 13N - 24$

Table 4.2: Required number of operations for the single-QR version of the proposed algorithm.

This latter requires 3 multiplications and 2 additions per square norm for a total of $\frac{3}{2}N$ multiplications and N additions. As stated previously the computation of matrix **R** requires $\frac{4}{3}N^3$ multiplications and $\frac{4}{3}N^3$ additions. So the total is: $\frac{2}{3}(N^4 - 2N^3) + \frac{3}{2}N$ multiplications and $\frac{2}{3}(N^4 - 2N^3) + N$ additions.

2. SUBOPT2: The computation of the square Frobenius norm of each column of matrix **H** requires N multiplications and $N - 1$ additions. Because we have N such columns the total number of required multiplications is N^2. Because we operate in column pairs we need to perform $\frac{N}{2}$ extra additions for a total of $\frac{2N^2-N}{2}$ additions.

It should be noted that the ordering operations are only performed once per frame and as a consequence they form part of the preprocessing phase. Table 4.3 summarizes the number of operations required by the optional ordering algorithms.

Ordering Method	Multiplications	Additions	Flops
SUBOPT1	$\frac{2}{3}(N^4 - 2N^3) + \frac{3}{2}N$	$\frac{2}{3}(N^4 - 2N^3) + N$	$\frac{16}{3}(N^4 - 2N^3) + 11N$
SUBOPT2	N^2	$\frac{2N^2-N}{2}$	$8N^2 - N$

Table 4.3: Required number of operations for the ordering algorithms

4.6.2 Complexity of the Multiple-QR Version

We first derive the complexity of the multiple-QR version with no iteration. Essentially the multiple-QR version is composed of N_{qr} single-QR versions in parallel along with a combining operation in between. So the complexity calculation can be based on Table 4.2. We just have to account for the cost of coherent combining which requires only $N(N_{qr} - 1)$ additions. A single division operation is required the complexity of which was not included in our calculations for two reasons. First, the algorithms we are going to compare ours to ignore the cost of division as it is only rarely used (as is the case here). Second, when N_{qr} is a power of 2, the division operation can simply be achieved by a simple right-shift operation. The complexity of the multiple-QR version is summarized in Table 4.4.

Complexity of Iterative Detection

Because Alamouti processing is not needed for the first stage we only perform a total of $N_{it}N_{qr}(\frac{N-2}{2})$ extra Alamouti RX processing where N_{it} is the number of iterations. This translates into $3N_{it}N_{qr}(\frac{N-2}{2})$ multiplications and $N_{it}N_{qr}(N-2)$ additions for Alamouti processing. The total additional cost for SIC is $N_{it}N_{qr}(\frac{N^2+N-6}{2})$ multiplications and $N_{it}N_{qr}(\frac{N^2+N-6}{2})$ additions. The additional cost of coherent combining is $N_{it}N(\frac{N_{qr}-1}{2})$ additions. The required number of extra payload processing operations is summarized in Table 4.5.

Stage	**Multiplications**	**Additions**	**Flops**
Comp. of **Q** and **R**	$\frac{8}{3}N_{qr}N^3$	$\frac{8}{3}N_{qr}N^3$	$\frac{64}{3}N_{qr}N^3$
RX vector proj.	$N_{qr}N^2$	$N_{qr}(N^2-N)$	$N_{qr}(8N^2-2N)$
SIC	$N_{qr}(\frac{N^2+N-6}{2})$	$N_{qr}(\frac{N^2+N-6}{2})$	$N_{qr}(4N^2+4N-24)$
Alamouti RX proc.	$\frac{3}{2}N_{qr}N$	$N_{qr}N$	$11N_{qr}N$
Coherent combining	0	$N(\frac{N_{qr}-1}{2})$	$N(N_{qr}-1)$
Total: Preprocessing	$\frac{8}{3}N_{qr}N^3$	$\frac{8}{3}N_{qr}N^3$	$\frac{64}{3}N_{qr}N^3$
Total: Payload processing	$N_{qr}(\frac{3}{2}N^2+2N-3)$	$N_{qr}(\frac{3}{2}N^2+\frac{1}{2}N-3)+N(\frac{N_{qr}-1}{2})$	$N_{qr}(12N^2+13N-24)+N(N_{qr}-1)$

Table 4.4: Required number of operations for the multiple-QR with no iterations

Stage	**Multiplications**	**Additions**	**Flops**
SIC	$\frac{1}{2}N_{it}N_{qr}(N^2+N-6)$	$\frac{1}{2}N_{it}N_{qr}(N^2+N-6)$	$N_{it}N_{qr}(4N^2+4N-24)$
Alamouti RX proc.	$\frac{3}{2}N_{it}N_{qr}(N-2)$	$N_{it}N_{qr}(N-2)$	$11N_{it}N_{qr}(N-2)$
Coherent combining	0	$\frac{1}{2}N_{it}N(N_{qr}-1)$	$N_{it}N(N_{qr}-1)$
Total:	$N_{it}N_{qr}(\frac{1}{2}N^2+2N-6)$	$N_{it}N_{qr}(\frac{1}{2}N^2+2N-5)-\frac{1}{2}N_{it}N$	$N_{it}N_{qr}(4N^2+16N-46)-N_{it}N$

Table 4.5: Extra payload processing required for iterative detection

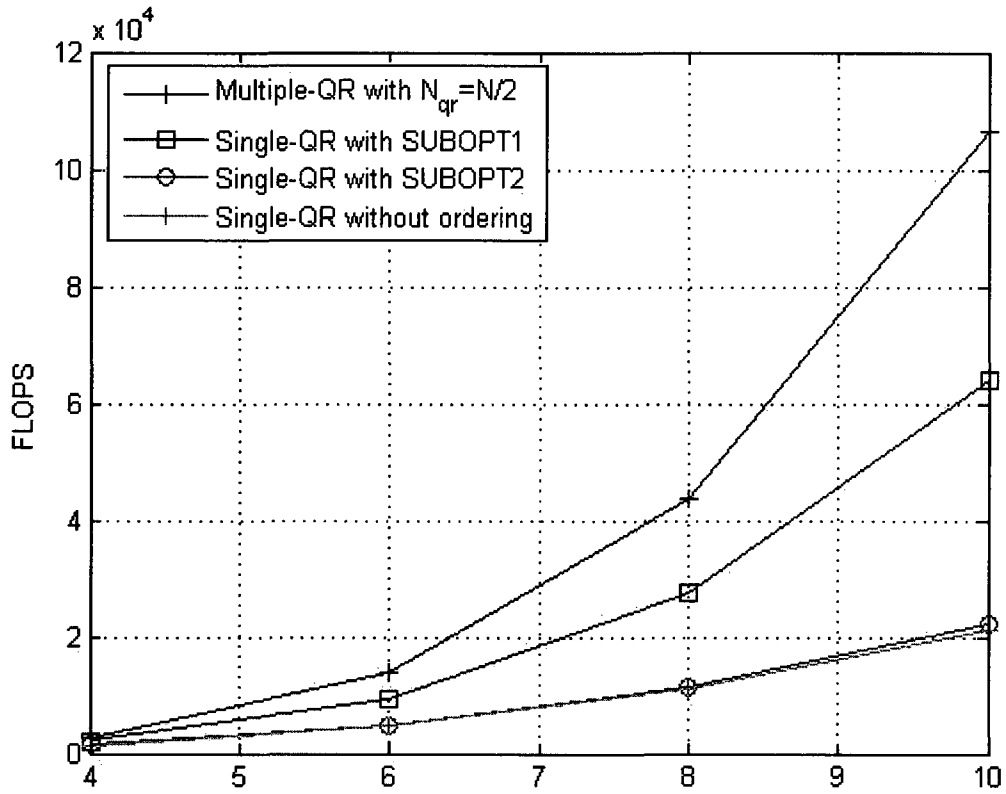

Figure 4.16: Preprocessing complexity comparison between different versions of the proposed algorithm.

4.6.3 Complexity Comparison Between the Different Versions of the Proposed Algorithm

In this section we compare the complexity of the single-QR version to that of the multiple-QR version. We compare the preprocessing and payload processing complexities separately.

Figure 4.16 shows a comparison of the preprocessing complexity. We clearly see that the complexity difference between the different versions grows with N. Also, as

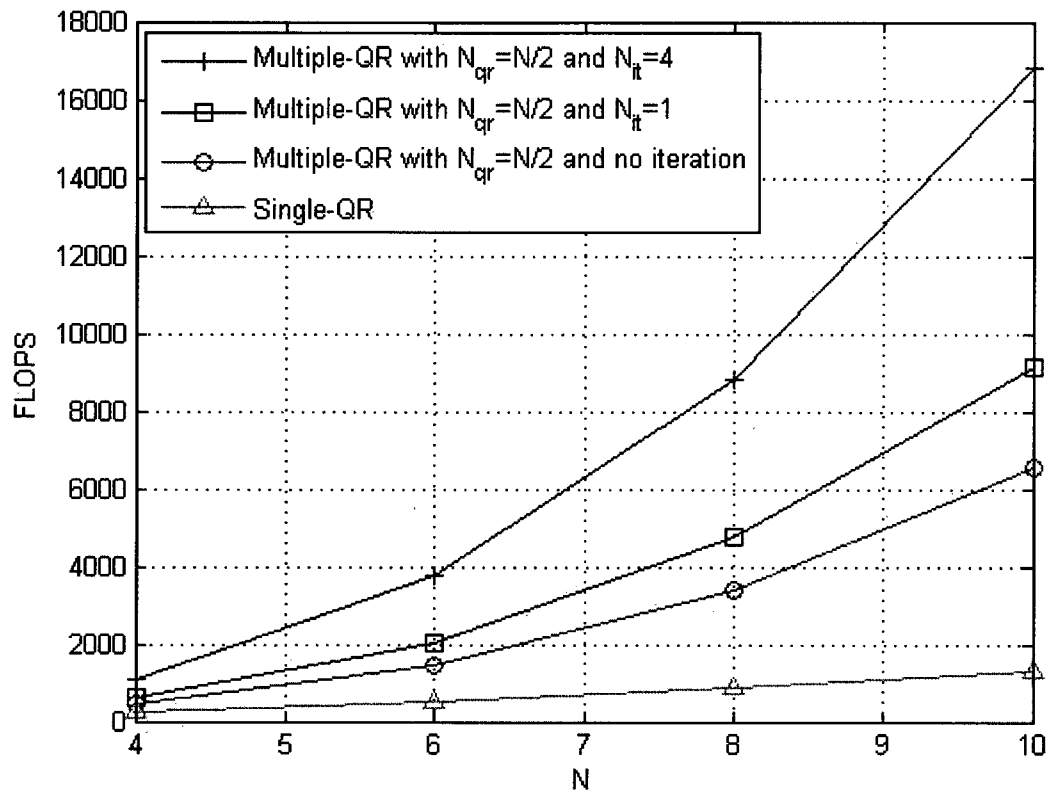

Figure 4.17: Payload complexity comparison between different versions of the proposed algorithm.

expected, the single-QR version with no ordering offers the lowest preprocessing complexity. The single-QR version with SUBOPT2 requires negligible extra complexity. Remember that the SNR gain we obtain with this ordering method is 1.2dB at 1% FER in the 8×8 case. The single-QR version with SUBOPT1 is 2.7 times more complex but provides a gain of 3.3dB. The complexity of the multiple-QR version with $N_{qr} = \frac{N}{2}$ is approximately four times more complex but provides a gain of 4.7dB.

Figure 4.17 compares payload complexity. We again see that the complexity difference between the different versions grows with N. We see that for the 8×8 case

the multiple-QR with $N_{qr} = \frac{N}{2}$ is about four times more complex than the single-QR version. The multiple-QR version with one turbo iteration requires only 40% more complexity. Remember that one turbo iteration provides a gain of 2.5dB at 1% FER. The multiple-QR version with four turbo iterations requires more than twice the complexity compared to the version with no iteration while only providing an incremental gain of of less than 1dB. So we conclude that the multiple-QR version with one turbo iteration provides a good compromise between complexity and performance.

