Precoding and Multiuser Scheduling in MIMO Broadcast Channels

Seung-Hwan Lee

A thesis submitted for the degree of Doctor of Philosophy.
The University of Edinburgh.
July 2007

Abstract

Multiple input multiple output (MIMO) techniques are the most promising technologies for next-generation wireless systems to achieve improved channel reliability as well as high spectral efficiencies. Based on the theoretical foundation, a number of studies focusing on practical implementation of MIMO techniques have been recently presented.

In real MIMO downlink scenarios, the number of users are usually greater than that of transmit antennas at a base station and the base station is likely to provide a variety of services to different quality-of-service (QoS) users. Therefore, a multiuser scheduling algorithm for the real MIMO downlink scenarios has to support a mixture of QoS users simultaneously by exploiting the performance gains of multiple antennas whilst maximizing the sum-rate capacity by selecting a user set for transmission according to performance criteria.

The main topic of this thesis is the design of QoS-guaranteed multiuser schedulers for MIMO systems. This should provide different QoS services to different users whilst satisfying the system-level requirements such as fairness among users, minimum data rate and delay constraints as well as trying to maximize the sum-rate capacity of MIMO channels. For this, this thesis first investigates the performance of MIMO transceiver techniques in terms of error rates and the sum-rate capacity with practical considerations to select a practically appropriate MIMO precoding technique. Then a QoS-aware sequential multiuser selection algorithm is proposed, which selects a user set sequentially from each QoS group in order to satisfy QoS requirements by trading off the transmit antennas between different QoS groups. Using a temporally-correlated MIMO channel model validated by channel measurements, a statistical channel state information (SCSI)-assisted multiuser scheduling algorithm is also proposed, which can minimize the effect of the temporal correlation on the sum-rate capacity. Finally, new metrics are proposed to support fairness among users in terms of throughput or delay whilst maximizing the sum-rate capacity. With these proposed algorithms, the objective of this thesis, to support a mixture of different QoS users simultaneously with fairness considerations whilst maximizing the sum-rate capacity by exploiting the advantages of MIMO techniques with practical implementation in mind, can be achieved.

Declaration of originality

I hereby declare that the research recorded in this thesis and the thesis itself was composed and originated entirely by myself in the School of Engineering and Electronics at The University of Edinburgh.

The first exception was the MIMO channel measurements in Chapter 5, which was conducted as part of a Mobile Virtual Centre of Excellence Programme. The measured data was obtained during the placement at Toshiba Telecommunications Research Laboratory in Bristol, UK in 2006.

The second exception was Figure 3.3 in Chapter 3, which was taken from [1].

Seung-Hwan Lee

Acknowledgements

I would like to thank the following people for their assistance during the course of this PhD:

- First of all, Dr. John Thompson, my supervisor, for his invaluable support and sincere guidance.
- Prof. Stephen McLaughlin, my second supervisor, for his support.
- Prof. Bernard Mulgrew, the internal and Prof. Alister Burr, the external for reading and checking this thesis at the PhD viva.
- Toshiba Research Europe Limited (TREL) Telecommunications Research Laboratory (TRL) in Bristol, UK, for its financial support.
- Dr. Darren McNamara for his technical discussions.
- People in the Institute for Digital Communications.
- People in Korea, especially Prof. Kyu-Tae Lee, Dr. Jin-Up Kim, Dr. Kyung-Seok Kim and Dr. Yeon-Woo Lee for encouraging me to start my PhD in the UK, and Prof. Heung-Gyoon Ryu, Prof. Chung-Ku Kang, Dr. Byung-Hwa Lee, Dr. Jong-Hun Han and Dr. In-Soon Jeong for their support.
- Last but not least, my parents and my family for their love.

Contents

List of figures

List of tables

Acronyms and abbreviations

ADC	Analogue to Digital Converter
AMC	Adaptive Modulation and Coding
AOA	Angle of Arrival
AOD	Angle of Departure
AWGN	Additive White Gaussian Noise
BE	Best Effort
BER	Bit Error Rate
BLAST	Bell Labs Layered Space Time
BS	Base Station
CDF	Cumulative Distribution Function
CDMA	Code Division Multiple Access
CR	Constant Rate
CSI	Channel State Information
DAC	Digital to Analogue Converter
DFE	Decision Feedback Equalizer
DPC	Dirty Paper Coding
D-BLAST	Diagonal Bell Labs Layered Space Time
FDMA	Frequency Division Multiple Access
FTP	File Transfer Protocol
HOL	Head-of-Line
HTTP	Hyper Text Transfer Protocol
ICSI	Instantaneous Channel State Information
i.i.d.	independent and identically distributed
LWDF	Largest Weighted Delay First
LRD	Long Range Dependence
MAC	Medium Access Control
MEA	Method of Equal Area
MIMO	Multiple Input Multiple Output
MIMO BC	Multiple Input Multiple Output Broadcast Channel

MIMO MAC	Multiple Input Multiple Output Multiple Access Channel
ML	Maximum Likelihood
M-LWDF	Modified Largest Weighted Delay First
MMSE	Minimum Mean Square Error
M-VCE	Mobile Virtual Centre Excellence
PDF	Probability Density Function
PF	Proportional Fair
QAM	Quadrature Amplitude Modulation
QF	QoS-aware Fair
QoS	Quality of Service
rms	root mean squared
RF	Radio Frequency
RL-EAFRP	Rate-Limited Extended Alternating Fractal Renewal Process
RT	Real Time
RX	Receiver
SCSI	Statistical Channel State Information
SDMA	Space Division Multiple Access
SER	Symbol Error Rate
SNR	Signal-to-Noise Ratio
SOS	Sum-of-sinusoids
TDD	Time Division Duplex
TDMA	Time Division Multiple Access
THP	Tomlinson-Harashima Precoding
TX	Transmitter
V-BLAST	Vertical Bell Labs Layered Space Time
XPD	Cross polarization discrimination
ZF	Zero Forcing
ZFBF	Zero Forcing Beamforming
ZF-DPC	Zero Forcing Dirty Paper Coding
ZMCSCG	Zero Mean Circular Symmetric Complex Gaussian

Nomenclature

a_i	QoS parameters for the LWDF scheduling rule
a_d^t	Constants for different delay requirements in the delay fairness metric
a_k^t	Constants for different throughput requirements in the throughput fairness metric
$a\bar{D}$	Average weighted delay in the exponential rule
$\mathbf{a}$	Input data vector
$\hat{\mathbf{a}}$	Estimated input data vector at the receiver
α_0	Constant for the OFF period of the Pareto distribution
α_1	Constant for the ON period of the Pareto distribution
α_l	Path loss exponent
α_t	Smoothing factor for the exponential moving average
α_{data}	Constant for the data length of the Pareto distribution
$\alpha^c(t)$	Envelope in the multiplicative channel model
$\alpha_k^c(t)$	Envelope of kth tap in the tapped delay line channel model
$\mathbf{B}$	Lower triangular matrix by the QR decomposition
$\hat{\beta}$	Asymptotic slope
c_{ki}	Gains for the SOS random process
C	Speed of light
C_{DPC}	Sum-rate capacity of DPC
C_{MAX}	Maximum sum-rate capacity in MIMO broadcast channels
C_{THP}	Sum-rate capacity of THP
C_{ZFBF}	Sum-rate capacity of ZFBF
C_{ZF-DPC}	Sum-rate capacity of ZF-DPC
$d_i(t)$	HOL delays
D	Decorrelation distance
D_A	Antenna spacing
$\bar{D}(t)$	Average of HOL delays of all users
$\bar{D}_{RT}$	Average HOL delay of RT users
D_{TH}	Delay threshold

δ	Dynamic range of the soft trade-off scheme
$\Delta\mathbf{H}$	Measurement noise matrix
η_i	Maximum probability of exceeding the delay threshold
ε	Variable used in fairness metrics to prevent zero denominators
$f(\mathbf{s})$	Probability distribution of the vector $\mathbf{s}$
f_c	Carrier frequency
f_d^k	Doppler spreads
f_D	Maximum Doppler shift
f_{ki}	Frequencies for the SOS random process
$\mathbf{F}$	Beamforming matrix at the transmitter
$\mathbf{F}^H$	Beamforming matrix at the receiver
$\mathbf{g}_k$	Rayleigh-distributed fading between the transmit antenna and user k
$\mathbf{G}$	Scaling matrix
G_t	Transmit power gain
G_r	Received power gain
γ_{ij}	Fraction of transmission for user j in the subgroup i
γ_k	Effective channel gain of ZFBF
Γ_{THP}	Power increase in THP due to the modulo operation
Γ_{SNR}	SNR gap
$h_{ij}(t)$	Time-varying impulse response between transmit antenna j and receive antenna i
H	Hurst parameter
$H(\mathbf{y})$	Differential entropy of the vector $\mathbf{y}$
$H(\mathbf{y}\|\mathbf{s})$	Conditional differential entropy of the vector $\mathbf{y}$ given the knowledge of the vector $\mathbf{s}$
$H(\mathbf{z})$	Differential entropy of the vector $\mathbf{z}$
$\mathbf{H}$	MIMO channel matrix
$\mathbf{H}_0$	Perfectly estimated channel matrix
$\mathbf{H}_w$	$i.i.d.$ MIMO channel matrix
$\mathbf{H}_m$	Uncorrelated estimation error matrix
$\hat{\mathbf{H}}$	Estimated Channel matrix
I_{QF}	Fairness Index
$\mathbf{I}$	Identity matrix
$I(\mathbf{s};\mathbf{y})$	Mutual information between vectors $\mathbf{s}$ and $\mathbf{y}$

K	Number of users in a MIMO broadcast channel
K_i	Number of users in ith subgroup
K_{BE}	Number of BE users in a MIMO broadcast channel
K_{CR}	Number of CR users in a MIMO broadcast channel
K_{RT}	Number of RT users in a MIMO broadcast channel
λ_c	Wavelength
λ_d	Mean arrival rate of Poisson data
λ_i	Eigenvalues
m_s	Area mean
m_d	Delay slope in the delay fairness metric
m_t	Throughput slope in the throughput fairness metric
M_k	Modulation orders for spatial channels
M_R	Number of receive antennas
M_S	Number of pre-assigned transmit antennas to RT users
M_T	Number of transmit antennas
M_{BE}	Number of transmit antennas for BE users
M_{CR}	Number of transmit antennas for CR users
M_{RT}	Number of transmit antennas for RT users
μ	Water level in the waterfilling algorithm
$\mu_d^t(t)$	Delay fairness metric
$\mu_k^t(t)$	Throughput fairness metric
N_0	Single-sided power spectral density of AWGN
N_G	Number of subgroups whose fairness requirements are identical
N_S	Number of slots in a frame
N_{sos}	Number of sinusoids for the SOS random process
p_k	Arbitrary symbol due to the modulo operation
P_0	Packet error rate
P	Total transmit power
P_d	Packet drop probability
P_k	Transmit power allocated to user k
$\tilde{P}_k$	Transmit power allocated to user k by ZFBF
P_r	Received power
P_t	Transmitted power

$\mathbf{P}$	Permutation matrix
$\mathbf{P}(n)$	Autocorrelation matrix at slot index n
$\mathbf{P}_{opt}$	Optimum permutation matrix
$\pi(k)$	user k considering all permutations
$\mathbf{Q}_{MMSE}$	Matrix of MMSE equalizer
r	Rank of the MIMO channel
r_0^d	Reference distance
r_k^d	Distance between the base station and user k
r'	Number of spatial channels for the waterfilling algorithm
$r_i(t)$	Data rates
$\bar{r}_i(t)$	Exponential moving average of past throughput for user i
R_a	Packet arrival rate
R_{BE}	Throughput of BE users
R_{cell}	Cell radius
R_{CR}	Throughput of CR users
R_{RT}	Throughput of RT users
R_{TH}	Throughput threshold
R_{TOTAL}	Throughput of all users
$\bar{R}(t)$	Average of $\bar{r}_i(t)$ of all users
$\mathbf{R}_k$	Spatial correlation matrix of the MIMO channel by the Kronecker model
$\mathbf{R}_r$	Spatial correlation matrix at the receiver
$\mathbf{R}_t$	Spatial correlation matrix at the transmitter
$\mathbf{R}_{MIMO}$	Receive processing matrix
$\mathbf{R}_{ss}$	Input covariance matrix
ρ	Temporal correlation
$\mathbf{s}$	Transmit signal vector
S	User set
S_{BE}	BE user set
S_{CR}	CR user set
S_{max}	User set obtained by the sum-rate maximization rule
S_{RT}	RT user set
S_{BE}^t	BE user set with throughput fairness among users

S_{RT}^{d}	RT user set with delay fairness among users
SNR_0	Median of the mean SNR of all users
σ_e^2	Variance of measurement noise
σ_e^2	Variance of measurement noise
σ_i^2	Variance of interference
σ_s	Shadow standard deviation
σ_z^2	Variance of complex AWGN
T_f	Frame period
T_i	Delay threshold for user i in LWDF
T_P	Cut-off limit of the Pareto distribution
T_S	Slot period
$\mathbf{T}_{MIMO}$	Transmit processing matrix
τ	Delay
θ_{ki}	Phases for the SOS random process
$\theta^c(t)$	Phase in the multiplicative channel model
$\theta_k^c(t)$	Phase of kth tap in the tapped delay line channel model
U_{BE}	BE user set
U_{CR}	CR user set
U_{RT}	RT user set
v_k	Mobile velocities
w_k	Shadowing variables
W	Signal bandwidth
$\zeta_k(t)$	Sum-of sinusoids random process
$\mathbf{y}$	Received signal vector
$\mathbf{z}$	Zero mean circular symmetric complex Gaussian noise vector

Chapter 1
Introduction

1.1 Motivation

After pioneering works on the multiple-input multiple-output (MIMO) capacity in [7] and [8] were presented, various researches on MIMO areas have been performed for next generation wireless communication systems. These include multiuser MIMO capacity, MIMO channel models, MIMO transceiver techniques, multiuser scheduling algorithms and practical implementation issues related to MIMO techniques.

It is well known that the single user capacity of a MIMO system increases linearly with the minimum number of transmit and receive antennas as long as the environment has sufficiently rich scattering. This work has been extended to MIMO multiple access channels (MACs or uplink, *i.e.*, channels from mobile terminals to a base station) and MIMO broadcast channels (BCs or downlink, *i.e.*, channels from a base station to mobile terminals), where the multiuser capacity is defined by the capacity region due to the infinite ways to share the transmit power among users. Although there is a duality between the capacity region of these two MIMO channels, calculating the capacity of the MIMO BC is much more difficult than for the MIMO MAC because the MIMO BC is nondegraded. However, the maximum sum-rate capacity of the MIMO BC can be achieved by using the dirty paper coding principle. Dirty paper coding (DPC) provides a theoretical background for practical nonlinear MIMO precoding techniques such as Tomlinson-Harashima precoding (THP) for the MIMO BC. Because a base station in the MIMO BC knows the transmit signal at the transmitter, it can minimize the interference among users by intelligently designing the transmit signal using the DPC principle.

In general, MIMO techniques can provide not only diversity gain which increases the reliability of transmission but multiplexing gain which boosts data rate by exploiting the space dimension due to multiple antennas both at the transmitter and receiver. This diversity-multiplexing trade-off in multiple antennas channels can be applied to MIMO BC scenarios, where the performance gain due to multiple antennas is used to balance the overall system performance with individual requirements.

There are two aspects on the overall system performance in real MIMO BC scenarios. One aspect is to maximize the sum-rate capacity of the MIMO BC in view of the physical layer. The maximum sum-rate capacity of the MIMO BC can be achieved by the optimal power allocation and encoding order of DPC as long as the number of transmit and receive antennas are identical. When the number of users are larger than that of transmit antennas, which is likely to be in real MIMO BC scenarios, a multiuser selection procedure should be considered. Because the wireless channel is random and each user is likely to have independent fading, the multiuser selection procedure should exploit the multiuser diversity gain by selecting a group of users with the best channel conditions. The sum-rate maximization rule selects a user set of the most orthogonal users who can maximize the sum-rate capacity as well as the multiuser diversity gain among all possible choices of user set. The other aspect is to satisfy different system-level requirements from higher network layers. These include fairness among users, different quality-of-service (QoS) requirements such as minimum data rate and delay constraints. Higher layer protocols consider the physical layer as limited resources such as power and bandwidth, so that they seek the optimal algorithm for allocating these limited resources for a desirable system performance. Therefore, a multiuser scheduling algorithm in the real MIMO BC scenarios combines the multiuser selection procedure considering wireless channel conditions with the resource allocation procedure satisfying system-level requirements in order to maximize the overall system performance, which is related to the cross-layer optimization problem. The spatial gain due to multiple antennas gives one more degree of freedom available, so that the multiuser scheduling algorithm should exploit the spatial gain in order to satisfy the system-level requirements whilst maximizing the sum-rate capacity of MIMO broadcast channels.

1.2 Contribution

Figure 1.1 illustrates the system model for this thesis.

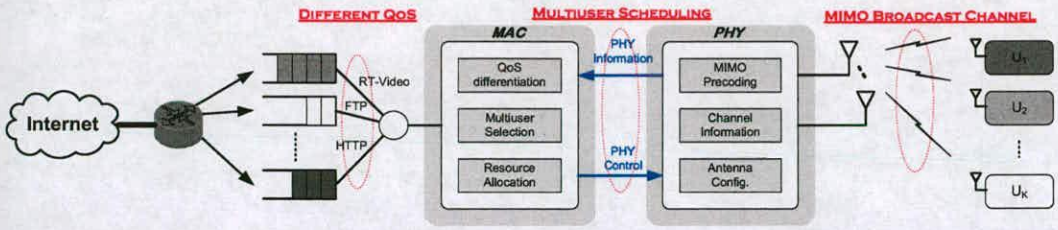


Figure 1.1: *The system model for this thesis*

The objective of this thesis can by summarized by the following statement:

> **To propose a multiuser scheduling algorithm for MIMO broadcast channels, which can support a mixture of different QoS users simultaneously with fairness considerations whilst maximizing the sum-rate capacity by exploiting the advantages of MIMO techniques with practical implementation in mind.**

This can be split into five key areas:

- To investigate the performance of MIMO precoding techniques in terms of the sum-rate capacity as well as error rates.
- To propose a multiuser selection algorithm, which can support different QoS users simultaneously as well as maximizing the multiuser diversity gain.
- To propose a MIMO broadcast channel model, which is realistic with reasonable complexity from the viewpoint of multiuser scheduling.
- To propose a multiuser scheduling algorithm, which exploits statistical channel state information to minimize performance degradation.
- To propose a multiuser scheduling algorithm, which considers fairness among users with QoS differentiation whilst maximizing the sum-rate capacity

1.3 Thesis Structure

This thesis is structured as follows:

Chapter 2 will provide some background of this thesis. This includes the characteristics of the wireless channel, the definition of the MIMO channel and existing literature reviews for the modelling of the MIMO channel, and the capacity of the MIMO channel. It will also describe the basic concept of multiuser scheduling in wireless channels.

Chapter 3 will briefly describe MIMO transceiver techniques and compare their performance in terms of error rates with several practical considerations such as a transmit power constraint and channel estimation errors. This comparison will be performed in an independent and identically distributed ($i.i.d.$) single user MIMO channel.

Chapter 4 will briefly introduce the multiuser MIMO capacity and compare MIMO precoding techniques in terms of the sum-rate capacity in an $i.i.d.$ MIMO broadcast channel. It will also propose a new QoS-aware sequential multiuser selection algorithm, which can support different QoS users simultaneously with QoS constraints by selecting user sets sequentially from the highest QoS user group and changing the number of pre-assigned transmit antennas allocated to different QoS groups dynamically to improve the probability of QoS-guaranteed transmission as well as maximizing the sum-rate capacity.

Chapter 5 will propose a temporally correlated MIMO broadcast channel model, which is simple enough to be used in any system level simulator with ease but realistic enough to describe the key phenomena of real MIMO broadcast channels in view of a multiuser scheduling algorithm. The proposed model will be validated by comparing with the MIMO channel measurements. Chapter 5 will also propose a new statistical channel state information (SCSI)-assisted multiuser scheduling algorithm, which can minimize the effect of the temporal correlation. The proposed algorithm exploits statistical channel state information to minimize the mismatch of channel estimates, so that it can improve the sum-rate capacity in time-varying MIMO broadcast channels.

Chapter 6 will briefly review conventional scheduling algorithms and propose a new QoS-guaranteed multiuser scheduling algorithm, which can support different QoS users with the consideration of fairness among users in terms of throughput or delay, and QoS differentiation between different QoS groups. The proposed algorithm exploits the QoS-aware sequential

multiuser selection algorithm proposed in Chapter 4 for supporting different QoS groups simultaneously. It can further improve the sum-rate capacity by using the soft antenna trade-off scheme, which can adapt to the time-varying channel conditions gracefully to improve the outage probability of QoS violations. It uses the weighted sum-rate maximization rule to find a user set for transmission, whose weight vectors are determined by degree of fairness among users in the same QoS group.

Finally, Chapter 7 will draw conclusions from the work that has been described in this thesis. Suggestions for relevant further work will also be presented.

Chapter 2

Background

This chapter will provide some fundamental theories about the wireless channel, the MIMO channel, the MIMO capacity and a brief description of the concept of scheduling. These are essential to understand this thesis. It will also present common notations used throughout this thesis.

2.1 The Wireless Channel

A signal propagating through the wireless channel arrives at the destination through a number of different paths called multipath. These paths arise from scattering, reflection and diffraction by objects in the environment or refraction in the medium. The different propagation mechanisms influence path loss and fading models differently. We refer to all these distorting mechanisms as "scattering". The signal power drops off due to three effects: mean path loss, macroscopic fading and microscopic fading. The path loss comes from inverse square low power loss and the propagation channel environment, which is range dependent. Long term fading results from a blocking effect by buildings and natural features and is also known as shadowing. Short term fading results from the constructive and destructive combination of multipath and is also known as fast fading. Figure 2.1 shows the combined path loss, long term fading and short term fading wireless channel. Multipath propagation results in the spreading of the signal in different dimensions. These are delay spread, Doppler spread and angle spread [9].

It is useful to note that most of contents in this section are based on [9], [10] and [11]. More detained discussion about the wireless channel can be found in some classic textbooks [12] [13].

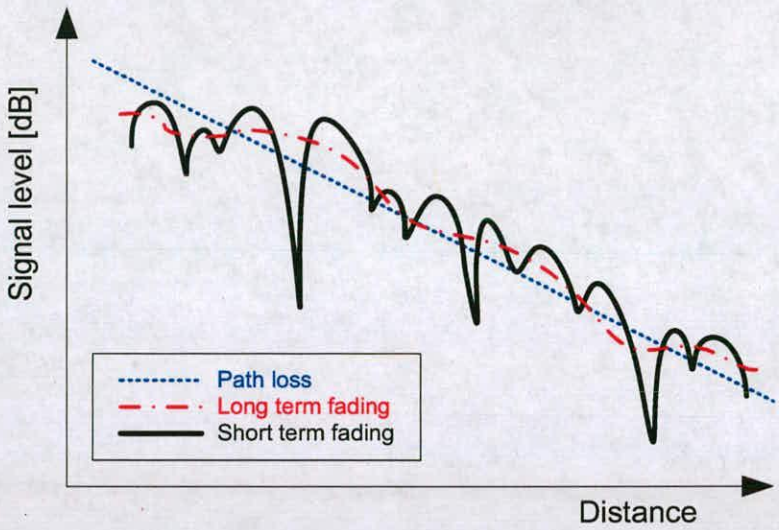


Figure 2.1: *The Combined path loss, long term fading and short term fading wireless channel*

2.1.1 Path Loss

In free space, due to the inverse square law power loss, the received signal at a distance r^d is given by

$$P_r = P_t \left(\frac{\lambda_c}{4\pi r^d}\right)^2 G_t G_r \quad [\text{Watts/m}^2], \tag{2.1}$$

where P_t and P_r denote the transmitted and received power respectively, λ_c denotes the wave length of the carrier frequency, G_t and G_r are the power gains of the transmit and receive antennas, respectively.

For real environments, the following path loss model is commonly used for system design.

$$\mathrm{P}_r = P_t \left(\frac{r_0^d}{r^d}\right)^{\alpha_l}, \tag{2.2}$$

where r_0^d denotes a reference distance and α_l denotes the path loss exponent, which varies from 2.5 to 5 according to the channel environment.

2.1.2 Fading

2.1.2.1 Long Term Fading

Long term fading is caused by blockage from objects such as buildings or natural features in the signal path. It has been shown that its probability density distribution (PDF) can be modelled by a log-normal distribution

$$p(x) = \frac{1}{\sqrt{2\pi}\sigma_s} \exp\left\{-\frac{(\log x - m_s)^2}{2\sigma_s}\right\}, \quad x > 0, \tag{2.3}$$

where σ_s denotes the shadow standard deviation and m_s denotes the area mean. The areas mean is obtained by averaging the received signal strength over an area that is large enough to average over the shadowing effects [14]. In this case, the area mean is equal to the distance dependent path loss. The decorrelation distance D is defined as the distance to which the correlation coefficient reduces to e^{-1}. This has different value according to shadowing environments. In [15], the measurement results showed that D=8.3m and D=503.9m for the urban and suburban environments respectively.

2.1.2.2 Short Term Fading

Short term fading corresponds to the rapid fluctuations of the received signal. It is caused by the scattering of the signal off objects between the transmitter and the receiver. If we assume that fading is caused by the superposition of a large number of independent scattered components, the received signal can be assumed to be an independent zero mean complex Gaussian process. The envelope of the received signal has a Rayleigh density function given by

$$p(x) = \begin{cases} \frac{x}{\sigma^2} \exp\left\{-\frac{x^2}{2\sigma^2}\right\}, & x \geq 0 \\ 0, \quad x < 0 \end{cases}, \tag{2.4}$$

where σ^2 is the variance of x.

2.1.2.3 Doppler Spread - Time Selectivity

Time-varying fading due to scatterers or transmitter/receiver motion results in the Doppler spread. The Fourier transform of the time autocorrelation of the channel response to a con-

tinuous wave tone is defined as the Doppler power spectrum. The Doppler spread is defined as the root mean square bandwidth of the Doppler power spectrum. Coherence time is typically defined as the time lag for which the signal autocorrelation coefficient reduces to 0.7. The coherence time can be approximated as the inverse of the Doppler spread [9].

2.1.2.4 Delay Spread - Frequency Selectivity

In a multipath propagation environment, several time-shifted and scaled versions of the transmitted signal arrive at the receiver. The span of path delays is called the delay spread. If the channel is time-varying, the delay spread also varies with time. In this case, the rms (root mean squared) delay spread is defined by the standard deviation of the delay spread. The delay spread causes frequency-selective fading when the signal bandwidth is larger than the coherence bandwidth. Frequency selective fading can be characterized in terms of the coherence bandwidth, which is the frequency lag for which the channel's autocorrelation coefficient reduces to 0.7. The coherence bandwidth can be approximated as the inverse of the delay spread. If the channel has frequency nonselective or flat fading, it can be modelled as the multiplicative channel model [16].

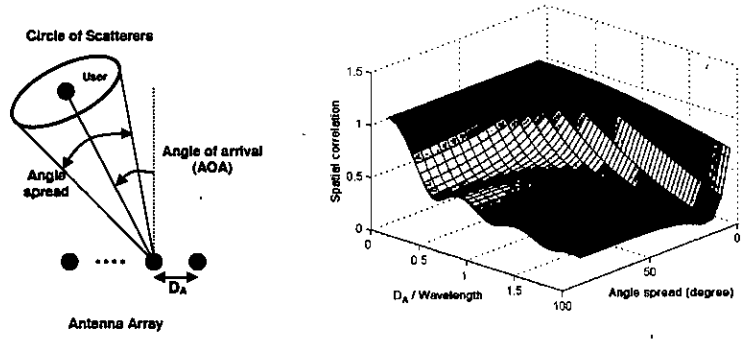
$$y(t) = \alpha^c(t)e^{j\theta^c(t)}s(t), \tag{2.5}$$

where $\alpha^c(t)$ and $\theta^c(t)$ denote the envelope and the phase of the channel respectively and $s(t)$ and $y(t)$ are the transmitted and received signal respectively. Note that this thesis assumes frequency nonselective fading, where the signal bandwidth is narrower than the coherence bandwidth. This assumption can be justified if an orthogonal frequency division multiplexing (OFDM) is used with MIMO techniques. In this case, each subcarrier of the MIMO-OFDM system experiences the frequency nonselective fading.

If the channel has frequency selective fading, it can be modelled as the tapped delay line channel model as

$$y(t) = \sum_{k=1}^{L} \alpha_k^c(t)e^{j\theta_k^c(t)}s(t - \frac{k}{W}), \tag{2.6}$$

where W denotes the signal bandwidth, L denotes the number of taps, $\alpha_k^c(t)$ and $\theta_k^c(t)$ denote the envelope and the phase of kth tap respectively.



(a) One-ring model *(b) Spatial correaltion*

Figure 2.2: *The one-ring model and its spatial correlation*

2.1.2.5 Angle Spread - Space Selectivity

Angle spread refers to the spread of angles of arrival/departure of the multipath components, which is usually defined as the rms angle spread. The angle spread causes space selective fading, which means that signal amplitude depends on the spatial location of the antenna. Space selective fading is characterized by the coherence distance, which is the spatial separation for which the signal autocorrelation coefficient reduces to 0.7. The coherence distance can be approximated as the inverse of the angle spread. The angle spread determines the characteristics of spatial correlation channel models. Figure 2.2 shows the one-ring model [17] and its spatial correlation. The power azimuth spectrum of the angle spread can be Gaussian, Laplace or uniform distribution [18]. In Figure 2.2, Gaussian distribution and AOA=45° are assumed. We notice that the spatial correlation decreases as the antenna spacing (D_A) and the angle spread increases. This implies that the MIMO channel can be spatially uncorrelated if the antenna spacing is more than half the wavelength of the carrier frequency and a large angle spread obtained from rich scattering around multiple antennas.

2.2 MIMO Channel

Consider a MIMO system with M_T transmit and M_R receive antennas, which is illustrated in Figure 2.3. Denoting $h_{ij}(t)$ as the complex time-varying impulse response between transmit antenna j ($j = 1, \cdots, M_T$) and the receive antenna i ($i = 1, \cdots, M_R$), the MIMO channel is

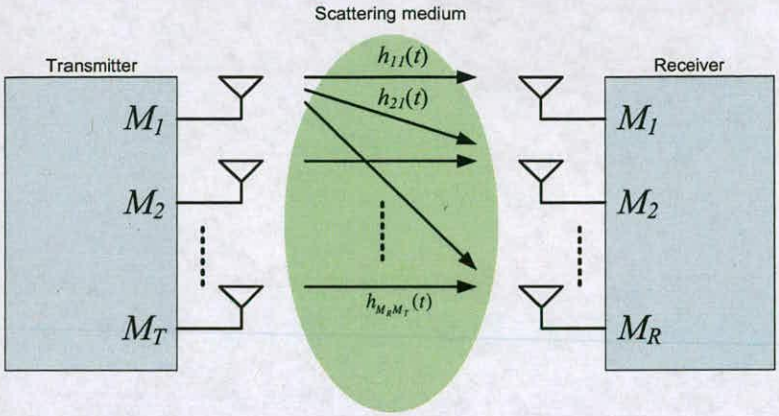


Figure 2.3: *The Multiple Input Multiple Output (MIMO) channel*

given by the $M_R \times M_T$ matrix $\mathbf{H}(t)$

$$\mathbf{H}(t) = \begin{bmatrix} h_{11}(t) & h_{12}(t) & \cdots & h_{1M_T}(t) \\ h_{21}(t) & h_{22}(t) & \cdots & h_{2M_T}(t) \\ \vdots & \vdots & \ddots & \vdots \\ h_{M_R1}(t) & h_{M_R2}(t) & \cdots & h_{M_RM_T}(t) \end{bmatrix}. \tag{2.7}$$

With the narrowband assumption in Section 2.1.2, the relationship between the input and the output for the MIMO channel can be expressed as

$$\mathbf{y}(t) = \mathbf{H}(t)\mathbf{s}(t) + \mathbf{z}(t), \tag{2.8}$$

where $\mathbf{y}(t) \in \mathbb{C}^{M_R \times 1}$ denotes the received signal vector at time t, $\mathbf{s}(t) \in \mathbb{C}^{M_T \times 1}$ denotes the transmit signal vector at time t and $\mathbf{z}(t) \in \mathbb{C}^{M_R \times 1}$ denotes the zero mean circular symmetric complex Gaussian (ZMCSCG) noise vector with variance σ_z^2 at time t.

If the MIMO channel response is unchanged for a certain time period of interest ($i.e.$, frequency nonselective block fading), the equation (2.8) reduces to

$$\mathbf{y} = \mathbf{H}\mathbf{s} + \mathbf{z}, \tag{2.9}$$

If the entries of the channel matrix $\mathbf{H}$ have ZMCSCG variables with unit variance, it is called an $i.i.d.$ channel $\mathbf{H}_w$, which is spatially white. However, the real MIMO channel is not spatially

white due to practical impairments such as spatial correlation, temporal correlation, keyhole effect and so on. The keyhole effect happens in certain propagation scenarios where all multipath components go through a single keyhole, so that the MIMO channel capacity decreases even though the signals at the antenna elements are uncorrelated. However, it has been shown that the keyhole effect might not occur frequently in practice [19] and was not observed in the MIMO channel measurement [20].

A lot of MIMO channel models have been developed to describe real MIMO environments. Most of them are geometry-based spatial channel models like the one-ring model illustrated in Figure 2.2. The channel parameters are modelled by finite [21] or infinite scatterers between the transmit and receive antennas. The geometrical distributions of scatterers as well as their number determine the characteristics of spatial channel models. In [22], an overview of spatial channel models has been presented. A survey of channel and radio propagation models for wireless MIMO systems was presented in [23]. Most of spatial channel models assumed frequency flat fading. Several papers have explored frequency selective fading channels for wideband MIMO channels based on the tapped-delay line models [24] [25]. The spatial fading correlation can be simply expressed as a function of AOA, beamwidth, and antenna spacing [26]. The probability density function of AOA depends on the distribution of scatterers. If the spatial correlation coefficients at the transmit antennas are independent of those of the receive antennas, the correlations can be separable. Therefore, the $M_R M_T \times M_R M_T$ spatial correlation matrix of the MIMO channel $\mathbf{R}_k$ is the Kronecker product of the spatial correlation matrix at the transmitter and the receiver, which is given by

$$\mathbf{R}_k = \mathbf{R}_r \otimes \mathbf{R}_t, \tag{2.10}$$

where $\mathbf{R}_r$ is the $M_R \times M_R$ spatial correlation matrix at the receiver and $\mathbf{R}_t$ is the $M_T \times M_T$ spatial correlation matrix at the transmitter. This concept was validated in several papers by comparison with measurement results [19]. However, the Kronecker structure of the channel covariance matrix is suitable for arrays with a moderate number of antenna elements [25]. The Kronecker model fails in correlated environments where the spatial correlations of transmitter and receiver are not separable [27]. The MIMO channel with spatial correlation is given by [25] [28]

$$\mathbf{H} = \mathbf{R}_r^{1/2} \mathbf{H}_w \mathbf{R}_t^{1/2}. \tag{2.11}$$

However, these spatial channel models ignored the effect of temporal variations. Joint spatio-temporal models were presented in [24] and [29], where spatio-temporal correlations are modelled based on angle of departure, Rician K factor, keyhole, Doppler spread, delay spread, far clusters, shadowing, and path loss, in addition to conventional spatial model parameters. Rician K factor is defined as the ratio of signal power in dominant component (line-of-sight) over the scattered power (non line-of-sight). In [29], the effect of the Doppler frequency on the fading correlation has been considered. In [30], the effect of the Doppler frequency has been quantified in terms of the level crossing rate.

The effect of the spatial fading correlation on the MIMO capacity has been presented in [31] with the spatial correlation model of equation (2.11), in [32] with the one-ring model, in [33] with a ray-tracing propagation model and in [21] with finite scatterers channel model. All the results from these papers tell us that performance of MIMO systems decreases if there are spatial fading correlations.

It is useful to note that as the number of parameters considered increases, the complexity of the channel model also grows, making it difficult to implement and run the MIMO channel model efficiently despite its improved accuracy. Therefore, a simple but realistic MIMO downlink channel model will be proposed in Chapter 5, which is useful for investigating multiuser scheduling algorithms for the MIMO downlink channel.

2.3 MIMO Capacity

The data capacity for an additive white Gaussian noise (AWGN) channel was first introduced by C. E. Shannon in his paper "A mathematical theory of communication" [34]. It is defined as the maximum data rate that can be supported with an arbitrarily small probability of error.