Another important comparison to make is to compare the complexities required by the multiple-QR version with $N_{qr} = 2$ to that of the single-QR with SUBOPT1. We clearly see from Figure 4.8 and Figure 4.12 for the 4×4 case, and Figure 4.9 and Figure 4.13 for the 8×8 case, that these versions have almost similar performance at 1-10% FER. Yet while the preprocessing complexities are comparable as shown in Table 4.6, the payload processing complexity of the latter version is half of that required by the multiple-QR version one (see Table 4.7). We conclude that the use of the multiple-QR version becomes only necessary when better FER performance than the one provided by the single-QR version with SUBOPT1 is required. We have to mention that this conclusion does not take into consideration the parallelism advantage of the two-QR version: despite the fact that it is twice more complex than the single-QR version with SUBTOPT1 the two-QR version may be preferred in practice.

Algorithm	**N = 4**	**N = 6**	**N = 8**
single-QR with SUBOPT1	2029	9282	27395
multiple-QR with $N_{qr} = 2$	2731	9216	21845

Table 4.6: Comparison between the preprocessing complexities of single-QR with SUBOPT1 and multiple-QR with $N_{qr} = 2$

Algorithm	**N = 4**	**N = 6**	**N = 8**
single-QR with SUBOPT1	220	486	848
multiple-QR with $N_{qr} = 2$	444	978	1704

Table 4.7: Comparison between the payload processing complexities of single-QR with SUBOPT1 and multiple-QR with $N_{qr} = 2$

Chapter 5

Comparison of the Proposed Technique to other MIMO Detection Algorithms

In the previous chapter we presented the proposed algorithm but did not mention how it compares to other MIMO detection algorithms. In this chapter we compare the complexity and the performance of the proposed algorithm to other MIMO detection algorithms. Where comparison of the FER performance is always possible, a direct and fair complexity comparison is not always possible as we shall see later.

Since our proposed algorithm resorts to SIC, it makes perfect sense to compare it to other SIC-based algorithms such as VBLAST. We choose here to compare it to the best performing VBLAST, i.e. ordered MMSE-VBLAST and one of its variants. Also, because of the hybrid nature of the algorithm, we compare it to other hybrid algorithms.

5.1 Comparison to MMSE-VBLAST and One of its Variants

We will first compare the FER performance of these algorithms, followed by capacity comparison and then we compare their respective complexities. For the FER and capacity comparison, we will use the original MMSE-VBLAST. A reduced complexity variant of MMSE-VBLAST that was first proposed in [28], commonly known as the Square Root Algorithm for BLAST (SRAB), will only be discussed in the complexity section as the author claims that it has the same performance as that of the original MMSE-VBLAST. Unless otherwise mentioned, the multiple-QR version with two iterations will be used for the FER and complexity comparisons.

5.1.1 FER Comparison

In order to make a fair comparison, we obtain FER results for a wide range of rates. This is rather important as focusing on one rate does not tell us the overall picture as MIMO algorithms may have different multiplexing and diversity gains.

It is important to test the robustness of the algorithms against channel correlation and channel estimation errors. For channel correlation we will use the exponential correlation model discussed in Section 2.3.1. As for channel estimation error we will use the model discussed in Section 3.3. We will show simulation results for the 4×4 and 8×8 cases as representative examples. Table 5.1 lists the modulation format used for each rate.

As we have mentioned before, we have to keep in mind that practical SNR levels rarely exceeds 20dB for both indoor and outdoor applications (see [10] for example). In 4G cellular applications, where a frequency reuse factor of 1 is normally used,

Algorithm	$\mathbf{r_H = 4}$ for 4×4, $\mathbf{r_H = 8}$ for 8×8	$\mathbf{r_H = 8}$ for 4×4, $\mathbf{r_H = 16}$ for 8×8	$\mathbf{r_H = 12}$ for 4×4, $\mathbf{r_H = 24}$ for 8×8
MMSE-VBLAST	BPSK	QPSK	8-QAM
Proposed Algorithm	QPSK	16-QAM	64-QAM

Table 5.1: Modulation format used for MMSE-VBLAST and the proposed algorithm

interference from neighboring cells will have a detrimental effect especially on the User Equipment (UE) that is within the vicinity of the cell boundary. Real-world measurements conducted in [10] show that, in a fully-loaded LTE cell, 90% of UEs enjoy a maximum SINR of 20dB.

Figure 5.1 shows the FER curves for the 4×4 i.i.d. case with bandwidth efficiencies of 4, 8, 12 bit/s/Hz. It is evident that the diversity of MMSE-VBLAST depends on the rate and the operational SNR as its FER curve loses its sharpness with increasing rate and with increasing SNR. In fact, the diversity of MMSE-VBLAST at the high-rate-low-FER region is 1 as each decay requires 10dB increase in SNR. For a bandwidth efficiency of 4 bit/s/Hz our algorithm outperforms MMSE-VBLAST by 3.5dB at FER of 1%. For $r_H = 8$, MMSE-VBLAST outperforms our algorithm at FER of 10% by 0.7dB while our algorithm outperforms MMSE-VBLAST by 1.5dB and 6dB at FER of 1% and 0.1%, respectively. For $r_H = 12$, MMSE-VBLAST outperforms our algorithm at FER of 10% by 1.2dB while our algorithm outperforms MMSE-VBLAST by a significant 4dB and 10.25dB at FER of 1% and 0.1%, respectively. The reason that MMSE-VBLAST performance improves with increasing rate at high FER levels is that because it has a slightly better finite-SNR multiplexing gain compared to our algorithm. MMSE-VBLAST loses this advantage at low FER rates (and consequently

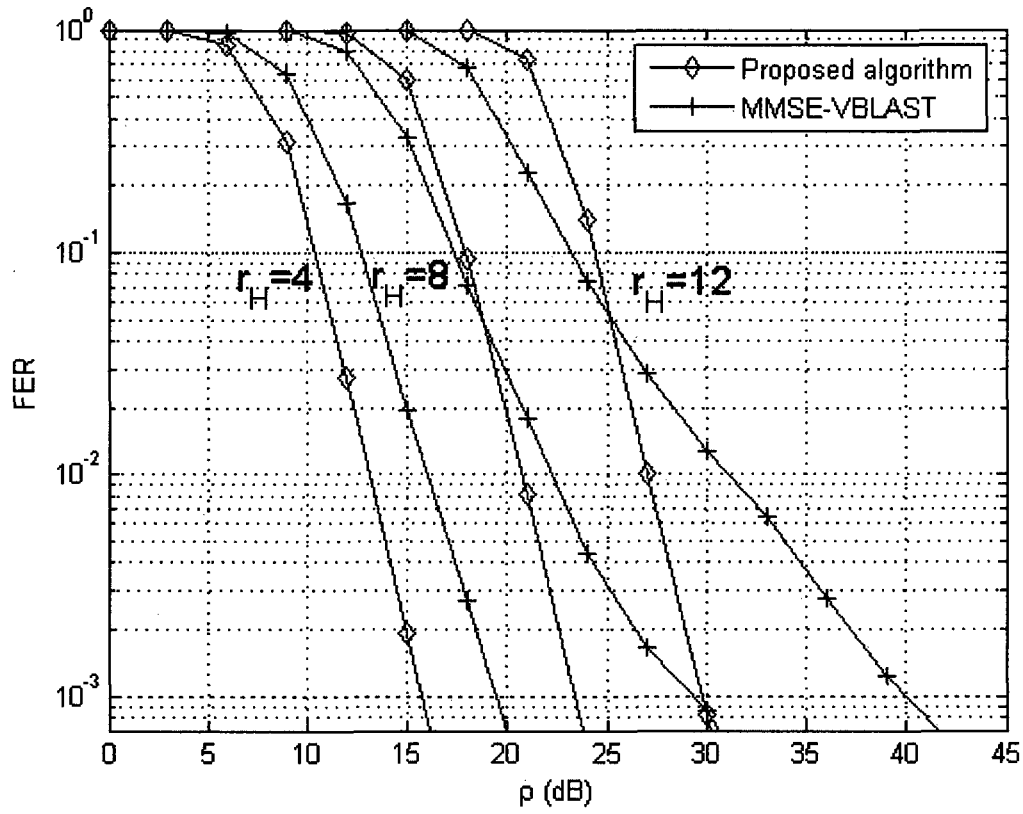

Figure 5.1: FER comparison between the multiple-QR version of our proposed algorithm and that of MMSE-VBLAST for varying bandwidth efficiencies. 4×4 i.i.d. channel

at high SNR where noise is negligible) because it loses the diversity gained from the noise canceling capability.

Figure 5.2 shows the FER for the 4×4 correlated MIMO channel with $\psi_t = 0.7$ and $\psi_r = 0.2$. These values are typical in outdoor downlink applications as correlation at the BS is much higher due to poor scattering. Compared to the i.i.d. case we clearly see that as the rate increases the MMSE-VBLAST curves start to lose their sharpness. Now the performance of our algorithm at 10% FER is similar to MMSE-VBLAST

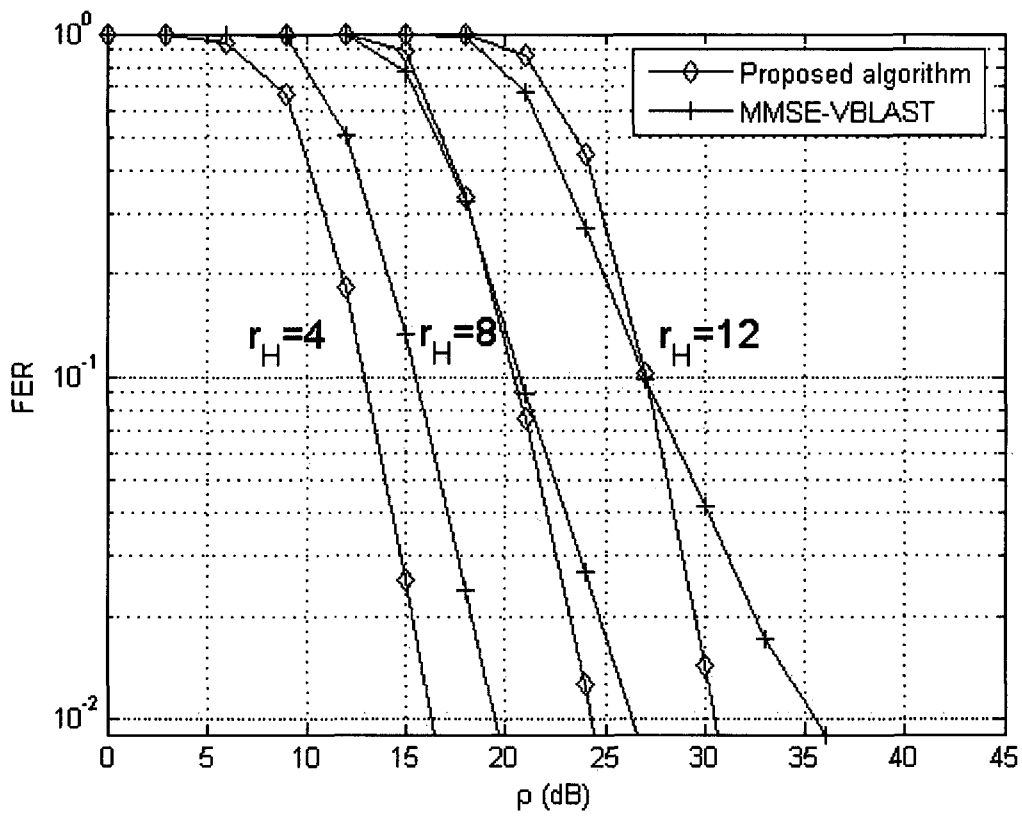

Figure 5.2: FER comparison between the multiple-QR version of our proposed algorithm and that of MMSE-VBLAST for varying bandwidth efficiencies. 4×4 correlated channel. $\psi_t = 0.7$ and $\psi_r = 0.2$

while still outperforming it significantly at low FER levels. This is a clear indication that our algorithm is more robust to channel correlation than MMSE-VBLAST. This is because the performance of the latter depends heavily on channel matrix inversion. The channel matrix becomes ill-conditioned as correlation increases.

Figure 5.3 shows the FER for the 4×4 case with 3% channel estimation error. We see that both algorithms react similarly to channel estimation error at low rates. The story is different at high-rate-low-FER regions. For example, for $r_H = 12$ and

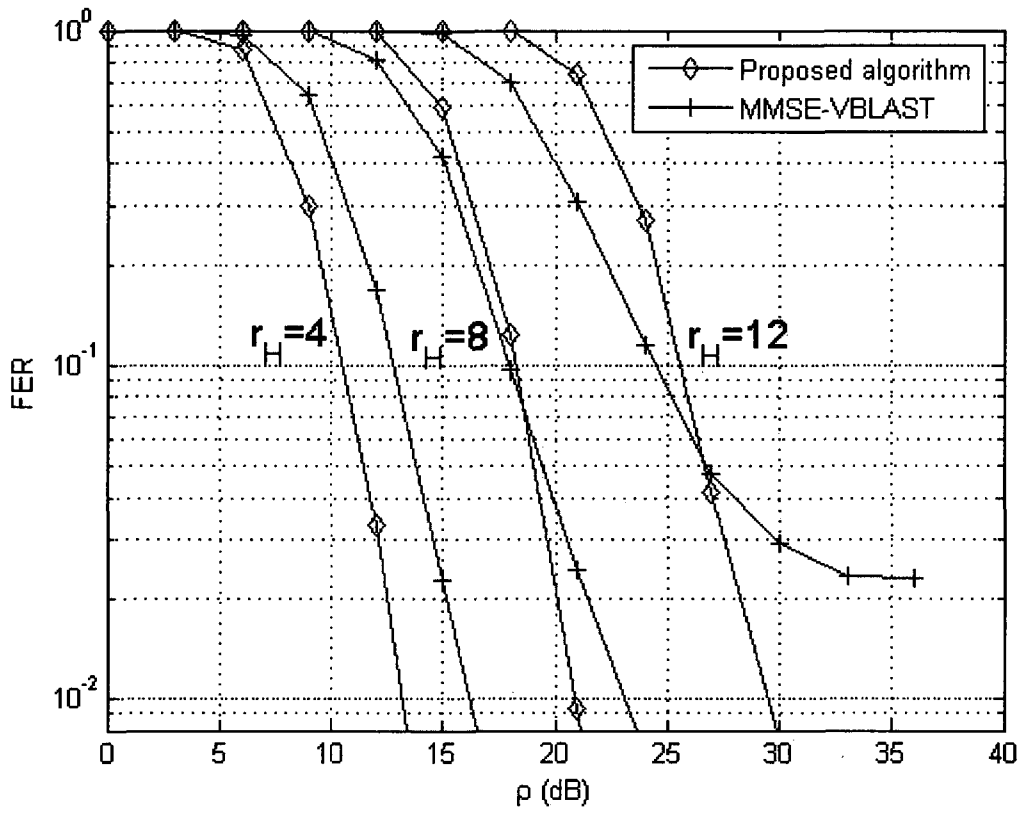

Figure 5.3: FER comparison between the multiple-QR version of our proposed algorithm and that of MMSE-VBLAST for varying bandwidth efficiencies. 4×4 channel with 3% channel estimation error.

1% FER, our algorithm suffers 2dB penalty compared to the ideal case while MMSE-VBLAST reaches an error floor at 2.5% FER.

Figure 5.4 shows the FER curves for the 8×8 i.i.d. case with bandwidth efficiencies of 8,16 and 24 bit/s/Hz. We now see that MMSE-VBLAST outperforms our algorithm for $r_H = 16$ and $r_H = 24$ for all FER. Our algorithm still outperforms MMSE-VBLAST for $r_H = 8$. This is due to the fact that MMSE-VBLAST's diversity at finite SNR increases with increasing N.

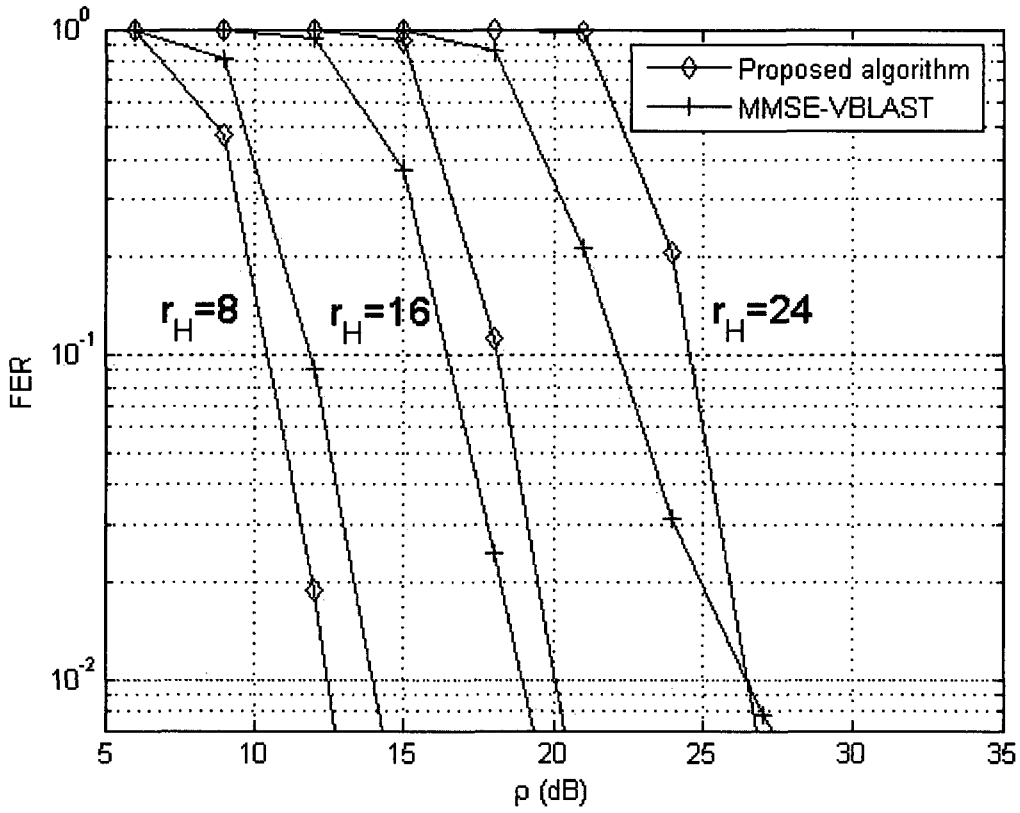

Figure 5.4: FER comparison between the multiple-QR version of our proposed algorithm and that of MMSE-VBLAST for varying bandwidth efficiencies. 8×8 i.i.d. channel

In the presence of correlation, as shown in Figure 5.5, the performance gap between both algorithms shrinks which again suggests that our algorithm is more robust to channel correlation especially at high-rate-low-FER regions. In fact, now our algorithm outperforms MMSE-VBLAST by 1.5dB at 1% FER for $r_H = 24$.

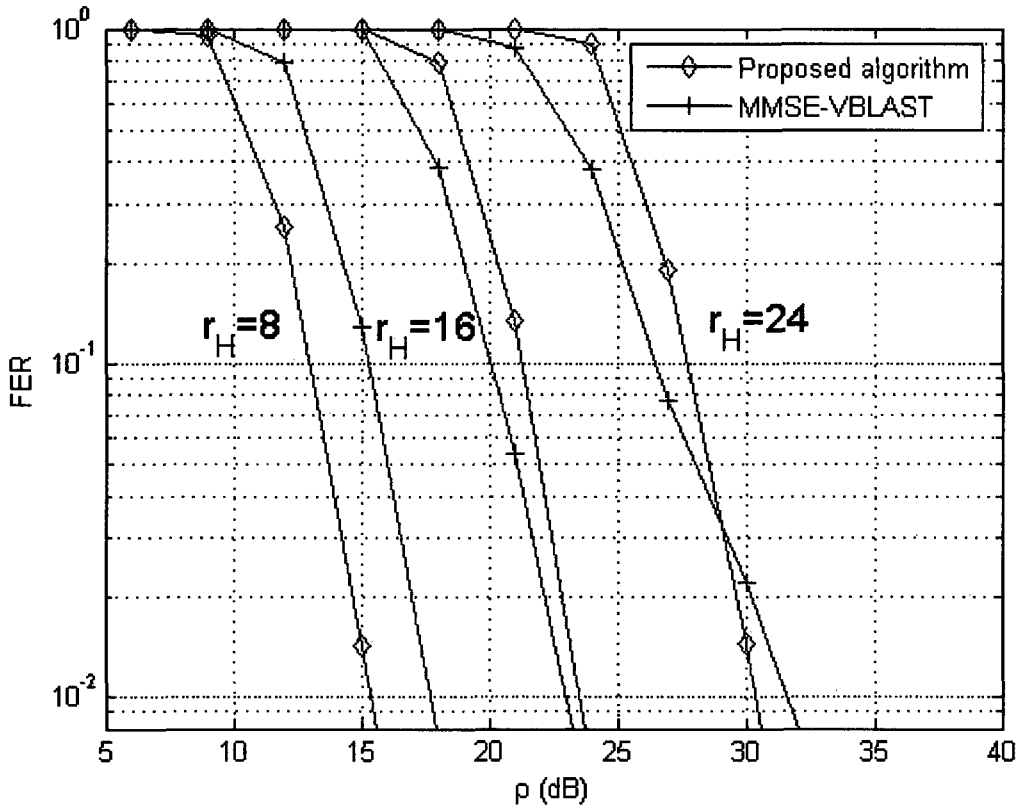

Figure 5.5: FER comparison between the multiple-QR version of our proposed algorithm and that of MMSE-VBLAST for varying bandwidth efficiencies. 8×8 correlated channel. $\psi_t = 0.7$ and $\psi_r = 0.2$

The effect of channel estimation error on the 8×8 case is shown in Figure 5.6. We see that a 3% channel estimation error has a negligible effect on both algorithms at low rates. The negative effect of channel estimation error on MMSE-VBLAST becomes evident at 1% FER and $r_H = 24$ where we start to see an error floor while our algorithm almost maintains its diversity as demonstrated by the unaffected sharpness of its FER curve.

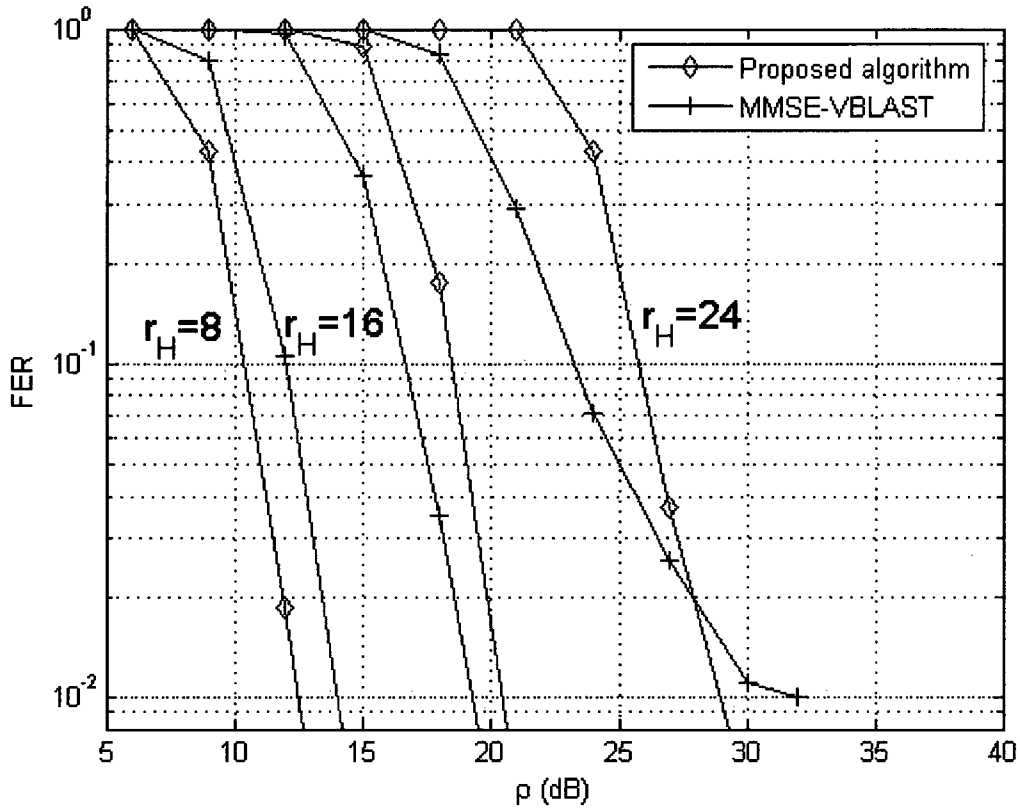

Figure 5.6: FER comparison between the multiple-QR version of our proposed algorithm and that of MMSE-VBLAST for varying bandwidth efficiencies. 8×8 channel with 3% channel estimation error.