The capacity of a deterministic MIMO channel is defined as the maximum mutual information between input vector $\mathbf{s}$ and output vector $\mathbf{y}$, given in equation (2.8), as

$$C = \max_{f(\mathbf{s})} I(\mathbf{s};\mathbf{y}), \tag{2.12}$$

where $f(\mathbf{s})$ is the probability distribution of the vector $\mathbf{s}$ and $I(\mathbf{s};\mathbf{y})$ is the mutual information between vectors $\mathbf{s}$ and $\mathbf{y}$. After some mathematical manipulation (See Appendix B), the single

user capacity of the deterministic MIMO channel is given by

$$C = \max_{\mathrm{Tr}(\mathbf{R}_{ss})=P} \log_2 \left| \mathbf{I} + \frac{\mathbf{H}\mathbf{R}_{ss}\mathbf{H}^H}{N_0} \right|, \tag{2.13}$$

where $\mathbf{R}_{ss} = E\left[\mathbf{s}\mathbf{s}^H\right]$ denotes the input covariance matrix and N_0 denotes the single-sided power spectral density of AWGN. The total transmit power P is given by

$$P = \mathrm{Tr}\left(\mathbf{R}_{ss}\right) = \sum_{i=1}^{M_T} P_i, \tag{2.14}$$

where P_i is the transmit power allocated to transmit antenna i.

If the transmitter does not know the channel state information (CSI), the total power can be equally divided among the transmit antennas as $\mathbf{R}_{ss} = \frac{P}{M_T}\mathbf{I}$. Then the capacity of the MIMO channel without channel knowledge at the transmitter is given by

$$C = \log_2 \left| \mathbf{I} + \frac{P}{M_T N_0}\mathbf{H}\mathbf{H}^H \right|. \tag{2.15}$$

When the CSI is known to the transmitter, by the singular value decomposition of $\mathbf{H}\mathbf{H}^H$, the capacity of the MIMO channel is given by

$$C = \sum_{i=1}^{r} \log_2 \left(1 + \frac{P_i \lambda_i}{N_0}\right), \tag{2.16}$$

where r ($r \leq \min(M_R, M_T)$) is the rank of the MIMO channel $\mathbf{H}$ and λ_i are the eigenvalues of $\mathbf{H}\mathbf{H}^H$.

The optimal P_i of any spatial channel can be obtained by the waterfilling algorithm [35]

$$P_i = \left(\mu \frac{\lambda_i}{N_0} - 1\right)^+, \tag{2.17}$$

where $(x)^+$ denotes $\max(x, 0)$ and the water level μ is selected to satisfy

$$\sum_{i=1}^{r'} \left(\mu - \frac{N_0}{\lambda_i}\right)^+ = P, \tag{2.18}$$

where r' ($r' \leq r$) denotes the number of spatial channels with the allocated power $P_i > 0$.

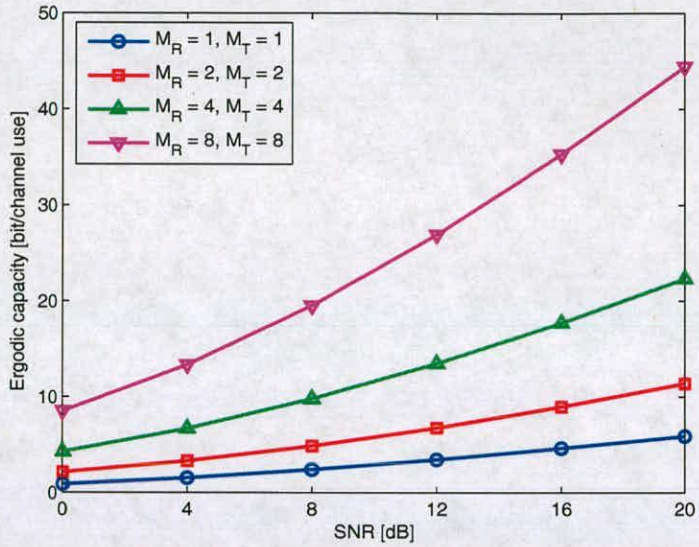


Figure 2.4: *The ergodic capacity of a single user MIMO channel*

The capacity of the MIMO channel can be increased by using the waterfilling algorithm but the capacity gap between the known and unknown channel cases reduces at high SNR values.

If the channel matrix $\mathbf{H}$ is random, the information rate associated with the MIMO channel is also a random variable. In order to deal with this, there are ergodic capacity and outage capacity measures for fading MIMO channels. The ergodic capacity is the ensemble average of the information rate over the distribution of the random elements of $\mathbf{H}$. The q % outage capacity is the information rate that is guaranteed for $(100 - q)$ % of channel realizations. Figure 2.4 and Figure 2.5 show the ergodic capacity and the 10% outage capacity of a single user MIMO channel for different antenna configurations with the waterfilling algorithm. We notice that the MIMO capacity grows linearly with the minimum number of transmit and receive antennas in terms of both the ergodic capacity and the 10% outage capacity in the single user MIMO channel.

Figure 2.6 shows the ergodic capacity of a 4×4 MIMO channel with and without channel knowledge at the transmitter. The ergodic capacity when the channel is known to the transmitter is higher than the ergodic capacity when the channel is unknown. However, the difference in the ergodic capacity decreases as SNR increases. This implies that an equal power allocation can have a similar performance to the waterfilling algorithm with a high SNR assumption.

Unlike the single user capacity, a multiuser MIMO capacity is characterized by a capacity

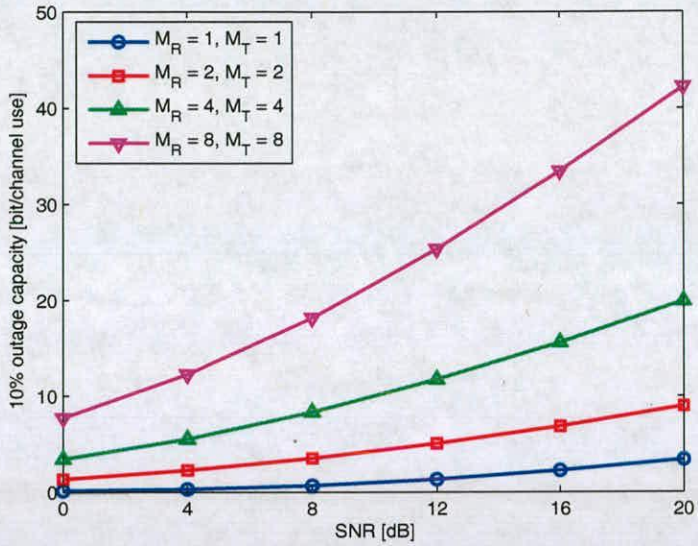


Figure 2.5: *The 10 % outage capacity of a single user MIMO channel*

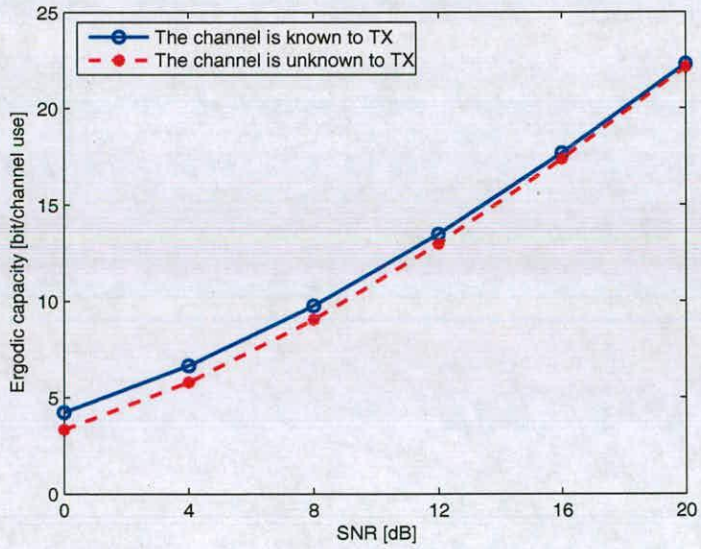


Figure 2.6: *The ergodic capacity of a* 4×4 *MIMO channel with and without channel knowledge at the transmitter*

region, which will be described later in Chapter 4. In multiuser MIMO scenarios, there may be a base station with multiple antennas and multiple users each with multiple antennas in a single cell. The base station transmits data to multiple users via downlink channels and receives data from multiple users via uplink channels. The MIMO downlink channel is called a MIMO broadcast channel and the MIMO uplink channel is called a MIMO multiple access channel. These concepts will be described in detail in Chapter 4. In order to obtain the maximum value of the multiuser MIMO capacity from the capacity region, there should be an optimal way of dealing with multiple users simultaneously as well as considering the wireless channels. In the next section, the concept of multiuser scheduling will be introduced to explain what could be the optimal way for maximizing the multiuser MIMO capacity, which is one of the main contributions of this thesis.

2.4 Multiuser Scheduling in Wireless Channels

Scheduling is defined as a process which gives a list of events and indicates in what order they will take place. Especially for wireless communications, it is concerned with how to allocate limited resources such as power and bandwidth efficiently for a desirable performance.

Round-robin (RR) scheduling is one of the simplest scheduling algorithms which assigns time slices to each user in equal portions and in order without priority consideration. However, it is inefficient when applied to a time-varying wireless channel. For example, consider two users in a downlink scenario where a base station transmits data to each user according to a certain scheduling algorithm. The maximum possible data rate for user 1 in a time slot is either 4 or 6 [bits/channel use] with equal probabilities 0.5. This is determined by the signal-to-noise ratio (SNR) of the time-varying wireless channel. For user 2, the maximum possible data rate in a time slot is either 2 or 8 [bits/channel use], also with equal probabilities. The base station can serve one user at a time, so that it should select one user according to certain scheduling algorithms. It is assumed that the channels for both users are independent. Because the RR scheduling allows each user to use the channel for transmission once per every two time slots, the average data rates of user 1 (R_1) and user 2 (R_2) with the RR scheduling are

$$R_1 = 0.5 \times (0.5 \times 4 + 0.5 \times 6) = 2.5[\text{bits/channel use}] \tag{2.19}$$

and

$$R_2 = 0.5 \times (0.5 \times 2 + 0.5 \times 8) = 2.5[\text{bits/channel use}], \tag{2.20}$$

respectively. Then, the sum-rate $(R_1 + R_2)$ is 5 [bits/channel use] if the RR scheduling is used.

Consider a channel aware scheduling algorithm, which only gives a transmission opportunity to a user in a relatively better channel condition. For this, the wireless channel information from the physical layer is sent to the medium access control layer, so that the channel aware scheduler can select a user in a cross-layer fashion. There are four possible combinations of the channel conditions of two users. When the possible data rate of user 1 is 4 [bits/channel use], the possible data rate of user 2 can be 2 or 8 [bits/channel use]. When the possible data rate of user 1 is 6 [bits/channel use], the possible data rate of user 2 can also be 2 or 8 [bits/channel use]. The channel aware scheduler selects user 1 when user 1 has 4 [bits/channel use] and user 2 has 2 [bits/channel use] or user 1 has 6 [bits/channel use] and user 2 has 2 [bits/channel use]. In this way, the average date rates using the channel aware scheduling are

$$R_1 = 0.25 \times 4 + 0.25 \times 6 = 2.5[\text{bits/channel use}] \tag{2.21}$$

and

$$R_2 = 0.25 \times 8 + 0.25 \times 8 = 4[\text{bits/channel use}]. \tag{2.22}$$

With the channel aware scheduling algorithm, the sum-rate is increased by 1.5 [bits/channel use] compared to the RR rule.

In real multiuser wireless communication scenarios, there are several factors to be considered such as quality-of service (QoS) and data queue status in addition to wireless channel conditions [36]. Different users may require different QoS in different channel conditions. Because the queue length is not infinite, a scheduling algorithm should consider the queue length so that queue overflow is minimized. Let P_d denote the packet dropping probability due to queue overflow and P_0 the packet error rate. A packet from a transmitter is correctly received by a receiver, only if it is not dropped from the queue (with probability $1 - P_d$) and correctly received through the wireless channel (with probability $1 - P_0$) [37]. In this case, the average throughput can be given by $R_a(1 - P_d)(1 - P_0)$, where R_a is the packet arrival rate on the queue at the transmitter. Therefore, in order to maximize the average throughput whilst satisfying different QoS requirements, a multiuser scheduler should consider the wireless channel conditions ('channel aware'), the queue status ('queue aware') and QoS ('service aware') si-

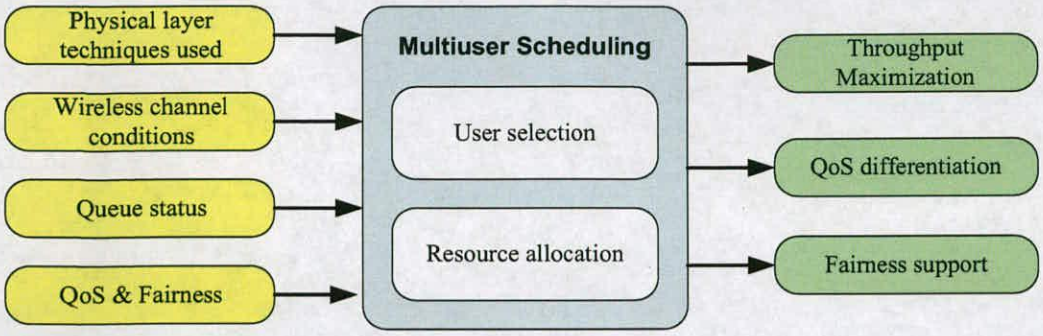


Figure 2.7: *The concept of multiuser scheduling*

multaneously. Note that, in this thesis, the packet dropping probability and the packet error rate are assumed to be negligible, so that throughput is determined by the packet arrival rate and the channel capacity. When the packet arrival rate is less than or equal to the channel capacity, it determines the average throughput. However, when the packet arrival rate is greater than the channel capacity, the average throughput is limited by the channel capacity.

Figure 2.7 illustrates the concept of multiuser scheduling for general wireless communication systems. In general, a multiuser scheduling consists of a user selection procedure and a resource allocation procedure. In the user selection procedure, the multiuser scheduling algorithm selects a single user or a set of multiple users according to a certain scheduling metric. When multiple antennas are used at the transmitter of MIMO systems, the multiuser scheduling algorithm can select multiple users simultaneously to increase the MIMO capacity or a single user to improve the channel reliability of the user. In the resource allocation procedure, the multiuser scheduling algorithm allocates limited resources such as time, bandwidth, power and space (if MIMO techniques are used) to the selected user set in an efficient way considering system level requirements such as QoS and fairness among users as well as channel conditions.

2.5 Notations

Throughput this thesis, we use boldface to denote matrices and vectors. For any general matrix $\mathbf{A}$, $\mathbf{A}^T$ denotes the transpose, $\mathbf{A}^H$ denotes the conjugate transpose, $\mathrm{Tr}(\mathbf{A})$ denotes the trace, $||A||$ denotes the vector norm, $\mathrm{diag}\{\lambda_i\}$ denotes a diagonal matrix with the (i, i) entry equal to λ_i and $\mathbf{I}$ denotes the identity matrix. $E[\cdot]$ denotes expectation. For any general set B, $|B|$ denotes the cardinality of the set. $\mathrm{MOD}[\cdot]$ denotes the modulo operation.

2.6 Conclusions

This chapter has briefly introduced the characteristics of a MIMO channel and its capacity. It also described the basic concept of multiuser scheduling in wireless channels. The MIMO capacity grows linearly with the minimum number of transmit and receive antennas. The multiuser scheduling consists of the user selection procedure and the resource allocation procedure, so that it maximizes throughput whilst supporting QoS differentiation and fairness among users in a cross-layer fashion. Multiuser scheduling combined with MIMO techniques is the main topic of this thesis.

In Chapter 3, MIMO transceiver techniques are compared in a single user $i.i.d.$ MIMO channel. From Chapter 4 to Chapter 6, topics related to multiuser scheduling algorithms will be presented. In Chapter 5 and Chapter 6, a temporally correlated MIMO channel will be considered to investigate the performance of multiuser scheduling algorithms in a realistic MIMO downlink scenarios.

Chapter 3
Comparison of MIMO Transceiver Techniques with Practical Considerations

3.1 Introduction

This chapter will briefly introduce MIMO transceiver techniques and compare their performance with several practical considerations in an $i.i.d.$ single user MIMO channel. There are two types in MIMO transceiver techniques: MIMO transmit techniques and MIMO receive techniques. The MIMO transmit techniques include beamforming, dirty paper coding (DPC), zero forcing dirty paper coding (ZF-DPC), Tomlinson-Harashima precoding (THP) and various vector precoding techniques. The MIMO receive techniques include maximum likelihood (ML) detector, linear equalizer, decision feedback equalizer (DFE) and Bell Labs layered space-time (BLAST) techniques. Among these transceiver techniques, the performance of ZF-DPC, THP, zero forcing beamforming (ZFBF) and vertical BLAST (V-BLAST) will be compared in terms of error rates with practical considerations such as a transmit power constraint and channel estimation errors.

ZF-DPC is a nonlinear suboptimal implementation of DPC. THP applies a modulo operation to ZF-DPC in order to prevent a possibly large power increase due to the pre-subtraction of ZF-DPC. However, THP suffers from the modulo loss. Unlike V-BLAST, both ZF-DPC and THP use a feedback filter similar to a DFE at the transmitter for interference pre-subtraction. The V-BLAST technique has the feedback filter of the DFE at the receiver for the successive detection and interference subtraction. Although V-BLAST has no power increase like ZF-DPC and no modulo loss like THP, it suffers from the error propagation at the receiver. Indeed, V-BLAST is not likely to be feasible for MIMO BC scenarios, especially with a single receive antenna for each user, because it requires receive antenna cooperation to perform successive detection efficiently at the receiver. Therefore, by comparing the performance of these MIMO transceiver techniques in terms of error rates, it will show why we use THP as a practical MIMO precoding

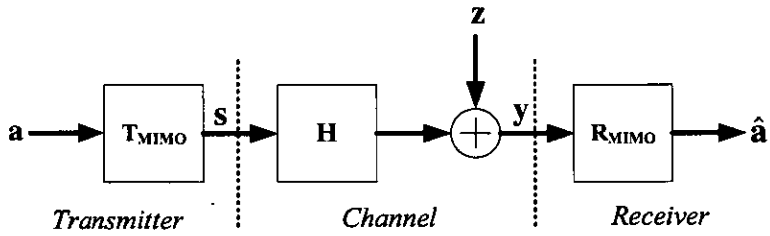


Figure 3.1: *The system model of a single user MIMO channel*

technique despite its suboptimality in the theoretical sense. Later chapters will use THP for developing multiuser scheduling algorithms for multiuser MIMO downlink wireless systems.

This chapter is organized as follows. Section 3.2 describes the system model used in this chapter. Section 3.3 gives brief descriptions of MIMO transceiver techniques. Section 3.4 compares the performance of MIMO transceiver techniques in terms of error rates with practical considerations. A summary for this chapter is presented in Section 3.5.

3.2 System Model

In this section, we focus on a full rank single-user block fading MIMO channel with an equal number of K transmit and receive antennas, which is illustrated in Figure 3.1. Refer to Figure 2.3 for the MIMO channel.

The $K \times 1$ data symbol vector $\mathbf{a} \in \mathbb{C}^{K\times 1}$, whose covariance matrix is $\mathbf{R}_{aa} = E\left[\mathbf{a}\mathbf{a}^H\right] = \sigma_a^2\mathbf{I}$, is the input to the transmit processing matrix $\mathbf{T}_{MIMO}$, whose function depends on MIMO transceiver technique being used. The $K \times 1$ transmit signal vector is $\mathbf{s} \in \mathbb{C}^{K\times 1}$, where each element of $\mathbf{s}$ has transmit power P_i and its covariance matrix $\mathbf{R}_{ss} = E\left[\mathbf{s}\mathbf{s}^H\right] = \sigma_s^2\mathbf{I}$. The $K \times K$ matrix $\mathbf{H} = [h_{rt}]$ consists of gain factors h_{rt} between transmit antenna t and receive antenna r which are $i.i.d.$ zero mean circular symmetric complex Gaussian (ZMCSCG) random variables with unit variance. The $K \times 1$ vector $z \in \mathbb{C}^{K\times 1}$ denotes complex AWGN with zero mean and variance σ_z^2. The total transmit power P per symbol period is constrained as

$$\mathrm{Tr}\left(\mathbf{R}_{ss}\right) \leq P. \tag{3.1}$$

Then the $K \times 1$ received signal $\mathbf{y} \in \mathbb{C}^{K \times 1}$ can be expressed as

$$\mathbf{y} = \mathbf{H}\mathbf{s} + \mathbf{z}. \tag{3.2}$$

After appropriate receive processing at the receiver by matrix $\mathbf{R}_{MIMO}$, the transmit data symbol vector is decoded as $\hat{\mathbf{a}}$.

A simple additive measurement noise model [38] for imperfect channel state information (CSI) is assumed as

$$\hat{\mathbf{H}} = \mathbf{H} + \Delta\mathbf{H}, \tag{3.3}$$

where the estimated channel $\hat{\mathbf{H}}$ consists of an exact channel matrix $\mathbf{H}$ and a measurement noise $\Delta\mathbf{H}$ of $K \times K$ matrix whose entries are often modelled as $i.i.d.$ ZMCSCG variables with variance σ_e^2. Unlike the perfect CSI case, the received signal includes the disturbance component $\Delta\mathbf{H}\mathbf{s}$, which is proportional to both the measurement noise and the transmit signal power.

3.3 MIMO Transceiver Techniques

In this section, we briefly present MIMO transceiver techniques. A detailed analysis of MIMO transceiver techniques in terms of error rates can be found in [39]. Among them, ZF-DPC, THP, ZFBF and V-BLAST are compared in the next section in terms of error rates with several practical considerations such as modulo loss, transmit power constraints and channel estimation errors.

3.3.1 MIMO Receive Techniques

For any MIMO receive technique, the data symbol vector $\mathbf{a}$ becomes the transmit signal vector $\mathbf{s}$ without any pre-processing ($\mathbf{T}_{MIMO} = \mathbf{I}$). In this case, the variance of the total transmit power of any MIMO receive technique is equal to $K\sigma_a^2$.

3.3.1.1 Maximum Likelihood Detector

The ML detector is the optimal receiver but it has exponential growth of complexity with the number of transmit antennas and the constellation size of the transmit symbols. This makes it

difficult to implement in practice. The ML detector searches over all possible vector symbols exhaustively and selects the vector $\hat{\mathbf{a}}$ according to

$$\hat{\mathbf{a}} = \arg\min_{\mathbf{s}} \|\mathbf{y} - \mathbf{H}\mathbf{s}\|^2 \tag{3.4}$$

In order to reduce the prohibitive complexity of the ML detector, the sphere decoding algorithm has been suggested [40]. The main idea of the sphere decoding algorithm is to reduce the search area from the entire vector space to within a certain hypersphere around the received signal $\mathbf{y}$.

3.3.1.2 Linear Equalizer

There are two types in the linear equalizer: Zero forcing (ZF) equalizer and minimum mean square error (MMSE) equalizer. Obviously, the output of ZF equalizer is given by

$$\hat{\mathbf{a}} = \mathbf{s} + \mathbf{H}^{+}\mathbf{z}, \tag{3.5}$$

where $\mathbf{H}^{+} = \left(\mathbf{H}^{T}\mathbf{H}\right)^{-1}\mathbf{H}^{T}$ is the pseudoinverse matrix of $\mathbf{H}$. The ZF equalizer decouples the matrix channel into K parallel scalar channels with additive noise. While the ZF equalizer completely eliminates the interference from other spatial channels at the expense of noise enhancement, the MMSE equalizer balances the residual interference and the noise enhancement by finding the matrix

$$\mathbf{Q}_{MMSE} = \arg\min_{\mathbf{Q}} E\left[\|\mathbf{Q}\mathbf{y} - \mathbf{s}\|^2\right]. \tag{3.6}$$

Using the principle of orthogonality

$$E\left[\left(\mathbf{Q}_{MMSE}\mathbf{y} - \mathbf{s}\right)\mathbf{y}^{H}\right] = \mathbf{0}, \tag{3.7}$$

$\mathbf{Q}_{MMSE}$ is easily derived as

$$\mathbf{Q}_{MMSE} = \left(\mathbf{H}^{H}\mathbf{H} + \frac{\sigma_z^2}{\sigma_s^2}\mathbf{I}\right)\mathbf{H}^{H}. \tag{3.8}$$

It is useful to note that the MMSE matrix $\mathbf{Q}_{MMSE}$ approaches the inverse matrix of ZF equalizer $\mathbf{H}^{+}$ as the signal to noise ratio becomes large.

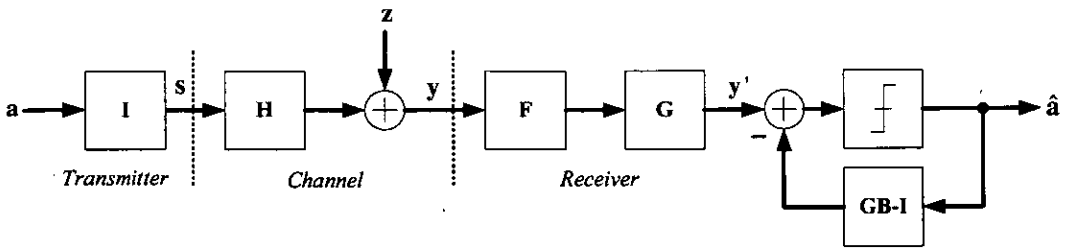


Figure 3.2: *The matrix form of the decision feedback equalizer(DFE)*

3.3.1.3 Decision Feedback Equalizer

The conventional decision feedback equalizer consists of two parts: a feedforward filter and a feedback filter. The feedforward filter is identical to the linear transversal equalizer. The feedback filter has the sequence of decisions on previously detected symbols as its input. For a DFE in MIMO channels, the channel $\mathbf{H}$ is decomposed into a unitary beamforming matrix $\mathbf{F}^H$ and a lower triangular matrix $\mathbf{B}$ by taking the QR decomposition as follows:

$$\mathbf{H} = \mathbf{F}^H\mathbf{B} \tag{3.9}$$

Figure 3.2 shows the general structure of the matrix DFE for MIMO channels. The unitary beamforming matrix $\mathbf{F}$ is the feedforward filter and the scaled lower triangular matrix $\mathbf{GB}$ becomes a part of the feedback filter, where the diagonal scaling matrix $\mathbf{G} = \text{diag}\left\{b_1^{-1}, \cdots b_K^{-1}\right\}$ determines the effective gain of all spatial channels. The received signal vector after the feedforward filter $\mathbf{F}$ and the scaling matrix $\mathbf{G}$ becomes

$$\begin{aligned} \mathbf{y}' &= \mathbf{GFy} \\ &= \mathbf{GBs} + \mathbf{GFz}. \end{aligned} \tag{3.10}$$

The scalar form of (3.10) for spatial channel k is given by

$$y'_k = s_k + \sum_{l=1}^{k-1} \frac{b_{kl}s_l}{b_{kk}} + z'_k, \tag{3.11}$$

where z'_k is a scaled AWGN for spatial channel k multiplied by the unitary matrix $\mathbf{F}$ and the diagonal scaling matrix $\mathbf{G}$. Because the unitary matrix $\mathbf{F}$ does not enhance or color the noise $\mathbf{z}$, z'_k has a variance of $\frac{\sigma_z^2}{b_{kk}^2}$. Because $\mathbf{a} = \mathbf{s}$, the output of the feedback filter for spatial channel

k is given by

$$\begin{aligned}\hat{a}_k &= y'_k - \sum_{l=1}^{k-1} \frac{b_{kl}\hat{a}_l}{b_{kk}} \\ &= a_k + \sum_{l=1}^{k-1} \frac{b_{kl}}{b_{kk}} (a_l - \hat{a}_l) + z'_k \end{aligned} \tag{3.12}$$

From (3.12), it is obvious that $(k-1)$ previously detected symbols are required for obtaining $\hat{a}_k$ in the feedback operation. However, the error propagation occurs when $a_l - \hat{a}_l \neq 0$. This results in performance degradation.

3.3.1.4 V-BLAST Receiver

Since the successive detection of V-BLAST [41] [42] is equivalent to the operation of a DFE, the V-BLAST receiver can be interpreted as a special case of the generalized DFE [43]. The main difference between the DFE and V-BLAST is the ordering of detections. The decoding order affects the channel gain, so that the performance can be improved by the optimal ordering. The conventional V-BLAST algorithm finds the optimal ordering of detections in the sequence of maximizing the signal to interference plus noise ratio (SINR) [44]. In Section 3.4, an optimal ordering for minimizing average error rates of all spatial channels will be presented.

3.3.2 MIMO Transmit Techniques

For MIMO transmit techniques, the data symbol vector $\mathbf{a}$ is filtered by the pre-processing matrix $\mathbf{T}_{MIMO}$ before it is transmitted. In this case, the total transmit power depends on the type of MIMO transmit technique used.

3.3.2.1 Linear Beamforming

With transmit beamforming [45], transmit data symbols are separated by different beamforming directions. Let $\mathbf{h}_k \in \mathbb{C}^{K\times 1}$ $\left(\mathbf{H} = [\mathbf{h}_1 \cdots \mathbf{h}_K]^T\right)$ and $\mathbf{w}_k \in \mathbb{C}^{K\times 1}$ $(\mathbf{W} = [\mathbf{w}_1 \cdots \mathbf{w}_K])$ be the column channel vector for spatial channel k and the beamforming weight column vector for

spatial channel k respectively. Then the received signal for spatial channel k is given by [46]

$$y_k = \mathbf{h}_k^T \mathbf{w}_k s_k + \sum_{j \neq k} \mathbf{h}_k^T \mathbf{w}_j s_j + z_k. \tag{3.13}$$

The receiver detects the transmit symbol s_k by treating the interference terms as an additive Gaussian noise. However, it is difficult to find the optimal weight vector, especially for large K. Therefore, suboptimal zero forcing beamforming (ZFBF) is widely used for practical purposes. In ZFBF, beamforming vectors are selected so that they satisfy the zero interference condition $\mathbf{h}_k \mathbf{w}_j = 0$ for $j \neq k$. In this case, the weight matrix $\mathbf{W}$ is simply the pseudoinverse of $\mathbf{H}$, which is given by $\left(\mathbf{H}^T\mathbf{H}\right)^{-1}\mathbf{H}^T$.

In order to make the variance of the total transmit power of ZFBF the same as the variance of V-BLAST $K\sigma_a^2$, the normalization factor should be used at the transmitter, which is of the form

$$g_{ZFBF} = \sqrt{\frac{K}{\mathrm{Tr}\left(\mathbf{H}^{+}\mathbf{H}^{+H}\right)}} \tag{3.14}$$

3.3.2.2 Dirty Paper Coding

In [47], Costa presented the principle of 'dirty-paper coding'; the capacity of a system with known interference at the transmitter is the same as if there were no interference present. No more power is required to cancel the interference.

Let assume $u \in \mathbb{R}$, $s \in \mathbb{R}$, $i \sim \mathcal{N}\left(0, \sigma_i^2\right)$ and $z \sim \mathcal{N}\left(0, \sigma_z^2\right)$ be the codeword, the transmit data, the interference with Gaussian distribution of zero mean and variance σ_i^2, and the Gaussian noise with zero mean and variance σ_z^2 respectively. The receiver does not know the interference i. The objective of DPC is to find an appropriate transmit data s considering the interference i for delivering the codeword u to the receiver to avoid the effect of the interference i. Then the received signal y is given by

$$y = s + i + z. \tag{3.15}$$

If the power of the transmit data satisfies the power constraint $|s|^2 \leq P$, the capacity of this system is

$$C = \log_2\left(1 + \frac{P}{\sigma_z^2}\right), \tag{3.16}$$

which is independent of σ_i^2. However, it is very difficult to implement in practical multiuser

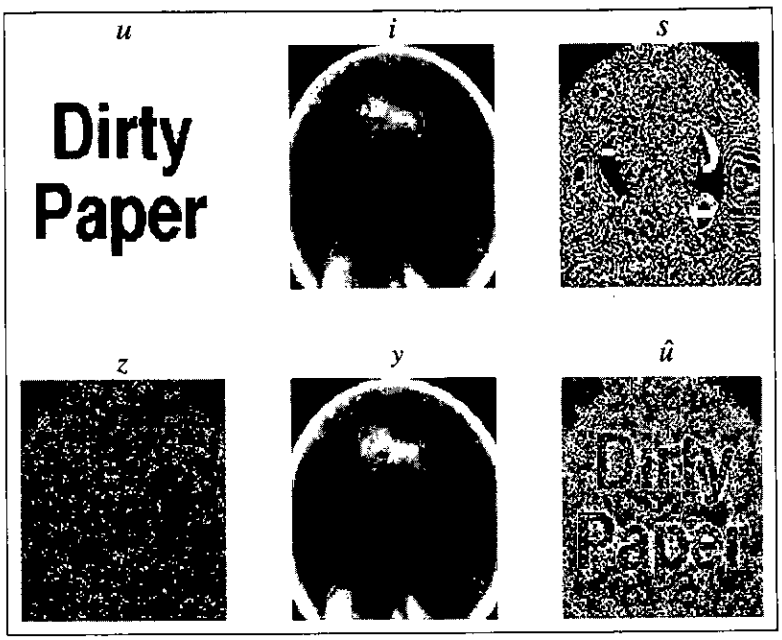


Figure 3.3: *An example of dirty paper coding (DPC) (from [1])*

MIMO systems due to its high computational burden of encoding and decoding, especially when the number users is large [46]. Therefore, practical precoding techniques such as ZF-DPC and THP have been proposed, which exploits the DPC principle. A good illustrative example for DPC can be found in [1]. In Figure 3.3, the transmitter sends s instead of u using DPC and the receiver obtains $\hat{u}$ despite the interference i after an appropriate signal processing.

3.3.2.3 Zero Forcing Dirty Paper Coding

Zero forcing dirty paper coding exploits the DPC principle [48]. It decomposes the MIMO channel $\mathbf{H}$ into a transmit beamforming matrix $\mathbf{F}$ and a lower triangular matrix $\mathbf{B}$ as

$$\mathbf{H} = \mathbf{BF}. \tag{3.17}$$

Note that the QR decomposition for V-BLAST performs the Gram-Schmidt orthogonalization on the columns of the channel while that for ZF-DPC does the Gram-Schmidt procedure along the rows of the channel [49]. The difference between these two procedures is the position of unitary matrix $\mathbf{F}$ and lower triangular matrix $\mathbf{B}$. Figure 3.4 shows the general structure of ZF-

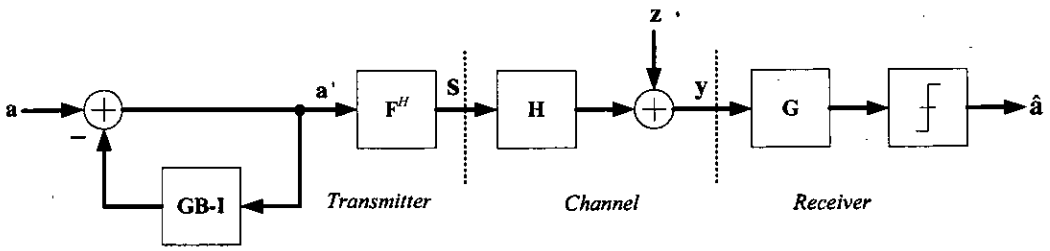


Figure 3.4: *The structure of ZF-DPC in MIMO channels*

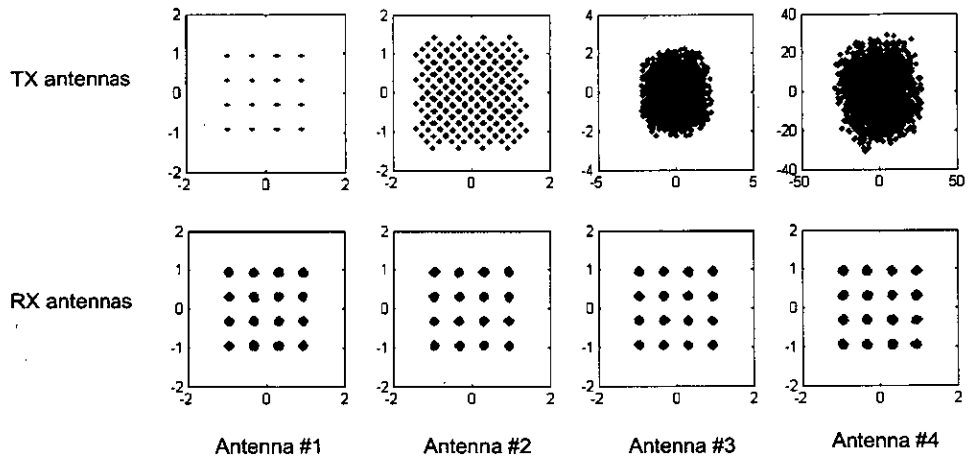


Figure 3.5: *Constellations of ZF-DPC at transmit and receive antennas, 16-QAM and SNR = 20dB*

DPC. We notice that the feedback filter in the DFE receiver is moved to the transmitter. Because of the lower triangular matrix $\mathbf{B}$, the first encoded data stream has no interference, the second encoded data stream experiences interference only from the first encoded data stream and so on. Any interference caused by data stream $j > i$ on each data stream i is forced to zero by the pre-subtraction at the transmitter. The output of the feedback filter at spatial channel k is given by

$$a'_k = a_k - \sum_{l=1}^{k-1} \frac{b_{kl} a_l}{b_{kk}}, \tag{3.18}$$

The second term in equation (3.18) is the pre-subtracted interference, which increases the transmit power of ZF-DPC. Because the transmitter knows the transmitted data symbol, no error propagation occurs unlike the receive technique such as DFE and V-BLAST.