From the above simulation results, it seems that, when compared to MMSE-VBLAST, our algorithm performs as good as or better than MMSE-VBLAST. This is especially true when we take into account practical channel imperfections such as correlation and channel estimation errors. Let us not forget that throughout we assumed that MMSE-VBLAST has perfect knowledge of noise variance. An error in estimating this parameter, which is unavoidable in practice, will most likely lead to further degradation of its FER performance.

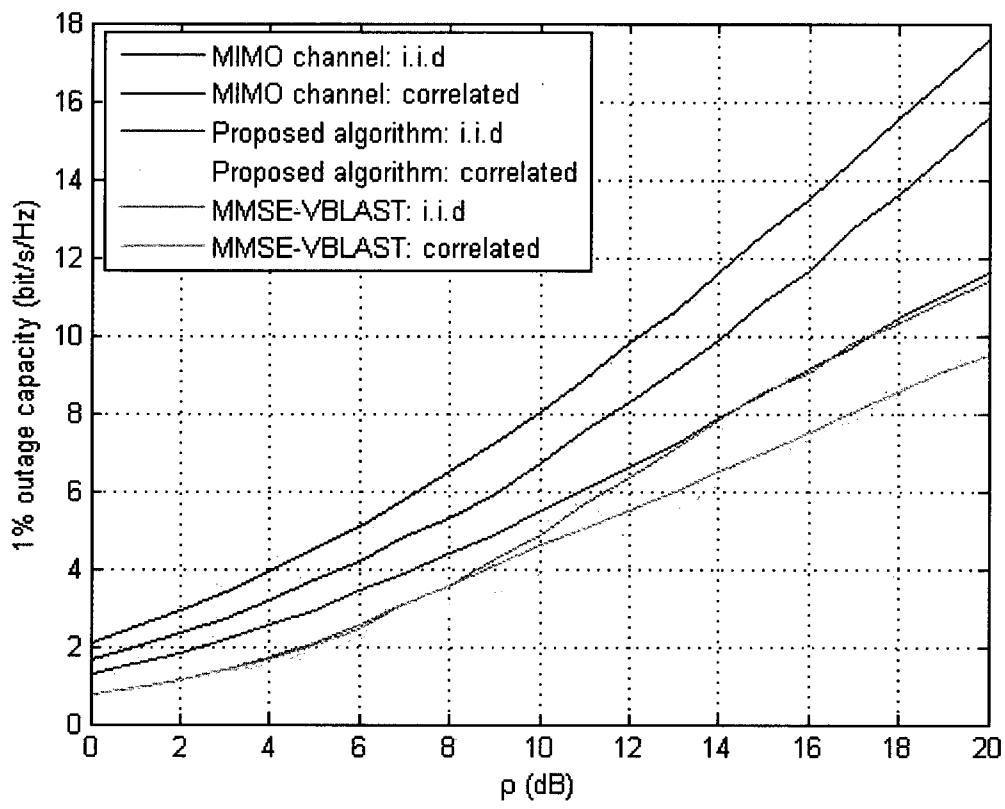

Figure 5.7: 1% outage Capacity comparison between the proposed algorithm and MMSE-VBLAST for 4×4 MIMO. Exponential correlation model was used with $\psi_t = 0.7$ and $\psi_r = 0.2$

5.1.2 Capacity Comparison

Capacity comparison allows us to look at the performance of the various MIMO algorithms once capacity-achieving codes are used as outer codes.

We know that, at infinite SNR, algorithms that maximize multiplexing gain such as MMSE-VBLAST only require 3dB increase in SNR for every N bit/s/Hz increase in capacity. The story at practical SNR levels is completely different as we shall see

next.

Figure 5.7 and Figure 5.8 compare the 1% outage capacities of MMSE-VBLAST to that of our proposed algorithm for the 4×4 and 8×8 cases, respectively. Capacities for i.i.d. and correlated channels are plotted. The following interesting observations can be made.

First, these figures confirm that MMSE-VBLAST does not attain the full MIMO capacity when the channel is quasi-static. Also, while the infinite-SNR multiplexing gain of MMSE-VBLAST is N, the finite-SNR multiplexing gain, as manifested by the slope of the capacity curve, is smaller. For instance, a 3dB increase in SNR from 14dB to 17dB gives us approximately $\frac{N}{2}$ increase in capacity for both the 4×4 and 8×8 cases. The multiplexing gain of our algorithm almost reaches its infinite-SNR value of $\frac{N}{2}$ at that same range. This proves what we stated in the previous chapter that comparing the various space-time algorithms using the idealistic infinite-SNR multiplexing-diversity tradeoff curves may be somehow misleading.

Second, in the 4×4 i.i.d. case our algorithm outperforms MMSE-VBLAST for $0 < \rho < 14$ while offering the same capacity for 14dB $< \rho <$ 20dB. In the 8×8 i.i.d. case MMSE-VBLAST offers better capacity for $\rho >$ 5dB. This is explained by the fact that the finite-SNR diversity of MMSE-VBLAST improves because of noise canceling capability.

Third, when correlation is taken into account MMSE-VBLAST outperforms our algorithm slightly at medium-SNR range ($4 < \rho < 17$) in the 4×4 channel case. In the 8×8 correlated case, MMSE-VBLAST outperforms our algorithm by approximately 4 bit/s/Hz across the entire SNR range. Again this is attributed to the fact that MMSE-VBLAST's diversity at finite SNR improves with increasing number of antennas.

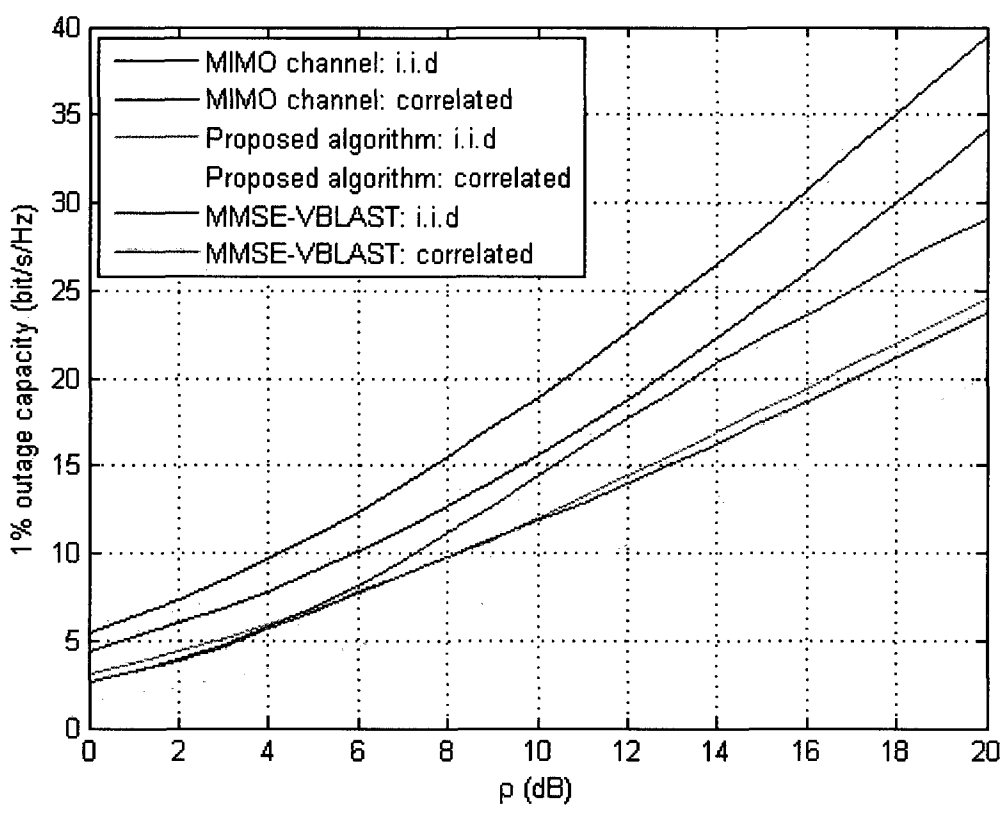

Figure 5.8: 1% outage Capacity comparison between the proposed algorithm and MMSE-VBLAST for 8×8 MIMO for i.i.d. and correlated channel. Exponential correlation model was used with $\psi_t = 0.7$ and $\psi_r = 0.2$

Note however that in order to reach this capacity the use of more powerful channel codes is required compared to our algorithm. This is because the performance of MMSE-VBLAST is much worse at 1% FER. Note also that a more complete capacity comparison will look at the effect of channel estimation error. From the FER curves we can expect that channel estimation error will have a detrimental effect on MMSE-VBLAST's channel capacity especially at high rates where the required SNR is high (because of the FER error floor that appears at high SNR).

Finally, since practical SNR rarely exceeds 20dB in many applications of interest, the capacity curves corresponding to the algorithms (versus the actual MIMO channel) tell us that we can employ capacity-achieving codes to achieve no more than 3-3.5 bit/s/Hz/subchannel efficiency at 20dB. This is in contrast to the theoretical MIMO capacity which offers 3.8-4.8 bit/s/Hz/subchannel at the same SNR.

5.1.3 Complexity Comparison

When comparing complexities in quasi-static scenarios, it is important to differentiate between preprocessing and payload complexities. One or the other may dominate the overall complexity depending on the frame size and the channel estimation frequency. In fixed or low mobility applications, where the channel can be considered invariant over a long period of time, the payload complexity will dominate the total complexity. In this case, complexity reduction should focus on this phase of processing. On the other hand, when the channel changes frequently, such as in high-mobility applications that must be supported by 4G systems, channel estimation is dominated by the preprocessing complexity.

As we shall see later the preprocessing complexity of MMSE-VBLAST is $O(N^4)$ and its payload processing complexity is $O(N^2)$. A lower complexity version of MMSE-VBLAST, called the square-root algorithm for BLAST (SRAB), was proposed in [28]. This algorithm reduces the preprocessing complexity to $O(N^3)$ while maintaining the same payload complexity of $O(N^2)$. The reduction of complexity in SRAB is achieved through the use of unitary transformations[1] to invert the channel matrix. It is unclear whether the proposed modifications introduce performance loss as the author is silent on this issue.

[1]The author resorts to the implicit use of the QR decomposition

Algorithm	**Multiplications**	**Additions**	**Flops**
MMSE-VBLAST	$\frac{27}{4}N^4$	$\frac{27}{4}N^4$	$54N^4$
SRAB	$\frac{29}{3}N^3$	$\frac{29}{3}N^3$	$\frac{232}{3}N^3$

Table 5.2: Summary of the preprocessing complexity of MMSE-VBLAST and SRAB.

Table 5.2 summarizes the preprocessing complexity of MMSE-VBLAST and SRAB. The preprocessing complexity of MMSE-VBLAST is $\frac{27}{4}N^4$ multiplications and $\frac{27}{4}N^4$ additions. SRAB has a complexity of $\frac{29}{3}N^3$ multiplications and $\frac{29}{3}N^3$ additions. The payload processing for both is equal to $2N^2$ multiplications and $2N^2$ additions. These expressions for preprocessing and payload processing complexities were taken from [28]. The author used flops as the unit for the above expressions. But his definition of a flop is one multiplication and one addition which justifies the expressions listed above.

A fair complexity comparison can only be made within the context of a real-world radio air interface. We here choose the 3GPP-LTE air interface [51] as a good representative example of a 4G system which uses OFDMA for multiple access on the downlink (DL). Note that because OFDM is used, each subcarrier undergoes flat fading. We will exclusively consider DL as reducing complexity on the terminal side is more crucial than reducing complexity at the basestation.