Figure 3.5 shows the constellations of a 4×4 MIMO configuration with 16-QAM and SNR = 20dB, which can visualize the effect of pre-subtraction on the constellations. We notice that as

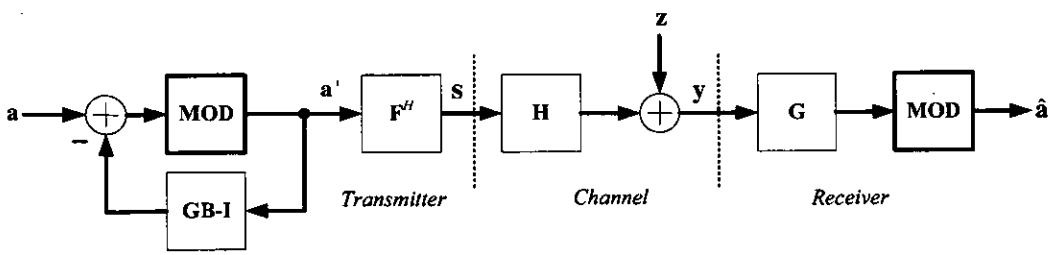


Figure 3.6: *The structure of THP in MIMO channels*

the antenna index increases, the transmit power increases due to the pre-subtraction. However, the transmit power increase of ZF-DPC is cancelled through the MIMO channel. The variance of the total transmit power of ZF-DPC is given by

$$\mathrm{Tr}\left(\mathrm{R_{ss}}\right) = \left(K + \sum_{k=1}^{K}\sum_{l=1}^{k-1}\frac{b_{kl}^2}{b_{kk}^2}\right)\sigma_a^2 \tag{3.19}$$

where the index k corresponds to the antenna number in Figure 3.5.

In order to make the variance of the total transmit power of ZF-DPC the same as the variance of V-BLAST $K\sigma_a^2$, the normalization factor should be used at the transmitter, which is of the form

$$g_{ZF-DPC} = \sqrt{\frac{K}{K + \sum_{k=1}^{K}\sum_{l=1}^{k-1}\frac{b_{kl}^2}{b_{kk}^2}}} \tag{3.20}$$

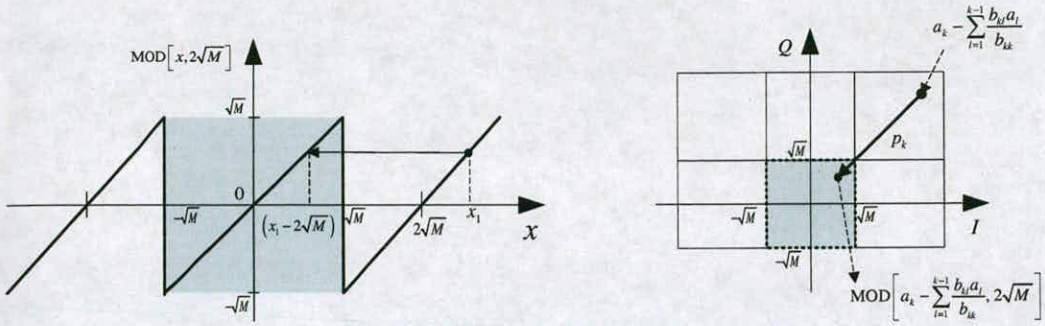
3.3.2.4 Tomlinson-Harashima Precoding

As with ZF-DPC, the output symbols of the feedback filter are successively generated from the input data symbol vector a. However, THP uses modulo operators MOD[·] to avoid a possibly large increase in transmit power of ZF-DPC [50]. Figure 3.6 shows the general structure of THP.

The output of the feedback filter at spatial channel k is given by

$$a_k' = a_k + p_k - \sum_{l=1}^{k-1}\frac{b_{kl}a_l}{b_{kk}}, \tag{3.21}$$

where p_k is an arbitrary symbol due to the modulo operation, which moves a_k' into a certain modulo boundary region $\mathfrak{A}$ [49]. See Figure 3.7(a). For example, when M-QAM is used for



(a) Modulo function *(b) Modulo operation for M-QAM*

Figure 3.7: *The modulo function and its operation for M-QAM*

a_k, the symbol p_k is an appropriate integer multiple of $2\sqrt{M}$, which means the constellation is bounded by the square region of width $2\sqrt{M}$. Figure 3.7 (b) illustrates how the modulo operator moves the the symbol p_k to be bounded by the square region of width $2\sqrt{M}$.

At the receive antennas, due to the symbol p_k, there are periodic replicas of the original constellation, which makes it necessary to use modulo operators at the receiver. Figure 3.8 shows the constellations of a 4×4 MIMO configuration with 16-QAM and SNR = 20dB, which can visualize the effect of the modulo operation on the constellations. The transmit signal from TX antenna 1 has no pre-subtracted interference. However, as the antenna index increases, the transmit symbols are uniformly distributed within the boundary region, unlike the case of ZF-DPC illustrated in Figure 3.5. Large antenna index implies much pre-subtracted interference. Due to the symbol p_k at the transmitter, the constellations at the receiver have a more expanded signal set. However, the modulo operator at the receiver removes the effect of the symbol p_k and recovers the original constellations.

Although THP prevents the large power increase of ZF-DPC, it still suffers from three kinds of losses: receive modulo loss, transmit power loss and shaping loss [51] [52]. The receive modulo loss is due to the existence of more neighbours at the edge of original constellations at the receiver. The transmit power loss is due to the extension of the original constellations up to the modulo boundary at the transmitter. The shaping loss is due to the cubic shape of the M-QAM constellation, which generates a capacity loss of 1.53dB from the Shannon capacity for spherical Gaussian distributed signals. If we assume the transmitted symbols of M-QAM THP are uniformly distributed over the boundary region of $2\sqrt{M}$, the power increase calculates

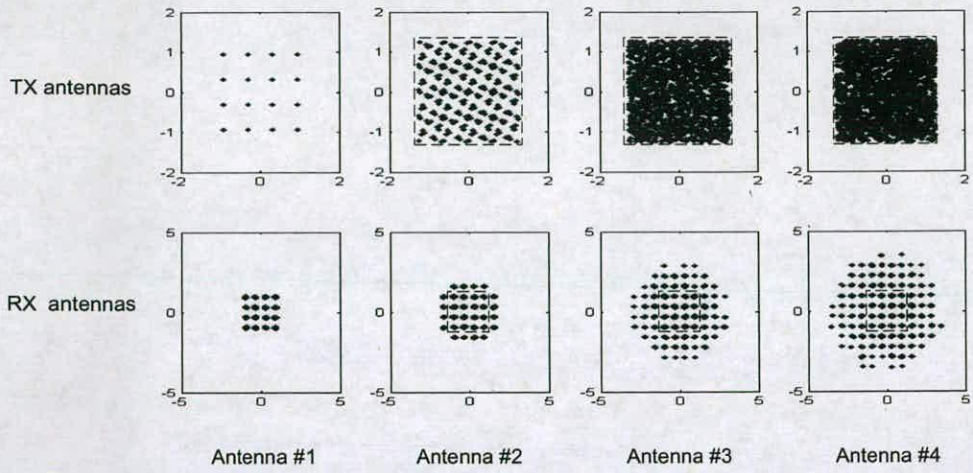
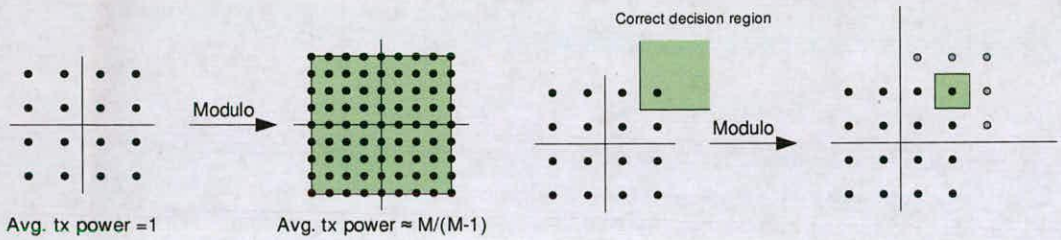


Figure 3.8: *Constellations of THP at transmit and receive antennas, 16-QAM and SNR = 20dB*



(a) The power loss at the transmitter due to THP

(b) The modulo loss at the receiver due to THP

Figure 3.9: *The transmit power loss and the receive modulo loss of THP*

to [53]

$$\Gamma_{THP} = \frac{M}{M-1} \tag{3.22}$$

Figure 3.9 illustrates the power loss at the transmitter and the modulo loss at the receiver. Note that when M increases, this loss becomes negligible. Indeed, if SNR is high, the effect of the receive modulo loss due to more neighbors at the receiver can be ignored [54]. Assuming M-QAM is used, the variance of the total transmit power of THP is given by

$$\mathrm{Tr}\left(\mathbf{R}_{ss}\right) = \sum_{k=1}^{K} \frac{M_k}{M_k - 1} \sigma_a^2, \tag{3.23}$$

where M_k is the modulation order for spatial channel k.

In order to make the variance of the total transmit power of THP the same as the variance of V-BLAST $K\sigma_a^2$, the normalization factor should be used at the transmitter, which is of the form

$$g_{THP} = \sqrt{\frac{K}{\sum_{k=1}^{K} \frac{M_k}{M_k - 1}}} \tag{3.24}$$

3.3.2.5 Vector Precoding

As discussed in Section 3.3.2.4, the shaping loss of 1.53dB exists when THP is used [55]. In order to minimize the shaping loss, the spherical shape of constellation with an optimal Gaussian code should be used. This can be performed by considering high-dimensional lattice codes and a vector quantization instead of complex symbols and the modulo operation respectively. For an excellent overview of linear Gaussian channels and the shaping loss, see [56].

Several vector precoding methods have been studied. In [57], a combined precoding and signal shaping algorithm have been proposed, which applies a vector signal shaping algorithm instead of the modulo operation. A trellis precoding technique has been presented in [51], which uses a vector quantization at the transmitter to minimize the shaping loss. In [58] and [59], a vector-perturbation technique has been presented to reduce the energy of transmit signal. Although these vector precoding techniques can achieve an additional shaping gain of 1.53dB compared to THP, these are difficult for the practical implementation. Indeed, because the objective of this thesis is to develop multiuser scheduling algorithms with practical considerations in MIMO broadcast channels, simple precoding techniques like THP are sufficient for the analysis of the developed scheduling algorithms. Therefore, no further discussion about vector precoding techniques will be presented.

3.4 Comparison of MIMO Transceiver Techniques

In this section, we compare ZF-DPC, THP, ZFBF and V-BLAST in terms of error rates in order to investigate the effect of practical impairments such as the modulo loss of THP, the transmit power constraint and channel estimation errors. This comparison will show that THP, despite its modulo loss, is an appropriate MIMO precoding technique for practical implementation of real MIMO BC wireless systems. Figure 3.10 illustrates a general MIMO transceiver for this comparison and Table 3.1 shows its configuration according to MIMO transceiver types.

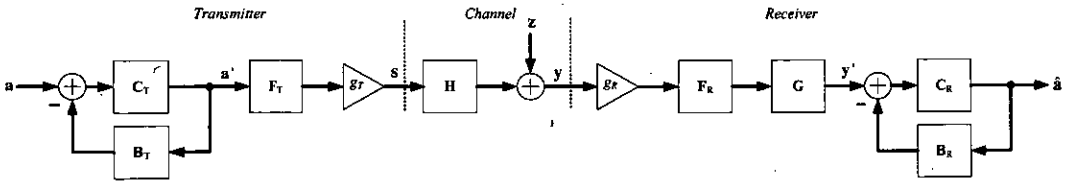


Figure 3.10: *A block diagram of the generalized MIMO transceiver*

	Block	ZF-DPC	THP	ZFBF	V-BLAST
TX	$\mathbf{C}_T$	$\mathbf{I}$	MOD[·]	$\mathbf{I}$	$\mathbf{I}$
	$\mathbf{B}_T$	$\mathbf{GB}-\mathbf{I}$	$\mathbf{GB}-\mathbf{I}$	$\mathbf{0}$	$\mathbf{0}$
	$\mathbf{F}_T$	$\mathbf{F}^H$	$\mathbf{F}^H$	$\mathbf{H}^+$	$\mathbf{I}$
	g_T	$\sqrt{\frac{K}{K+\sum_{k=1}^{K}\sum_{l=1}^{k-1}\frac{b_{kl}^2}{b_{kk}^2}}}$	$\sqrt{\frac{K}{\sum_{k=1}^{K}\frac{M_k}{M_k-1}}}$	$\sqrt{\frac{K}{\mathrm{Tr}(\mathbf{H}^+\mathbf{H}^{+H})}}$	1
RX	$\mathbf{C}_R$	$\mathbf{I}$	MOD[·]	Slicer	Slicer
	$\mathbf{B}_R$	$\mathbf{0}$	$\mathbf{0}$	$\mathbf{0}$	$\mathbf{GB}-\mathbf{I}$
	$\mathbf{F}_R$	$\mathbf{I}$	$\mathbf{I}$	$\mathbf{I}$	$\mathbf{F}$
	g_R	$\frac{1}{g_T}$	$\frac{1}{g_T}$	$\frac{1}{g_T}$	1

Table 3.1: *Configuration of the generalized MIMO transceiver*

In order to compare those MIMO transceiver techniques, several assumptions are made. First, we assume $K = 4$ and M-QAM signal sets with cardinality $M = 2^2, \cdots, 2^8$, which corresponds to integer rates ranging from 2 to 8 [bits/symbol]. Secondly, the total data rate (R_{TOTAL}) is fixed to 16 [bits/channel use]. In this case, the total data rate R_{TOTAL} is distributed to each spatial channel according to the channel gains, which is called 'loading', determined by the diagonal components of the lower triangular matrix $\mathbf{B}$. Because the noise variance of each spatial channel is identical for single user multi-antenna scenarios, the signal to noise ratio at the receiver is determined by the channel gain factor b_{kk}. Therefore, the data rate R_k for spatial channel k is determined by

$$R_k = \frac{R_{TOTAL}}{K} + \frac{1}{K} \cdot \log_2 \left(\frac{|b_{kk}|^{2K}}{\prod_{l=1}^{K} |b_{ll}|^2} \right). \tag{3.25}$$

These rates are quantized to integers from 2 to 8. In order to compensate for rate quantization errors and to make the same error rates for all spatial channels, a residual transmit power control is performed before transmission.

$$P_k = \frac{P \cdot 2^{R_{Qk}}}{\sum_{i=1}^{K} 2^{R_{Qi}}}, \tag{3.26}$$

where R_{Qk} denotes the quantized data rate of spatial channel k. This power control algorithm adjusts the power ratio of each spatial channel, whilst not increasing the total transmit power. The rate allocation algorithm of equation (3.25) and the residual power control algorithm of equation (3.26) are discussed in [60], so that it is recommended to see the reference for more detailed information. Thirdly, as discussed in Section 3.3.1.4, the decoding order of V-BLAST affects the system performance. Therefore, an optimal permutation matrix $\mathbf{P}_{opt}$ for the ordering is applied to improve the error rate performance. Similar to V-BLAST, the performance of MIMO precoding techniques is also affected by the ordering. The selection of the permutation matrix depends on performance criteria. The following permutation matrix is used in order to maximize the minimum spatial channel gain as in [61] because the smallest spatial channel gain dominates the overall error rate.

$$\mathbf{P}_{opt} = \arg\max_{\mathbf{P}} \min \left\{ |b_{11}|^2, ..., |b_{KK}|^2 \right\} \tag{3.27}$$

Note that the permutation matrix $\mathbf{P}_{opt}$ is applied when the QR decomposition of the channel matrix $\mathbf{H}$ is performed. For ZF-DPC and THP, the permuted channel matrix $\mathbf{HP}$ is decomposed as

$$\mathbf{HP}_{opt} = \mathbf{BF}. \tag{3.28}$$

For V-BLAST, the permutated channel matrix $\mathbf{HP}$ is decomposed as

$$\mathbf{HP}_{opt} = \mathbf{F}^H\mathbf{B}. \tag{3.29}$$

If the performance criterion is to maximize the capacity, the optimum permutation matrix can be given by [62]

$$\mathbf{P}_{opt} = \arg\max_{\mathbf{P}} \sum_{k=1}^{K} \frac{1}{|b_{kk}|^2}. \tag{3.30}$$

Lastly, we only focus on the ZF solution though there is some improvement by using the MMSE criterion at low SNR. For more information on the use of the MMSE criterion, refer to [54] and [63].

For the fair comparison, we use an effective SNR $\left(\sigma_s^2/\sigma_z^2\right)$, the signal variance with the transmit power constraint applied, divided by the noise variance, instead of SNR. This is because the transmit power and received SNR is different according to the normalization factor of the MIMO transceiver technique. For the receive V-BLAST technique, the effective SNR is iden-

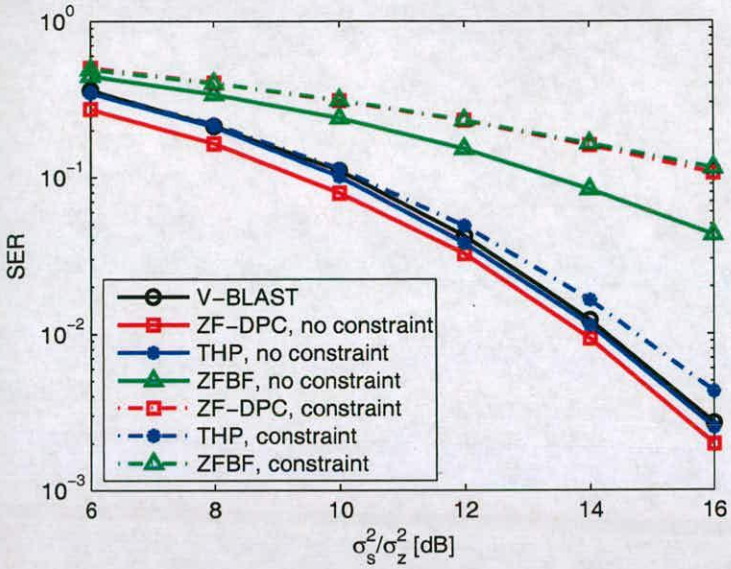


Figure 3.11: *The error rate performance of different MIMO transceiver techniques with fixed sum-rate loading and no channel estimation error* $K = 4$, $R_{TOTAL} = 16$, *and* $\sigma_e^2 = 0$

tical to SNR at the receiver. But the transmit power of precoding techniques consists of the transmit signal and the pre-subtracted interference, which reduces the overall SNR at the receiver.

3.4.1 Effect of Transmit Power Constraint

Figure 3.11 compares the average symbol error rates of different MIMO transceiver techniques with and without the transmit power constraint. When the transmit power constraint is applied, the transmit power of all MIMO transceiver techniques is set to be $K\sigma_a^2$. In the case of no transmit power constraint (label: no constraint), ZF-DPC performs best because it has neither modulo loss nor error propagation. Since the effect of the modulo loss in THP and that of the error propagation in V-BLAST is almost the same, the performance difference between THP and V-BLAST is negligible without the transmit power constraint. Linear ZFBF performs the worst among MIMO transceiver techniques. However, with the transmit power constraint (label: constraint), the performance of the precoding technique is degraded because the effective SNRs of all spatial channels at the receiver decrease due to the transmit power constraint compared to the V-BLAST technique. Since the power increase of ZF-DPC is much larger than that of THP, the performance of ZF-DPC is much worse than that of THP with the transmit

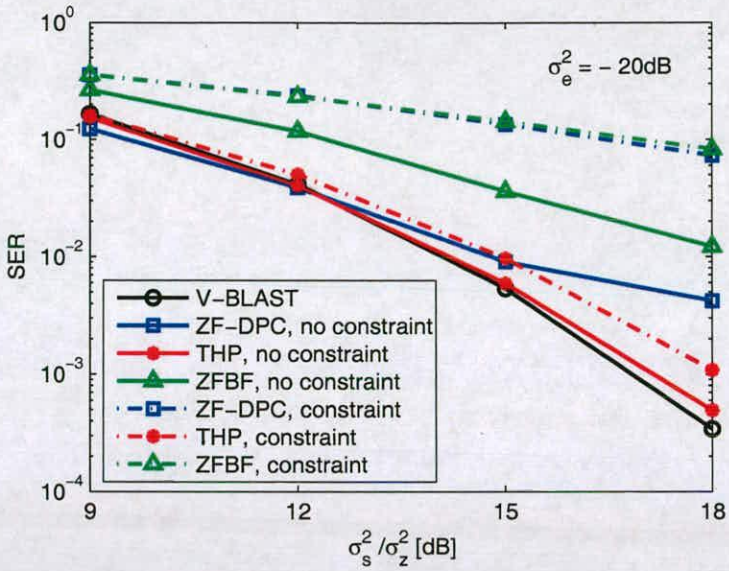


Figure 3.12: *The error rate performance of different MIMO transceiver techniques with channel estimation error $K = 4$, $R_{TOTAL} = 16$, and $\sigma_e^2 = -20dB$.*

power constraint. Although the modulo operation of THP prevents a possibly large increase in power, THP still has the modulo loss, which causes the decrease of effective SNR. Hence, the performance of THP is slightly worse than that of V-BLAST with the transmit power constraint.

3.4.2 Effect of Channel Estimation Errors

In Figure 3.12 and Figure 3.13, we compare the average symbol error rates of MIMO transceiver techniques with measurement noise variance $\sigma_e^2 = -20$dB and $\sigma_e^2 = -15$dB relative to the transmit power respectively. Because the pre-subtraction boosts the effect of channel measurement noise, the precoding technique is more sensitive to the effect of imperfect CSI than the V-BLAST technique regardless of the transmit power constraint. In contrast to the case of perfect CSI, THP performs slightly worse than V-BLAST with imperfect CSI without the transmit power constraint at high SNR. The amount of performance degradation of the precoding technique is proportional to both SNR and measurement noise variance. Between MIMO precoding techniques, ZF-DPC is the most sensitive to the effect of imperfect CSI because the total amount of spatial interference is proportional to transmit power in the additive measurement noise MIMO channel. The performance degradation of ZF-DPC with imperfect CSI is so severe that ZF-DPC is not likely to be appropriate for practical dirty paper coding.

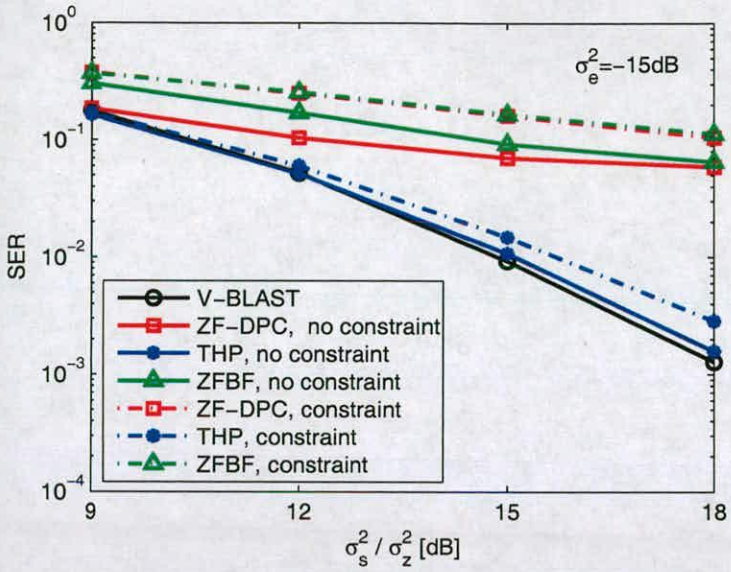


Figure 3.13: *Average SER performance of different MIMO transceiver techniques with channel estimation error $K = 4$, $R_{TOTAL} = 16$, and $\sigma_e^2 = -15dB$.*

3.5 Summary

This chapter has briefly introduced MIMO transceiver techniques. The performance of MIMO transceiver techniques such as ZF-DPC, THP, ZFBF and V-BLAST has been compared in terms of error rates. Nonlinear MIMO transceiver techniques such as ZF-DPC, THP and V-BLAST techniques have similar complexities and require the ordering of encoding or decoding for improving the performance based on the DFE structure [64].

THP performs slightly worse than V-BLAST with practical considerations such as the transmit power constraint and channel estimation errors due to the modulo loss. However, V-BLAST is not likely to be feasible for multiuser MIMO BC scenarios with single receive antenna for each user because it requires antenna cooperation for the successive detection at the receiver. Therefore, V-BLAST is not appropriate for the objective of this thesis. However, by comparing the error performance of V-BLAST with MIMO precoding techniques, the effect of the modulo loss in THP and the large power increase in ZF-DPC can be evaluated in terms of error rates. Between nonlinear MIMO precoding techniques, THP, despite its modulo loss, is much more feasible for the practical implementation than ZF-DPC in terms of the error rate performance because the large power increase of ZF-DPC makes its performance too sensitive to practical impairments. Nonlinear THP performs better than linear ZFBF regardless of practical consid-

erations in terms of error rates.

In the next chapter, we will analyse these MIMO precoding techniques, combined with several multiuser selection algorithms, in terms of the sum-rate capacity in an $i.i.d.$ MIMO broadcast channel.

Chapter 4
QoS-aware Sequential Multiuser Selection in MIMO Broadcast Channels

4.1 Introduction

In Chapter 3, MIMO precoding techniques have been introduced and compared in terms of error rates with practical considerations such as the transmit power constraint and channel estimation errors. This comparison has been performed in a single user MIMO channel.

This chapter will propose a new QoS-aware sequential multiuser selection algorithm. Indeed, it will compare MIMO precoding techniques in terms of the sum-rate capacity. MIMO receive techniques will not be considered from this chapter because they are not likely to be feasible for multiuser MIMO BC scenarios due to difficulty in receive antenna cooperation among users.

In real MIMO BC scenarios, each user may request different types of service such as the unsolicited grant service (UGS) for the T1/E1 application, real-time (RT) video, file transfer protocol (FTP) , email and hypertext transfer protocol (HTTP), each with different QoS requirements. For example, RT video is sensitive to delay but FTP is not sensitive to delay as much as RT video. For the T1/E1 application, a minimum data rate should always be guaranteed. If different QoS requirements are prioritized according to their service characteristics, T1/E1 (constant rate: CR) service may have the highest priority followed by RT video. HTTP (Best Effort: BE) may have the lowest priority among these services.

Two important issues related to QoS parameters are as follows [65]:

- How can a mixture of different QoS services, for example, such as RT and BE services be supported simultaneously with satisfying delay constraints of RT users whilst maximizing the sum-rate capacity?
- How can fairness among users with QoS differentiation be achieved whilst minimizing the degradation of the overall performance?

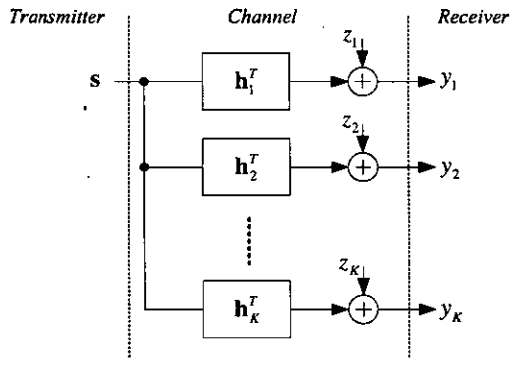


Figure 4.1: *The system model of a multiuser MIMO broadcast channel*

This chapter will propose a new QoS-aware sequential multiuser selection algorithm in order to answer the first question. In Chapter 6, by considering fairness among users in terms of throughput or delay, and applying antenna trade-offs between different QoS groups, the answer to the second question will be presented.

This chapter is organized as follows. Section 4.2 describes the system model used in this chapter. Section 4.3 explains the concept of the multiuser MIMO capacity and presents the sum-rate capacity of MIMO precoding techniques. Section 4.4 proposes a QoS-aware sequential multiuser selection algorithm. Numerical results are presented in Section 4.5. A summary for this chapter is presented in Section 4.6.

4.2 System Model

Consider an $i.i.d.$ block fading MIMO broadcast channel with M_T transmit antennas at a base station (BS) and $K(K \geq M_T)$ users each with a single receive antenna ($M_R = 1$), which is illustrated in Figure 4.1. Let $\mathbf{h}_k \in \mathbb{C}^{M_T \times 1}$ denote the channel between the transmit antenna array and the receive antenna for user k, whose elements are complex Gaussian random variables with zero mean and unit variance. Then the MIMO BC can be represented as

$$y_k = \mathbf{h}_k^T \mathbf{s} + z_k, \qquad k = 1, \cdots, K, \tag{4.1}$$

where $\mathbf{s} \in \mathbb{C}^{M_T \times 1}$ is the transmit signal vector with a covariance matrix $\mathbf{R}_{ss} = E[\mathbf{s}\mathbf{s}^H] = \sigma_s^2 \mathbf{I}$ and a power constraint $\mathrm{Tr}\,(\mathbf{R}_{ss}) \leq P$, the scalar y_k is the received signal for user k and z_k is the complex additive white Gaussian noise (AWGN) with zero mean and variance σ_z^2 for user

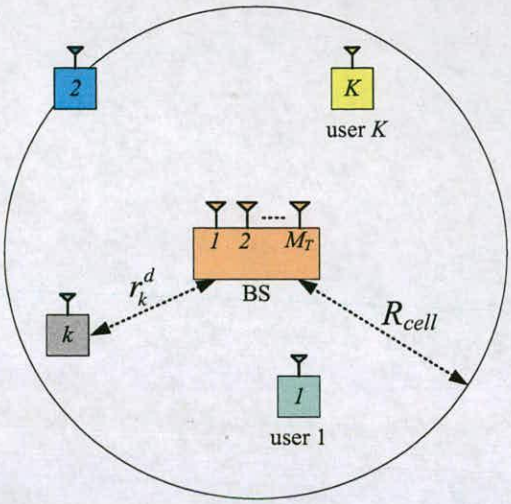


Figure 4.2: *The geometric model of the multiuser MIMO broadcast channel*

k. For simplicity, $\sigma_z^2 = 1$ in the remainder of this thesis. The transmit signals are assumed to experience path loss, log-normal shadowing and Rayleigh fading. In this case, the channel $\mathbf{h}_k$ can be represented as [66] [67]

$$\mathbf{h}_k = \sqrt{SNR_0 \left(\frac{r_k^d}{R_{cell}}\right)^{-\alpha_l}} \omega_k \cdot \mathbf{g}_k, \tag{4.2}$$

where SNR_0 denotes the median of signal-to noise-ratio (SNR) of all users averaged over fading, r_d^k denotes the distance between user k and the base station, R_{cell} denotes the cell radius, α_l denotes the path loss exponent, ω_k denotes a shadowing variable with standard deviation σ_s. The vector $\mathbf{g}_k = [g_{k1}\ g_{k2}\ \cdots\ g_{kM_T}]^T$ represents the Rayleigh-distributed fading between the transmit antenna array and user k, whose components are $i.i.d.$ zero mean circular symmetric complex Gaussian (ZMCSCG) random variables with unit variance. Figure 4.2 shows the geometric system model.

4.3 Multiuser MIMO Capacity

In Chapter 2, we derived the capacity of a single user MIMO channel

$$C = \max_{\mathrm{Tr}(\mathbf{R}_{ss})=P} \log_2 \left| \mathbf{I} + \frac{\mathbf{H}\mathbf{R}_{ss}\mathbf{H}^H}{N_0} \right|, \tag{4.3}$$

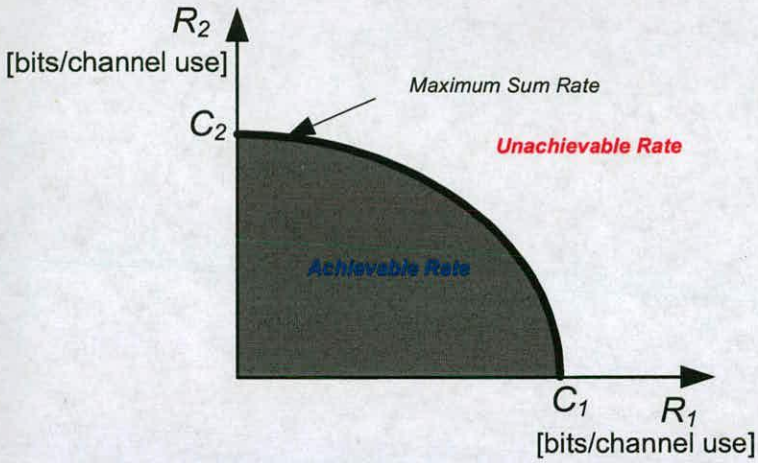


Figure 4.3: *The capacity region for a two user case*

However, the multiuser capacity is not defined by a single value. This is because there are infinitely many ways to share the transmit power among multiple users. In this case, all possible achievable rate vectors consist of a rate region. Mathematically, the union of all achievable rate vectors is called the capacity region of a multiuser system [11].

Figure 4.3 shows a simple example of the capacity region for a two user case. Let P, P_1 and P_2 be the total transmit power, the allocated power to the first user and the allocated power to the second user respectively ($P_1 + P_2 \leq P$). Rates of the first user (R_1 [bits/channel use]) and the second user (R_2 [bits/channel use]) are determined by P_1 and P_2. When $P_1 = P$, all transmit power is assigned to the first user and the corresponding capacity becomes C_1. When $P_1 + P_2 = P$, the maximum sum-rate capacity (solid line) is achieved. When $P_1 + P_2 < P$, the rate vector is located at a certain point within the achievable rate region. With the total power constraint ($P_1 + P_2 \leq P$), the system cannot achieve any sum-rate in the unachievable rate region. More transmit power is needed to expand the achievable rate region. For an overview of the capacity region, see [68].

There are two kinds of multiuser MIMO channels [68]: the MIMO multiple access channel (MIMO MAC) and the MIMO broadcast channel (MIMO BC), which are illustrated in Figure 4.4. In the MIMO MAC, a group of users transmit their data to one receiver via uplink channels. In this case, a multiple access scheme such as Time Division Multiple Access (TDMA), Code Division Multiple Access (CDMA) or Space Division Multiple Access (SDMA) is required to share the common wireless channel efficiently. In the MIMO BC, one transmitter sends its data

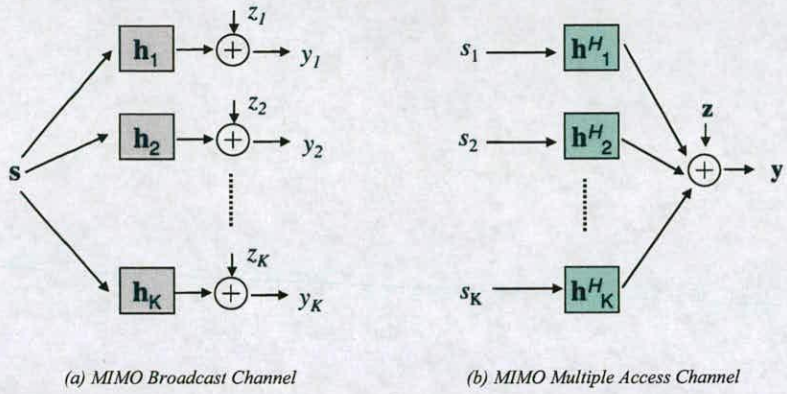


Figure 4.4: *System models of the MIMO BC and the MIMO MAC*

to many receivers via downlink channels. This means that all users have to share the bandwidth and transmit power by certain performance criteria. An introduction to the multiuser MIMO downlink channels is found in [69] and [70].

The capacity of the MIMO MAC has been derived in [71]. However, the capacity of the MIMO BC is difficult to find due to the lack of a general theory on nondegraded broadcast channels. It is said that the multiple antenna broadcast channel is nondegraded when users receive different strength signals from different transmit antennas, so that the users cannot be absolutely ranked by their channel strength [68]. See [35] for a detailed definition of degradedness. There is a duality between the capacity region of the MIMO BC and that of the MIMO MAC. For more detailed information on the duality, see [72]. However, the achievable rate region of the MIMO BC can be obtained by applying the DPC result at the transmitter [48] [73]. Based on the result of multiple receive antenna case in [74], denoting $\pi(k)$ as user k considering all permutations of the encoding order by π, an achievable rate of user $\pi(k)$ with a single receive antenna is given by

$$r\left(P_{\pi(k)}\right) = \log_2 \frac{\left(1 + \sum\limits_{j \geq k} P_{\pi(j)} \mathbf{h}_{\pi(k)}^H \mathbf{h}_{\pi(k)}\right)}{\left(1 + \sum\limits_{j > k} P_{\pi(j)} \mathbf{h}_{\pi(k)}^H \mathbf{h}_{\pi(k)}\right)} \tag{4.4}$$

where $P_{\pi(j)}$ is the input power allocated to user $\pi(j)$. With a transmit power constraint $\sum\limits_{k=1}^{K} P_k \leq P\,(P_k \geq 0)$, the dirty paper region $C_{DPC}\,(P, \mathbf{H})$ is determined by a set of all possible points in a K-dimensional vector space, which is mathematically defined as the convex hull of the union of all such rate vectors over all permutations. In mathematics, the convex hull for

a set of points X in a vector space is the minimal convex set containing X. An object is called convex if, for every pair of points within the object, every point on the straight line segment that joins them is also within the object. For example, a solid cube is convex, but anything that is hollow or has a dent in it is not convex.

$$C_{DPC}(P, \mathbf{H}) \equiv Co\left(\bigcup_{\pi, P_k} r(P_{\pi(k)})\right). \tag{4.5}$$

In this case, the maximum sum-rate capacity is the surface of the convex hull, which is the upper bound of the dirty paper region. Readers interested in the details of the the information-theoretic multiuser capacity are referred to [75], [76], [77] and [78].

In general, the number of users in real MIMO BC scenarios are likely to be greater than that of transmit antennas at the base station. This generates a multiuser selection problem, which selects a user set for transmission according to certain performance criteria. Because all users in a wireless channel are likely to have independent fading with rich scattering, any user selection algorithm should utilize this characteristics well. Obviously, if one user should be selected at a time in order to maximize the capacity, the optimal selection would be to select a user with the best channel condition. This exploits the multiuser diversity gain. The multiuser diversity is a form of selection diversity, which selects one user with the best channel condition at a certain time among multiple users, each with independent fading [79]. However, if the base station serves multiple users simultaneously by transmitting different data through different spatial channels similar to spatial multiplexing in a single user MIMO channel, it can increase the multiuser capacity compared to the case of using all transmit antennas for one user at a time [75]. If all users are orthogonal to each other, the optimal multiuser selection may be to select users with the largest individual spatial channel gains up to the number of transmit antennas. In reality, unfortunately, it is not likely that all users are orthogonal. This necessitates a joint consideration of multiple users in obtaining a user set for transmission. When nonlinear MIMO precoding techniques exploiting the DPC principle such as ZF-DPC [48] and THP [49] are used, any interference from other users can be successively cancelled by a certain encoding order provided perfect channel state information (CSI) is available at the transmitter. However, linear MIMO precoding techniques like zero forcing beamforming (ZFBF) cannot cancel the interference. They only use orthogonal channel directions by suppressing other users' interference. Therefore, the capacity of nonlinear MIMO precoding techniques is usually better than that of linear MIMO precoding techniques. However, as the number of users goes to infinity,

the performance of linear MIMO precoding techniques approaches that of DPC [46]. This is because the probability of finding orthogonal users among all users may increase as the number of users becomes very large [80].

When the number of users K is larger than the number of transmit antennas M_T in MIMO broadcast channels, a multiuser selection algorithm finds a user set S ($S \subset \{1, \cdots, K\}, |S| \leq M_T$), which satisfies certain performance criteria by considering all possible choices of the user set S. If the performance criterion is to maximize the sum-rate capacity, the maximum sum-rate capacity can be obtained by

$$C_{MAX} = \max_S C(S), \tag{4.6}$$

where $C(S)$ is determined by the type of MIMO precoding technique used. Equation (4.6) is called the *sum-rate maximization rule*. Throughout this thesis, this rule will be used despite its relatively large computational complexity compared to the techniques in [46], [80] and [81]. This is because the sum-rate maximization rule can be the most intuitive multiuser selection algorithm, which can be applied to any MIMO precoding technique with any QoS requirements. A complexity-reducing technique can be developed later based on the QoS-guaranteed multiuser scheduling algorithm proposed in this thesis. In Chapter 6, the weighted sum-rate maximization rule [46] will be used, where a weight vector representing a certain performance criterion is considered along with the achievable rate vector in selecting a user set.

If DPC is used for the MIMO precoding technique, from (4.5), $C(S) = C_{DPC}(P, \mathbf{H}(S))$, where $\mathbf{H}(S)$ is the MIMO channel matrix composed by the selected user set S. If ZF-DPC is used for the MIMO precoding technique, $C(S)$ is given by

$$C_{ZF-DPC}(S) = \max_S \sum_{k \in S} \log_2 \left(1 + b_{kk}^2 P_k\right), \tag{4.7}$$
$$\text{subject to } P_k \geq 0, \sum_{k \in S} P_k \leq P.$$

where b_{kk} denotes the kth diagonal element of the lower triangular matrix $\mathbf{B}$ obtained by the QR decomposition of the channel matrix formed by the selected user set and P_k denotes the transmit power allocated to user k. Denoting the modulo loss of user k due to THP as Γ_{THP}^k,

the sum-rate capacity of THP is

$$C_{THP}(S) = \max_{S} \sum_{k \in S} \log_2 \left(1 + \frac{b_{kk}^2 P_k}{\Gamma_{THP}^k}\right), \tag{4.8}$$
$$\text{subject to } P_k \geq 0, \sum_{k \in S} P_k \leq P.$$

If ZFBF is used, the sum-rate capacity is of the form [46]

$$C_{ZFBF}(S) = \max_{S} \sum_{k \in S} \log_2 \left(1 + \gamma_k \tilde{P}_k\right), \tag{4.9}$$
$$\text{subject to } \tilde{P}_k \geq 0, \sum_{k \in S} \tilde{P}_k \leq P,$$

where γ_k is the effective channel gain, which is given by

$$\gamma_k = \frac{1}{\left[\left(\mathbf{H}(S)\,\mathbf{H}(S)^H\right)^{-1}\right]_{k,k}}. \tag{4.10}$$

$\tilde{P}_k = \gamma_k^{-1} P_k$ is the transmit power allocated to user k.

Note that the optimal P_k of any MIMO precoding techniques can be found by the waterfilling algorithm [46] similar to the single user case in Chapter 2

$$P_k = (\mu \gamma_k - 1)^+, \tag{4.11}$$

where $(x)^+$ denotes $\max(x, 0)$ and the water level μ is selected to satisfy

$$\sum_{k \in S} \left(\mu - \frac{1}{\gamma_k}\right)^+ = P. \tag{4.12}$$

For simplicity of implementation in this thesis, several assumptions can be made. First, although the maximum sum-rate capacity of THP can be achieved by the optimal transmit power allocation [82], these optimal strategies are ignored for simplicity. Therefore, an equal power allocation over spatial channels can be assumed. Secondly, the number of elements in a user set S is assumed to be the same as the number of transmit antennas ($|S| = M_T$). Then, the

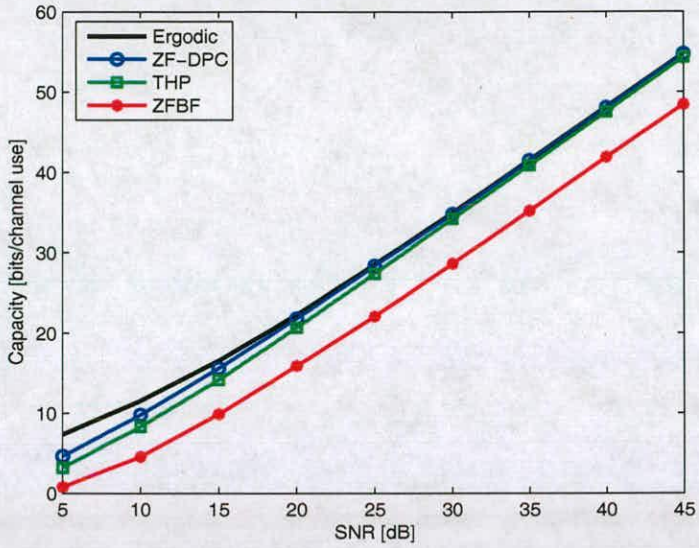


Figure 4.5: *Comparison of the sum-rate capacity of ZF-DPC, THP and ZFBF with the ergodic MIMO capacity,* $M_T = 4$ *and* $K = 4$

sum-rate capacity of ZF-DPC , THP and ZFBF reduces respectively to

$$C_{ZF-DPC}(S, M_T) = \sum_{k=1}^{M_T} \log_2 \left(1 + \frac{b_{kk}^2(t)P}{M_T}\right), \tag{4.13}$$

$$C_{THP}(S, M_T) = \sum_{k=1}^{M_T} \log_2 \left(1 + \frac{b_{kk}^2(t)P}{M_T \Gamma_{THP}^k}\right), \tag{4.14}$$

$$C_{ZFBF}(S, M_T) = \sum_{k=1}^{M_T} \log_2 \left(1 + \frac{\gamma_k P}{M_T}\right). \tag{4.15}$$

Figure 4.5 compares the sum-rate capacity of MIMO precoding techniques such as ZF-DPC, THP and ZFBF with the ergodic MIMO capacity in an $i.i.d.$ 4×4 MIMO channel. In order to account for the transmit power increase of THP due to the modulo loss in the equation (3.23), M-QAM is assumed to be used. We notice that as SNR increases, the modulo loss of THP can be ignored. This means that the sum-rate capacity of THP can be approximated as the sum-rate capacity of ZF-DPC with a high SNR assumption. Linear ZFBF performs more poorly than nonlinear MIMO precoding techniques such as ZF-DPC and THP. This is because nonlinear MIMO precoding techniques can cancel the interference from other users using the DPC principle. However, ZFBF only uses orthogonal channel directions, which reduces the

effective channel gains. This result is consistent with the result in [83].

There are several multiuser selection algorithms in addition to the sum-rate maximization rule. For example, the norm-based rule, originally proposed for antenna selection with single user MIMO systems (Refer to [84] and [85] for an overview of antenna selection), selects M_T users with the largest channel norm $\|\mathbf{h}_k\|^2$ in decreasing order, so that the sum of the selected individual channel norm is maximized. However, this does not consider multiple users jointly, so that the sum-rate capacity obtained by the norm-based rule is smaller than the sum-rate capacity by the sum-rate maximization rule as long as multiple users are not orthogonal. The round-robin (RR) is the simplest multiuser selection rule, which only gives equal transmission opportunity to all users without considering channel conditions, so that it cannot exploit the multiuser diversity gain.

4.4 QoS-aware Sequential Multiuser Selection Algorithm

This section proposes a new QoS-aware sequential multiuser selection algorithm, which can serve a mixture of different QoS users simultaneously whilst satisfying their QoS requirements as well as maximizing the sum-rate capacity in MIMO broadcast channels.

Multiple transmit and receive antennas can provide not only diversity gain for improving reliability [86] but multiplexing gain [7] for increasing data rate by exploiting the space dimension. In [87] and [88], the concept of spatial gain trade-off between multiplexing gain and diversity gain have been presented. This motivates the invention of a new antenna trade-off scheme between different QoS groups in MIMO broadcast channels. The higher priority group takes precedence in using multiple transmit antennas. The number of transmit antennas assigned to each QoS group can be determined by the wireless channel conditions and QoS status. The proposed QoS-aware multiuser selection algorithm exploits the concept of the antenna trade-off scheme to achieve QoS differentiation.

Unlike conventional opportunistic scheduling algorithms [2] [3] [4], which serve one user at a time, the proposed algorithm selects multiple users from different QoS groups sequentially and changes the number of transmit antennas assigned to each QoS group dynamically according to certain performance criteria. For example, let there be K_{RT} users in the RT user group with the delay constraint D_{TH} and K_{BE} users in the BE user group in a MIMO BC, where $K_{RT} + K_{BE} = K$. Further, let M_{RT} and M_{BE} be the number of pre-assigned transmit

antennas for RT and BE users respectively. Initially, a fixed number of transmit antennas is assumed to be assigned to each group, that is, M_{RT} for the RT user group and M_{BE} for the BE user group, where $M_{RT} + M_{BE} = M_T$.

The proposed QoS-aware sequential multiuser selection algorithm works as follows. When the average delay of the RT group is less than the delay threshold D_{TH}, the multiuser selection algorithm selects M_{RT} users from the RT user group, then it selects M_{BE} users from the BE user group. However, when the average delay of the RT user group is larger than the delay threshold D_{TH}, the multiuser selection algorithm selects M_T users from the RT user group and assigns all transmit antennas to the selected RT user set assuming $K_{RT} \geq M_T$. In this case, the BE user group has no chance of using its pre-assigned transmit antennas because the RT group occupies them to satisfy the delay constraint D_{TH}. This is called the *antenna trade-off scheme*, which changes the number of transmit antennas between QoS groups dynamically in order to satisfy QoS requirements whilst maximizing the sum-rate capacity.

The computational complexity decreases as the number of subgroups increases. However, this will reduce the sum-rate capacity due to the reduction in the user space for exploiting multiuser diversity in the sum-rate maximization rule. When the RT user group always use all transmit antennas, the number of all possible combinations of users for the sum-rate maximization rule with the sequential multiuser selection is always less than or equal to the case of no division

$$\binom{K_{RT}}{M_T} \leq \binom{K}{M_T}. \tag{4.16}$$

When the BE group always has a chance of using its pre-assigned transmit antennas, the number of all possible combinations for the sum-rate maximization rule with the sequential multiuser selection is also always less than or equal to the case of no division

$$\binom{K_{RT}}{M_{RT}} + \binom{K_{BE}}{M_{BE}} \leq \binom{K}{M_T}. \tag{4.17}$$

In real situations, the computational complexity when using the sum-rate maximization rule for the sequential multiuser selection depends on the multiuser MIMO configuration such as number of transmit antennas, number of QoS groups and users in these groups, QoS requirements, wireless channel conditions and so on. Figure 4.6 shows a simple example of the complexity reduction when user groups are equally divided into several equal-sized subgroups.

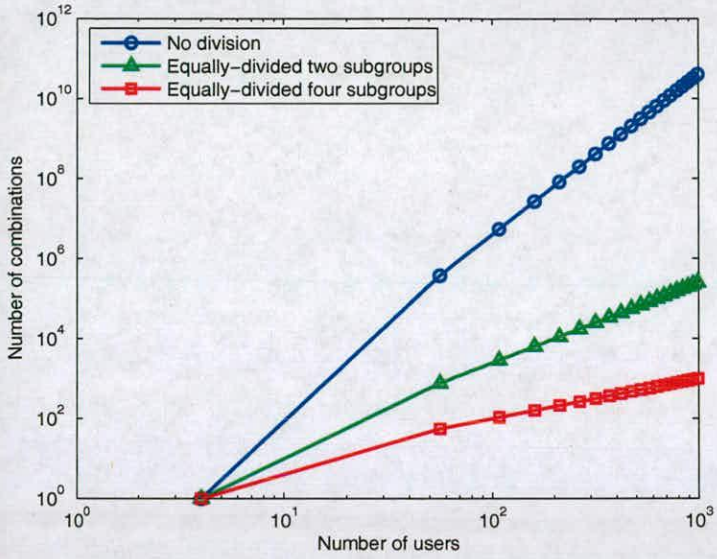


Figure 4.6: *Number of combinations for the sum-rate maximization rule with the sequential multiuser selection,* $M_T = 4$

4.4.1 Example I: Multiuser Selection with the Case of Real Time and Best Effort Users

Figure 4.7 shows the flow chart of the proposed QoS-aware sequential multiuser selection algorithm with the case of RT and BE users. The sum-rate maximization rule is assumed to be used. Assuming the base station selects a user set for transmission on a frame-by-frame basis, the proposed algorithm selects a user set for each frame according to the flow chart shown in Figure 4.7. In this case, the selected user set for each frame may be different from one another with the proposed algorithm. Let there be two QoS groups: the delay-constrained RT user set $U_{RT} := \{u_k \,|k = 1, \cdots, K_{RT}\}$ and the BE user set $U_{BE} := \{u_k \,|k = 1, \cdots, K_{BE}\}$. It is assumed that each user group has its priority: here RT users have higher priority than BE users. Let S_{RT} $(S_{RT} \subset U_{RT})$ and S_{BE} $(S_{BE} \subset U_{BE})$ be the selected RT user set and the selected BE user set respectively. The function $C(X, Y)$ used in this flow chart means the sum-rate capacity determined by the user set X and the number of transmit antennas Y based on (4.13), (4.14) or (4.15), which is determined by the type of the MIMO precoding technique used.

At first, the proposed algorithm compares the average HOL delay of RT users ($\bar{D}_{RT}$) with the delay threshold D_{TH}. If $\bar{D}_{RT} > D_{TH}$, all transmit antennas are assigned to the user set S_{RT} in order to improve the probability of supporting delay-guaranteed transmission ($|S_{RT}| = M_T$).

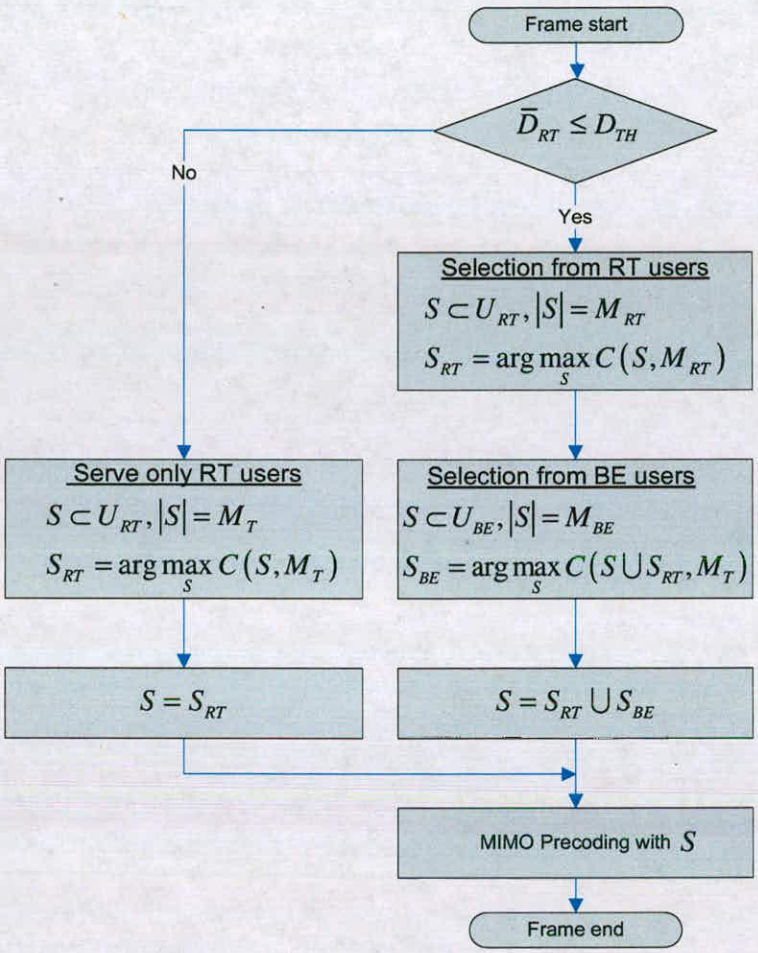


Figure 4.7: *The flow chart of the proposed QoS-aware sequential multiuser selection algorithm with the sum-rate maximization rule in the case of a mixture of RT and BE users based on (4.13), (4.14) or (4.15)*

In this case, there are $\binom{K_{RT}}{M_T}$ possible ways of selecting S_{RT} from U_{RT}. If the average HOL delay of RT users ($\bar{D}_{RT}$) does not exceed the delay threshold D_{TH}, the proposed algorithm finds user sets S_{RT} of M_{RT} users from U_{RT} and S_{BE} of M_{BE} users from U_{BE} sequentially. In this case, there are $\binom{K_{RT}}{M_{RT}} + \binom{K_{BE}}{M_{BE}}$ possible ways of selecting users because the 'already-selected' RT user set S_{RT} does not increase the combinations of possible user sets of S_{BE}. It only affects the selection of BE users. There are some restrictions in selecting BE users with the proposed algorithm if S_{RT} should be included in that procedure. Any BE user who has high cross correlation with users in S_{RT} might be excluded even if it has high individual channel gain. In other words, we need to find a BE user set S_{BE} which includes the effect of the previously selected RT user set S_{RT} is to select the 'most orthogonal' M_{BE} users with respect to the previously selected RT user set S_{RT} as well as BE users themselves. This restricts our choice of users and results in a further reduction of multiuser diversity gain. The performance degradation of the RT user set S_{RT} due to the addition of the BE user set S_{BE} may be negligible because users from these two sets are likely to be the most orthogonal users. Indeed, if the encoding order of nonlinear MIMO precoding techniques such as ZF-DPC and THP is such that the selected RT users are encoded first, the effect of BE users on the performance of the RT user set S_{RT} can be ignored due to interference cancellation in the DPC principle.

4.4.2 Example II: Multiuser Selection with the Case of Constant Rate, Real Time and Best Effort Users

Let there be three QoS groups in the MIMO BC: the rate-constrained CR user set $U_{CR} := \{u_k \,|k = 1, \cdots, K_{CR}\}$, the delay-constrained RT user set $U_{RT} := \{u_k \,|k = 1, \cdots, K_{RT}\}$ and the BE user set $U_{BE} := \{u_k \,|k = 1, \cdots, K_{BE}\}$, where K_{CR}, K_{RT} and K_{BE} are the number of CR, RT and BE users respectively and $K_{CR} + K_{RT} + K_{BE} = K$. It is assumed that each user group has its priority: CR users have the highest priority, RT the second highest and BE the lowest priority. Initially, each user group has its pre-assigned transmit antennas but there is a priority for their actual use in transmission. For example, when the throughput performance of CR users does not satisfy a certain rate constraint, RT users and BE users have no chance of using their pre-assigned transmit antennas for transmission because their priorities are lower than CR users. CR users occupy all transmit antennas in order to satisfy the rate constraint. When the throughput performance of CR users satisfies the rate constraint and the delay performance of RT users does not satisfy the delay constraint, only BE users have no chance of using their pre-assigned transmit antennas for transmission. In this case, CR and RT

users occupy all transmit antennas for transmission.

Let S_{CR} ($S_{CR} \subset U_{CR}$), S_{RT} ($S_{RT} \subset U_{RT}$) and S_{BE} ($S_{BE} \subset U_{BE}$) be the selected CR user set, the selected RT user set and the selected BE user set respectively and M_{CR}, M_{RT} and M_{BE} be the number of pre-assigned transmit antennas for CR, RT and BE users respectively and $M_{CR} + M_{RT} + M_{BE} = M_T$. Figure 4.8 shows the flow chart of the proposed QoS-aware sequential multiuser selection algorithm with the case of CR, RT and BE users. For simplicity, it is assumed that $K_{CR} \geq M_T$ and $K_{RT} \geq M_T$ in this flow chart. At first, the proposed algorithm selects S_{CR} using the sum-rate maximization rule from the CR group U_{CR}. There are $\binom{K_{CR}}{M_{CR}}$ possible ways of selecting S_{CR} from U_{CR} for the selection of CR users. If the sum-rate capacity of S_{CR} (R_{CR}) does not exceed the throughput threshold R_{TH}, all transmit antennas are assigned to CR users in order to improve the probability of supporting throughput-guaranteed transmission. Otherwise, the proposed algorithm compares the average HOL delay of RT users ($\bar{D}_{RT}$) with the delay threshold D_{TH}. If the average HOL delay of RT users ($\bar{D}_{RT}$) exceeds the delay threshold D_{TH}, the proposed algorithm finds a user set S_{RT} of $(M_T - M_{CR})$ users from RT users. In this case, all transmit antennas are assigned to CR and RT users. If the average HOL delay of RT users ($\bar{D}_{RT}$) does not exceed the delay threshold D_{TH}, the proposed algorithm finds user sets S_{RT} of M_{RT} users from U_{RT} and S_{BE} of M_{BE} users from U_{BE} sequentially.

4.5 Numerical Results

Throughout the simulations, it is assumed that the median SNR of all users $SNR_0 = 6$dB, the path loss exponent $\alpha_l = 4$, the cell radius $R_{cell} = 1$km and the shadow standard deviation $\sigma_s = 8$dB for channel parameters. ZF-DPC, THP and ZFBF are used for MIMO precoding techniques. The data arrivals in queues for RT and BE users are assumed to be independent Poisson processes with mean arrival rate $\lambda_d = 2$ [bits/channel use]. Similar to the case in Chapter 3, M-QAM signal sets with cardinality $M = 2^2, \cdots, 2^8$, which corresponds to integer rates ranging from 2 to 8 [bits/symbol/channel use] are used. It is also assumed that the received SNR of each user experiences saturation due to several practical impairments caused by antennas, radio frequency (RF) circuitry, analogue-to-digital converters (ADC), digital-to-analogue converters (DAC) and so on. We do not investigate the source of the saturation in this thesis but assume the received SNR is saturated at 40dB. In this case, 256-QAM will be sufficient to provide the maximum possible spectral efficiency for any user with very large channel gain.

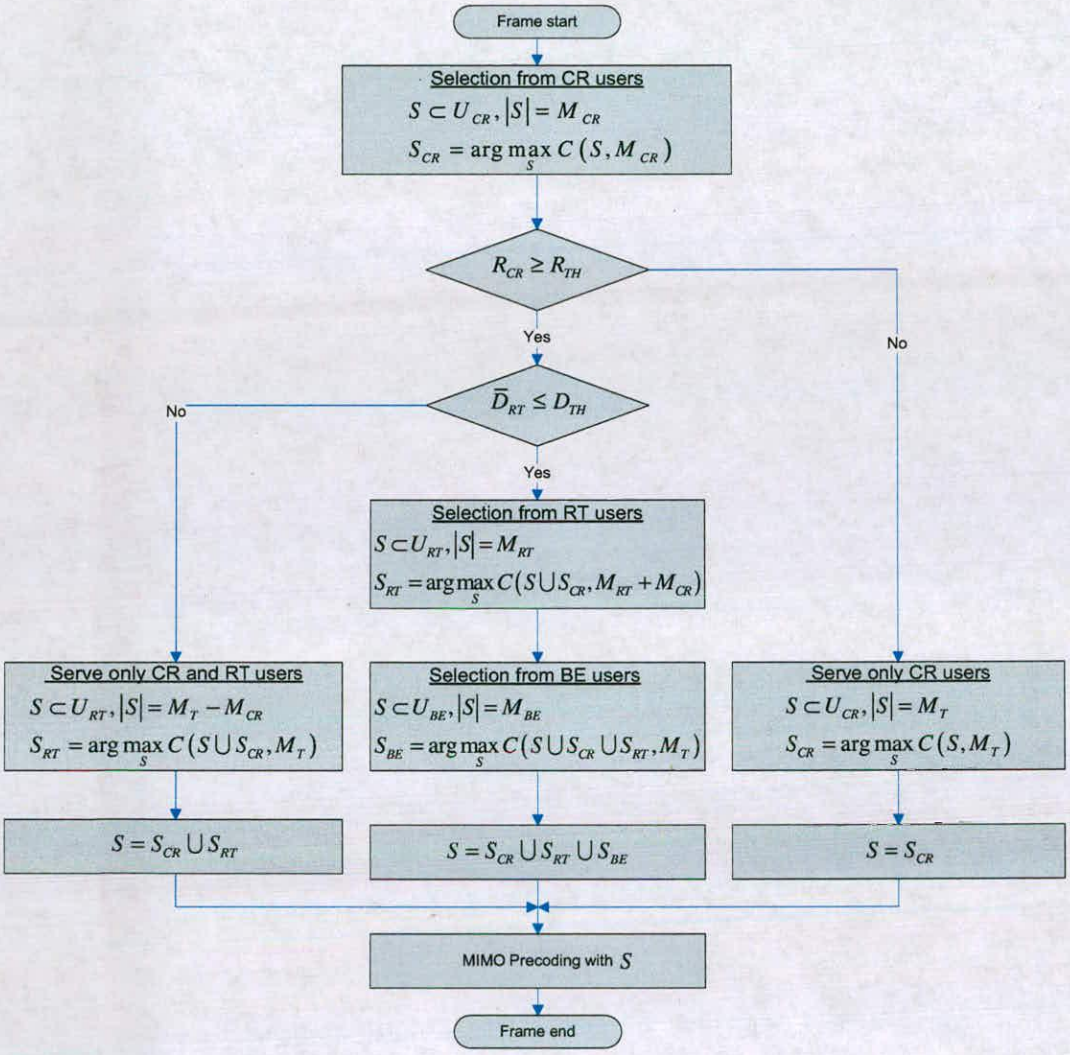


Figure 4.8: *The flow chart of the proposed QoS-aware sequential multiuser selection algorithm with CR, RT and BE users based on (4.13), (4.14) or (4.15)*

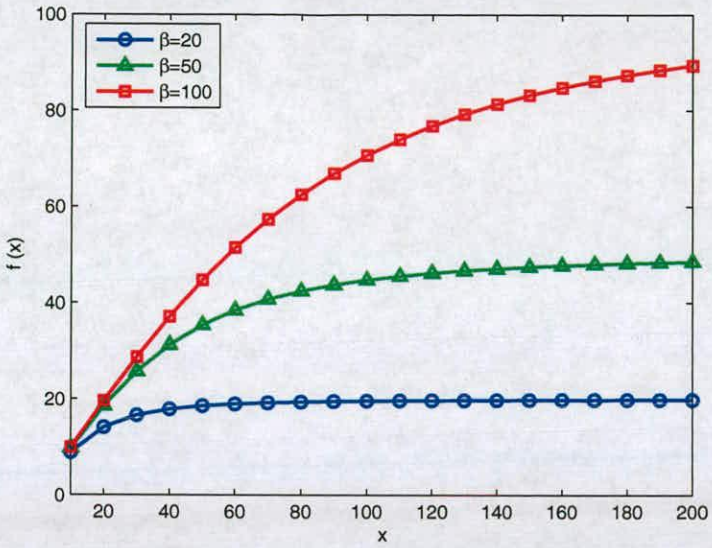


Figure 4.9: *The operation of the saturation function.*

A simple saturation function can be given by

$$f(x) = \sqrt{\frac{x^2}{1+\beta^{-2}x^2}}, \tag{4.18}$$

where β is a desired saturation value, and x is the linear signal-to-noise ratio determined by the MIMO channel matrix. Figure 4.9 shows the operation of the saturation function defined as (4.18). We notice that the input signal is saturated nearly from the value of β. However, it is useful to note that the saturation function used in this thesis may not describe the real saturation effect exactly. Further work should be performed to make a real model for the saturation effect.

4.5.1 Sum-rate Capacity in MIMO Broadcast Channels

THP and ZFBF are known to be the most practically feasible MIMO precoding techniques. THP is a nonlinear MIMO precoding technique using the DPC principle and ZFBF is a linear MIMO precoding technique applying the inverse channel matrix at the transmitter. Therefore, these two techniques are compared in this section.

Figure 4.10 shows the sum-rate capacity of THP and ZFBF with $M_T = 4$. We notice that user selection algorithms affect the sum-rate capacity. The sum-rate maximization rule shows the best performance regardless of the type of MIMO precoding techniques in terms of the sum-rate

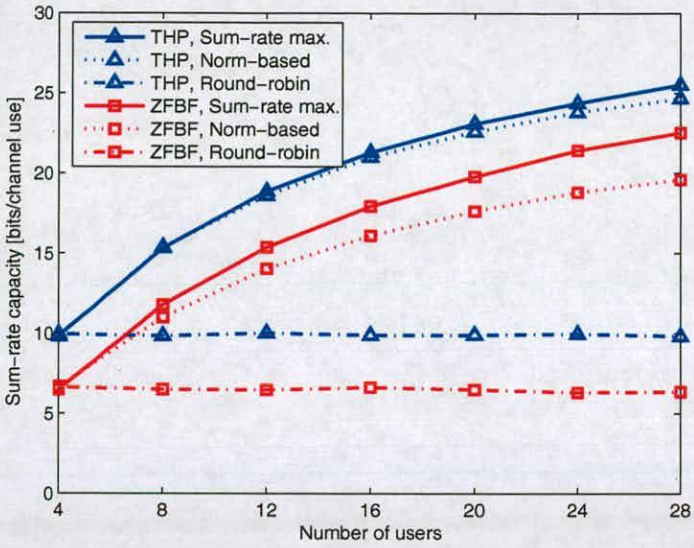


Figure 4.10: *The sum-rate capacity of THP and ZFBF with different user selection criteria,* $M_T = 4$, $SNR_0 = 6dB$, $\alpha_l = 4$, $R_{cell} = 1km$, $\sigma_s = 8dB$.

capacity because it optimizes the user set choice for that parameter. However, it requires much more computation than other selection rules. Because the norm-based rule does not consider interference among users, it has worse performance than the sum-rate maximization rule. The round-robin rule shows the poorest performance, which is unchanged regardless of the number of users. This is because the round-robin rule does not exploit the multiuser diversity when it selects a user set for transmission. The amount of degradation of ZFBF when using the norm-based rule is larger than that of THP. This implies that THP is more robust to the type of user selection criterion used than ZFBF. Similar comparisons of MIMO precoding techniques can be found in [46], [81] and [89], which have similar results as presented in this thesis.

4.5.2 QoS-aware Sequential Multiuser Selection for Real Time and Best Effort Users

In this section, we apply the proposed QoS-aware sequential multiuser selection algorithm to the case of supporting RT and BE users simultaneously with a delay constraint.

Due to the nature of random fading wireless channels, "QoS-guarantee" in wireless channels means statistical QoS, not deterministic QoS [90]. The deterministic QoS does not allow any QoS violations at all, so that it is very difficult to satisfy such a stringent QoS requirement. For-

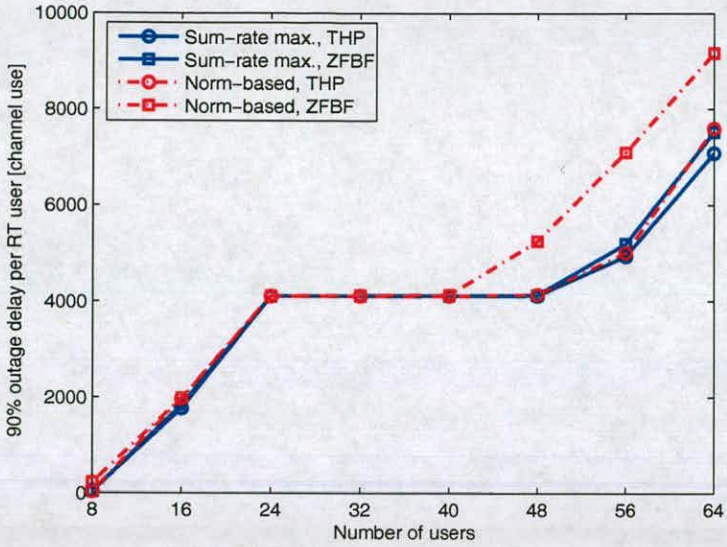


Figure 4.11: *The delay performance of THP and ZFBF with different user selection criteria,* $M_T = 4$, $SNR_0 = 6dB$, $\alpha_l = 4$, $R_{cell} = 1km$, $\sigma_s = 8dB$, $M_{RT} = 1$, $M_{BE} = 3$, $D_{TH} = 4096$ *and* $K_{RT} = 0.25K$.

tunately, in most real wireless systems, deterministic QoS conditions are not usually required. They can tolerate a small amount of instantaneous QoS violations (*i.e.*, by buffering), so that statistically-defined QoS can be satisfied. For example, a guaranteed-delay in wireless systems means that the average delay is within a target delay threshold with a certain outage probability. Therefore, for the delay performance, we assume 90% outage delay, where the head-of-line (HOL) delay of each user's queue is guaranteed for 90% of the channel realizations, so that 10% of delay values can exceed the delay threshold D_{TH}. Twenty five percent of users are assumed to be RT users ($K_{RT} = 0.25K$ and $K_{BE} = 0.75K$).

Figure 4.11 shows the delay performance of THP and ZFBF with different user selection criteria when using the QoS-aware sequential multiuser selection algorithm with $M_T = 4$, $M_{RT} = 1$ and $D_{TH} = 4096$ [channel uses]. We notice that the sum-rate maximization rule has better performance than the norm-based rule in terms of delay regardless of the MIMO precoding technique used. With the sum-rate maximization rule, the number of users satisfying the delay constraint of the RT users ranges up to about 48 users. If the norm-based rule is used, the delay performance of both MIMO precoding techniques are worse than the case of the sum-rate maximization rule. The amount of degradation of ZFBF with the norm-based rule is larger than that of THP with the norm-based rule because THP can actively remove mutual interference

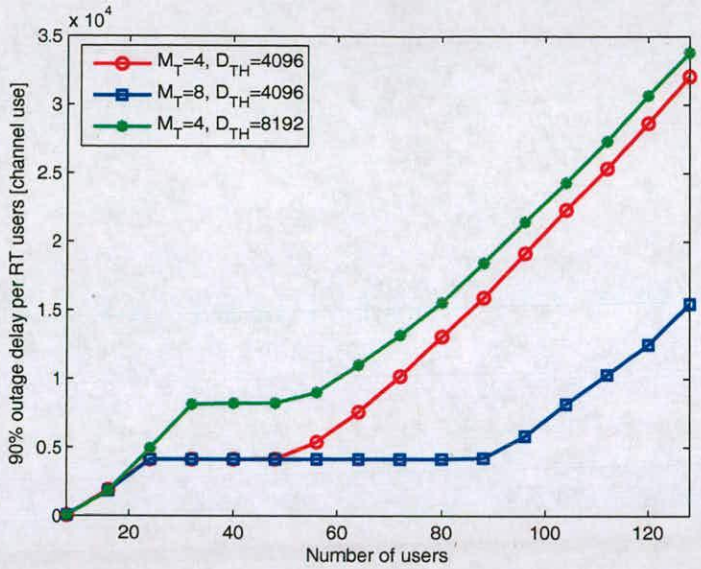


Figure 4.12: *The delay performance for different number of transmit antennas and delay thresholds with THP and the norm-based rule,* $SNR_0 = 6dB$, $\alpha_l = 4$, $R_{cell} = 1km$, $\sigma_s = 8dB$, $M_{RT} = 1$, $M_{BE} = 3$ *and* $K_{RT} = 0.25K$.

between users by the pre-subtraction process.