In LTE, a frame which is 10ms long consists of 10 subframes of 1ms duration. Each subframe consists of two 0.5ms time slots. A time slot contains 6 or 7 OFDM symbols depending on whether normal or extended cyclic prefix is used. Each OFDM symbol contains from 128 subcarriers for 1.4MHz bandwidth to 2048 subcarriers for

20MHz channel bandwidth. Multiple access is achieved by assigning each user a single or multiple blocks of Resource Elements (RE). Each RE consists of data and pilot bits modulated using QPSK, 16QAM or 64QAM depending on current channel conditions. The block(s) assigned to each user, called Resource Block(s) (RB), each contains 12 subcarriers and spans a single time slot. Each RB contains 4 pilot REs that are used to estimate the channel. Channel estimation (and interpolation) is done twice during a time slot. When multiple antennas are used, the locations of these pilot REs on the time-frequency grid is different for each antenna so that orthogonality can be easily achieved by inserting zeros at these same locations within the RB transmitted from the other antennas.

With this information, it is now straightforward to calculate the complexity required to process the entire RB. Since the channel has to be estimated twice and that each RB contains 12 subcarriers then we need 24 times of preprocessing complexity calculations. Depending on the array size and the number of OFDM symbols per slot (i.e. 6 or 7) the payload processing requirements will vary as the number of payload REs will be different. The total complexity C_{tot} required per RB can be calculated as follows:

$$C_{tot} = 24C_{pp} + (12N_{symb} - 4N)C_{pl} \tag{5.1.1}$$

where C_{pp} is the preprocessing complexity, C_{pl} payload complexity and N_{symb} is the number of OFDM symbols per slot.

Table 5.3 and Table 5.4 lists the number of FLOPS required by MMSE-VBLAST, SRAB and our proposed algorithm for $N = 4, 6, 8$ and $N_{symb} = 7, 6$, respectively. In both cases we notice that for $N = 4$, our algorithm requires about $1/3$ of the

Algorithm	**N = 4**	**N = 6**	**N = 8**
MMSE-VBLAST	349184	1714176	5361664
SRAB	136192	435456	1003520
Proposed Algorithm	125104	504216	1389280

Table 5.3: Complexity comparison between MMSE-VBLAST, SRAB and the proposed algorithm (with $N_{qr} = \frac{N}{2}$ and two iterations) to process one RB when $N_{symb} = 7$

Algorithm	**N = 4**	**N = 6**	**N = 8**
MMSE-VBLAST	346112	1707264	5349376
SRAB	133120	428544	991232
Proposed Algorithm	114592	469728	1310656

Table 5.4: Complexity comparison between MMSE-VBLAST, SRAB and the proposed algorithm (with $N_{qr} = \frac{N}{2}$ and two iterations) to process one RB when $N_{symb} = 6$

complexity required by MMSE-VBLAST. Compared to SRAB our algorithm requires 15% less complexity. The complexity gap between our algorithm and SRAB shrinks as N grows. For $N = 6$ both algorithms possess comparable complexities while for $N = 8$ SRAB's complexity is about 30% less than that required by our algorithm. It is evident that the complexity required by MMSE-VBLAST becomes prohibitive as N grows beyond four.

A more complete complexity analysis would also include the SNR and FER targets. Averaged over a practical SNR range of let's say 10-25dB and FER target of 10%, the complexity, and by consequence power consumption, of our algorithm will

certainly be much less. Taking the 8×8 case as an example one can resort to using lower complexity versions of our algorithm for SNR greater than 18dB to meet this FER target. In this case, only a subset of the QR-Alamouti-SIC processors will be active which also results in power savings. MMSE-VBLAST lacks this advantage as the algorithm has the same complexity regardless of the SNR and FER targets[2].

It should also be mentioned that the complexity comparison presented above ignores the implementation platform. If the above algorithms were to be implemented in hardware (such as ASIC or FPGA), the proposed algorithm has an advantage compared to MMSE-VBLAST and SRAB. This is because parallelism is inherent in the algorithm where the computations of all quantities (such the various $\mathbf{Q}$ and $\mathbf{R}$ matrices, the modified RX vectors $\tilde{\mathbf{y}}$, etc) can be done in parallel[3]. This enables high-speed symbol processing at a lower power consumption, a must-have feature for 4G battery-powered terminals.

5.2 Comparison to Hybrid Algorithms

In this section we give a brief description of the most well-known hybrid architectures that were proposed within the last decade. Exact complexity analysis is not possible as the authors do not focus much on this issue but rather focus on the improvement obtained. Nevertheless, where possible, we mention the order of complexity required by each.

The first architecture that combined a space-time code with SM-type of algorithm was proposed in [49]. This architecture is similar to ours except that a 2×2 STTC

[2]Slight preprocessing complexity reduction is possible by ignoring the ordering procedure

[3]This also makes our algorithm well suited for implementation using modern multi-core DSP processors

code is used instead of the Alamouti STBC. The other main difference is that group interference nulling (GIN) is achieved through matrix inversion. Decoding for each group is achieved with a Viterbi decoder. Realizing that the diversity is worst for the first detected stage, the author proposes the use of unequal power allocation at the transmitter so that power decreases proportionally with increasing decoding order. In the sequel we will refer to this algorithm as Hybrid Variant 1 or HV1 for short.

While STTC provides coding gain in addition to diversity gain when compared to Alamouti code, their complexity and delay requirements make them unattractive for practical implementations. Also, complexity of coding/decoding increases exponentially with increasing rates therefore restricting its use to low-rate applications. This exponential increase in complexity is a direct consequence of the need to increase the number of trellis states. Another disadvantage of HV1 is that the use of matrix inversion amplifies noise when nulling the group interference therefore limiting the overall performance.

Complete complexity comparison with HV1 is not possible. But because it uses matrix inversion we can say that the lower bound on preprocessing complexity is $O(N^3)$. The complexity during payload processing is mainly dominated by ML decoding. For this latter, we need to use $\frac{N}{2}$ Viterbi decoders, i.e., one for each decoding stage. Compared to our simple linear processing at the receiver one can safely assume that the complexity of this algorithm is a lot higher than the complexity of our proposed algorithm.

In [47], a similar architecture to HV1, called Generalized Layered Space-Time Code (GLST), is proposed where 2×2 STTC is also used. There are two flavors to the proposed solution. One is called Horizontal GLST (HGLST) where no interleaver

is used at the transmitter and another called Diagonal GLST (DGLST) where an interleaver is used at the transmitter. The purpose of this spatial interleaver is to increase diversity of the TX groups. The authors also propose various solutions to improve on the performance of HV1. One solution consists of optimizing the power allocation procedure. This solution does not require additional complexity as it can be done offline. Another solution suggests using equal power allocation with reordering the decoding procedure based on channel energy. In this case, decoding starts with the group that has the strongest receive SNR. This further improvement requires additional preprocessing complexity on the order of $O(N^2)$.

When an interleaver with depth L_{int} is used at the TX (i.e. with DGLST), the computations of L_{int} nulling matrices at the RX become necessary (the authors use $L_{int} = 3$ in the simulations regardless of N). To improve performance even further, the authors propose the use of iterative detection and the use of frame-long interleavers/deinterleavers at the receiver.

While promising, the above solution has the same shortcomings of HV1 when it comes to the use of matrix inversion. In addition, its complexity is extremely high especially with DGLST where the use of L_{int} nulling matrices and $\frac{L_{int}N}{2}$ Viterbi decoders are needed. On top of the huge complexity required, their most performing solution (with iterative detection) also introduces huge system delay hence limiting its applications to frames that carry non real-time data (such as emails). Hereafter, we will call this solution Hybrid Variant 2 or HV2 for short.

The first hybrid architecture that departed from the use of STTC, due to their complexity and delay requirements, was proposed in [41]. The author replaces STTC with the Alamouti STBC. Optimum TX power allocation and rate allocation per

layer is derived based on minimizing the FER at $\rho = 24$dB and frame lengths of 2 and 4. GIN and GSIC are performed at the receiver to recover the TX symbols. It is important to note that the use of $\frac{N}{2}$ Alamouti codes is only a special case with this architecture.

The above solution, while flexible in providing incremental rates, with VBLAST and full STC being two extreme versions, it still uses matrix inversion to achieve GIN, therefore limiting the overall performance gain that can be obtained. Another simplifying and limiting assumption on the part of the authors is the use of frame sizes of 2 and 4. The authors resorted to the use of these short frame sizes in order to simplify the optimization procedure. This makes the algorithm applicability restricted to few applications where the frame size is very small. Hereafter, we will refer to this solution as Hybrid Variant 3 or HV3 for short.

In [14], an architecture called Quasi-Orthogonal Group Space-Time (QoGST) is proposed where the TX symbols are grouped together but left uncoded. Instead, inter-group space-time coding is applied. In contrast to the above algorithms, the receiver performs group space-time decoding before group detection. As usual, group detection is achieved through matrix inversion. The authors show that QoGST has a better diversity-multiplexing tradeoff than HV1.

Because the above solution resorts to group detection after performing space-time decoding, matrix inversion has to be performed on the equivalent expanded effective channel matrix. If we assume the use of the Alamouti-type TX coding which has a code length of 2, the preprocessing complexity of group detection is $O((2N)^3) = O(8N^3)$. This is equivalent to performing 8 times more matrix inversions.

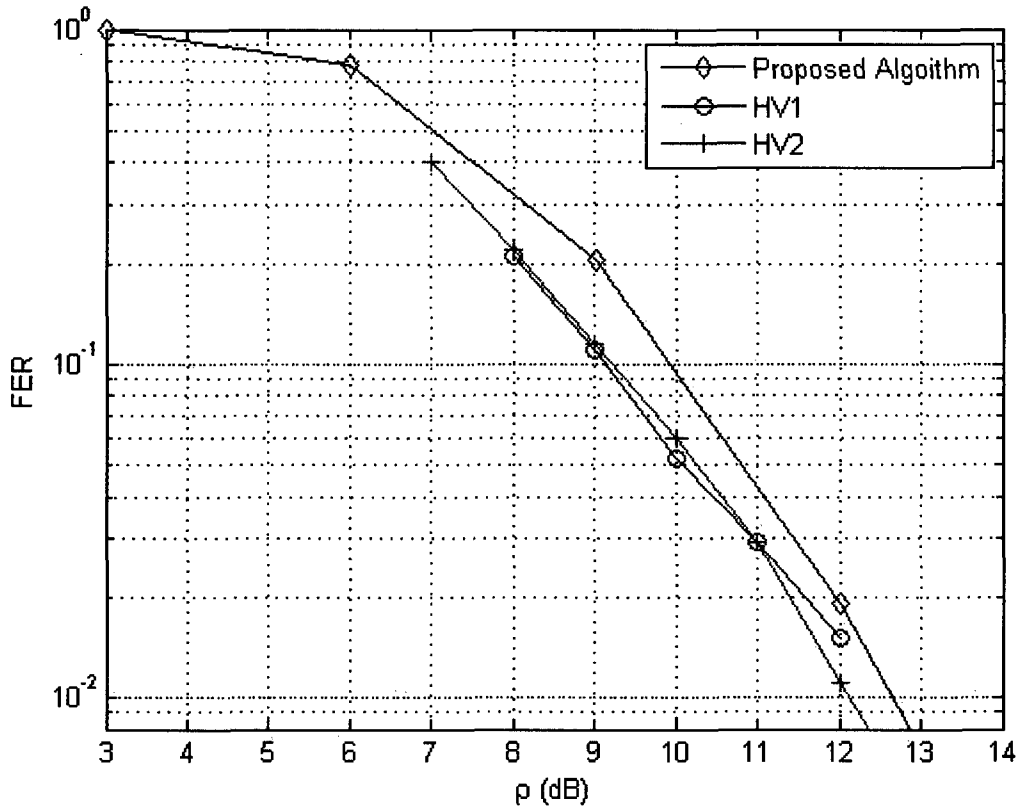

Figure 5.9: FER comparison between the multiple-QR version of our proposed algorithm and that of HV1 and HV2 for the same bandwidth efficiency of 4 bit/s/Hz

Said differently, QoGST has roughly the same complexity as the original MMSE-VBLAST for $N \leq 8$. Hereafter we will refer to QoGST as Hybrid Variant 4 or HV4 for short.