Figure 4.12 shows the delay performance of different number of transmit antennas and delay thresholds with the proposed QoS-aware sequential multiuser selection algorithm. THP and the norm-based rule are used for this plot. The proposed algorithm can satisfy any delay threshold D_{TH} for a certain range of the number of users by changing transmit antennas dynamically to satisfy the delay constraint. When $M_T = 8$, the proposed algorithm can support about two times as many users as the case of $M_T = 4$ whilst maintaining the HOL delay of the RT user near $D_{TH} = 4096$ [channel uses].

Figure 4.13 shows the throughput performance of RT and BE users when using the QoS-aware sequential multiuser selection algorithm. We notice that the total throughput R_{TOTAL} decreases when there is the delay constraint D_{TH} applied at the transmitter. As the number of users increases, the throughput of RT users (R_{RT}) increases in order to keep the delay within the delay threshold D_{TH}. In this case, the throughput of BE users (R_{BE}) decreases because the RT users take all the transmit antennas more frequently to guarantee their delay performance, which results in less chance of transmission for BE users.

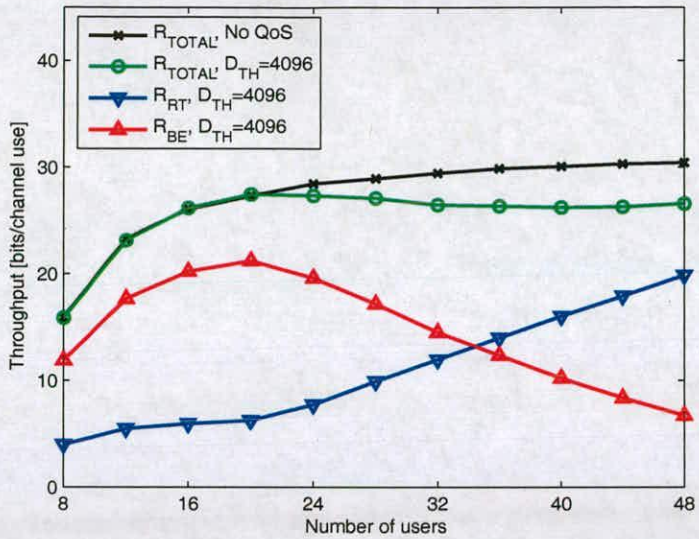


Figure 4.13: *The throughput performance of RT and BE users,* $M_T = 4$, $SNR_0 = 6dB$, $\alpha_l = 4$, $R_{cell} = 1km$, $\sigma_s = 8dB$, $M_{RT} = 1$, $M_{BE} = 3$, $D_{TH} = 4096$ *and* $K_{RT} = 0.25K$.

4.5.3 QoS-aware Sequential Multiuser Selection for Constant Rate, Real Time and Best Effort Users

In this section, we apply the proposed QoS-aware sequential multiuser selection algorithm to the case of supporting CR, RT and BE users simultaneously with a delay constraint. Unlike Section 4.5.2, the saturation effect and M-QAM modulation are not considered in this section. Instead, ZF-DPC is assumed to be used. It is also assumed that $K_{CR} = 2$, $M_{CR} = 2$ (*i.e.*, no CR user selection and the number of CR users is fixed), $M_{RT} = 1$, $M_{BE} = 1$ and $D_{TH} = 1024$ [channel uses].

Figure 4.14, Figure 4.15 and Figure 4.16 show the throughput of CR, RT and BE users with different throughput threshold values (R_{TH}) against the number of non-CR users.

When there are no QoS constraints (label: no QoS) such as R_{TH} and D_{TH}, that is, all CR, RT and BE users have transmission opportunities through their pre-assigned transmit antennas, the throughput is maximized as presented in Figure 4.14. However, when $D_{TH} = 1024$ [channel use] and $R_{TH} = 4$ [bits/channel use], the throughput of BE users begins to decrease over about 15 non-CR users as presented in Figure 4.15. This is because few transmission opportunities are given to the lowest priority BE user group in order to support higher priority CR and RT

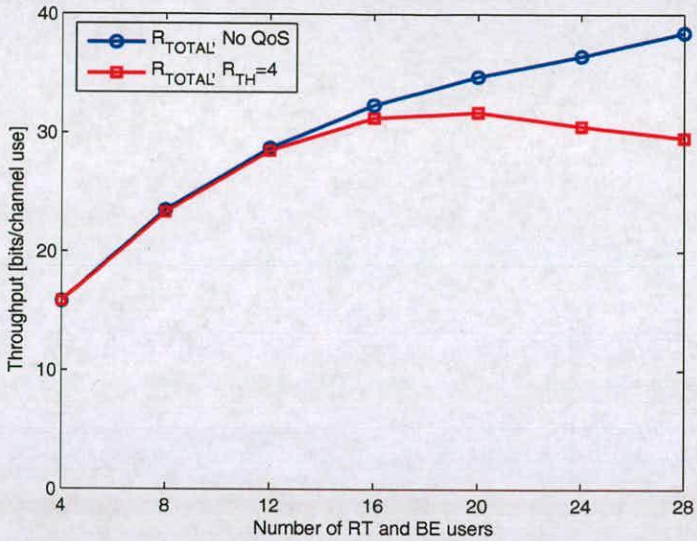


Figure 4.14: *Throughput performance of CR, RT and BE users with different throughput threshold values,* $M_T = 4$, $SNR_0 = 6dB$, $\alpha_l = 4$, $R_{cell} = 1km$, $\sigma_s = 8dB$, $K_{CR} = 2$, $M_{CR} = 2$, $M_{RT} = 1$, $D_{TH} = 1024$, $M_{BE} = 1$, $\lambda_d = 3$ *and* $K_{RT} = 0.25K$.

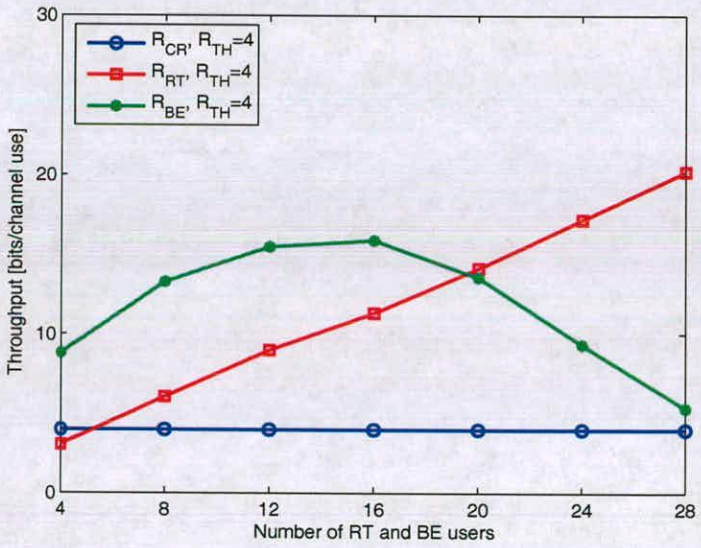


Figure 4.15: *Throughput performance of CR, RT and BE users with different throughput threshold values,* $M_T = 4$, $SNR_0 = 6dB$, $\alpha_l = 4$, $R_{cell} = 1km$, $\sigma_s = 8dB$, $K_{CR} = 2$, $M_{CR} = 2$, $M_{RT} = 1$, $D_{TH} = 1024$, $M_{BE} = 1$, $\lambda_d = 3$ *and* $K_{RT} = 0.25K$.

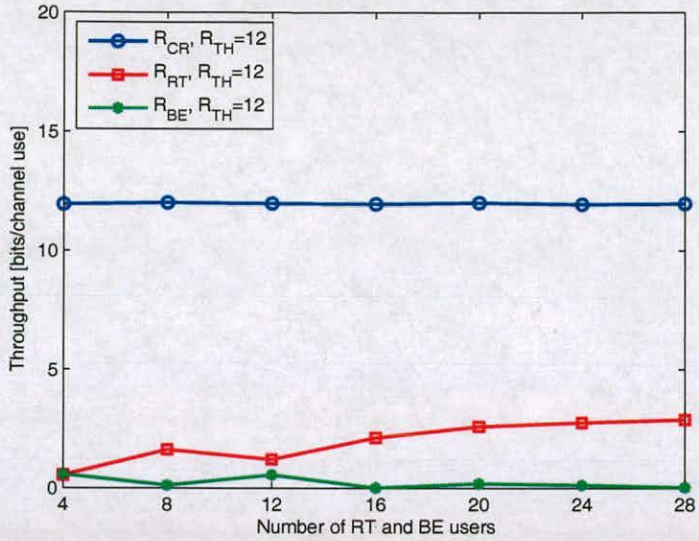


Figure 4.16: *Throughput performance of CR, RT and BE users with different throughput threshold values,* $M_T = 4$, $SNR_0 = 6dB$, $\alpha_l = 4$, $R_{cell} = 1km$, $\sigma_s = 8dB$, $K_{CR} = 2$, $M_{CR} = 2$, $M_{RT} = 1$, $D_{TH} = 1024$, $M_{BE} = 1$, $\lambda_d = 3$ *and* $K_{RT} = 0.25K$.

user group. When $R_{TH} = 12$ [bits/channel use], the throughput of non-CR users is very small because the algorithm usually assigns all transmit antennas to the CR user group to satisfy the guaranteed throughput as presented in Figure 4.16. For rate-constrained CR users, the rate constraints within the achievable rate region is always satisfied regardless of the number of RT and BE users, who have lower priorities than CR users. In this case, the throughput performance of RT users, not to mention of BE users, is worse than the case of no QoS constraints, and the throughput of CR users becomes the total throughput.

Figure 4.17 shows the average HOL delay of RT users with different throughput threshold values (R_{TH}) against number of the non-CR users. We notice that when $R_{TH} = 4$ [bits/channel use], the delay threshold $D_{TH} = 1024$ [channel uses] is guaranteed up to about 32 non-CR users. However, when $R_{TH} = 16$ [bits/channel use], the delay constraint of RT users cannot be guaranteed regardless of the number of non-CR users because the highest priority CR users take most of the transmission opportunities to attain the throughput threshold.

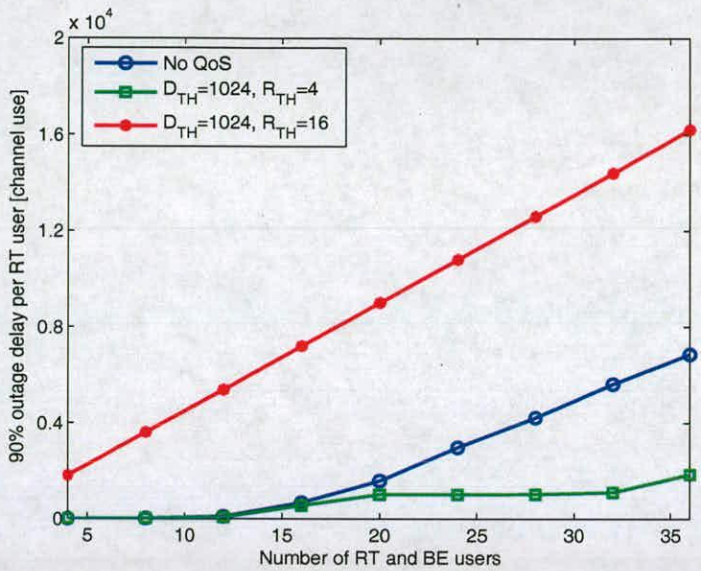


Figure 4.17: *Delay performance of RT users with different throughput threshold* R_{TH}, $M_T = 4$, $SNR_0 = 6dB$, $\alpha_l = 4$, $R_{cell} = 1km$, $\sigma_s = 8dB$, $K_{CR} = 2$, $M_{CR} = 2$, $M_{RT} = 1$, $D_{TH} = 1024$, $M_{BE} = 1$, $\lambda_d = 3$ *and* $K_{RT} = 0.25K$.

4.6 Summary

This chapter has proposed a new QoS-aware sequential multiuser selection algorithm for MIMO broadcast channels with different types of QoS users. The concept of multiuser MIMO capacity has been briefly introduced and sum-rate capacities of several MIMO precoding techniques such as ZF-DPC, THP and ZFBF have been compared.

The proposed QoS-aware sequential multiuser selection algorithm can support a mixture of QoS users by selecting users sequentially and applying MIMO precoding techniques. It can also improve the probability of supporting QoS-guaranteed transmissions using the antenna trade-off scheme. This will be investigated in more detail in Chapter 6. Furthermore, it can maximize the sum-rate capacity with QoS differentiation by using the sum-rate maximization rule in selecting users from each QoS group.

Among MIMO precoding techniques, nonlinear THP performs better than linear ZFBF in terms of the sum-rate capacity despite the modulo loss. It has been shown that the modulo loss can be ignored at high SNRs compared to ZF-DPC. This justifies that THP is an appropriate MIMO precoding technique for real MIMO BC scenarios in terms of the sum-rate capacity, along with the results in Chapter 3, where THP was compared with other MIMO transceiver techniques

in terms of error rates with practical considerations. In Chapter 5 and Chapter 6, we focus on multiuser scheduling design using THP as an example.

Before developing a QoS-guaranteed multiuser scheduling algorithm exploiting the proposed QoS-aware sequential multiuser selection algorithm in Chapter 6, a simple MIMO broadcast channel model will be proposed and compared with the MIMO channel measurements in the next chapter. Based on the proposed channel model, a new multiuser scheduling algorithm , which can improve the throughput performance by exploiting statistical channel state information in selecting a user set for transmission, will be proposed.

Chapter 5
Statistical CSI-assisted Multiuser Scheduling in Temporally-correlated MIMO Broadcast Channels

5.1 Introduction

The performance of MIMO precoding techniques have been investigated in terms of error rates in a single user MIMO channel in Chapter 3 and in terms of the sum-rate capacity in a multiuser MIMO broadcast channel in Chapter 4. However, this analysis assumed *i.i.d.* block fading MIMO channels, which is not likely to match real propagation environments.

This chapter will propose a temporally correlated MIMO broadcast channel model. There are a lot of factors that should be considered for the modelling of real MIMO broadcast channels but it is not only difficult but too complex to include all these factors such as spatial correlation, temporal correlation, polarization, keyhole, delay spread, shadowing and path loss in a realistic simulation model. Some parameters may have little effect despite the complexity to model them accurately, so that the proposed channel model is simple enough to be used in any system level simulator with ease but realistic enough to describe the key phenomena of temporally correlated MIMO broadcast channels. In order to validate the proposed channel model, a comparison with the MIMO channel measurements will be performed.

Based on the proposed channel model, a new SCSI-assisted multiuser scheduling algorithm will be suggested. The SCSI-assisted multiuser scheduling algorithm can minimize the effect of the temporal correlation. The proposed algorithm utilizes SCSI for the selection of a user set, which can minimize the mismatch of channel estimates in order to improve the sum-rate capacity.

This chapter is organized as follows. Section 5.2 proposes a temporally correlated MIMO broadcast channel model. Section 5.3 briefly describes the MIMO channel measurements. Section 5.4 compares the characteristics of the proposed channel model and the MIMO channel

measurements. Section 5.5 proposes a new SCSI-assisted multiuser scheduling algorithm. Numerical results for the proposed SCSI-assisted multiuser scheduling algorithm are presented in 5.6. A summary for this chapter is presented in Section 5.7.

5.2 Simulation Model for Real MIMO Broadcast Channels

In reality, MIMO channels are correlated both in space due to insufficient spacing between multiple antenna elements and in time due to Doppler spread effects. For a multiuser scheduler in a MIMO broadcast channel, the temporal correlation is more crucial than the spatial correlation. The reason is as follows: If any MIMO precoding technique is used in the MIMO BC, CSI at the transmitter is essential for transmit processing. It is likely that there is a time difference between obtaining CSI and using that CSI at the transmitter. Because the channel is time-varying due to the Doppler spread, there always exist mismatches of the channel estimate due to that difference. This degrades the performance of any MIMO precoding techniques. However, if the multiuser scheduler exploits the statistical CSI (SCSI) such as mean and covariance from the past instantaneous CSI (ICSI), it can allocate users to different time slots more efficiently when considering the temporal correlation, which can minimize the performance degradation due to imperfect CSI.

Consider a temporally correlated MIMO BC with M_T transmit antennas at a base station (BS) and $K(K \geq M_T)$ users each with a single receive antenna ($M_R = 1$). Let $\mathbf{h}_k(t) \in \mathbb{C}^{M_T \times 1}$ denote the channel at time t between the transmit antenna array and the receive antenna for user k. Then the MIMO BC at time t can be represented as

$$y_k(t) = \mathbf{h}_k^T(t)\mathbf{s}(t) + z_k(t), \qquad k = 1, \cdots, K, \tag{5.1}$$

where $\mathbf{s}(t) \in \mathbb{C}^{M_T \times 1}$ is the transmit signal vector with a covariance matrix $\mathbf{R}_{ss} = E[\mathbf{s}(t)\mathbf{s}(t)^H] = \sigma_s^2\mathbf{I}$ and a transmit power constraint $\mathrm{Tr}\,(\mathbf{R}_{ss}) \leq P$, the scalar $y_k(t)$ is the received signal for user k and $z_k(t)$ is the complex additive white Gaussian noise (AWGN) with zero mean and unit variance for user k at time t.

The transmit signals are assumed to experience path loss, log-normal shadowing and Rayleigh

fading. In this case, the channel $\mathbf{h}_k(t)$ can be represented as [66] [67]

$$\mathbf{h}_k(t) = \sqrt{SNR_0 \left(\frac{r_k^d(t)}{R_{cell}}\right)^{-\alpha_l}}\, \omega_k(t) \cdot \mathbf{g}_k(t), \tag{5.2}$$

where SNR_0 denotes the median SNR of all users averaged over fading, $r_k^d(t)$ denotes the distance between user k and the base station, R_{cell} denotes the cell radius, α_l denotes the path loss exponent and $\omega_k(t)$ denotes a shadowing of user k at time t. The vector $\mathbf{g}_k(t) = [g_{k1}(t)\ g_{k2}(t)\ \cdots\ g_{kM_T}(t)]^T$ represents the Rayleigh-distributed fading between the transmit antenna array and user k at time t, whose components are $i.i.d.$ zero mean circular symmetric complex Gaussian (ZMCSCG) random variables with unit variance. For simplicity, all users are assumed to be located at the same distance from the base station and experience only log-normal shadowing and Rayleigh fading due to the Doppler spread f_d^k, which is determined by the carrier frequency f_c and the velocity v_k as $f_d^k = \frac{v_k f_c}{C}$.

The effect of shadowing is modelled as a log-normal process with standard deviation σ_s [91]

$$\omega_k(t) = 10^{[\sigma_s \zeta_k(t) + m_s]/20}, \tag{5.3}$$

where m_s is the area mean determined by shadowing and $\zeta_k(t)$ is a sum-of-sinusoids (SOS) random process for user k at time t, which is given by

$$\zeta_k(t) = \sum_{i=1}^{N_{sos}} c_{ki} \cos\left(2\pi f_{ki} t + \theta_{ki}\right). \tag{5.4}$$

In equation (5.4), the gains c_{ki} and the frequencies f_{ki} are non-zero, real-valued constant values for user k, N_{sos} is the number of sinusoids and the phases θ_{ki} are $i.i.d.$ uniform random variables on the interval $(0,\ 2\pi]$ for user k. There are several methods to determine the parameters c_{ki} and f_{ki}. In this thesis, the method of equal areas (MEA) [91] is used. Then, the following equations are derived for these parameters

$$c_{ki} = \sqrt{\frac{2}{N_{sos}}},\ f_{ki} = \frac{v_k}{2\pi D} \tan\left[\frac{\pi(i-0.5)}{2N_{sos}}\right], \tag{5.5}$$

where D denotes the decorrelation distance determined by shadowing environments. For Rayleigh fading, the sum-of-sinusoids statistical model [92] is applied. Many conventional Rayleigh fading models have been developed based on the Jakes model [93]. However, the Jakes model is

Item	Parameters
Structure	Circular cell with radius R_{cell}
Antenna Configuration	M_T TX antennas for a BS and K users with a single RX antenna
Path Loss	α_l
Long Term Fading	Log-normal shadowing (σ_s, D)
Short Term Fading	Rayleigh
Doppler Spread	f_d
Delay Spread	None
Temporal Correlation	$\rho(\tau) = J_0(2\pi f_d \tau)$
Spatial Correlation	None
Cross Polarization	None
Saturation Effect	None

Table 5.1: *Parameters for the proposed channel model for MIMO broadcast channels*

a deterministic model because it has no random variables in its equation, which has limitations when modelling $i.i.d.$ random multiple antenna matrix entries. The Rayleigh fading variable $g_{km}(t)$ of the mth transmit antenna for user k at time t is given by

$$\begin{aligned} g_{km}(t) &= \sqrt{\frac{1}{N_{sos}}} \left\{ \sum_{i=1}^{N_{sos}} \cos(\psi_{i,km}) \cos\left[2\pi f_d^k t \cdot \cos\left(\frac{2\pi i - \pi + \theta_{km}}{4N_{sos}}\right) + \phi_{km}\right] \right. \\ &\left. + j \sum_{i=1}^{N_{sos}} \sin(\psi_{i,km}) \cos\left[2\pi f_d^k t \cdot \cos\left(\frac{2\pi i - \pi + \theta_{km}}{4N_{sos}}\right) + \phi_{km}\right] \right\}, \end{aligned} \tag{5.6}$$

where θ_{km}, ϕ_{km} and $\psi_{i,km}$ are mutually $i.i.d.$ uniform random variables over $(-\pi, \pi]$ for all i, m and k. With this model, the temporal correlation of any user with delay τ can be given by $\rho(\tau) = m_s J_0(2\pi f_d \tau)$, where J_0 is the zeroth order Bessel function of the first kind [94] [95]. Table 5.1 summarizes the proposed channel model.

5.3 MIMO Channel Measurements

The MIMO channel measurements were conducted in the city centre of Bristol, UK as part of a Mobile Virtual Centre of Excellence (M-VCE) programme in 2005 [6] [96] [97] [98]. This campaign used realistic antenna deployments, where two pairs of dual polarized transmit antennas at the base station and two pairs of crossed dipoles for reception were used at different locations in a dense urban cellular environment. The transmitting site was located on the roof of a building 33m above ground level and with a coverage area of one 65° sector. MIMO channel

Parameter	Setting
Carrier Frequency	2GHz
Measurement Bandwidth	20MHz
Number of Transmit Antennas	4
Number of Receive Antennas	4
Timing Grid for Sampling the Channel	6.144ms
Measurement Duration	6s

Table 5.2: *Measurement parameters for MIMO channel sounding [6]*

soundings were made with a user walking or sitting within the coverage area. The base station had maximum output power of 43dBm in the 20MHz band, centred at 2GHz with two panel antennas with the frequency range between 1900 and 2170MHz. Parameters for the MIMO channel soundings are given in Table 5.2.

5.4 Validation of the Proposed Channel Model

In this section, numerical results for the validation of the proposed channel model are presented. Throughout the simulations for the validation, it is assumed that the number of transmit antennas $M_T = 4$, the number of users $K = 4$, $SNR_0 = 0$dB and the number of sinusoids for the generation of Rayleigh fading and shadowing $N_{sos} = 128$, which is sufficiently large enough to have good agreement with desired properties [91] [92]. In order to model the large scale environment of the MIMO channel measurements, the shadow standard deviation of the measured data from 19 different locations was estimated. Assuming no significant path loss effect, the shadow standard deviation is roughly estimated as 6.2dB by mapping points with the received power and its distance from the base station. This result seems reasonable since in [15], the measured values for standard deviation of shadowing are 4.3dB and 7.5dB for urban and suburban environments respectively. The decorrelation distance D for shadowing might be several tens of meters for that urban environment and it is assumed to be $D = 30$m because the simulated data with that value matches better with the measured data than any other values considered in terms of the distribution of channel eigenvalues. In [15], the decorrelation distance for the urban environment and the suburban environment are given as 8.3m and 503.9m respectively. We also assume the carrier frequency $f_c = 2$GHz and the Doppler spread $f_d = 8$Hz, which corresponds to walking pace for all users. The time period between discrete channel realizations is set to be 6.144ms, which is the same as time grid used in the measurements. By

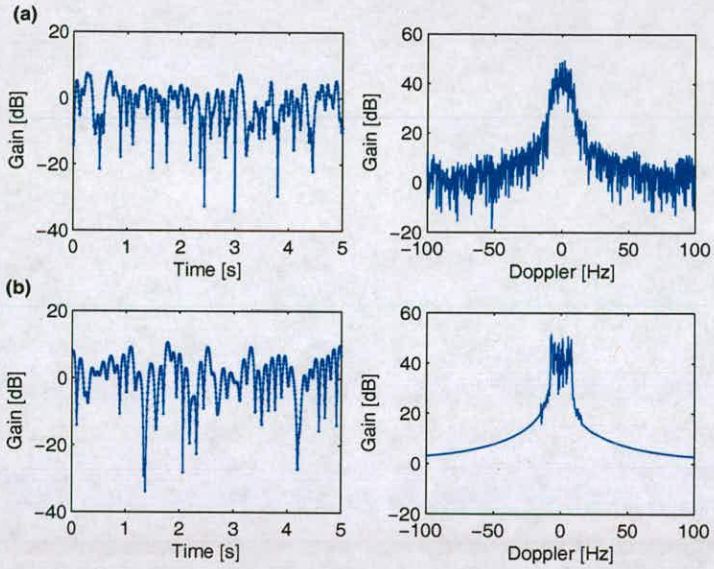


Figure 5.1: *Time and frequency representation of (a) the measured data and (b) the simulated data*

normalizing each column vector of the channel matrix over the time axis separately, the effect of cross polarization discrimination (XPD), defined as a mean level difference between two orthogonal polarized components, can be ignored.

Figure 5.1 shows a comparison between the measured data and the simulated data both in the time domain and in the Doppler domain. We notice that the Doppler spread of the walking reference is about 8Hz, which corresponds to a velocity of about 4.3km/h at the 2GHz centre frequency. The Doppler spread of the simulated data is set to be the same as that of the measured data for comparison. In this case, we observe that the temporal variations of the measured data and the simulated data seem to match well.

Figure 5.2 shows the cumulative distribution function (CDF) of the eigenvalues of the measured data and the simulated data. In this plot, it is assumed that there are four users, each with a single receive antenna and four transmit antennas at the base station, which forms a 4×4 MIMO broadcast channel. We notice that the measured data and the simulated data match well. The small differences may be due to spatial correlations between transmit antennas.

Figure 5.3 shows the effect of the spatial correlation between receive antennas. Denoting the 4×4 MIMO channel matrix as $\mathbf{H}$, every point in each curve is obtained by averaging all non-diagonal entries of the matrix $\mathbf{H}^H\mathbf{H}$. Then the CDF is constructed with these points from all

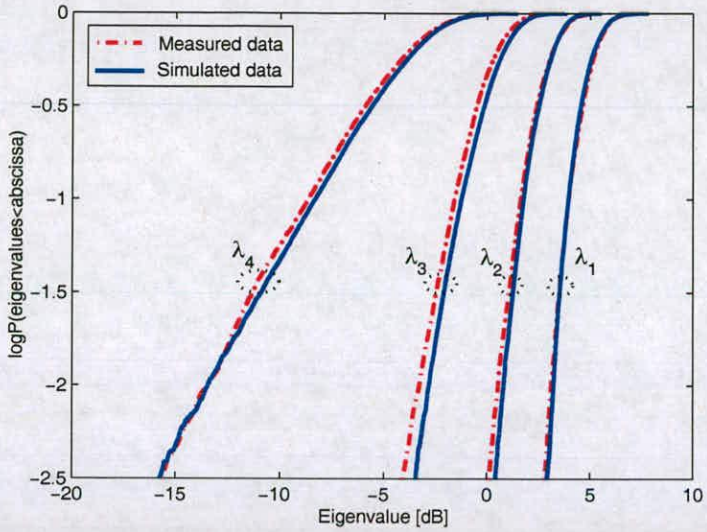


Figure 5.2: *Comparison of the eigenvalues between the measured data and the simulated data with a* 4×4 *channel matrix* $M_T = 4$, $K = 4$ *and* $M_R = 1$.

channel realizations. We notice that the case of two pairs of crossed dipoles at the receiver ($M_R = 4$) suffers from non-zero cross correlations ($0.1 \sim 0.2$) due to insufficient spacing of receive antennas. However, if the spatial correlation at the receiver is removed, namely, four different antennas from four different users ($K = 4, M_R = 1$) to emulate MIMO BC conditions, the cross correlation become small enough to be ignored compared to the case of four receive antennas from one user ($M_R = 4$). If an antenna spacing at the base station is designed so that spatial correlation is minimized, the proposed model can be used for emulating real MIMO broadcast channels with a single receive antenna for each user in terms of spatial correlation characteristics.

5.5 Statistical CSI-assisted Multiuser Scheduling

In this section, we investigate the sum-rate capacity of THP considering scheduling delay and propose a new SCSI-assisted multiuser scheduling algorithm using THP to minimize the effect of scheduling delay in the temporally correlated MIMO broadcast channel proposed in Section 5.2. The discrete time index n will be used in this section because CSI is obtained discretely in real MIMO BC scenarios. Figure 5.4 shows the system model for the SCSI-assisted multiuser scheduling algorithm with THP.

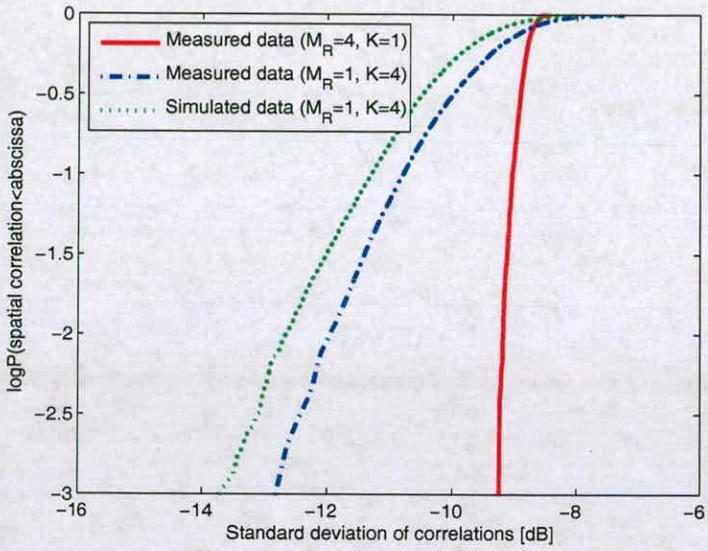


Figure 5.3: *CDFs of the standard deviation of cross correlations with different antenna configuration with a* 4×4 *MIMO channel matrix,* $M_T = 4$

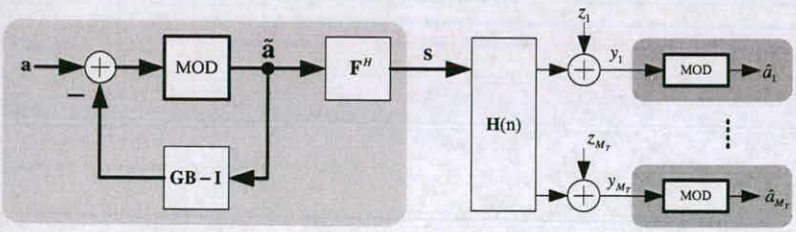


Figure 5.4: *The system model for the SCSI-assisted multiuser scheduling algorithm with THP*

5.5.1 Sum-rate Capacity of THP with Scheduling Delay

For simplicity of implementation, several assumptions were made in Chapter 4. In addition to these assumptions, if the target bit-error rate (BER) of the system is very small (*i.e.*, BER $\leq 10^{-6}$) and high signal-to-noise ratio (SNR) operation is assumed for all users, the modulo loss can be ignored [54] except for the shaping loss of 1.53dB, which can be achieved by using multidimensional lattice codes rather than M-QAM. Then, the sum-rate capacity of THP at time n can be approximated as

$$C_{THP}(S,n) = \sum_{k=1}^{M_T} \log_2\left(1 + \frac{b_{kk}^2(n)P}{M_T}\right). \tag{5.7}$$

In order to investigate the effect of scheduling delay on the sum-rate capacity of THP, it is assumed that one downlink frame of the MIMO BC with time period T_f consists of N_S ($N_S > 1$) time slots with time period T_s. Then the ICSI is updated once per every frame at the base station. Denoting the maximum Doppler shift as f_D, the frame period should be selected such that [99]

$$T_f < \frac{1}{2f_D}. \tag{5.8}$$

ICSI can be obtained by using either a feedback channel or the reciprocity principle in the case of time division duplex (TDD) systems. There are several sources for imperfect ICSI [39] but it is assumed that the ICSI is perfect at the beginning of every frame [94] for simplicity. In this case, the only imperfection is the mismatch of the channel estimate due to scheduling delay. This mismatch results in performance degradation, which will generally increases through the time slots of each frame. The amount of degradation for user k is determined by the temporal correlation, which is a function of the Doppler spread f_d^k and multiples of the slot period nT_s. Based on the results in [94], an equivalent time-varying channel matrix reflecting scheduling delay within the frame at slot index n can be written as

$$\mathbf{H}(n) = \mathbf{P}(n)\mathbf{H_0} + \sqrt{\mathbf{I} - \mathbf{P}(n)\mathbf{P}^H(n)}\mathbf{H_m},\ n = 1, \ldots, N_S, \tag{5.9}$$

where $\mathbf{P}(n) = \text{diag}\{\rho_1(nT_s), \ldots, \rho_{M_T}(nT_s)\}$ denotes the autocorrelation matrix of the user set S, $\sqrt{\mathbf{I} - \mathbf{P}(n)\mathbf{P}^H(n)}$ represents the amplitude increase in the channel estimation error due to scheduling delay, $\mathbf{H_0}$ is the perfectly estimated channel matrix at the beginning of each frame and $\mathbf{H_m}$ is an uncorrelated estimation error matrix which has the same statistical characteristics

as the estimated channel $\mathbf{H_0}$. Equation (5.9) implies that any decrease in ICSI becomes a delay-induced channel estimation error due to scheduling delay.

From equation (5.1), the received signal vector via the equivalent MIMO BC becomes

$$\mathbf{y}(n) = \mathbf{P}(n)\mathbf{H}_0\mathbf{s} + \sqrt{\mathbf{I} - \mathbf{P}(n)\mathbf{P}^H(n)}\mathbf{H}_m\mathbf{s} + \mathbf{z}(n), \tag{5.10}$$

where $\mathbf{y}(n) \in \mathbb{C}^{M_T \times 1}$ and $\mathbf{z}(n) \in \mathbb{C}^{M_T \times 1}$ are vector forms of the received signal and complex AWGN respectively.

In equation (5.10), the first term denotes the desired signal and the second term denotes the additive measurement noise due to the delay-induced estimation error. After the modulo operation at the receiver, the output data vector becomes

$$\begin{aligned}
\hat{\mathbf{a}} &= \mathrm{MOD}_{RX}[\mathbf{P}(n)\mathbf{H}_0\mathbf{s}] \\
&= \mathrm{MOD}_{RX}[\mathbf{P}(n)\mathbf{B}\mathbf{F} \cdot \mathbf{F}^H \mathrm{MOD}_{TX}[(\mathbf{B}\mathbf{G})^{-1}\mathbf{a}]\,] \\
&= \mathrm{MOD}_{RX}[\mathbf{P}(n)\mathbf{B} \cdot (\mathbf{B}\mathbf{G})^{-1}(\mathbf{a}+\mathbf{p})] \\
&= \mathrm{MOD}_{RX}[\mathbf{P}(n)\mathbf{G}^{-1}(\mathbf{a}+\mathbf{p})] \\
&= \mathbf{P}(n)\mathbf{G}^{-1}\mathbf{a},
\end{aligned} \tag{5.11}$$

where $\mathrm{MOD}_{TX}[\cdot]$ and $\mathrm{MOD}_{RX}[\cdot]$ are the modulo operations at the transmitter and receiver respectively. The vector $\mathbf{p}$ denotes an arbitrary symbol vector due to the modulo operation at the transmitter. (See Section 3.3.2.4). The modulo operation at the receiver subtracts the arbitrary symbol vector added at the transmitter. Because $\mathbf{F}$ is a unitary matrix, $\mathbf{F}^H\mathbf{F} = \mathbf{I}$ in the above equation.

The variance of the estimate noise, $\sqrt{\mathbf{I} - \mathbf{P}(n)\mathbf{P}^H(n)}\mathbf{H}_m\mathbf{s}$, is given by

$$\left(\mathbf{I} - \mathbf{P}(n)\mathbf{P}^H(n)\right)E[(\mathbf{H}_m\mathbf{s})(\mathbf{H}_m\mathbf{s})^H] \overset{(a)}{=} M_T\left(\mathbf{I} - \mathbf{P}(n)\mathbf{P}^H(n)\right)\mathbf{R}_{ss}, \tag{5.12}$$

where (a) follows from the fact that $E\left[\mathbf{H}_m\mathbf{H}_m^H\right] = M_T\mathbf{I}$. Because $\sigma_s^2 \simeq \sigma_a^2$ with THP [100], the effective SNR of user k at slot index n is given by

$$SNR_k(n) = \frac{b_{kk}^2\rho_k^2(n)\sigma_a^2}{M_T(1-\rho_k^2(n))\sigma_a^2 + 1}. \tag{5.13}$$

The effective SNR is obtained by the channel gain b_{kk}^2 and the temporal variation determined

by the autocorrelation function $\rho_k(n)$. If there is no temporal correlation, the effective SNR reduces to the normal SNR value $b_{kk}^2\sigma_a^2$.