5.2.1 FER Comparison

In this section we borrow the curves reported by the authors of HV1, and HV2. It is not possible to compare our curves to HV3 and HV4. because the author of HV3

presented figures that are so small that accurate comparison is not possible. As for HV4 no FER curves were presented for equal number of TX and RX antennas and as such fair comparison is not possible.

The comparison with HV1 and HV2 is straightforward as the authors used frame sizes of 130-symbol long. Figure 5.9 shows the results for the 4×4 i.i.d. case while Figure 5.10 shows the results for the 8×8 i.i.d. case for bandwidth efficiencies of 4 bit/s/Hz and 8 bit/s/Hz, respectively. The multiple-QR version of our algorithm with $N_{qr} = \frac{N}{2}$ and $N_{it} = 4$ was used.

First we note that the performance of HV1 and HV2 are almost similar in both cases. Note that HV2 uses a 16-state STC while HV1 uses 32-state STC. So HV2 achieves the same performance as HV1 at a much lower complexity.

In the 4×4 case both HV1 and HV2 outperform our proposed algorithm by 0.8dB at FER of 10%. This difference shrinks to 0.5dB at 1% FER. This is an indication that our proposed algorithm has a better diversity performance. This is clearly seen in the 8×8 case where our proposed algorithm outperforms HV2 (or HV1) by 0.7dB and 1.6dB at 10% and 1% FER, respectively.

The above results can be explained as follows. Because 2×2 STTCs provide constant coding gain in addition to diversity gain, when compared to the Alamouti STBC, they easily outperform our algorithm for small N. This coding gain remains constant while the diversity gain in our algorithm increases with increasing N. This explains why our algorithm performs better when $N = 8$ versus when $N = 4$.

It is remarkable that our algorithm achieves almost similar or better performance than HV1 (or HV2) despite the fact that HV1 (or HV2) uses STTCs as component codes. This is due to the fact that HV1 (or HV2) resorts to matrix inversion for

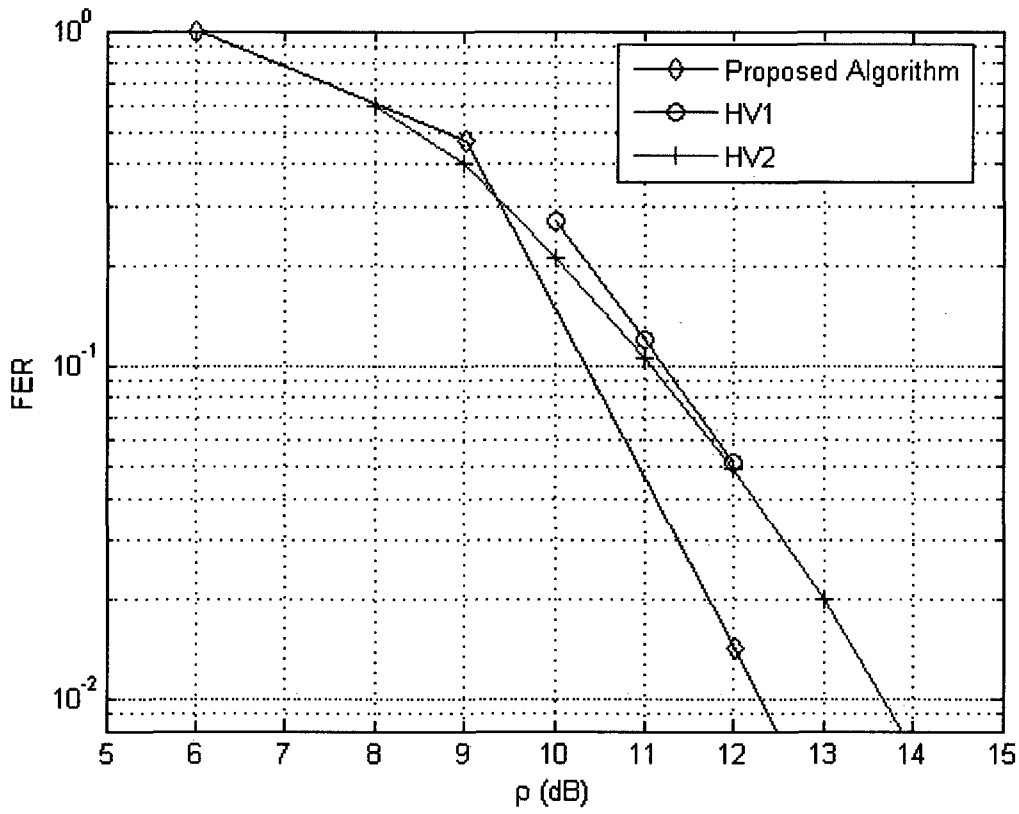

Figure 5.10: FER comparison between the multiple-QR version of our proposed algorithm and that of HV1 and HV2 for the same bandwidth efficiency of 8 bit/s/Hz

GIN which has the effect of amplifying the noise, therefor limiting the coding gain normally obtained with STTCs.

5.3 Application of the Proposed Algorithm to Multi-User MIMO

So far in our study we considered MIMO as a solution to increase data rates belonging to the same user. We have mentioned little about the feasibility of implementing

MIMO in practical systems.

While the use of multiple RF chains may not pose a problem on the BS side, where we can afford complexity and power is available in abundance, this poses a major challenge for hand-held terminals which are power-limited and generally have small form-factor which make the integration of multiple antennas on these devices extremely challenging. This mostly explains why most 4G hand-held terminals are currently equipped with no more than 2 antennas. This practical limitation may explain the huge interest in applying single-user MIMO (SU-MIMO) techniques to multi-user MIMO (MU-MIMO), where the addition of multiple antennas can be used to increase cell capacity instead. A good introduction to MU-MIMO can be found in [45].

Figure 5.11 illustrates a possible MU-MIMO setup where the BS is equipped with 8 antennas. The BS communicates with 4 user equipments (UE) each having 2 antennas. The various UE's communicate with the BS simultaneously, in time and frequency, hence increasing the cell capacity by four.

A major challenge in MU-MIMO is user scheduling, synchronization and power control. This problem is no different than the ones faced by CDMA systems and hence the vast research developed for CDMA systems over the last two decades can be applied to MU-MIMO. One other advantage that MU-MIMO has over SU-MIMO is that the correlation between the different UE's is close to zero due to the physical separation between the various UE's.

5.3.1 Application of our Algorithm in the Uplink

In the uplink (UL) direction our algorithm can be used with no modification (of course after proper synchronization and possibly slow power control). In the context

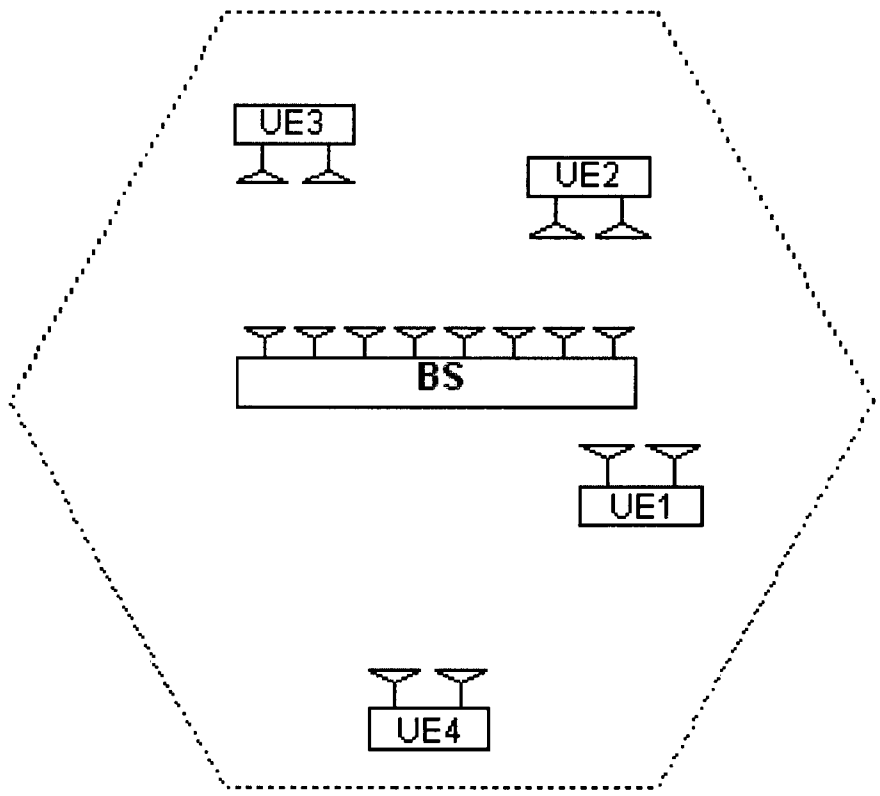

Figure 5.11: Illustration of MU-MIMO where BS is equipped with 8 antennas while each UE has 2 antennas.

of MU-MIMO our proposed algorithm has three major advantages. They are:

1. **Parallelism**: The multiple-QR version of our proposed algorithm can be implemented in parallel. Remember that each QR decomposition operates on a different column-permuted channel matrix. The associated processing with each is also independent. This is rather a crucial feature as the BS can only cope with the ever-increasing high data-rate requirements of modern wireless systems through the use of multi-core DSP processors (or multiple cores implemented in ASIC).

2. **Parametrization, near-far-problem and cell edge**: The different UE's have generally different requirements depending on the the type of traffic being transmitted (voice, video, digital data) and/or depending on how far they are from the BS. For example, if the scheduling algorithm decides (on a packet-by-packet basis) to group voice packets belonging to different UE's together, a strong possibility because of stringent delay requirements, then only one QR-Alamouti-SIC processor will be activated and the single-QR version may be sufficient. This in turn leads to reduced power consumption. Because diversity increases after each stage, the UE closest to the basestation can be decoded in the first stage followed by the next closest and so on until we decode the farthest (may be located at cell edge) in the last decoding stage. Another possibility for the scheduling algorithm is to prioritize decoding based on the type of traffic with voice being decoded first followed by video and then digital data.

3. **Flexibility**: Another advantage of the proposed algorithm is that the various UE's need not all have the same number of antennas. This flexibility that is inherent in the proposed algorithm essentially allows a tradeoff between per-user throughput and cell capacity. As an example of sacrificing cell capacity for greater per-user throughput, UE1 can have 4 antennas, UE2 can have 2 and UE3 can have 2, decreasing the cell capacity gain from 4 to 3 but now UE1's throughput is increased by a factor of 2. Other combinations are possible even as little as one antenna per user. This is possible as the structure of matrix $\mathbf{R}$ allows us to use simple backward substitution to solve for symbols transmitted using a single antenna, albeit at a cost of reduced diversity for these symbols. It should be mentioned that the scheduler in this case must leave the decoding

of these symbols to later stages where the diversity is theoretically highest. For example, UE1, UE2 and UE3 can have two antennas each and UE4 and UE5 (whose symbols are decoded during the last two stages) can have one antenna each therefore increasing the cell capacity from 4 to 5.

5.3.2 Application of our Algorithm in the Downlink

In the downlink our proposed algorithm can not be applied directly as multi-stream interference can not be resolved by the smaller number of UE antennas. But if we assume the use of a TDD link and the channel coherence time is much larger than the UL/DL time slot so that the channel can be considered constant over two consecutive UL and DL time slots (plus the round-trip time), then the channel coefficients estimated at the BS can be used to form beams in the direction of each UE. In other words, we can pre-cancel the multi-stream interference seen by each UE.

It is also important to mention that in order to achieve the above a lower triangular version of the QR decomposition, here called the LD decomposition, should be used [45]. The matrix $\mathbf{L}$ is lower triangular and the matrix $\mathbf{D}$ is orthogonal. We can easily verify that $\mathbf{L} = \mathbf{R}^{+}$ and $\mathbf{D} = \mathbf{Q}^{+}$, where matrices $\mathbf{Q}$ and $\mathbf{R}$ in this case are obtained by performing QR decomposition on $\mathbf{H}^{+}$.