Since $\sigma_a^2 = \frac{P}{M_T}$, the sum-rate capacity with THP is written as

$$C_{THP}(S, \mathbf{H}(n)) = \sum_{k=1}^{M_T} \log_2 \left(1 + \frac{b_{kk}^2 \rho_k^2(n) P}{M_T \left\{ (1 - \rho_k^2(n))P + 1 \right\}} \right). \tag{5.14}$$

5.5.2 SCSI-assisted Multiuser Scheduling Algorithm

Although the base station obtains ICSI once per every frame, it can utilize the previous ICSI for obtaining SCSI such as the mean and covariance due to channel stationarity [94]. This SCSI can be used for estimating the temporal correlation of the MIMO BC. In this case, the multiuser scheduling can use SCSI in selecting a user set on a slot-by-slot basis between frame starts when precise ICSI is not available . The SCSI-assisted sum-rate maximization rule at slot index $n(n = 1, \ldots, N_S)$ finds a user set $S_{max}(n)$ according to

$$S_{\max}(n) = \arg \max_{S \subset \{1, \cdots, K\}, |S| = M_T} C_{THP}\left(S, \mathbf{H}(n)\right). \tag{5.15}$$

The SCSI-assisted sum-rate maximization rule uses the equivalent estimated channel model (5.9) for the multiuser selection procedure. Any user with a rapid decrease in the temporal correlation may be assigned to the first a few slots in the frame. This may improve the total throughput in return for increased complexity. This is because the multiuser selection procedure has to be performed on a slot-by-slot basis if the proposed algorithm is applied. The selected user set for each time slot may be different from one another if the SCSI-assisted multiuser scheduling algorithm is applied on a slot-by-slot basis. However, the normal multiuser selection algorithm without SCSI on a frame-by-frame basis using the channel matrix $\mathbf{H}(0) = \mathbf{H}_0$ would allocate the same choice of users to all slots. Figure 5.5 shows the flow chart of the proposed algorithm, which also demonstrates the difference between the SCSI-assisted multiuser scheduling algorithm and the normal multiuser selection algorithm.

When the temporal variations of the MIMO BC are small compared to the coherence time, the proposed SCSI-assisted multiuser scheduling algorithm has little improvement from the normal sum-rate maximization algorithm, which is performed on a frame-by-frame basis. However, if the temporal variations are large, the performance improvement by the proposed algorithm is evident. This will be shown in the next section.

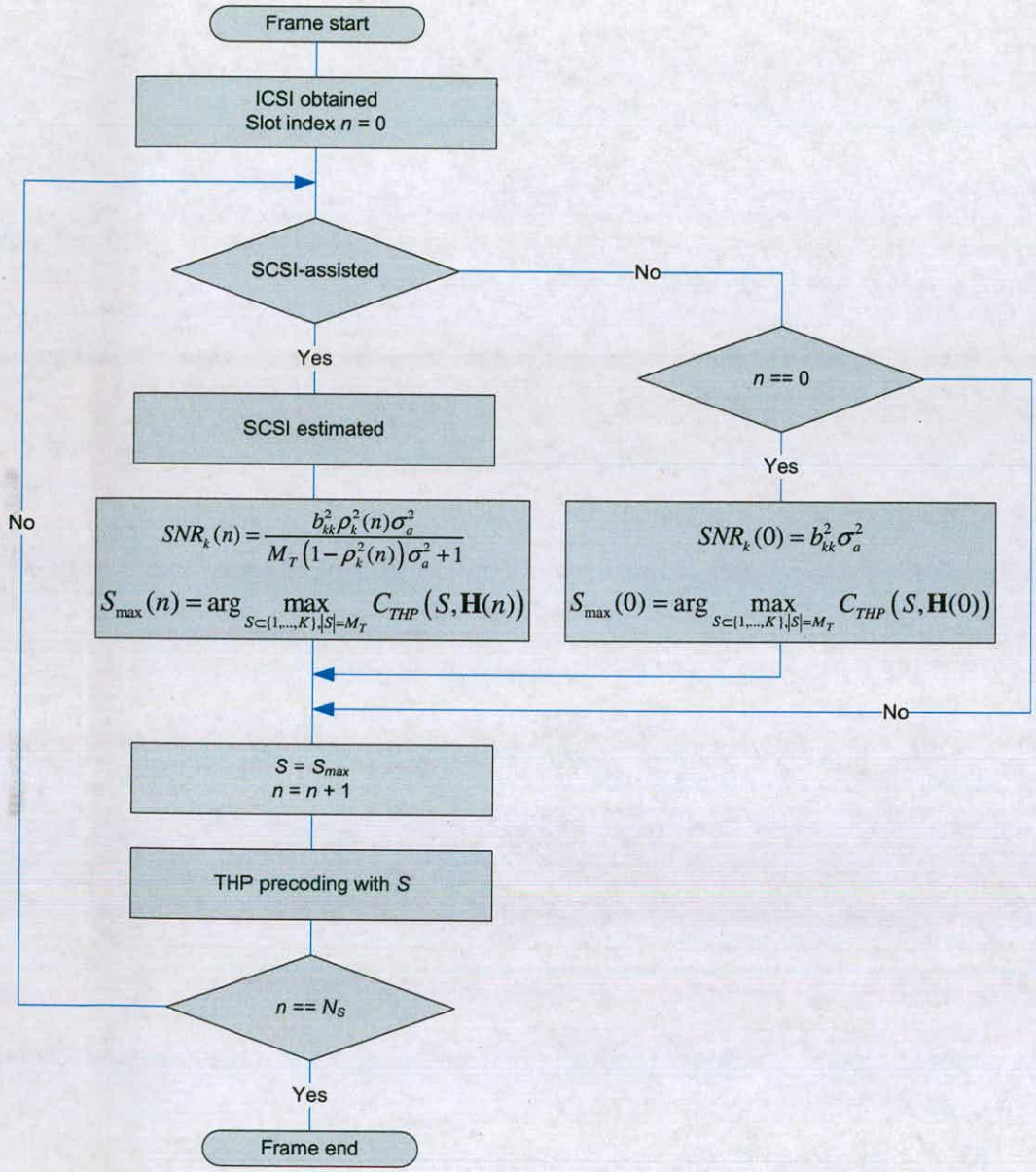


Figure 5.5: *The flow chart of the SCSI-assisted multiuser scheduling algorithm with THP based on (5.13) and (5.15)*

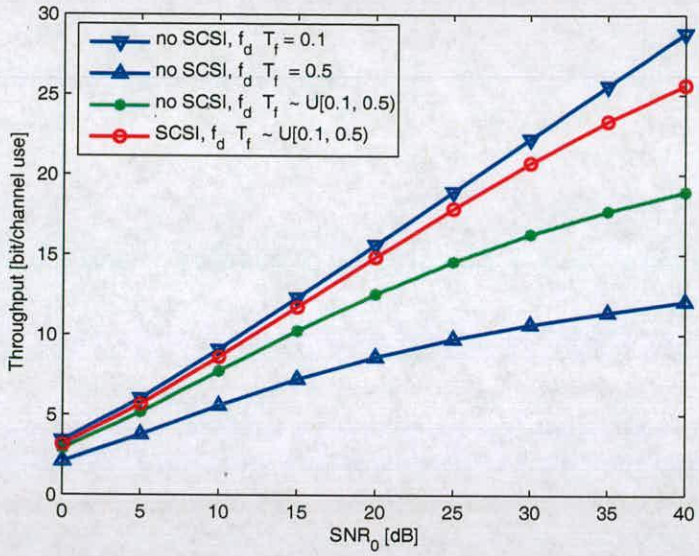


Figure 5.6: *The performance of the SCSI-assisted multiuser scheduling algorithm against SNR_0 with different relative Doppler spread configuration, $T_f = 2ms$, $M_T = 2$ and $K = 32$.*

5.6 Numerical Results

In this section, numerical results for the proposed SCSI-assisted multiuser scheduling algorithm are presented. Throughout the simulations, we assume the number of transmit antennas $M_T = 2$, the number of users $K = 32$ and the number of sinusoids $N_s = 5$. The relative Doppler spread values of all users, $f_d T_f$, are assumed to be $i.i.d.$ uniform random variables over $[0.1, 0.5)$. The estimated standard deviation for shadowing, 6.2dB, in Section 5.4 is not used in this section because it is not a typical value for emulating the urban environment and no corresponding decorrelation distance is obtained from the channel measurements. Instead, the standard deviation $\sigma_s = 4.3$dB and the decorrelation distance $D = 8.3$m are assumed for the urban environment [15].

Figure 5.6 shows the performance of the SCSI-assisted multiuser scheduling algorithm against the median SNR of all users SNR_0 with different relative Doppler spread configuration. We assume $T_f = 2$ms in this plot. When the relative Doppler spreads of all users are identical, the throughput decreases as the relative Doppler spread increases. This is because the temporal correlation for all users decreases more rapidly as the relative Doppler spread increases. However, when the relative Doppler spreads of all users have a uniform distribution over $[0.1, 0.5)$, the proposed SCSI-assisted multiuser scheduling algorithm performs better than the multiuser

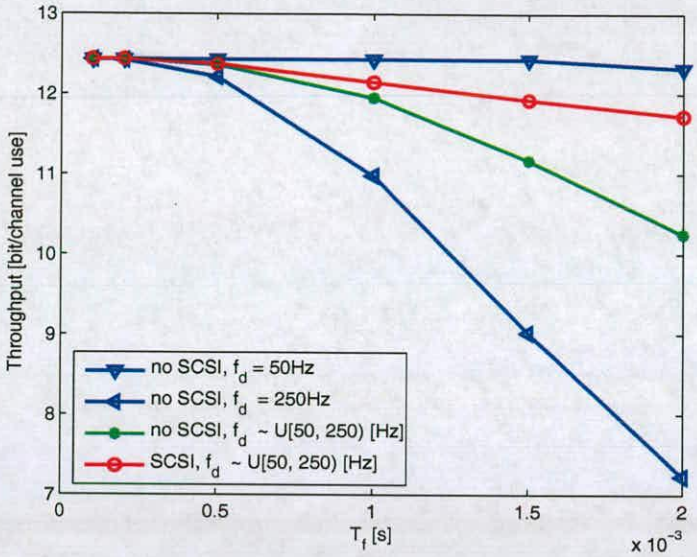


Figure 5.7: *The performance of the SCSI-assisted multiuser scheduling algorithm against frame period with different relative Doppler spread configuration,* $SNR_0 = 15dB$, $M_T = 2$ *and* $K = 32$.

scheduling algorithm which does not use SCSI and applies a frame-basis multiuser selection. The performance difference between these two schemes is evident as the SNR_0 increases because the delay-induced channel estimation error produces an irreducible error floor in the high SNR region.

Figure 5.7 shows the performance of the SCSI-assisted multiuser scheduling algorithm against the frame period T_f. We assume SNR_0 = 15dB for all users in this plot. When the frame period is small enough compared to the coherence time of the channel, the performance improvement of the proposed algorithm is negligible. However, as the frame period increases, the performance advantage of using SCSI becomes evident. The proposed SCSI-assisted multiuser scheduling algorithm performs better than the multiuser scheduling without SCSI considerations when $T_f > 0.2$ms. When $T_f = 2$ms, the throughput gain due to the proposed algorithm is about 1.5 [bits/channel use]. Note that the amount of performance improvement and complexity increase of the proposed algorithm depends on the channel conditions and the number of slots in the frame structure.

5.7 Summary

This chapter has proposed a temporally correlated MIMO broadcast channel model, which has reasonable complexity but can describe realistic temporally correlated MIMO broadcast channels with multiple antennas for a base station and a single receive antenna for each user. The proposed model was validated by comparing with the MIMO channel measurements. Numerical results showed that if there is only temporal correlation in MIMO broadcast channels, the proposed channel model can sufficiently represent the real MIMO broadcast channels. Based on the proposed channel model, a new SCSI-assisted multiuser scheduling algorithm has been proposed. Due to the temporal variations of the MIMO BC, mismatch in the channel estimate occurs, which affects the reliability of CSI at the transmitter. Because the degree of mismatch is different among users if their Doppler spreads are unequal, the proposed algorithm can minimize such mismatches. This is achieved by allocating users with high Doppler spreads to the front part of a frame. For this, the proposed algorithm makes use of SCSI as well as ICSI for the allocation of users to the spatio-temporal dimension. By doing this, it can minimize the effect of scheduling delay due to the temporal correlation to improve the sum-rate capacity of THP in the MIMO BC.

In the next chapter, a new QoS-guaranteed multiuser scheduling algorithm will be proposed based on the temporally correlated MIMO broadcast channel described in this chapter. Indeed, it will exploit the QoS-aware sequential multiuser selection algorithm in Chapter 4 for supporting different QoS users in the MIMO BC.

Chapter 6
QoS-guaranteed Multiuser Scheduling in MIMO Broadcast Channels

6.1 Introduction

Alongside the choice of MIMO precoding technique, multiuser scheduling is one of the most important issues of the MIMO BC for satisfying different QoS requirements such as throughput, delay constraint and fairness among users. A number of studies focusing on combined MIMO precoding techniques and multiuser scheduling methods for the MIMO BC have been presented. However, most of them do not consider different QoS requirements such as different data rates, delay constraints and different fairness requirements among users.

In this chapter, a new QoS-guaranteed multiuser scheduling algorithm will be proposed for supporting a mixture of different QoS users simultaneously whilst satisfying fairness among users in realistic MIMO BC scenarios. This will be useful when a base station is providing a variety of services to different users with different QoS requirements and fairness considerations. For this, the proposed algorithm includes a new metric to allow different levels of throughput or delay fairness among users in the same QoS group. It also uses a more advanced soft antenna trade-off scheme than the one used as part of the sequential multiuser selection algorithm in Chapter 4 for differentiation between different QoS groups. With the proposed antenna trade-off scheme, a higher priority group takes precedence in using multiple antennas for satisfying its QoS requirement. The number of transmit antennas allocated to each QoS group can be determined by the wireless channel conditions and the QoS requirements. After determining the number of transmit antennas allocated to each QoS group using the antenna trade-off scheme, the proposed scheduling algorithm finds user sets starting from the highest QoS group sequentially. It uses the QF scheduling algorithm from each QoS group, so that the final user set for transmission consists of users who can maximize the sum-rate capacity as well as satisfying fairness requirements.

The proposed QoS-guaranteed multiuser scheduling algorithm has the following features.

- The proposed algorithm can provide fairness among users in the same QoS group by the QoS-aware Fair (QF) scheduling algorithm. This fairness either relates to delay or throughput. This can be performed by applying new weight vectors to the weighted sum-rate maximization rule. The weight vectors represent the degree of fairness among users and affect the selection of a user set used in the weighted sum-rate maximization rule.
- The proposed algorithm can support different QoS classes simultaneously using the QoS-aware sequential multiuser selection algorithm proposed in Chapter 4 with an advanced soft antenna trade-off scheme, which gracefully adapts to the time-varying wireless channel conditions.
- The proposed algorithm can trade degree of fairness among users with the total throughput. This can be performed by adjusting the sensitivity of the weight vectors. It can provide more throughput in return for a decrease in fairness among users and vice versa.

This chapter is divided into two parts. The first part briefly introduces conventional scheduling algorithms. The second part proposes a QoS-guaranteed multiuser scheduling algorithm in detail. This chapter is organized as follows. Section 6.2 briefly overviews conventional scheduling algorithms. Section 6.3 proposes a QoS-guaranteed multiuser scheduling algorithm, which includes a QoS-aware fair scheduling algorithm for the same QoS group and antenna trade-off schemes between different QoS groups. A summary for this chapter is presented in Section 6.4.

6.2 Overview of the Scheduling Algorithms

The scheduling problem is one of the most important issues in multiuser wireless systems. The objective of any scheduling algorithm for wireless systems may be as follows: A scheduler should allocate limited resources such as bandwidth, power and frequency in the most efficient way for achieving a desirable performance. This can be performed by jointly considering the physical layer and higher network layers in an integrated framework. The most well-known and simple cross-layer resource allocation is link adaptation techniques such as adaptive modulation and coding (AMC) [101] [102]. The medium access control (MAC) layer determines the AMC parameters according to the wireless channel conditions in the physical layer.

For an overview on the cross-layer optimization, see [65] [103] and [104].

In general, the service requirements in the higher network layers such as fairness, data rate and

delay do not always coincide with the wireless channel conditions. This is because the fading in the wireless channels are independent of the service requirements. This makes it difficult to find the optimal trade-off point between the individual requirements and the overall system performance. Therefore, lots of scheduling algorithms have been suggested to find the optimal solution. Among these scheduling algorithms, this section introduces the most basic scheduling algorithms such as proportional fair (PF) [2], largest weighted delay first (LWDF) [3], modified largest weighted delay first (M-LWDF) [4] and the exponential rule [5]. However, these opportunistic selection rules do not consider the selection of multiple users simultaneously. This motivates the research on multiuser scheduling algorithms which select and serve multiple users simultaneously.

Overviews of scheduling algorithms with QoS considerations for wireless systems are found in [105] and [106].

6.2.1 Proportional Fair Scheduling [2]

The PF scheduling rule is designed to satisfy long-term throughput fairness among users by considering the channel conditions and the amount of past throughput simultaneously whilst exploiting multiuser diversity. The PF scheduler selects a user with index k at time t according to

$$k(t) = \arg \max_{i \in \{1, \cdots, K\}} \frac{r_i(t)}{\bar{r}_i(t)}, \quad . \tag{6.1}$$

where $r_i(t)$ is the data rate of user i at time t and $\bar{r}_i(t)$ is the exponential moving average of past throughput of user i, which is updated as

$$\bar{r}_i(t+1) = \left(1 - \frac{1}{\alpha_t}\right) \bar{r}_i(t) + \frac{1}{\alpha_t} r_i(t), \tag{6.2}$$

where α_t is the smoothing factor, which determines the fairness window size of the scheduler. Large values of α_t means a large time window. As α_t increases, the average period for long-term throughput also increases.

6.2.2 Largest Weighted Delay First Scheduling [3]

The LWDF scheduling rule selects a user with index k at time t according to

$$k(t) = \arg \max_{i \in \{1, \cdots, K\}} a_i d_i(t), \tag{6.3}$$

where $d_i(t)$ is the HOL delay of user i and $a_i (a_i > 0)$ are constants, which are determined by the probabilistic QoS parameters as

$$a_i = -\frac{\log(\eta_i)}{T_i}, \tag{6.4}$$

where T_i and η_i are the delay threshold and the maximum probability of exceeding that delay threshold respectively. They define the probabilistic QoS requirement of user i as follows.

$$\Pr\{d_i > T_i\} \leq \eta_i. \tag{6.5}$$

Any user with strict QoS requirement (*i.e.*, small T_i and η_i) has large value of a_i, which results in high priority in the selection procedure.

6.2.3 Modified Largest Weighted Delay First Scheduling [4]

The M-LWDF scheduling rule combines the PF scheduling rule and the LWDF scheduling rule in order to consider both throughput fairness and delay simultaneously.

$$k(t) = \arg \max_{i \in \{1, \cdots, K\}} \frac{a_i r_i(t) d_i(t)}{\bar{r}_i(t)}. \tag{6.6}$$

With this selection rule, the scheduler can support different QoS requirements by the probabilistic QoS definitions as well as considering throughput fairness among users simultaneously. For example, BE users may have small a_i compared to RT users if they have lower priority than RT users. Any BE user with a lower past throughput than other users may have a chance of transmission to keep throughput fairness among users.

6.2.4 The Exponential Rule [5]

The exponential rule uses the exponential function in dealing with the delay term in the M-LWDF scheduling rule, which is given by

$$k(t) = \arg \max_{i \in \{1,\cdots,K\}} \frac{a_i r_i(t)}{\bar{r}_i(t)} \exp\left(\frac{a_i d_i(t) - a\bar{D}}{1 + \sqrt{a\bar{D}}}\right), \tag{6.7}$$

where $a\bar{D} = \frac{1}{N}\sum_{i=1}^{N} a_i d_i(t)$ and N denotes the number of slots considered. For reasonable values of a_i, this selection rule tries to equalize the weighted delays $a_i d_i(t)$ when their differences are large. If any user has a large weighted delay than the other users by more than order $\sqrt{a\bar{D}}$, the exponential term becomes very large, which makes the PF term relatively very small. On the other hand, with small weighted delay differences, the exponential term is close to one and the selection rule acts as the PF scheduling rule.

6.3 QoS-guaranteed Multiuser Scheduling Algorithm

In this section, a new QoS-guaranteed multiuser scheduling algorithm is proposed and analyzed. The proposed algorithm consists of two parts: the QoS-aware fair (QF) scheduling algorithm for the same QoS users and an antenna trade-off scheme for different QoS users. The QF scheduling algorithm can provide fairness among users in terms of throughput or delay respectively. It can also adjust the degree of fairness by trading fairness among users with the total throughput. The antenna trade-off scheme can support different QoS groups simultaneously with QoS differentiation whilst keeping fairness among users. When the weighted sum-rate maximization rule [107] is used, the proposed algorithm can achieve the sum-rate capacity whilst satisfying these user specific requirements.

In the next section, a channel model will be presented, which was already described in Chapter 5. Based on the proposed channel model, new multiuser scheduling algorithms will be proposed in the remainder of this chapter.

6.3.1 System Model

Consider a temporally-correlated fading MIMO BC with M_T transmit antennas at a base station and K ($K \geq M_T$) users each with a single receive antenna ($M_{RT} = 1$). Let $\mathbf{h}_k(t) \in \mathbf{C}^{M_T \times 1}$

denote the channel at time t between the transmit antenna array and the receive antenna for user k. Then the MIMO BC at time t can be represented as

$$y_k(t) = \mathbf{h}_k^T(t)\mathbf{s}(t) + z_k(t), \qquad k = 1, \cdots, K, \tag{6.8}$$

where $\mathbf{s}(t) \in \mathbb{C}^{M_T \times 1}$ denotes the transmit signal vector with a covariance matrix $\mathbf{R}_{ss} = E[\mathbf{s}(t)\mathbf{s}(t)^H] = \sigma_s^2\mathbf{I}$ and a transmit power constraint $\mathrm{Tr}(\mathbf{R}_{ss}) \leq P$, the scalar $y_k(t)$ is the received signal for user k and $z_k(t)$ is the complex additive white Gaussian noise (AWGN) with zero mean and unit variance for user k at time t.

The transmit signals are assumed to experience path loss, log-normal shadowing and Rayleigh fading. In this case, the channel $\mathbf{h}_k(t)$ can be expressed as

$$\mathbf{h}_k(t) = \sqrt{SNR_0 \left(\frac{r_k^d(t)}{R_{cell}}\right)^{-\alpha_l}}\, \omega_k(t) \cdot \mathbf{g}_k(t), \tag{6.9}$$

where SNR_0 denotes the median of the mean SNR of all users averaged over fading, $r_k^d(t)$ denotes the distance between user k and the base station, R_{cell} denotes the cell radius, α_l denotes the path loss exponent and $\omega_k(t)$ denotes a shadowing of user k at time t. The vector $\mathbf{g}_k(t) = [g_{k1}(t)\ g_{k2}(t)\ \cdots\ g_{kM_T}(t)]^T$ represents the Rayleigh-distributed fading between the transmit antenna array and user k at time t. The statistical Rayleigh fading model [92] generates $i.i.d.$ circular symmetric complex Gaussian random variables for the elements of $\mathbf{g}_k(t)$ with zero mean and unit variance. All users are assumed to be located at the same distance from the base station and experience frequency non-selective Rayleigh fading due to the Doppler spread f_d^k.

6.3.2 QoS-aware Fair (QF) Scheduling Algorithm

6.3.2.1 Fairness metrics for QF scheduling

In real MIMO BC scenarios, a base station is likely to provide a variety of services to different users, each with different QoS requirements. In this case, throughput fairness among users does not mean the allocation of the same amount of bandwidth to all users. In order to support such heterogeneous user channels, define the scaled deviation as

$$\Delta x = \frac{x - \bar{x}}{x}, \tag{6.10}$$

where x is the observation and $\bar{x}$ is the required value of x. It can represent the relative degree of fairness regardless of the resource being considered. For example, there are two users with throughput requirements R_1 and R_2 respectively. It is assumed that the throughput requirement of the first user X_1 is two times as large as that of the second user X_2, namely, $R_2 = 2R_1$. If the data rates (the observations) of X_1 and X_2 are $0.5R_1$ and R_1 respectively at a certain instant, the scaled deviations with respect to the throughput requirements (the required values) are $\Delta x_1 = (0.5R_1 - R_1)/(0.5R_1) = -1.0$ for X_1 and $\Delta x_2 = (R_1 - 2R_1)/R_1 = -1.0$ for X_2. In this case, throughput fairness is said to be satisfied in terms of the relative achievement.

In order to apply the concept of the scaled deviation to the QF criterion, an exponential function taking the argument Δx is used. According to the type of resource, either $\exp(\Delta x)$ or $\exp(-\Delta x)$ can be used as elements of the weight vector for a weighted sum-rate maximization rule. In the case of throughput fairness, $\exp(-\Delta x)$ is used so that any user with relatively smaller throughput than other users has a large weight value. On the contrary, in the case of delay fairness, $\exp(\Delta x)$ is used so that any user with relatively larger delay than other users has a large weight value.

Prior to applying the scaled deviation to fairness cases, let us introduce the other parameter m, which changes the slope of the exponential function with a form of $\exp(m\Delta x)$ or $\exp(-m\Delta x)$. Later, this slope parameter will be used for controlling the degree of fairness. Figure 6.1 shows the characteristics of the scaled exponential function $\exp(m\Delta x)$ with $\bar{x} = 1$. When $m = 0.1$, the output of the scaled exponential function is saturated near one when $x > 1$ regardless of the scaled deviation. However, as m increases, the output of the scaled exponential function grows rapidly when $x > 1$. This makes a small change in x cause a significant increase in the output.

For throughput fairness, define the throughput fairness metric $\mu_k^t(t)$ for user k at time t as

$$\mu_k^t(t) = \exp\left(m_t \frac{\bar{R}(t) - a_k^t \bar{r}_k(t)}{\varepsilon + a_k^t \bar{r}_k(t)}\right), \tag{6.11}$$

where m_t ($m_t \geq 0$) is the throughput slope of the exponential function, which determines the sensitivity of (6.10) to throughput fairness as might be expected from Figure 6.1. Smaller values of m_t mean less strict throughput fairness among users and typically gives only average throughput fairness among users. Indeed, when $m_t = 0$, the weighted sum-rate maximization problem reduces to the normal sum-rate maximization problem without any fairness considera-

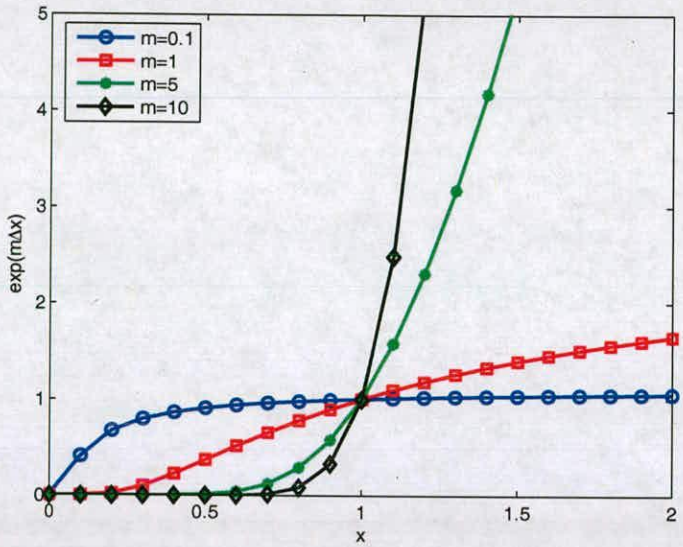


Figure 6.1: *The characteristics of the scaled exponential function* $\exp(m\Delta x)$ *with the slope parameter* m, $\bar{x} = 1$

tion. The scalars a_k^t are constants to allow different throughput requirements among users. The variable ε is an appropriate small value for ensuring the denominator is nonzero. The value $\bar{r}_k(t)$ is the exponential moving average of the past throughput for user k, which is updated as

$$\bar{r}_k(t+1) = \left(1 - \frac{1}{\alpha_t}\right)\bar{r}_k(t) + \frac{1}{\alpha_t}R_k(t), \tag{6.12}$$

where $\alpha_t(\alpha_t > 0)$ is the smoothing factor and $R_k(t)$ is the data rate of user k at time t. The scalar $\bar{R}(t)$ in (6.11) is defined as

$$\bar{R}(t) = (1/K)\sum\nolimits_{k=1}^{K}\bar{r}_k(t). \tag{6.13}$$

According to equation (6.11), any user with $a_k^t\bar{r}_k(t) > \bar{R}(t)$ has a value less than one as its weight. In this case, that user might be excluded from a selected user set for transmission by the weighted sum-rate maximization rule to meet the throughput fairness constraint among users despite its high channel gain.

For delay fairness, define the delay fairness metric $\mu_k^d(t)$ for user k at time t as

$$\mu_k^d(t) = \exp\left(m_d\frac{a_k^d d_k(t) - \bar{D}(t)}{\varepsilon + a_k^d d_k(t)}\right), \tag{6.14}$$

where m_d ($m_d \geq 0$) is the delay slope of the exponential function, which determines the sensitivity of delay fairness as in the case of throughput fairness. Again, small values of m_d means less strict delay fairness among users. The scalars a_k^d are constants to allow different delay requirements among users, which can be defined as in (6.4). The scalar $d_k(t)$ denotes the head-of-line (HOL) delay in the queue of user k and $\bar{D}(t)$ is defined as

$$\bar{D}(t) = (1/K)\sum_{k=1}^{K} d_k(t). \tag{6.15}$$

6.3.2.2 Weighed Sum-rate Maximization Rule with THP

From equation (5.7) in Chapter 5, the sum-rate capacity of THP at time t was given by

$$C_{THP}(S,t) = \sum_{k=1}^{M_T} \log_2\left(1 + \frac{b_{kk}^2(t)P}{M_T}\right). \tag{6.16}$$

From equations (6.11), (6.14) and (6.16), denoting $U := \{u_k \,|k = 1, \cdots, K\}$ as the total user set, the weighted sum-rate maximization rule considering throughput and delay fairness with THP at time t can be given as (6.17) and (6.18) respectively.

$$S_{\max}^t(t) = \arg \max_{S \subset U, |S| = M_T} \sum_{k=1}^{M_T} \mu_k^t(t) \log_2\left(1 + \frac{b_{kk}^2(t)P}{M_T}\right). \tag{6.17}$$

$$S_{\max}^d(t) = \arg \max_{S \subset U, |S| = M_T} \sum_{k=1}^{M_T} \mu_k^d(t) \log_2\left(1 + \frac{b_{kk}^2(t)P}{M_T}\right). \tag{6.18}$$

In order to evaluate the performance of the proposed algorithm, the PF scheduling and the LWDF scheduling are used for throughput fairness and delay fairness respectively. The weighting used for the existing PF scheduling rule for user k is the inverse of the exponential moving average of past throughput, so that the weighted sum-rate maximization rule for throughput fairness with the PF algorithm is given by

$$S_{\max}^t(t) = \arg \max_{S \subset U, |S| = M_T} \sum_{k=1}^{M_T} \frac{1}{\bar{r}_k(t)} \log_2\left(1 + \frac{b_{kk}^2(t)P}{M_T}\right). \tag{6.19}$$

For delay fairness with the existing LWDF scheduling rule, the weighted sum-rate maximiza-

tion rule is of the form

$$S^t_{\max}(t) = \arg \max_{S \subset U, |S|=M_T} \sum_{k=1}^{M_T} a_k^d d_k(t) \log_2 \left(1 + \frac{b_{kk}^2(t)P}{M_T}\right). \tag{6.20}$$

In equation (6.4), the constants a_k^d are defined by probabilistic QoS requirements. It is useful to note that the time index t will be omitted to simplify equations from the next section.

If there are two or more different QoS groups, the QoS-guaranteed multiuser scheduling algorithm determines the number of transmit antennas allocated to each QoS group. Then, the QF scheduling algorithm is applied at each stage to find a user set for that QoS group. In this case, the number of selected users from each QoS group is the number of transmit antennas allocated to each QoS group, which is determined by the antenna trade-off scheme introduced in the next section.

6.3.3 Antenna Trade-off between Different QoS Groups

In order to support different QoS users simultaneously in the MIMO BC, an antenna trade-off scheme for different QoS groups proposed in Chapter 4. In this section, the advanced soft antenna trade-off scheme is proposed and investigated in more detail. The improvement is obtained using the advanced soft antenna trade-off scheme by changing the number of transmit antennas allocated to each QoS group gracefully according to the time-varying channel conditions. We assume there are two QoS groups in the MIMO BC: a best effort (BE) user set $U_{BE} := \{u_k \,|k = 1, \cdots, K_{BE}\}$ and a delay-constrained real-time (RT) user set $U_{RT} := \{u_k \,|k = 1, \cdots, K_{RT}\}$, where K_{BE} and K_{RT} are the number of BE and RT users respectively and $K_{BE} + K_{RT} = K$. It is assumed that RT users have a higher priority than BE users in terms of QoS requirements. The objective of any multiuser scheduling algorithm in this configuration is to satisfy the delay constraint of RT users whilst maximizing the total throughput of all users. A simple multiuser scheduling algorithm may reserve a certain fixed number of transmit antennas to each group for supporting BE and RT users simultaneously. A more advanced way for serving different QoS groups is to change the number of transmit antennas assigned to each QoS group according to certain criteria. This can increase the degree of adaptability to fading MIMO channels. For this adaptive antenna trade-off scheme, each QoS group has its set of pre-assigned transmit antennas but there is a priority for their actual use in transmission. RT users have higher priority than BE users in using multiple transmit antennas.

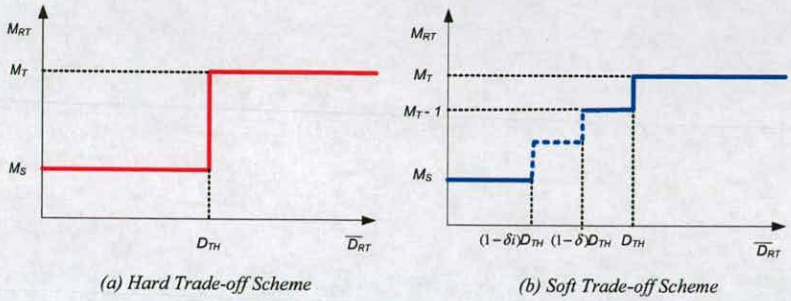


Figure 6.2: *The number of transmit antennas allocated to RT users for different antenna trade-off schemes*

For example, when the average HOL delay of RT users is within a certain delay threshold, RT and BE users use their pre-assigned groups of antennas for transmission. When the average HOL delay of RT users exceeds the delay threshold, RT users occupy a higher number of transmit antennas according to the antenna trade-off scheme, then BE users use any remaining transmit antennas for transmission. If there is no antenna trade-off between QoS groups ('fixed scheme'), the number of transmit antennas allocated to each group is always fixed.

There are two methods for trading off transmit antennas: the hard trade-off and soft trade-off schemes. The hard trade-off scheme, which is already proposed in Chapter 4 assigns all transmit antennas to RT users when the average delay of RT users exceeds the delay threshold. When this happens, BE users have no chance of using their pre-assigned antennas for transmission. The hard trade-off scheme determines the number of transmit antennas for RT users, M_{RT} according to

$$M_{RT} = M_S + (M_T - M_S)\,\mathrm{u}\left[\bar{D}_{RT} - D_{TH}\right], \tag{6.21}$$

where $\bar{D}_{RT}$ denotes the average HOL delay of RT users, D_{TH} denotes the delay threshold of RT users, M_S ($M_S < M_T$) denotes the pre-assigned groups of antennas for RT users and $\mathrm{u}\{\cdot\}$ denotes the unit step function. With the soft trade-off scheme, M_{RT} varies from the pre-assigned antennas M_S up to the total number of transmit antennas M_T determined by

$$M_{RT} = M_S + \sum_{i=0}^{M_T - M_S - 1} \mathrm{u}\left[\bar{D}_{RT} - (1-\delta i)\,D_{TH}\right], \tag{6.22}$$

where δ determines the dynamic range of the soft trade-off scheme. When $\delta = 0$, the equation (6.22) becomes the same as that for the hard trade-off scheme. Figure shows the number of transmit antennas allocated to RT users for different antenna trade-off schemes. This

graceful adaptation by the soft trade-off scheme has better performance than the hard trade-off scheme in terms of the outage probability. BE users have the remaining transmit antennas $M_{BE} = M_T - M_{RT}$ for their transmission. After the number of transmit antennas for each QoS group is determined, the multiuser scheduling algorithm selects a user set from each QoS group sequentially using the QoS-aware sequential multiuser selection algorithm proposed in Section 4.4 of Chapter 4. It is assumed that the number of selected BE and RT users are the same as the number of determined transmit antennas for BE and RT users respectively. For the selection of RT users, delay fairness among RT users is employed. Similarly, for the selection of BE users, throughput fairness among BE users is considered. In other words, the proposed QoS-guaranteed multiuser scheduling algorithm in this configuration not only considers throughput and delay fairness among BE and RT users respectively but includes the delay constraint of RT users. For this, the proposed multiuser scheduling algorithm uses the *antenna trade-off scheme* for determining the number of transmit antennas for each QoS group and the *weighted sum-rate maximization rule* for selecting an optimal user set with fairness considerations. The weighted sum-rate maximization rule with delay fairness finds a user set S_{RT}^d among all possible user sets S_{RT} ($S_{RT} \subset U_{RT}, |S_{RT}| = M_{RT}$) according to

$$S_{RT}^d = \arg\max_{S_{RT}} \sum_{k=1}^{M_{RT}} \mu_k^d \log_2 \left(1 + \frac{b_{kk}^2 P}{M_{RT}}\right). \tag{6.23}$$

For the selection of BE users, the weighted sum-rate maximization rule with throughput fairness finds a user set S_{BE}^t among all possible user sets S_{BE} ($S_{BE} \subset U_{BE}, |S_{BE}| = M_{BE}$) and the 'already selected' RT user set S_{RT}^d according to

$$S_{BE}^t = \arg\max_{S_{BE}} \left\{ \sum_{k=1}^{M_{BE}} \mu_k^t \log_2 \left(1 + \frac{b_{kk}^2 P}{M_{BE}}\right) + C\left(S_{RT}^d\right) \right\}, \tag{6.24}$$

where $C\left(S_{RT}^d\right)$ is the weighted sum-rate capacity of the already selected RT user set S_{RT}^d, which is of the form

$$C\left(S_{RT}^d\right) = \sum_{i \in S_{RT}^d} \mu_i^d \log_2 \left(1 + \frac{b_{ii}^2 P}{M_{RT}}\right). \tag{6.25}$$

After user sets are selected from each QoS group is determined, THP is performed on the final user set $S_{\max}$ ($S_{\max} = S_{RT}^d \cup S_{BE}^t$) for transmission.