Note that all intensive baseband processing (channel estimation, LD decomposition, etc) is performed at the basestation and all UE's only perform simple Alamouti payload processing. Also, because of interference pre-cancellation, each UE enjoys the full diversity of $4q$ ($q = 1, 2, ...\frac{N}{2}$). Again the role of the scheduling algorithm is crucial. For example, the first coding stage can be used to transmit data to the UE closest to the BS, followed by the next closest and so on. Alternatively, the order of coding and decoding can be based on the type of transmitted traffic.

Chapter 6

Conclusion

In this thesis we presented a parallel $N \times N$ MIMO architecture and associated algorithms that combined the benefits offered by diversity and multiplexing. The proposed algorithm groups the transmitted symbols in groups of two and codes each group using the simple Alamouti STBC. At the receiver side, one or more QR decompositions along with successive interference cancellation and iterative detection are used to decode the received symbols. Compared to a pure spatial multiplexing architecture (such as V-BLAST), the proposed one has the effect of increasing the diversity order of each processed symbol by at least a factor of four while the overall multiplexing gain is only decreased by a factor of two.

In Chapters 4 and 5, we presented a detailed description of our algorithm and compared its outage capacity and FER performance to that of the best performing V-BLAST algorithm, i.e. ordered MMSE-VBLAST. We showed through computer simulations that, at practical SNR levels below 20dB, the outage capacity of the proposed architecture is very similar to that offered by MMSE-VBLAST especially when spatial correlation is taken into account. This is despite the fact that our proposed algorithm has an infinite-SNR maximum multiplexing gain of only $\frac{N}{2}$ compared to

N for MMSE-VBLAST. This is because, at finite-SNR levels, the multiplexing gain offered by our algorithm is very similar to that offered by MMSE-VBLAST. In addition, the high diversity gain of $2N$ offered by our algorithm contributes to capacity enhancement.

Through FER comparison we showed that our algorithm has a huge SNR advantage compared to MMSE-VBLAST especially at FER $\leq 1\%$. We also showed that in the presence of moderate correlation or channel estimation errors, our algorithm is more robust than MMSE-VBLAST. It should be emphasized that we assumed throughout our study that MMSE-VBLAST has a perfect noise estimate which can never be the case in practice. As a result further FER degradation with MMSE-VBLAST is expected in practice especially at low-to-moderate SNR levels where the noise power dominates the performance.

We compared the complexity of our algorithm to MMSE-VBLAST and SRAB, a lower complexity variant of MMSE-VBLAST. We did this using the 3GPP-LTE radio interface as a real-world example. We showed that for $N \leq 6$ our algorithm requires less or comparable complexity than SRAB. For $N = 8$, our algorithm is only 30% more complex than SRAB. We also mentioned that the parallelism inherent in the algorithm outweighs this increase in complexity and makes our algorithm a strong candidate for high-speed low-power communications required in 4G and beyond-4G systems. Another advantage of the proposed algorithm, from an implementation point of view, is its use of the well studied QR decomposition and the Alamouti OSTBC; extensive knowledge and intellectual property exist regarding these two algorithms.

We also compared our algorithm to other well-known hybrid algorithms. We showed through two examples that the FER performance of our algorithm is slightly

worse for the 4×4 case while offering a better performance in the 8×8. We mentioned that these two competitive solutions use STTC's and as such their payload decoding complexity is much higher especially for high rates close to capacity.

While the results presented in this thesis dealt strictly with flat fading, they can be easily applied to selective fading scenarios where OFDM can be used to decouple the frequency-selective channel into multiple frequency-flat channels. Also, because of stringent delay requirements in 4G systems, one can replace the time index with frequency index resulting in the use of space-frequency Alamouti code instead of the time-based Alamouti code. In fact, this code is already part of some 4G standards such as 3GPP-LTE [8].

Suggested Future Research

We believe the performance of the algorithm presented in this thesis can be improved in many ways:

- The use of soft interference cancellation instead of hard interference cancellation should reduce the effect of the error propagation problem further. This should improve the FER performance of the multiple-QR version of the proposed algorithm.

- The multiplexing performance of the algorithm can be improved in one of two ways: by replacing the Alamouti code with full-diversity full-rate space-time codes such as the tilted-QAM or Golden codes [12] or by leaving some of the TX symbols uncoded. Decoding of these uncoded symbols can simply be done through simple backward substation. The former solution has the advantage of increasing the maximum multiplexing gain of our algorithm from $\frac{N}{2}$ to N while

maintaining the same diversity advantage. Unfortunately, this solution requires extremely higher payload complexity as only the ML receiver can take advantage of the full multiplexing gain provided by these optimum codes [12]. The second solution, while it only increases the maximum multiplexing gain from $\frac{N}{2}$ to $\frac{N+N_{us}}{2}$, where N_{us} is the number of uncoded symbols, is more attractive as the decoding of the uncoded symbols can be done through simple backward substitution which has only $O(N_{us}^2)$ payload complexity. In this case, we have to ensure that these symbols are only decoded at later stages when diversity is highest. To provide enough diversity (i.e. of at least four), this technique should only be applied for symbols s_q where $q \geq 3$, restricting its use for applications where $N \geq 5$. It should be noted that in this case the decoding process should be divided in two parts: one comprising of the coded symbols and the other one comprising of the uncoded symbols. For optimal performance, decoding should happen in two stages: the decoding of the uncoded symbols should start only after applying iterative detection to the coded symbols.

Throughout this thesis we have assumed the existence of non line-of-sight (NLOS) channel between transmitter and receiver. In practice, the BS may under certain conditions be considered to have a direct line-of-sight (LOS) to the UE. Further study in this area is warranted. We believe that our algorithm will fare well in these type of scenarios as it was found that the Alamouti code performs relatively well in LOS scenarios [38]. Lastly, the effect of mobility on the performance of the algorithm should as well be investigated.

Appendix A

TX and RX Correlation Coefficient Matrices for Typical Indoor, Urban and Rural Environments

The following table gives the values of the correlation coefficient matrices that correspond to Table 2.1

Scenario	TX Correlation Coefficient Matrix $\mathbf{\Psi}_t$				RX Correlation Coefficient Matrix $\mathbf{\Psi}_r$			
Indoor	+1.0000	−0.0621	−0.0448	+0.0868	+1.0000	−0.0621	−0.0448	+0.0868
	−0.0621	+1.0000	−0.0621	−0.0448	−0.0621	+1.0000	−0.0621	−0.0448
	−0.0448	−0.0621	+1.0000	−0.0621	−0.0448	−0.0621	+1.0000	−0.0621
	+0.0868	−0.0448	−0.0621	+1.0000	+0.0868	−0.0448	−0.0621	+1.0000
Urban	+1.0000	−0.0149	−0.0910	+0.1249	+1.0000	+0.4371	−0.1995	−0.0263
	−0.0149	+1.0000	−0.0149	−0.0910	+0.4371	+1.0000	+0.4371	−0.1995
	−0.0910	−0.0149	+1.0000	−0.0149	−0.1995	+0.4371	+1.0000	+0.4371
	+0.1249	−0.0910	−0.0149	+1.0000	−0.0263	−0.1995	+0.4371	+1.0000
Rural	+1.0000	+0.8962	+0.6228	+0.2775	+1.0000	+0.9543	+0.8248	+0.6322
	+0.8962	+1.0000	+0.8962	+0.6228	+0.9543	+1.0000	+0.9543	+0.8248
	+0.6228	+0.8962	+1.0000	+0.8962	+0.8248	+0.9543	+1.0000	+0.9543
	+0.2775	+0.6228	+0.8962	+1.0000	+0.6322	+0.8248	+0.9543	+1.0000

Appendix B

Various Reviews and Proofs

B.1 Householder Transformation

Householder transformation consists of converting an arbitrary column vector $\mathbf{x} = [x_1 \ x_2 \ \dots \ x_N]^T$ into a transformed vector $\tilde{\mathbf{x}} = [\tilde{x}_1 \ 0 \ \dots \ 0]^T$ whose elements below the first element are all zero and $\tilde{x}_1 = ||\mathbf{x}||$.

The above transformation is achieved through the use of a Householder matrix $\mathbf{Z}$ which is equal to:

$$\mathbf{Z} = I - 2\mathbf{u}\mathbf{u}^H \tag{B.1.1}$$

where

$$\mathbf{u} = a \begin{bmatrix} x_1 - \tilde{x}_1 \\ x_2 \\ \vdots \\ x_N \end{bmatrix} \text{ with } a = \frac{1}{\sqrt{2\tilde{x}_1(\tilde{x}_1 - x_1)}} \tag{B.1.2}$$

Using the above vector transformation we can transform a matrix $\mathbf{X}$ of size $N \times N$ into an upper triangular form by applying the householder method successively: we

start by finding a householder matrix $\mathbf{Z}^1$ based on the first column of $\mathbf{X}$. We then repeat the procedure on a submtarix of reduced dimension located under the first column and row of $\mathbf{X}^1 = \mathbf{Z}^1\mathbf{X}$. After $N-1$ householder transforms the matrices $\mathbf{Q}$ and $\mathbf{R}$ are obtained as follows:

$$\mathbf{Q} = \mathbf{Z}^1\mathbf{Z}^2 \ldots \mathbf{Z}^{N-1} \tag{B.1.3}$$

$$\mathbf{R} = \mathbf{Z}^{N-1}\mathbf{Z}^{N-2} \ldots \mathbf{Z}^1\mathbf{X} \tag{B.1.4}$$

B.2 Sufficient Number of QR Decompositions for Optimum Ordering

The proof starts by realizing that $|^p r_{ij}|^2 = |^l r_{ij}|^2$ $(p \neq l$ and $i <= j)$ if and only if $^p\mathbf{h}_j = {}^l\mathbf{h}_j$ and $^p\mathbf{h}_i = {}^l\mathbf{h}_i$. This is a direct consequence of how every r_{ij} is computed using the Householder transformation described in the previous section. For example, given permutations $(1,2,3,4)$ and $(2,3,1,4)$ only $|^1 r_{44}|^2 = |^2 r_{44}|^2$ (here $i = j$) while permutations $(1,2,3,4)$ and $(2,1,3,4)$ satisfy $|^1 r_{44}|^2 = |^2 r_{44}|^2$ and $|^1 r_{34}|^2 = |^2 r_{34}|^2$. Using this result we can then see that:

$$||^p\mathbf{\Gamma}_q||^2 = ||^l\mathbf{\Gamma}_q||^2 \tag{B.2.1}$$

as long as the corresponding column-pair in matrix $\mathbf{H}$ is the same.

If we want to find the permutation that has the optimal ordering, i.e. ordering based on decreasing $||^l\mathbf{\Gamma}_q||^2$ $(q = 1, 2, \ldots \frac{N}{2})$, we would normally have to compute all possible $\frac{N}{2}!$ decompositions. Because of the above result the number of permutations can be substantially reduced. To see this let us take a closer look at the number of required permutations N_{qr} when $N = 4, 6$. This will allow us to generalize for cases

when $N \geq 8$. For simplicity, and because a pair of columns are grouped together, we will use a single index to refer to a particular group. For example, when $N = 4$, permutations $(1, 2, 3, 4)$ and $(3, 4, 1, 2)$ can be equivalently described by group numbers, i.e. $(1, 2)$ and $(2, 1)$ where the index 1 represents group $(1, 2)$ and index 2 represents group $(3, 4)$. Note that only the explicit computation of matrix $\mathbf{R}$ is required.

Case When $N = 4$

In this case simple we only have 2 possible permutations so $N_{qr} = 2$

Case When $N = 6$

In this case we have a total of 6 possible permutations namely:

$$\begin{matrix} 1 & 2 & 3 \\ 2 & 1 & 3 \\ 3 & 1 & 2 \\ 1 & 3 & 2 \\ 2 & 3 & 1 \\ 3 & 2 & 1 \end{matrix}$$

Because of the useful property in (B.2.1) and the fact that each group number appears twice in the last column, we only have to compute 3 permutations to find the permutation that has the highest $||^p\mathbf{\Gamma}_1||^2$ $(p = 1, 2, \ldots 6)$. Once found our attention is restricted to this group of permutations that have the same optimum group index in the last column. In this particular case we only have another permutation. So we calculate the QR decomposition for this other permutation and compare the values of $|^l\mathbf{\Gamma}_2|^2$ where $l = k, k+1$ with $k \in \{1, 2, \ldots 6\}$. So in total we need to compute only $N_{qr} = 3 + 1 = 4$ decompositions.

Case When $N \geq 8$

Similar analysis that was adopted above can be applied and we can easily find that when $\frac{N}{2}$ is even we have $N_{qr} = \frac{1}{16}N^2 + \frac{1}{4}N + 1$ and when $\frac{N}{2}$ is odd we have $N_{qr} = \frac{1}{16}N^2 + \frac{1}{4}N + \frac{9}{4}$.

B.3 Sufficient Number of QR Decompositions for Suboptimum Ordering

Resorting to the analysis presented above one can easily see that the last column contains $\frac{N}{2}$ different group numbers. Therefor choosing the highest $||^p\mathbf{\Gamma}_1||^2$ with $p = 1, 2, \ldots \frac{N}{2}$ requires the computation of only $\frac{N}{2}$ QR decompositions. Note that only the explicit computation of matrix $\mathbf{R}$ is required.

Bibliography

[1] S. M. Alamouti, *A simple transmit diversity technique for wireless communications*, IEEE J. Sel. Areas Commun.

[2] A. Algans and P.E. Mogensen, *Experimental analysis of the joint statistical properties of azimuth spread, delay spread, and shadow fading*, IEEE J. Sel. Areas Commun. **20** (2002), no. 3, 523–531.

[3] M. Arar and A. Yongacoglu, *Efficient detection algorithm for 2N×2N MIMO systems using alamouti code and QR decomposition*, IEEE Commun. Lett. **10** (2006), no. 12, 819–821.

[4] K. Azarian and H. El Gamal, *The throughputreliability tradeoff in block-fading MIMO channels*, IEEE Trans. inform. Theory **53** (2007), no. 2, 288–501.

[5] L. Barbero and J. Thompson, *Performance of the complex sphere decoder in spatially correlated MIMO channels*, IEEE IET Commun. **1** (2007), no. 1, 122–130.

[6] ———, *Extending a fixed-complexity sphere decoder to obtain likelihood information for Turbo-MIMO systems*, IEEE Trans. Veh. Technol. **57** (2008), no. 5, 2804–2814.

[7] J.C. Belfiore, G. Rekaya, and E. Viterbo, *The golden code: a* 2×2 *full-rate space-time code with non-vanishing determinants*, IEEE Trans. Inform. Theory **51** (2005), no. 4, 1432–1436.

[8] J. Berkmann, C. Carbonelli, F. Dietrich, C. Drewes, and W. Xu, *On 3G LTE terminal implementation - standard, algorithms, complexities and challenges*, Wireless Commun. and Mobile Comput. Conf., (2008).

[9] R. Böhnke and K. Kammeyer, *Exact outage probability of V-BLAST with ordered MMSE-SIC detection*, Biennnial Symp. Commun. (2006).

[10] G. Boudreau, J. Panicker, N. Guo, R. Chang, N. Wang, and S. Vrzic, *Interference coordination and cancellation for 4G networks*, IEEE Commun. Mag. **47** (2009), no. 4, 74–81.

[11] G. Foschini C. B. Papadias, *On the capacity of certain space-time coding schemes*, EURASIP Journal on Applied Signal Processing (2002).

[12] B. Clerckx C. Oestges, *MIMO wireless communications*, Academic Press, 2007.

[13] Won-Joon Choi, R. Negi, and J.M Cioffi, *Combined ML and DFE decoding for the V-BLAST system*, IEEE ICC **3** (2000).

[14] L. Dai, S. Sfar, and K.B. Letaief, *A quasi-orthogonal group space-time architecture to achieve a better-diversity tradeoff*, IEEE Trans. Wireless. Commun. **6** (2007), no. 4, 1295–1307.

[15] M.O. Damen, H. Gamal, and G. Caire, *On maximum-likelihood detection and the search for the closest lattice point*, IEEE Trans. Inform. Theory **49** (2003), no. 10, 2389–2402.

[16] M.O. Damen, A. Tewfik, and J.C. Belfiore, *A construction of a space-time code based on number theory*, IEEE Trans. Inform. Theory **48** (2002), no. 3, 753–760.

[17] O. Damen, A. Chkeif, and J. Belfiore, *Lattice code decoder for space-time codes*, IEEE Commun. Lett. **4** (2000), no. 5, 161–163.

[18] P. Dayal and M.K. Varanasi, *An optimal two transmit antenna space-time code and its stacked extenions*, IEEE Trans. Inform. Theory **51** (2005), no. 12, 4348–4355.

[19] G. D.Golden, G. J.Foschini, R. A.Valenzuela, and P. W.Wolniansky, *Detection algorithm and initial laboratory results using the V-BLAST spacetime communication architecture*, Electron. Lett., **35** (1999), no. 1, 14–15.

[20] R. B. Ertel, P. Cardieri, K. W. Sowerby, T. S. Rappaport, and J. H. Reed, *Overview of spatial channel models for antenna array communication systems*, IEEE Pers. Commun. (1998).

[21] G.J. Foschini et al, *Analysis and performance of some basic space-time architectures*, IEEE J. Sel. Areas Commun. **21** (2003), no. 3, 281–320.

[22] G. J. Foschini and M. J. Gans, *On limits of wireless communications in a fading environment when using multiple antennas*, Wireless Pers. Commun. **6** (1998), no. 3, 311335.

[23] G.J Foschini, *Layered space-time architecture for wireless communication in a fading environment when using multiple antennas*, Bell Labs Technical Journal **1** (1996), no. 2, 41–59.

[24] C. V. Loan G. H. Golub, *Matrix computations*, JHU Press, 1996.

[25] K. Hoffmann G. Hämmerlin and L. Schumaker, *Numerical mathematics*, Springer, 1991.

[26] D. Gesbert, Helmut Bölcskei, D.A. Gore, and Arogyaswami J. Paulraj, *Outdoor wireless channel models: Models and performance prediction*, IEEE Trans. Commun. **50** (2002), no. 12, 1926–1934.

[27] D. Gesbert, M. Shafi, and D. Shiu, *From theory to practice: An overview of MIMO space-time coded wireless systems*, IEEE Journal **21** (2003), no. 3, 281–302.

[28] B. Hassibi, *An efficient square-root algotithm for BLAST*, IEEE ICASSP 2000 **2** (2000).

[29] B. Hassibi and B. M. Hochwald, *High-rate codes that are linear in space and time*, IEEE Trans. Inform. Theory **48** (2002), no. 7, 1804–1824.

[30] Y. Jiang, X. Sheng, and J. Li, *Asymptotic performance analysis of V-BLAST*, IEEE GLOBECOM **6** (2005).

[31] T. Kailath and A. H. Sayed, *Fast reliable algorithms for matrices with structure*, Society for Industrial Mathematics, 1999.

[32] K. Kumar, G. Caire, and A. Moustakas, *The diversity-multiplexing tradeoff of linear MIMO receivers*, IEEE Inform. Theory Workshop (2007).

[33] J. Ling, D. Chizhik, D. Samardzija, and R.A. Valenzuela, *Peer-to-peer MIMO radio channel measurements in a rural area*, IEEE J. Sel. Areas Commun. **6** (2007), no. 9, 3229–3237.

[34] S. Loyka and F. Gagnon, *On the outage error rate analysis of the ordered V-BLAST*, IEEE Trans. Wireless Commun. **7** (2008), no. 10, 3679–3685.

[35] S. Loyka and G. Levin, *On finite-snr diversity-multiplexing tradeoff*, IEEE GLOBECOM (2007).

[36] T. L. Marzetta, *BLAST training: Estimating channel characteristics for high capacity space-time wireless*, Proc. 37th Annual Allerton Conf. on Commun., Control, and Computing (1999).

[37] R. Narasimhan, *Finite-snr diversity-multiplexing tradeoff for correlated rayleigh and rician MIMO channels*, IEEE Trans. Inform. Theory **52** (2006), no. 9, 3995–3979.

[38] S. Parker, M. Sandell, and M. Lee, *The performance of space-time codes in office environments*, Veh. Tech. Conf. **1** (2003).

[39] A. J. Paulraj, D.A. Gore, R.U. Nabar, and H. Bölcskei, *An overview of MIMO communications: A key to gigabit wireless*, Proc. of the IEEE **92** (2004), no. 2, 198–218.

[40] A. J. Paulraj and T. Kailath, *Increasing capacity in wireless broadcast systems using distributed transmission/directional reception*, U.S. Patent 5345599, 1994.

[41] N. Prasad and M. K. Varanasi, *Optimum efficiently decodable layered space-time block codes*, Conf. Record of the Thirty-Fifth Asilomar Conf. on Signal, System and Comp. **1** (2001).

[42] S. Sezginer, H. Sari, and E. Biglieri, *A comparison of full-rate full-diversity 2x2 space-time codes for wimax systems*, IEEE 10th Intern. Symp. Spread Spect. Tech. and App. (2008).

[43] H. Shin and J.H. Lee, *Capacity of multiple-antenna fading channels: spatial fading correlation, double scaterring and keyhole*, IEEE Trans. Inform. Theory **49** (2003), no. 10, 2636–2647.

[44] B. Sklar, *Rayleigh fading channels in mobile digital communication systems part I: Characterization*, IEEE Commun. Mag. (1997).

[45] Q. Spencer, C. Peel, A. Swindlehurst, and M. Haardt, *An introduction to the multi-user MIMO downlink*, IEEE Commun. Mag. **42** (2004), no. 10, 60–67.

[46] C. Studer, A. Burg, and H. Bölcskei, *Soft-output sphere decoding: Algorithms and VLSI implementation*, IEEE J. Sel. Areas Commun. **26** (2008), no. 2, 290–300.

[47] M. Tao and R. S. Cheng, *Generalized layered space-time codes for high data rate wireless communications*, IEEE Trans. Wireless Commun. **3** (2004), no. 4, 1067–1075.

[48] V. Tarokh, H. Jafarkhani, and A. R. Calderbank, *Space-time block codes from orthogonal designs*, IEEE Trans. Inform. Theory **45** (1999), no. 5, 1456–1467.

[49] V. Tarokh, A. Naguib, N. Seshadri, and A. R. Calderbank, *Combined array processing and space-time coding*, IEEE Trans. Inform. Theory **45** (1999), no. 4, 1121–1128.

[50] V. Tarokh, N. Seshadri, and A. R. Calderbank, *Space-time codes for high rate wireless communication: performance criterion and code construction*, IEEE Trans. Inform. Theory **44** (1998).

[51] TR 36.211 V8.6.0 Technical Specification Group Radio Access Network, *Physical channels and modulation*, Tech. report, 3GPP, 2009.

[52] TR 36.913 V8.0.1 Technical Specification Group Radio Access Network, *Requirements for further advancements for evolved universal terrestrial radio access (E-UTRA)*, Tech. report, 3GPP, 2009.

[53] I.E. Telatar, *Capacity of multi-antenna gaussian channels*, European Trans. Telecom. **10** (1999), no. 6, 585–595.

[54] D. Tse and P. Viswanath, *Fundamentals of wireless communication*, Cambridge University Press, 2005.

[55] A. Tulino and S. Verdu, *Random matrix theory and wireless communications*, Now Publishers Inc., 2004.

[56] B. Vucetic and J. Yuan, *Space-time coding*, John Wiley and Sons Ltd., 2003.

[57] D. Wübben, R. Böhnke, J. Rinas, V. Kühn, and K. D. Kammeyer, *MMSE extension of V-BLAST based on sorted QR decomposition*, IEEE Electron. Lett. **37** (2001), no. 22, 13481350.

[58] H. Yao and G.W. Wornell, *Structured space-time block codes with optimal diveristy-multiplexing tradeoff and minimum delay*, Proc. Globecom- IEEE Global Telecommun. Conf. **4** (2003).

[59] K. Yu and A. Alexiou, *Impact of channel estimation errors on various spatial-temporal transmission schemes*, Mobile and Wireless Commun. Summit **1** (2007), no. 5, 1–5.

[60] L. Zheng and D. Tse, *Diversity and multiplexing: A fundemental tradeoff in multiple-antenna channels*, IEEE Trans. Inform. Theory **47** (2003), no. 5, 1073–1096.