6.3.4 Numerical Results

In this section, numerical results using MATLAB for the proposed algorithm are presented. Throughout the simulations, we assume the number of transmit antennas $M_T = 4$, the median SNR of all users $SNR_0 = 10$dB and the distance for user k, $r_k^d = R_{cell}$. It is also assumed that the carrier frequency f_c = 2GHz and the velocity v_k = 5.4km/h for all users. For shadowing, standard deviation $\sigma_s = 4.3$dB and decorrelation distance D = 8.3m are assumed to emulate urban environments [15]. The data arrivals in queues for all users are assumed to be independent Poisson processes with arrival rate λ_d.

Several fairness indices have been suggested for the measurement of fairness for different resource allocation schemes. The Jain Fairness Index [108] [109] is one quantitative measure of fairness. Denoting the number of users as K, the Jain Fairness Index is given by

$$I_{Jain} = \frac{\left(\sum_{k=1}^{K} \gamma_k\right)^2}{K \sum_{k=1}^{K} \gamma_k^2}, \tag{6.26}$$

where γ_k is the fraction of transmission resource allocated to user k. In this thesis, a modified fairness index based on the Jain Fairness Index is proposed in order to consider different fairness requirements as

$$I_{QF} = \frac{1}{N_G} \sum_{i=1}^{N_G} \frac{\left(\sum_{j=1}^{K_i} \gamma_{ij}\right)^2}{K_i \sum_{j=1}^{K_i} \gamma_{ij}^2}, \tag{6.27}$$

where N_G is the number of subgroups with different fairness requirements, K_i ($\sum_{i=1}^{N_G} K_i = K$) is the number of users in the subgroup i and γ_{ij} is the fraction of transmission resource allocated to user j in the subgroup i, which satisfies $\sum_{j=1}^{K_i} \gamma_{ij} = 1$. When perfect fairness is achieved, $I_{QF} = 1$. I_{QF} decreases from one as degree of unfairness increases. When perfect unfairness (*i.e.*, all resources are allocated to one user) is achieved, $I_{QF} = 1/K$.

6.3.4.1 Throughput Fairness with the QF Scheduling Algorithm

Figure 6.3 and Figure 6.4 compare the throughput and the fairness index of the PF algorithm and the proposed QF algorithm for different smoothing factors α_t against the throughput slope m_t respectively. The number of users $K = 8$ and the constants a_k^t are assumed to be one. We notice that PF-THP and QF-THP have the same fairness index when the average throughout of each scheme is identical (cross-over point). For example, when $\alpha_t = 10$, these two algorithms

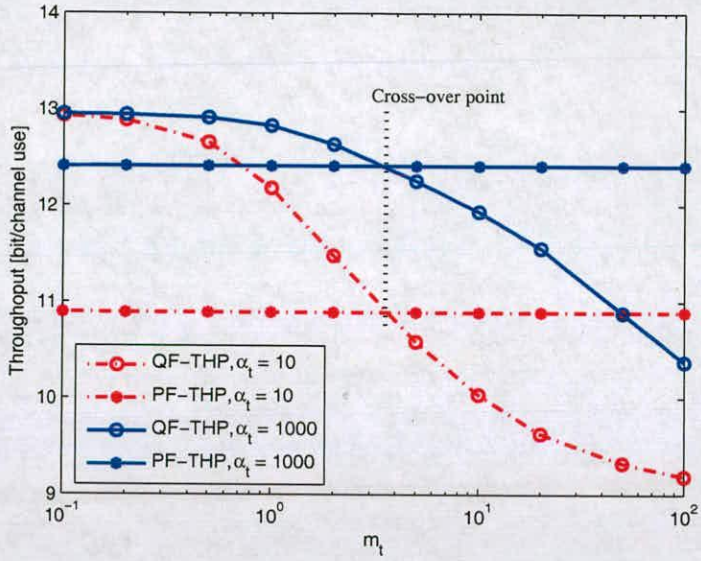


Figure 6.3: *The comparison between the PF algorithm and the proposed QF algorithm in terms of the average throughput for different smoothing factors against the throughput slope m_t, $M_T = 4$, $SNR_0 = 10dB$, f_c = 2GHz, v_k = 5.4km/h, σ_s = 4.3dB, $D = 8.3m$, $K = 8$ and $a_k^t = 1$.*

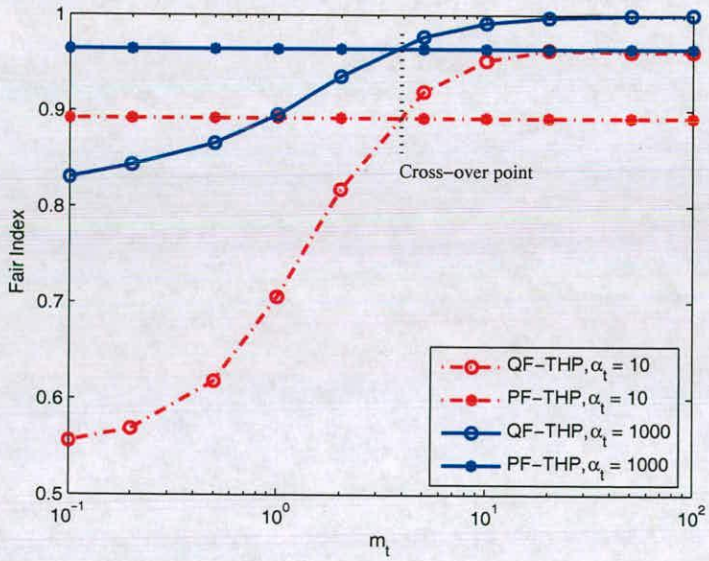


Figure 6.4: *The comparison between the PF algorithm and the proposed QF algorithm in terms of fairness index for different smoothing factors against the throughput slope m_t, $M_T = 4$, $SNR_0 = 10dB$, f_c = 2GHz, v_k = 5.4km/h, σ_s = 4.3dB, $D = 8.3m$, $K = 8$ and $a_k^t = 1$.*

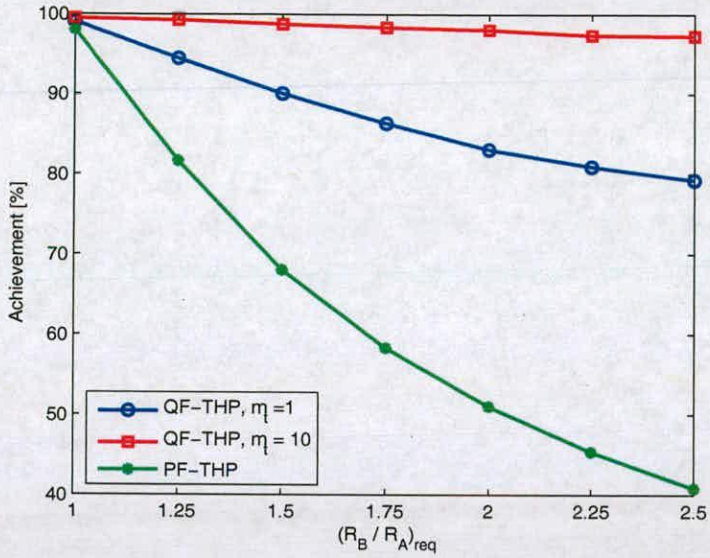


Figure 6.5: *The percentage achievement of different throughput requirements with the PF algorithm and the proposed QF algorithm,* $M_T = 4$, $SNR_0 = 10dB$, f_c = *2GHz,* v_k = *5.4km/h,* σ_s = *4.3dB,* $D = 8.3m$, $K = 8$ *and* $\alpha_t = 1000$.

have the same average throughput of about 10.9 [bits/channel use] with fairness index of about 0.90. This means that the performances of these two schemes are basically the same in terms of the average throughput and fairness. However, the proposed QF algorithm is able to control the degree of fairness by trading the average throughput for fairness among users. The larger m_t is, the higher the throughput fairness among users and the lower the throughput. The smoothing factor α_t determines the time window size for fairness. Larger time windows give more flexibility than smaller window sizes in selecting users due to the multiuser diversity. This increases the average throughput of all users. We notice that the case of $\alpha_t = 1000$ has larger average throughput and fairness index than the case of $\alpha_t = 10$. The percentage achievement is defined as

$$\frac{(R_B/R_A)_{out}}{(R_B/R_A)_{req}} \times 100 \ [\%], \tag{6.28}$$

where $(R_B/R_A)_{req}$ denotes the ratio of the required throughput of group B to group A and $(R_B/R_A)_{out}$ denotes the ratio of the actual throughput of group B to group A. Figure 6.5 shows the performance of the proposed QF algorithm for different throughput requirements. It is assumed that half of the users (group A) have a required throughput R_A and the others (group B) have a required throughput R_B. The smoothing factor $\alpha_t = 1000$ is also assumed. The smoothing factor $\alpha_t = 1000$ is also assumed. Because the PF algorithm does not consider

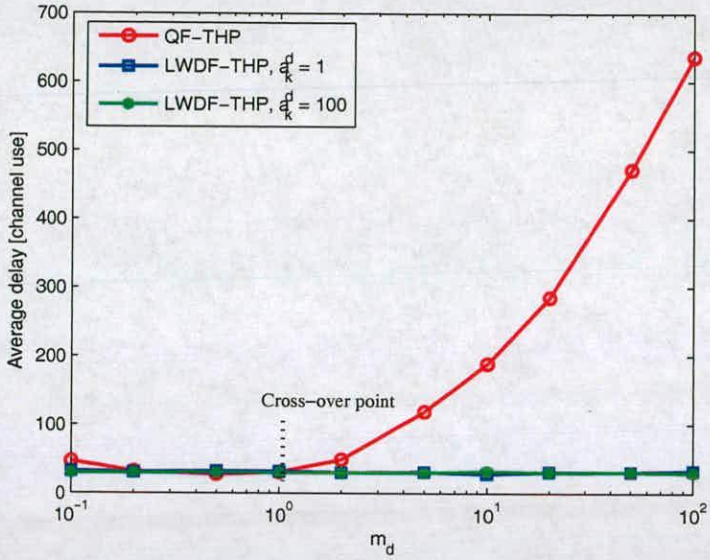


Figure 6.6: *The comparison between the LWDF algorithm and the proposed QF scheduling algorithm in terms of the average delay against the delay slope* m_d, $M_T = 4$, $SNR_0 = 10dB$, f_c = *2GHz*, v_k = *5.4km/h*, σ_s = *4.3dB*, $D = 8.3m$, $\lambda_d = 1.25$ *and* $K = 8$.

different throughput requirements among users, it performs the worst in terms of the percentage achievement to the required throughput ratio. A large throughput slope m_t means strict fairness among users, so that the case of $m_t = 10$ with QF-THP performs better than the case of $m_t = 1$ in terms of the degree of fairness, which corresponds to the trend of the results in Figure 6.4.

6.3.4.2 Delay Fairness with QF Scheduling Algorithm

Figure 6.6 and Figure 6.7 compare the average delay and the fairness index of the LWDF algorithm and the proposed QF algorithm for different smoothing factors α_d against the delay slope m_d respectively. The number of users $K = 8$ and the constants a_k^d are assumed to be one. The arrival rate $\lambda_d = 1.25$ [bits/channel use] is assumed for all users. Unlike the LWDF algorithm, the proposed QF algorithm can adjust the degree of fairness by changing the delay slope m_d. The average delay and the fairness index of the LWDF algorithm is the same regardless of the constants a_k^d. Similar to the case of throughput fairness, the performance of these two algorithms in terms of the average delay and the fairness index is the same when $m_d \simeq 1$. In this case,the fairness index is 0.82 and the average delay is about 25 [channel uses]. The delay fairness metric increases the degree of fairness by increasing the average delay of all

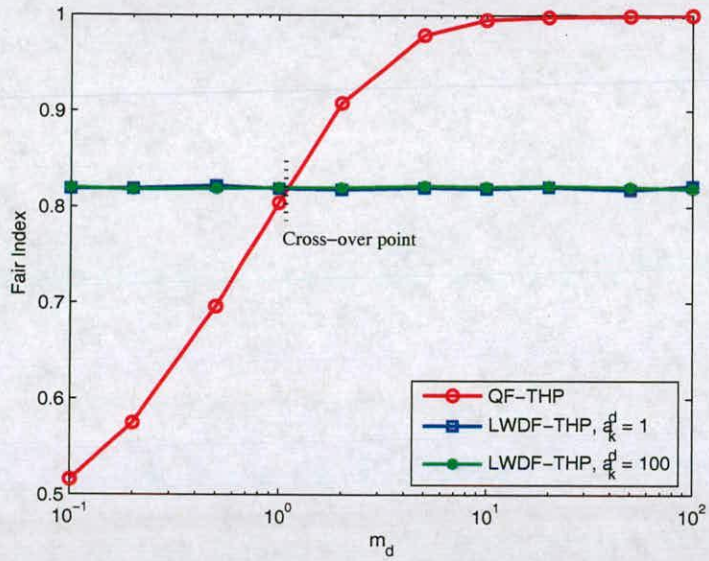


Figure 6.7: *The comparison between the LWDF algorithm and the proposed QF scheduling algorithm in terms of fairness index against the delay slope* m_d, $M_T = 4$, $SNR_0 = 10dB$, f_c = *2GHz*, v_k = *5.4km/h*, σ_s = *4.3dB*, $D = 8.3m$ $\lambda_d = 1.25$ *and* $K = 8$.

users. This results in the decrease of the average throughput of all users.

Figure 6.8 and Figure 6.9 show the effect of changing the degree of delay fairness on both the throughput and the delay performance respectively. As expected, the case of more strict delay fairness ($m_d = 10$) performs worst in terms of the throughput and delay performance. However, it achieves the highest degree of fairness in return for the performance degradation. The case of less strict delay fairness ($m_d = 0.1$) performs better than any other cases because it gives more flexibility than the case of $m_d = 10$ in selecting users for transmission. For example, when the arrival rate per user is 1.75 [bits/channel use], the average throughput when $m_d = 0.1$ is larger than the average throughput when $m_d = 10$ by about 1.2 [bits/channel use] if the proposed algorithm is used.

Figure 6.10 shows the performance of the proposed QF algorithm for different delay requirements. It is assumed that half of the users (group A) have a required delay D_A and the others (group B) have a required delay D_B. We notice that the proposed QF algorithm can satisfy the different delay requirements up to the ratio $(D_B/D_A)_{req}$ = 2.5 when $m_d = 10$. When $m_d = 1$, the proposed QF algorithm performs worse than the LWDF algorithm. However, when the delay slope m_d becomes large, the average delay of all users also increases, which results in a decrease of the average throughput of all users.

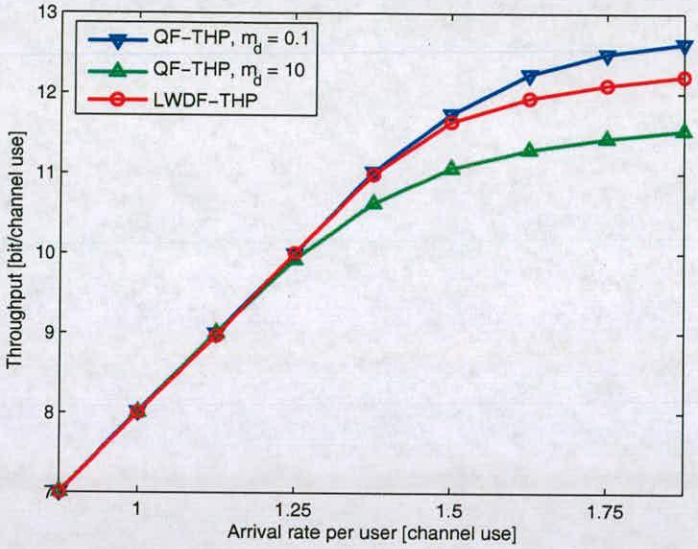


Figure 6.8: *Throughput performance of the LWDF algorithm and the proposed QF algorithm against different arrival rates per user* , $M_T = 4$, $SNR_0 = 10dB$, f_c = 2GHz, v_k = 5.4km/h, σ_s = 4.3dB, $D = 8.3m$, $K = 8$ *and* $a_k^d = 1$.

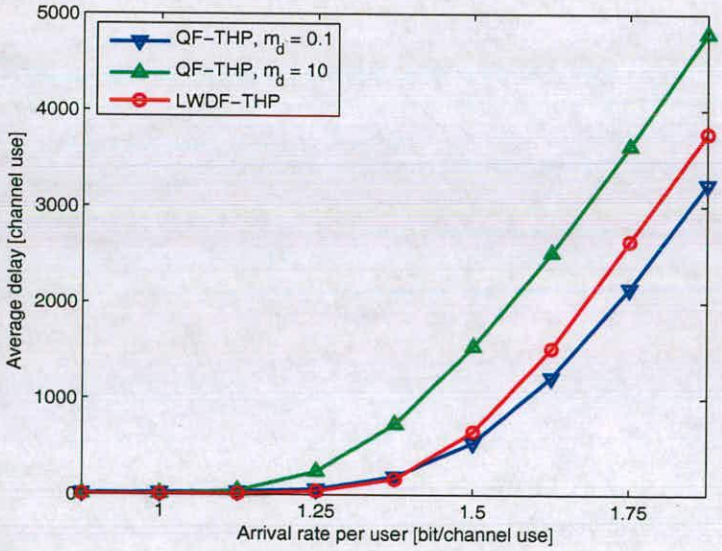


Figure 6.9: *Delay performance of the LWDF algorithm and the proposed QF algorithm against different arrival rates per user* , $M_T = 4$, $SNR_0 = 10dB$, f_c = 2GHz, v_k = 5.4km/h, σ_s = 4.3dB, $D = 8.3m$, $K = 8$ *and* $a_k^d = 1$.

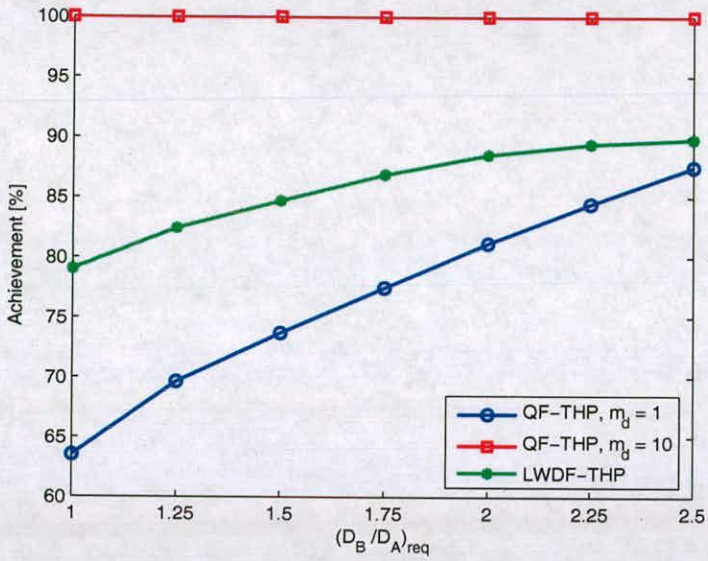


Figure 6.10: *The percentage achievement of different delay requirements with the LWDF algorithm and the proposed QF algorithm,* $M_T = 4$, $SNR_0 = 10dB$, f_c = *2GHz,* v_k = *5.4km/h,* σ_s = *4.3dB,* $D = 8.3m$ *and* $K = 8$.

6.3.4.3 Antenna Trade-off between Different QoS Groups

For the performance analysis of the proposed antenna trade-off scheme, it is assumed that there are 4 RT users ($K_{RT} = 4$) and 12 BE users ($K_{BE} = 12$) for each QoS group. The pre-assigned transmit antenna to RT users M_S is assumed to be one. For throughput fairness, $\alpha_t = 1000$ is also assumed. All users in BE and RT groups have the same throughput and delay requirements respectively. For the delay threshold for RT users, $D_{TH} = 50$ [channel use] is assumed. The parameter $\delta = 0.1$ is assumed for the dynamic range of the soft trade-off scheme.

Figure 6.11 and Figure 6.12 show the outage probability of the average delay of RT users exceeding the delay threshold D_{TH} and the average throughput with fairness considerations for different transmit antenna configurations respectively. We notice in Figure 6.11 that the soft trade-off scheme performs best among all transmit antenna configurations in terms of the outage probability because it can change the number of transmit antennas assigned to RT users gracefully according to the channel conditions. This enables the multiuser scheduling algorithm to minimize the outage probability of the average delay of RT users exceeding the delay threshold D_{TH}. The hard trade-off scheme performs worse than the soft trade-off scheme but better than the fixed scheme in terms of the outage probability. However, the fixed scheme performs best in terms of the throughput of all users (R_{TOTAL}) and that of BE users (R_{BE}) because it can al-

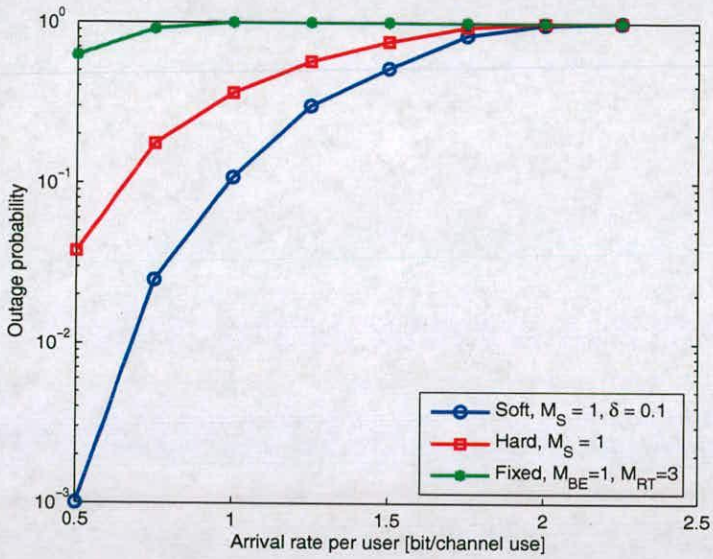


Figure 6.11: *The outage probability of the average delay of RT users exceeding the delay threshold D_{TH} for different antenna configurations, $M_T = 4$, $SNR_0 = 10dB$, f_c = 2GHz, v_k = 5.4km/h, σ_s = 4.3dB, $D = 8.3m$, $K_{RT} = 4$, $K_{BE} = 12$, $D_{TH} = 50$, $\alpha_t = 1000$, $a_k^t = 1$,$a_k^d = 1$ and $\delta = 0.1$.*

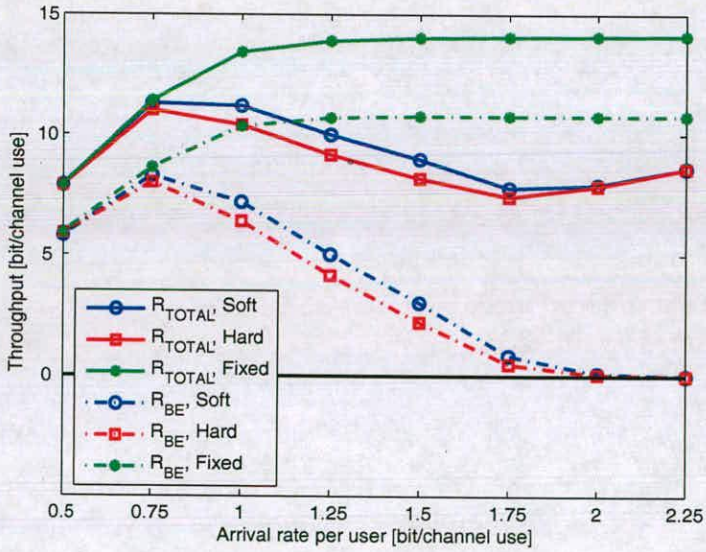


Figure 6.12: *The throughput performance for different antenna configuration, $M_T = 4$, $SNR_0 = 10dB$, f_c = 2GHz, v_k = 5.4km/h, σ_s = 4.3dB, $D = 8.3m$, $K_{RT} = 4$, $K_{BE} = 12$, $D_{TH} = 50$, $\alpha_t = 1000$, $a_k^t = 1$,$a_k^d = 1$ and $\delta = 0.1$.*

ways include BE users regardless of the delay status of RT users. This maximizes the multiuser diversity gain. Nevertheless, the throughput results of the fixed scheme are misleading because it cannot satisfy the delay constraints of RT users unlike the other antenna trade-off schemes. Between antenna trade-off schemes, the soft trade-off scheme performs also better than the hard trade-off scheme in terms of the throughput. When the arrival rate per user is greater than about 2 [bits/channel use], the throughput of the soft trade-off scheme is almost the same as that of the hard trade-off scheme. This is because the number of transmit antennas selected by the soft trade-off scheme is likely to be the same as that by the hard trade-off scheme when the average delay of RT users is much larger than the delay threshold. Indeed, the throughput of BE users with antenna trade-off schemes is almost zero when the arrival rate is greater than 2 [bits/channel use]. This is because RT users have higher priority and they take more transmit antennas than the pre-assigned value M_S very frequently to satisfy the delay constraint. This results in less chance of transmission for BE users.

6.4 Summary

This chapter has proposed a new QoS-guaranteed multiuser scheduling algorithm, which consists of a QoS-aware Fair (QF) scheduling for selecting users in the same QoS group with throughput or delay fairness among users. It also introduced a more advanced antenna trade-off scheme based on the antenna trade-off scheme proposed in Chapter 4 for supporting different QoS users simultaneously with QoS differentiation. For the selection of a user set with the proposed QF scheduling algorithm, the weighted sum-rate maximization rule is exploited so that the selected user set satisfies the different fairness requirements whilst maximizing the sum-rate capacity. The exponential function with a scaled fairness deviation as its argument is used for fairness among users in terms of throughput and delay respectively. The proposed QF scheduling algorithm can also control the degree of fairness by adjusting the slope of the exponential fairness metric. In serving BE and RT users with delay constraints simultaneously, the antenna trade-off scheme performs better than the fixed antenna scheme in terms of both the delay performance of RT users and the throughput performance of all users. Between antenna trade-off schemes, the soft trade-off scheme shows better performance in terms of the outage probability and the throughput due to the graceful adaption to the time-varying channel conditions.

Chapter 7
Conclusions

This chapter concludes this thesis. In Section 7.1, the contents of individual chapters are reviewed. Section 7.2 compares the objective of this thesis and the achievement. Some suggestions for further research are presented in Section 7.3

7.1 Review of Thesis Contents

Chapter 2 provided some background knowledge relevant to this thesis including the wireless channel, the MIMO channel and its capacity, and the basic concept of multiuser scheduling.

Chapter 3 compared the performance of MIMO transceiver techniques such as ZF-DPC, THP, ZFBF and V-BLAST in terms of error rates. THP performs slightly worse than V-BLAST with practical considerations such as the transmit power constraint and channel estimation errors due to the modulo loss. However, V-BLAST is not likely to be feasible for multiuser MIMO BC scenarios with a single receive antenna for each user because it requires antenna cooperation for successive detection at the receivers. Therefore, V-BLAST is not appropriate for the objective of this thesis. Among nonlinear MIMO precoding techniques, THP, despite its modulo loss, is much more feasible for practical implementation than ZF-DPC in terms of the error rate performance because the large power increase of ZF-DPC makes its performance too sensitive to practical impairments. This comparison showed that THP despite the modulo loss can be a practical MIMO precoding technique for real MIMO BC wireless systems.

Chapter 4 proposed a new QoS-aware sequential multiuser selection algorithm for MIMO broadcast channels with different QoS users. The proposed QoS-aware sequential multiuser selection algorithm can support different QoS users simultaneously by selecting user sets sequentially from the highest QoS group and changing the number of pre-assigned transmit antennas allocated to different QoS groups to improve the probability of QoS-guaranteed transmission as well as maximizing the sum-rate capacity. Numerical results showed that the proposed algorithm can support best effort and real time users simultaneously whilst satisfying the delay

constraint of real time users. It can also support constant rate, real time and best effort users simultaneously with the throughput constraint of constant rate users and the delay constraint of real time users. Indeed, the concept of multiuser MIMO capacity was briefly introduced and a comparison of MIMO precoding techniques such as ZF-DPC, THP and ZFBF in terms of the sum-rate capacity was performed. Numerical results showed that the modulo loss of THP could be ignored at high SNRs compared to ZF-DPC. Nonlinear THP performs better than linear ZFBF in terms of the sum-rate capacity despite the modulo loss. This justifies that THP is an appropriate MIMO precoding technique for real MIMO BC scenarios in terms of the sum-rate capacity, along with the results in Chapter 3, where THP was compared with other MIMO transceiver techniques in terms of error rates.

Chapter 5 proposed a temporally correlated MIMO BC simulation channel model, which has reasonable complexity but can describe realistic temporally-correlated MIMO broadcast channels with multiple antennas for a base station and a single receive antenna for each user. The proposed channel model was validated by comparing with the MIMO channel measurements. Numerical results showed that if there is only temporal correlation in MIMO broadcast channels, the proposed channel model can sufficiently represent the real MIMO broadcast channels. Based on the proposed channel model, a new SCSI-assisted multiuser scheduling algorithm was also proposed. Due to the temporal variation, mismatch in the channel estimate occurs, which affects the reliability of CSI at the transmitter. Because the degree of mismatch is different among users if their Doppler spreads are different, the proposed algorithm can minimize the effect of the channel mismatch by allocating users with high Doppler spreads to the front part of a frame using the statistical channel state information. By doing this, it can minimize the effect of scheduling delay due to the temporal variation, so that it can improve the sum-rate capacity.

Chapter 6 proposed a new QoS-guaranteed multiuser scheduling algorithm, which can support different QoS users simultaneously with the consideration of fairness among the same QoS users and QoS differentiation between different QoS groups. By selecting users sequentially, the proposed algorithm can support a mixture of QoS users whilst satisfying the system-level requirements. With the antenna trade-off scheme, the proposed algorithm can improve the probability of QoS-guaranteed transmission. For this, the more advanced soft trade-off scheme based on the hard trade-off scheme proposed in Chapter 4 was proposed, which performs better in terms of the outage probability by gracefully adapting to the time-varying channel conditions.

By using the weighted sum-rate maximization rule, the proposed algorithm can maximize the sum-rate capacity whilst satisfying fairness among users. Indeed, it can trade the degree of fairness with the total throughput by changing the weight vectors.

7.2 The Objectives and the Achievements of this Thesis

This section compares the five key areas related to the objective of this thesis presented in Chapter 1 and the results achieved in this thesis.

The five key areas and their corresponding achievements are as follows:

1. To investigate the performance of MIMO precoding techniques in terms of the sum-rate capacity as well as error rates.
    - The performance of THP was compared with MIMO transceiver techniques such as ZF-DPC, ZFBF and V-BLAST in an *i.i.d.* single user MIMO channel in terms of error rates in Chapter 3, and compared with MIMO precoding techniques such as ZF-DPC and ZFBF in an *i.i.d.* multiuser MIMO broadcast channel in terms of the sum-rate capacity in Chapter 4. The modulo loss of THP can be ignored with a high SNR assumption and exact CSI at the transmitter.
2. To propose a multiuser selection algorithm, which can support different QoS users simultaneously as well as maximizing the multiuser diversity gain.
    - A QoS-aware sequential multiuser selection algorithm can support a mixture of different QoS users simultaneously by selecting user sets from the highest QoS group sequentially and changing the number of transmit antennas allocated to each QoS group dynamically to increase QoS-guaranteed transmission. By applying the sum-rate maximization rule, the proposed algorithm find a user set of the most orthogonal users, which can maximize the sum-rate capacity.
3. To propose a MIMO broadcast channel model, which is realistic with reasonable complexity from the viewpoint of multiuser scheduling.
    - A comparison between the proposed temporally correlated MIMO broadcast channel model and the MIMO channel measurements showed that the proposed channel model can represent real temporally-correlated MIMO broadcast channels with

multiple antennas at a base station and a single antenna for each user with reasonable complexity.

4. To propose a multiuser scheduling algorithm, which exploits statistical channel state information to minimize performance degradation.

   - A SCSI-assisted multiuser scheduling algorithm can minimize the effect of scheduling delay to maximize the sum-rate capacity. Mismatch of channel estimates in temporally correlated MIMO broadcast channels decreases the sum-rate capacity but it can be minimized, unlike spatial correlation, if the proposed algorithm allocates users with high Doppler spreads to the first part of the time frame using statistical channel state information.

5. To propose a multiuser scheduling algorithm, which consider fairness among users with QoS differentiation whilst maximizing the sum-rate capacity

   - A QoS-guaranteed multiuser scheduling algorithm selects users sequentially from the highest QoS group considering fairness among users in the same QoS group. The weighted sum-rate maximization rule finds a user set of satisfying fairness in terms of throughput or delay whilst maximizing the sum-rate capacity. The spatial gain trade-off between different QoS groups can differentiate QoS services, so that higher priority QoS group has higher probability of QoS guaranteed transmission.

7.3 Suggestions for Further Work

The work of this thesis has considered the QoS-guaranteed multiuser scheduling algorithm combined with MIMO precoding techniques in realistic MIMO broadcast channels. However, there are a number of points which have not been addressed here and which merit much more work:

- In real MIMO BC scenarios, different services are likely to be based on the Internet, so that conventional traffic models may not describe their characteristics well. In Appendix D, the characteristics of real traffic models have been investigated. The self-similarity of traffic data degrades the delay performance due to burstiness. Therefore, a multiuser scheduling algorithm minimizing this effect could be investigated

- Although the weighted sum-rate maximization rule used in this thesis can achieve the sum-rate capacity whilst satisfying fairness among users, the computational complexity of the rule increases rapidly as the number of users grows. Therefore, any complexity reducing technique could be investigated.

- The channel model used in this thesis only considered the temporal correlation with a single receive antenna for each user. However, more realistic MIMO broadcast channel model could be considered by combining with spatial channel models including the saturation effect to simulate real MIMO broadcast channels with multiple antennas both for the base station and users.

- Although the channel state information at the transmitter is essential to any MIMO precoding techniques for the DPC principle, this thesis assumed only mismatch of channel estimates due to scheduling delay. Therefore, the effect of imperfect CSI could be included.

Appendix A
Publications

The author of this thesis has the following publications:

- S. Lee and J. S. Thompson, "Comparison of MIMO Transceiver Techniques with Practical Considerations," *12th European Wireless Conference EW2006, Athens, Greece*, April 2006.
- S. Lee and J. S. Thompson, "QoS-guaranteed User Selection in Multiuser MIMO Downlink Channels," *Accepted in 65th Semi-Annual IEEE Vehicular Technology Conference VTC2007-Spring, Dublin, Ireland*, April 2007.
- S.Lee and J. S. Thompson, "Trade-offs of Spatial gain for QoS-guaranteed Services in the MIMO Broadcast Channels," *Accepted in IEEE International Conference on Communications ICC2007, Glasgow, UK*, June 2007

In Preparation:

- S. Lee and J. S. Thompson, "A Simple Channel Model for Investigation of Multiuser Scheduling in MIMO Broadcast Channels," *67th Semi-Annual IEEE Vehicular Technology Conference VTC2008-Spring, Marina Bay, Singapore*, May. 2008.
- S. Lee and J. S. Thompson, "Statistical CSI-assisted Multiuser Scheduling in MIMO Broadcast Channels," *Submitted to IEEE Communication Letters.*
- S. Lee and J. S. Thompson, "QoS-guaranteed Multiuser Scheduling in MIMO Broadcast Channels," *Submitted to IEEE Transactions on Wireless Communications.*

Appendix B

The Capacity of the MIMO Channel

The capacity of a deterministic MIMO channel is defined as the maximum mutual information between input vector $\mathbf{s}$ and output vector $\mathbf{y}$ as

$$\begin{aligned} C &= \max_{f(\mathbf{s})} I(\mathbf{s};\mathbf{y}) \\ &= H(\mathbf{y}) - H(\mathbf{y}|\mathbf{s}), \end{aligned} \tag{B.1}$$

where $f(\mathbf{s})$ is the probability distribution of the vector $\mathbf{s}$, $I(\mathbf{s};\mathbf{y})$ is the mutual information between vectors $\mathbf{s}$ and $\mathbf{y}$, $H(\mathbf{y})$ is the differential entropy of the vector $\mathbf{y}$ and $H(\mathbf{y}|\mathbf{s})$ is the conditional differential entropy of the vector $\mathbf{y}$ given the knowledge of the vector $\mathbf{s}$. The maximum of $I(\mathbf{s};\mathbf{y})$ over input probability distribution is obtained when the input vector $\mathbf{s}$ is zero mean circular symmetric complex Gaussian (ZMCSCG) variable. Since the vector $\mathbf{s}$ and $\mathbf{z}$ are independent, $H(\mathbf{y}|\mathbf{s}) = H(\mathbf{z})$ [110]. Therefore the equation (B.1) simplifies to

$$\max_{f(\mathbf{s})} I(\mathbf{s};\mathbf{y}) = H(\mathbf{y}) - H(\mathbf{z}). \tag{B.2}$$

The differential entropy of vector $\mathbf{y}$ is [8]

$$\begin{aligned} H(\mathbf{y}) &= E\left[-\log_2 p(\mathbf{y})\right] \\ &= \log_2 |\pi \mathbf{R}_{yy}| + (\log_2 e)\, E\left[\mathbf{y}^H \mathbf{R}_{yy}^{-1} \mathbf{y}\right] \\ &= \log_2 |\pi \mathbf{R}_{yy}| + (\log_2 e)\, \mathrm{Tr}\left(E\left[\mathbf{y}\mathbf{y}^H\right] \mathbf{R}_{yy}^{-1}\right) \\ &= \log_2 |\pi \mathbf{R}_{yy}| + (\log_2 e)\, \mathrm{Tr}\left(\mathbf{I}\right) \\ &= \log_2 |\pi e \mathbf{R}_{yy}|, \end{aligned} \tag{B.3}$$

where $\mathbf{R}_{yy}$ is the covariance matrix of $\mathbf{y}$ and the probability distribution, $p(\mathbf{y})$ is

$$p(\mathbf{y}) = \det\left(\pi \mathbf{R}_{yy}\right)^{-1} \exp\left(-\mathbf{y}^H \mathbf{R}_{yy}^{-1} \mathbf{y}\right). \tag{B.4}$$

However, $\mathbf{R}_{yy}$ satisfies

$$\begin{aligned}\mathbf{R}_{yy} &= E\left[\mathbf{y}\mathbf{y}^H\right] \\ &= E\left[(\mathbf{H}\mathbf{s}+\mathbf{z})(\mathbf{H}\mathbf{s}+\mathbf{z})^H\right] \\ &= \mathbf{H}E\left[\mathbf{s}\mathbf{s}^H\right]\mathbf{H}^H + N_0\mathbf{I} \\ &= \mathbf{H}\mathbf{R}_{ss}\mathbf{H}^H + N_0\mathbf{I}. \end{aligned} \tag{B.5}$$

Therefore,

$$H(\mathbf{y}) = \log_2\left|\pi e\mathbf{H}\mathbf{R}_{ss}\mathbf{H}^H + \pi e N_0\mathbf{I}\right|. \tag{B.6}$$

Similarly,

$$H(\mathbf{z}) = \log_2\left|\pi e N_0\mathbf{I}\right|. \tag{B.7}$$

Then, $I(\mathbf{s};\mathbf{y})$ in the equation (B.1) becomes

$$I(\mathbf{s};\mathbf{y}) = \log_2\left|\mathbf{I} + \frac{\mathbf{H}\mathbf{R}_{ss}\mathbf{H}^H}{N_0}\right|. \tag{B.8}$$

The capacity of the MIMO channel is of the form

$$C = \max_{\mathrm{Tr}(\mathbf{R}_{ss})=P} \log_2\left|\mathbf{I} + \frac{\mathbf{H}\mathbf{R}_{ss}\mathbf{H}^H}{N_0}\right|,. \tag{B.9}$$

Appendix C
SNR Gap for Practical Coding Techniques

The Shannon capacity of an additive white Gaussian noise channel is given by

$$C = \log_2(1 + \text{SNR}). \tag{C.1}$$

This ideal result assumes infinite length of codebook to achieve very small bit error rates (BERs). However, uncoded QAM used in this thesis is far from capacity achieving in practice. This means that an extra amount power, called SNR gap, Γ_{SNR}, is required to achieve the Shannon capacity when uncoded QAM is used [52] [111]. The SNR gap Γ_{SNR} is a function of the BER and specific coding method. For example, for uncoded QAM with the target BER 10^{-3} and 10^{-6}, $\Gamma_{SNR} = 5.23$dB and $\Gamma_{SNR} = 8.8$dB respectively. When an ideal channel coding is used, Γ_{SNR} equals to 0dB. The capacity including the SNR gap is of the form

$$C = \log_2\left(1 + \frac{\text{SNR}}{\Gamma_{SNR}}\right). \tag{C.2}$$

Figure C.1 shows the effect of the SNR gap on the ergodic capacity of the MIMO channel.

This implies that the capacity results in this thesis have to be scaled according to the coding gain of practical channel coding techniques.

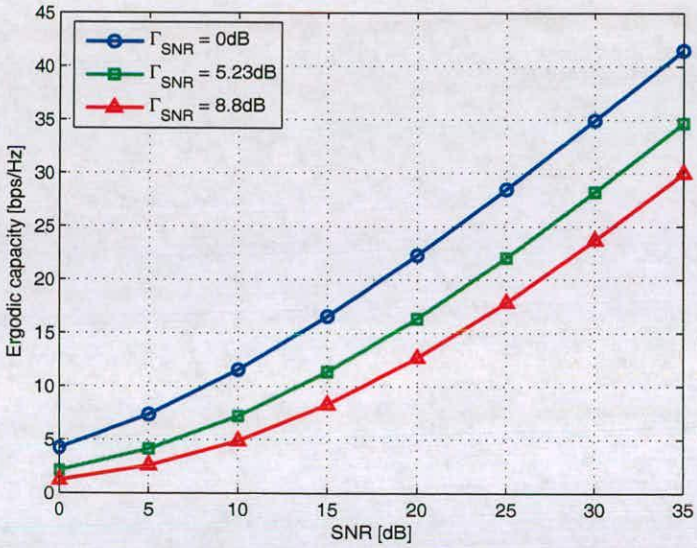


Figure C.1: *The ergodic capacity with different SNR gap values,* $M_T = 4$ *and* $M_R = 4$.

Appendix D
Real Traffic Models

Because different services in real MIMO BC scenarios are likely to be based on the Internet, the conventional telephony traffic models are not appropriate for real traffic data models. Most traditional traffic models developed for telephone systems are based on Markov models [112] [113]. They assume exponentially distributed inter-arrival times, which makes mathematical analysis simple. However, they cannot describe the behaviour of real traffic of data networks. An overview of traffic models in broadband networks is found in [114].

Empirical studies of traffic measurements demonstrated that real network traffic data has self-similarity [115] [116], which cannot be modelled by the conventional Markov models. Self-similarity implies that the traffic is similar to itself on all different time scales [117]. Self-similarity ('burstiness') preserves a certain property of a process irrespective of scaling in time. Long range dependence (LRD) means that there are statistically significant correlations across large time scales. LRD processes are referred to as asymptotically self-similar [118]. They have statistically significant correlations across large time scales. Long range dependence and self-similarity are the main properties of Internet traffic.

Self-similarity can be characterized by the Hurst parameter. The degree of self-similarity can be found by several methods such as the variance-time plot, the R/S plot, the periodiagram method and the Wavelet analysis [117] [118]. Traffic data is said to be self-similar if the Hurst parameter H is such that $0.5 < H < 1$.

There are several methods for constructing self-similar processes. Most of self-similar processes use heavy-tailed distribution to represent the burstiness of traffic data. The M/Pareto model overlaps many Poisson processes to make a burst process [119]. The number of Poisson processes has the Pareto distribution, which is given by cumulative distribution function

$$F(x) = 1 - \left(\frac{k}{x}\right)^{\alpha}, \tag{D.1}$$

where the constant $k > 0$, $0 < \alpha < 2$ and $x \geq k$. The constant α determines the decrease rate of the Pareto tail. For example, when α is large, the decrease rate is high. When plotted on a

log-log scale, as x increases, $1 - F(x)$ appears as a straight line with slope $-\alpha$. The ON/OFF model [116] overlaps many $i.i.d.$ alternating ON/OFF processes, where each ON and OFF duration has heavy-tailed distribution. The rate-limited extended alternating fractal renewal process (RL-EAFRP), which is based on the ON/OFF model, limits the data rate by using a cut-off Pareto distribution for the length of the generated data packet [120]. However, it does not limit the ON/OFF period. In this case, any user may experience an infinite OFF period, which should be excluded from an active user set by a multiuser scheduler. Users without data in their queues more than a certain period of time will be classified into an inactive user set, so that the multiuser scheduler only need to consider the active user set for its multiuser selection algorithm. Hence, we propose the modified RL-EAFRP traffic model, which uses the cut-off Pareto distribution for generating the length of data packets as well as the ON/OFF period, so that all users considered can be included in the active user set whilst maintaining self-similarity. The one parameter cut-off Pareto distribution is given by [121] [122]

$$F_T(x) = \frac{1 - \left(\frac{1}{1+x}\right)^{\alpha}}{1 - \left(\frac{1}{1+T_P}\right)^{\alpha}}, \tag{D.2}$$

where T_P is the cut-off limit and $x \geq 0$.

For the ON period, the OFF period and the length of data, α_1, α_0 and α_{data} are used respectively for α in the equation (D.2). The simulator can generate the cut-off Pareto-distributed random numbers easily by the inverse CDF method [123].

Figure D.1 shows examples of traffic data generated by the proposed cut-off Pareto traffic model. The cut-off limit of the ON/OFF periods are assumed to be 10,000 throughout this appendix. The parameter settings (T_P, α_1, α_0, α_{data}) of the examples are (200, 1.25, 1.75, 0.75) for RT video, (1500, 0.75, 0.75, 0.75) for FTP and (1500, 1.75, 0.75, 0.75) for HTTP.

Figure D.2 shows the performance degradation when the heavy-tailed traffic data is used. The plot was obtained with the delay fairness among users as in Chapter 6 with $M_T = 4$, $K = 8$, $m_d = 1$, $\alpha_1 = 1.25$, $\alpha_t = 1.25$ and $T_P = 10$. Small α_0 means large burstiness. Therefore, as α_0 decreases, the performance degradation becomes evident.

Figure D.3 shows the variance-time plot of the heavy-tailed traffic data with different α_0 and Poisson data with the same parameters as in Figure D.2. If the asymptotic slope, $\hat{\beta}$ ranges between -1 and 0 in the variance-time plot, this suggests self-similarity with the estimate of the

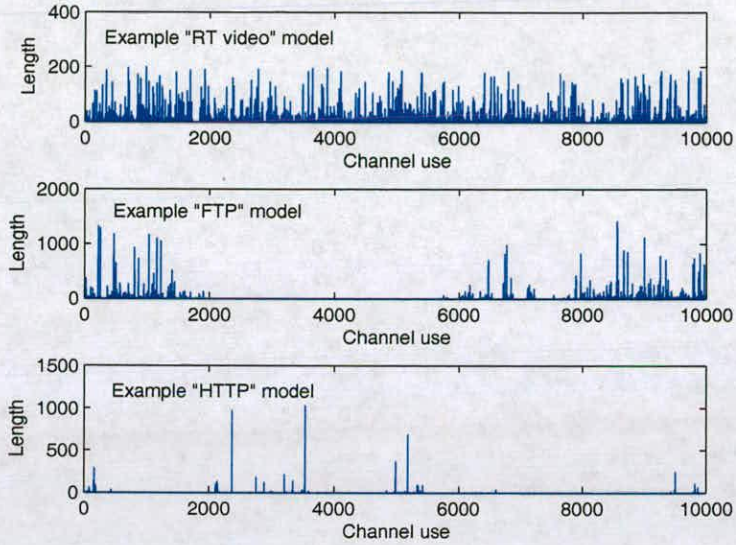


Figure D.1: *Examples of heavy-tailed traffic data*

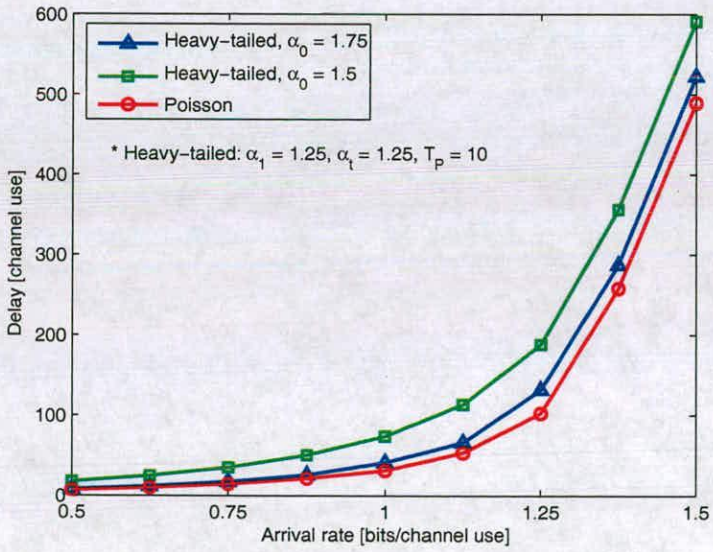


Figure D.2: *Delay performance with different traffic models,* $M_T = 4$, $K = 8$, $m_d = 1$, $\alpha_1 = 1.25$, $\alpha_t = 1.25$ *and* $T_P = 10$.

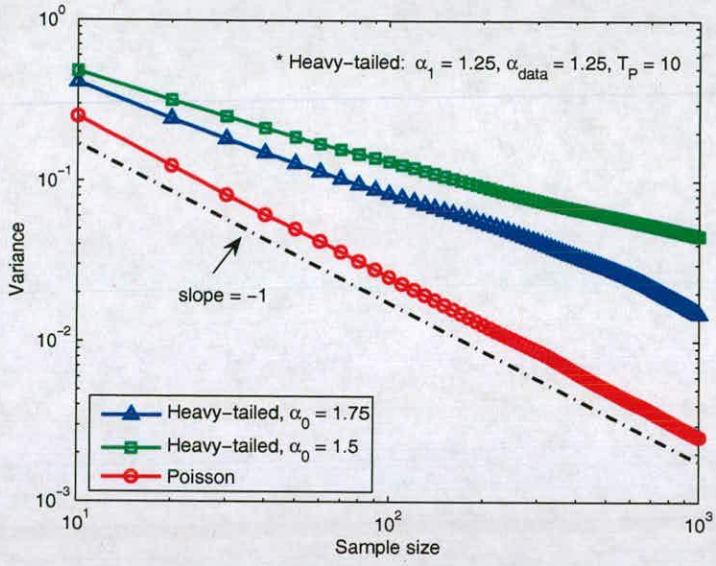


Figure D.3: *The variance-time plot of heavy-tailed traffic data with different* α_0 *and Poisson data* , $\alpha_1 = 1.25$, $\alpha_{data} = 1.25$ *and* $T_P = 10$.

Hurst parameter $H = 1 + \frac{\hat{\beta}}{2}$. We notice that the Poisson data has $\hat{\beta} = -1$, which corresponds to $H = 0.5$. This means that the Poisson data has no self-similarity. However, the heavy-tailed traffic data have self-similarity with the Hurst parameter between $0.5 < H < 1$.

References

[1] C. Peel, "On Dirty-Paper coding," *IEEE Signal Processing Magazine*, vol. 20, no. 3, pp. 112–113, 2003.

[2] A. Jalali, R. Padovani, R. Pankaj, Q. Inc, and C. San Diego, "Data throughput of CDMA-HDR a high efficiency-high data rate personal communication wireless system," *IEEE 51st Vehicular Technology Conference Proceedings, VTC 2000-Spring*, vol. 3, pp. 1854–1858, 2000.

[3] A. Stolyar and K. Ramanan, "Largest weighted delay first scheduling: large deviations and optimality," *The Annals of Applied Probability*, vol. 11, no. 1, pp. 1–48, 2001.

[4] M. Andrews, K. Kumaran, K. Ramanan, A. Stolyar, R. Vijayakumar, and P. Whiting, "CDMA data QoS scheduling on the forward link with variable channel conditions," *Bell Labs Technical Memorandum*, April 2000.

[5] S. Shakkottai and A. Stolyar, "Scheduling for multiple flows sharing a time-varying channel: The exponential rule," *Analytic Methods in Applied Probability*, vol. 207, no. 2, pp. 185–202, 2002.

[6] M. Hunukumbure and M. Beach, "MIMO channel measurements and analysis with prototype user devices in a 2GHz urban cell," *17th IEEE International Symposium on Personal, Indoor and Mobile Radio Communications, PIMRC 2006*, 2006.

[7] G. Foschini and M. Gans, "On Limits of wireless communications in a fading environment when using multiple antennas," *Wireless Personal Communications*, vol. 6, no. 3, pp. 311–335, 1998.

[8] I. Telatar, "Capacity of multi-antenna Gaussian channels," *European Transactions on Telecommunications*, vol. 10, no. 6, pp. 585–595, 1999.

[9] A. Paulraj, R. Nabar, and D. Gore, *Introduction to Space-Time Wireless Communications*. Cambridge University Press, 2003.

[10] A. Paulraj and C. Papadias, "Space-time processing for wireless communications," *IEEE Signal Processing Magazine*, vol. 14, no. 6, pp. 49–83, 1997.

[11] A. Goldsmith, *Wireless Communications*. Cambridge University Press, 2005.

[12] J. Proakis, *Digital Communications*. McGraw-Hill, 1983.

[13] T. Rappaport, *Wireless Communications: Principles and Practice*. IEEE Press Piscataway, NJ, USA, 1996.

[14] G. Stuber, *Principles of Mobile Communication*. Kluwer Academic Publishers Norwell, MA, USA, 1996.

[15] M. Gudmundson, "Correlation model for shadow fading in mobile radio systems," *Electronics Letters*, vol. 27, no. 23, pp. 2145–2146, 1991.

[16] E. Biglieri, J. Proakis, S. Shamai, and D. di Elettronica, "Fading channels: information-theoretic and communications aspects," *IEEE Transactions on Information Theory*, vol. 44, no. 6, pp. 2619–2692, 1998.

[17] W. Lee, "Effects on correlation between two mobile radio base-station antennas," *IEEE Transactions on Vehicular Technology*, vol. 22, no. 4, pp. 130–140, 1973.

[18] J. Fuhl, A. Molisch, and E. Bonek, "Unified channel model for mobile radio systems with smart antennas," *IEE Proceedings Radar, Sonar and Navigation*, vol. 145, no. 1, pp. 32–41, 1998.

[19] J. Kermoal, L. Schumacher, K. Pedersen, P. Mogensen, and F. Frederiksen, "A stochastic MIMO radio channel model with experimental validation," *IEEE Journal on Selected Areas in Communications*, vol. 20, no. 6, pp. 1211–1226, 2002.

[20] A. Molisch, "A generic model for MIMO wireless propagation channels in macro-and microcells," *IEEE Transactions on Signal Processing*, vol. 52, no. 1, pp. 61–71, 2004.

[21] A. Burr, "Capacity bounds and estimates for the finite scatterers MIMO wireless channel," *IEEE Journal on Selected Areas in Communications*, vol. 21, no. 5, pp. 812–818, 2003.

[22] R. Ertel, P. Cardieri, K. Sowerby, T. Rappaport, and J. Reed, "Overview of spatial channel models for antenna array communication systems," *IEEE Personal Communications*, vol. 5, no. 1, pp. 10–22, 1998.

[23] P. Almers, E. Bonek, A. Burr, N. Czink, M. Debbah, V. Degli-Esposti, H. Hofstetter, P. Kyosti, D. Laurenson, G. Matz, A. F. Molisch, C. Oestges, and H. Ozcelik, "Survey of Channel and Radio Propagation Models for Wireless MIMO Systems," *EURASIP Journal on Wireless Communications and Networking*, vol. 2007, pp. 1–19, 2007.

[24] H. Xu, D. Chizhik, H. Huang, and R. Valenzuela, "A generalized space-time multiple-input multiple-output (MIMO) channel model," *IEEE Transactions on Wireless Communications*, vol. 3, no. 3, pp. 966–975, 2004.

[25] K. Yu, M. Bengtsson, B. Ottersten, D. McNamara, P. Karlsson, and M. Beach, "Modeling of wide-band MIMO radio channels based on NLoS indoor measurements," *IEEE Transactions on Vehicular Technology*, vol. 53, no. 3, pp. 655–665, 2004.

[26] J. Salz and J. Winters, "Effect of fading correlation on adaptive arrays in digital mobile radio," *IEEE Transactions on Vehicular Technology*, vol. 43, no. 4, pp. 1049–1057, 1994.

[27] H. Ozcelik, M. Herdin, W. Weichselberger, J. Wallace, and E. Bonek, "Deficiencies of'Kronecker'MIMO radio channel model," *Electronics Letters*, vol. 39, no. 16, pp. 1209–1210, 2003.

[28] D. Gesbert, H. Bolcskei, D. Gore, and A. Paulraj, "Outdoor MIMO wireless channels: models and performance prediction," *IEEE Transactions on Communications*, vol. 50, no. 12, pp. 1926–1934, 2002.

[29] G. Byers and F. Takawira, "Spatially and temporally correlated MIMO channels: modeling and capacity analysis," *IEEE Transactions on Vehicular Technology*, vol. 53, no. 3, pp. 634–643, 2004.

[30] P. Petrus, J. Reed, and T. Rappaport, "Geometrical-based statistical macrocell channel model for mobileenvironments," *IEEE Transactions on Communications*, vol. 50, no. 3, pp. 495–502, 2002.

[31] A. van Zelst and J. Hammerschmidt, "A single coefficient spatial correlation model for multiple-input multiple-output (MIMO) radio channels," *Proceedings of 27th General Assembly of the International Union of Radio Science*, pp. 1–4, 2002.

[32] D. Shiu, G. Foschini, M. Gans, J. Kahn, Q. Inc, and C. Santa Clara, "Fading correlation and its effect on the capacity of multi-element antenna systems," *IEEE Transactions on Communications*, vol. 48, no. 3, pp. 502–513, 2000.

[33] C. Chuah, D. Tse, J. Kahn, and R. Valenzuela, "Capacity scaling in MIMO wireless systems under correlated fading," *IEEE Transactions on Information Theory*, vol. 48, no. 3, pp. 637–650, 2002.

[34] C. Shannon, "A mathematical theory of communication (parts I and II)," *Bell System Technical Journal*, vol. 27, no. Part I, pp. 379–423, 1948.

[35] T. Cover and J. Thomas, *Elements of Information Theory*. Wiley New York, 1991.

[36] M. Andrews, K. Kumaran, K. Ramanan, A. Stolyar, P. Whiting, and R. Vijayakumar, "Providing quality of service over a shared wireless link," *IEEE Communications Magazine*, vol. 39, no. 2, pp. 150–154, 2001.

[37] T. View, "Queuing with adaptive modulation and coding over wireless links: cross-Layer analysis and design," *IEEE Transactions on Wireless Communications*, vol. 4, no. 3, pp. 1142–1153, 2005.

[38] T. Yoo and A. Goldsmith, "Capacity of fading MIMO channels with channel estimation error," *IEEE International Conference on Communications*, vol. 2, 2004.

[39] C. Windpassinger, *Detection and Precoding for Multiple Input Multiple Output Channels*. PhD thesis, Dissertation University Erlangen, June 2004.

[40] U. Fincke and M. Pohst, "Improved Methods for Calculating Vectors of Short Length in a Lattice, Including a Complexity Analysis," *Mathematics of Computation*, vol. 44, no. 170, pp. 463–471, 1985.

[41] P. Wolniansky, G. Foschini, G. Golden, and R. Valenzuela, "V-BLAST: an architecture for realizing very high data rates overthe rich-scattering wireless channel," *International Symposium on Signals, Systems, and Electronics*, pp. 295–300, 1998.

[42] G. Golden, C. Foschini, R. Valenzuela, and P. Wolniansky, "Detection algorithm and initial laboratory results using V-BLASTspace-time communication architecture," *Electronics Letters*, vol. 35, no. 1, pp. 14–16, 1999.

[43] G. Ginis and J. Cioffi, "On the relation between V-BLAST and the GDFE," *IEEE Communications Letters*, vol. 5, no. 9, pp. 364–366, 2001.

[44] M. Debbah, B. Muquet, M. de Courville, M. Muck, S. Simoens, and P. Loubaton, "A MMSE successive interference cancellation scheme for a newadjustable hybrid spread OFDM system," *IEEE 51st Vehicular Technology Conference Proceedings, VTC 2000-Spring Tokyo*, vol. 2, 2000.

[45] B. Van Veen and K. Buckley, "Beamforming: a versatile approach to spatial filtering," *IEEE Signal Processing Magazine*, vol. 5, no. 2, pp. 4–24, 1988.

[46] T. Yoo and A. Goldsmith, "On the optimality of multiantenna broadcast scheduling using zero-forcing beamforming," *IEEE Journal on Selected Areas in Communications*, vol. 24, no. 3, pp. 528–541, 2006.

[47] M. Costa, "Writing on dirty paper.," *IEEE Transactions on Information Theory*, vol. 29, no. 3, pp. 439–441, 1983.

[48] G. Caire and S. Shamai, "On the achievable throughput of a multiantenna Gaussian broadcast channel," *IEEE Transactions on Information Theory*, vol. 49, no. 7, pp. 1691–1706, 2003.

[49] C. Windpassinger, R. Fischer, T. Vencel, and J. Huber, "Precoding in multiantenna and multiuser communications," *IEEE Transactions on Wireless Communications*, vol. 3, no. 4, pp. 1305–1316, 2004.

[50] H. Harashima and H. Miyakawa, "Matched-Transmission Technique for Channels With Intersymbol Interference," *IEEE Transactions on Communications*, vol. 20, no. 4, pp. 774–780, 1972.

[51] W. Yu, D. Varodayan, and J. Cioffi, "Trellis and convolutional precoding for transmitter-based interference presubtraction," *IEEE Transactions on Communications*, vol. 53, no. 7, pp. 1220–1230, 2005.

[52] C. Fung, W. Yu, and T. Lim, "Precoding for the Multi-Antenna Downlink: Multiuser SNR Gap and Optimal User Ordering," *IEEE Transactions on Communications*, vol. 55, no. 1, pp. 188–197, 2007.

[53] R. Fischer, *Precoding and Signal Shaping for Digital Transmission*. Wiley-Interscience, 2002.

[54] R. Wesel and J. Cioffi, "Precoding and the MMSE-DFE," *Conference on Signals, Systems and Computers*, vol. 2, pp. 1144–1148, 1994.

[55] R. Wesel and J. Cioffi, "Achievable rates for Tomlinson-Harashima precoding," *IEEE Transactions on Information Theory*, vol. 44, no. 2, pp. 824–831, 1998.

[56] G. Forney Jr and G. Ungerboeck, "Modulation and coding for linear Gaussian channels," *IEEE Transactions on Information Theory*, vol. 44, no. 6, pp. 2384–2415, 1998.

[57] R. Fischer, C. Stierstorfer, and C. Windpassinger, "Precoding and signal shaping for transmission over MIMO channels," *Proc. 2003 Canadian Workshop on Inf. Theory*, pp. 83–87.

[58] C. Peel, B. Hochwald, and A. Swindlehurst, "A vector-perturbation technique for near-capacity multiantenna multiuser communication-part I: channel inversion and regularization," *IEEE Transactions on Communications*, vol. 53, no. 1, pp. 195–202, 2005.

[59] B. Hochwald, C. Peel, and A. Swindlehurst, "A Vector-Perturbation Technique for Near-Capacity Multiantenna Multiuser Communication-Part II: Perturbation," *IEEE Transactions on Communications*, vol. 53, no. 3, pp. 537–544, 2005.

[60] R. Fischer and J. Huber, "A new loading algorithm for discrete multitone transmission," *Global Telecommunications Conference*, vol. 1, pp. 724–728, 1996.

[61] T. Vencel, C. Windpassinger, and R. Fischer, "Sorting in the V-BLAST Algorithm and Loading," *Proceedings of CSN 2002, Malaga, Spain*, pp. 304–309, 2002.

[62] C. Windpassinger, T. Vencel, and R. Fischer, "Precoding and loading for BLAST-like systems," *IEEE International Conference on Communications*, vol. 5, pp. 3061–3065, 2003.

[63] M. Joham, J. Brehmer, and W. Utschick, "MMSE Approaches to Multiuser Spatio-Temporal Tomlinson-Harashima Precoding," *Proceedings of ITG SCC*, vol. 4, pp. 387–394, 2004.

[64] O. Simeone, Y. Bar-Ness, and U. Spagnolini, "Linear and nonlinear preequalization/equalization for MIMO systems with long-term channel state information at the transmitter," *IEEE Transactions on Wireless Communications*, vol. 3, no. 2, pp. 373–378, 2004.

[65] S. Shakkottai, T. Rappaport, and P. Karlsson, "Cross-layer design for wireless networks," *IEEE Communications Magazine*, vol. 41, no. 10, pp. 74–80, 2003.

[66] S. Catreux, P. Driessen, L. Greenstein, I. Wireless, and C. San Jose, "Data throughputs using multiple-input multiple-output (MIMO) techniques in a noise-limited cellular environment," *IEEE Transactions on Wireless Communications*, vol. 1, no. 2, pp. 226–235, 2002.

[67] O. Shin and K. Bok, "Antenna-assisted round robin scheduling for MIMO cellular systems," *IEEE Communications Letters*, vol. 7, no. 3, pp. 109–111, 2003.

[68] A. Goldsmith, S. Jafar, N. Jindal, and S. Vishwanath, "Capacity limits of MIMO channels," *IEEE Journal on Selected Areas in Communications*, vol. 21, no. 5, pp. 684–702, 2003.

[69] Q. Spencer, C. Peel, A. Swindlehurst, and M. Haardt, "An introduction to the multi-user MIMO downlink," *IEEE Communications Magazine*, vol. 42, no. 10, pp. 60–67, 2004.

[70] S. Jafar, G. Foschini, and A. Goldsmith, "PhantomNet: exploring optimal multicellular multiple antenna systems," *EURASIP Journal on Applied Signal Processing*, vol. 2004, no. 5, pp. 591–604, 2004.

[71] W. Yu, W. Rhee, S. Boyd, and J. Cioffi, "Iterative water-filling for Gaussian vector multiple-access channels," *IEEE Transactions on Information Theory*, vol. 50, no. 1, pp. 145–152, 2004.

[72] S. Vishwanath, N. Jindal, and A. Goldsmith, "Duality, achievable rates, and sum-rate capacity of Gaussian MIMO broadcast channels," *IEEE Transactions on Information Theory*, vol. 49, no. 10, pp. 2658–2668, 2003.

[73] H. JWeingarten, Y. Steinberg, and S. Shamai, "The Capacity region of the Gaussian multiple-input multiple-output broadcast channel," *IEEE Transactions on Information Theory*, vol. 52, no. 9, pp. 3936–3964, 2006.

[74] S. Vishwanath, N. Jindal, and A. Goldsmith, "On the capacity of multiple input multiple output broadcast channels," *IEEE International Conference on Communications, ICC 2002,*, vol. 3, pp. 1444–1450.

[75] N. Jindal and A. Goldsmith, "Dirty-paper coding versus TDMA for MIMO Broadcast channels," *IEEE Transactions on Information Theory*, vol. 51, no. 5, pp. 1783–1794, 2005.

[76] H. Viswanathan, S. Venkatesan, and H. Huang, "Downlink capacity evaluation of cellular networks with known-interference cancellation," *IEEE Journal on Selected Areas in Communications*, vol. 21, no. 5, pp. 802–811, 2003.

[77] W. Rhee and J. Cioffi, "On the capacity of multiuser wireless channels with multiple antennas," *IEEE Transactions on Information Theory*, vol. 49, no. 10, pp. 2580–2595, 2003.

[78] W. Yu and J. Cioffi, "Sum capacity of Gaussian vector broadcast channels," *IEEE Transactions on Information Theory*, vol. 50, no. 9, pp. 1875–1892, 2004.

[79] P. Viswanath, D. Tse, and R. Laroia, "Opportunistic beamforming using dumb antennas," *IEEE Transactions on Information Theory*, vol. 48, no. 6, pp. 1277–1294, 2002.

[80] C. Swannack, E. Uysal-Biyikoglu, and G. Wornell, "Low complexity multiuser scheduling for maximizing throughput in the MIMO broadcast channel," *Proc. Allerton Conf. Communications, Control and Computing*, 2004.

[81] T. Yoo and A. Goldsmith, "Sum-Rate Optimal Multi-Antenna Downlink Beamforming Strategy Based On Clique Search," *IEEE Global Telecommunications Conference, GLOBECOM'05*, vol. 3, pp. 1510–1514.

[82] N. Jindal, W. Rhee, S. Vishwanath, S. Jafar, and A. Goldsmith, "Sum power iterative water-filling for multi-antenna Gaussian broadcast channels," *IEEE Transactions on Information Theory*, vol. 51, no. 4, pp. 1570–1580, 2005.

[83] J. Lee and N. Jindal, "Dirty Paper Coding vs. Linear Precoding for MIMO Broadcast Channels," *Proceedings of Asilomar Conf. on Signals, Systems, & Comp*, 2006.

[84] A. Molisch and M. Win, "MIMO systems with antenna selection," *IEEE Microwave Magazine*, vol. 5, no. 1, pp. 46–56, 2004.

[85] S. Sanayei and A. Nosratinia, "Antenna selection in MIMO systems," *IEEE Communications Magazine*, vol. 42, no. 10, pp. 68–73, 2004.

[86] S. Alamouti, "A simple transmit diversity technique for wireless communications," *IEEE Journal on Selected Areas in Communications*, vol. 16, no. 8, pp. 1451–1458, 1998.

[87] L. Zheng and D. Tse, "Diversity and multiplexing: a fundamental tradeoff in multiple-antenna channels," *IEEE Transactions on Information Theory*, vol. 49, no. 5, pp. 1073–1096, 2003.

[88] D. Tse, P. Viswanath, and L. Zheng, "Diversity–multiplexing tradeoff in multiple-access channels," *IEEE Transactions on Information Theory*, vol. 50, no. 9, pp. 1859–1874, 2004.

[89] R. Zhang, J. Cioffi, and Y. Liang, "Throughput comparison of wireless downlink transmission schemes with multiple antennas," *IEEE International Conference on Communications, ICC 2005*, vol. 4, pp. 2700–2704, 2005.

[90] W. Ajib, D. Haccoun, and J. Frigon, "An efficient QoS-based scheduling algorithm for MIMO wireless systems," *IEEE 62nd Vehicular Technology Conference, VTC-2005-Fall*, vol. 3, pp. 1579–1583, 2005.

[91] M. Patzold and V. Nguyen, "A spatial simulation model for shadow fading processes in mobile radio channels," *15th IEEE International Symposium on Personal, Indoor and Mobile Radio Communications, PIMRC 2004*, vol. 3, pp. 1832–1838, 2004.

[92] Y. Zheng and C. Xiao, "Simulation models with correct statistical properties for Rayleigh fading channels," *IEEE Transactions on Communications*, vol. 51, no. 6, pp. 920–928, 2003.

[93] W. Jakes, *Microwave Mobile Communications, Piscataway*, NJ: IEEE Press, 1993.

[94] M. Vu and A. Paulraj, "A robust transmit CSI framework with applications in MIMO wireless precoding," *Proc. 39th Asilomar Conf. Signals Systems, Computers*, pp. 623–627, 2005.

[95] Q. Sun, D. Cox, H. Huang, and A. Lozano, "Estimation of continuous flat fading MIMO channels," *Wireless Communications and Networking Conference, 2002. WCNC2002. 2002 IEEE*, vol. 1, pp. 549–553, 2002.

[96] M. Beach, M. Hunukumbure, P. Rogers, and M. Oxley, "Device Charaterization, Trial Plan and Initial Measurement Analysis ," *Internal Report, Telecommunications Research Laboratory, Toshiba Research Europe Limited*, 2005.

[97] M. Beach, M. Hunukumbure, and P. Rogers, "An overview of the pre-trial and its results ," *Internal Report, Telecommunications Research Laboratory, Toshiba Research Europe Limited*, 2005.

[98] M. Beach, C. Williams, A. Nix, M. Hunukumbure, P. Rogers, E. Tameh, and D. Kitchener, "Requirements capture and development issues," *Internal Report, Telecommunications Research Laboratory, Toshiba Research Europe Limited*, 2005.

[99] S. Dragan, M. Narayan, and C. Dmitry, "Adaptive transmitter optimization in multiuser multiantenna systems: theoretical limits, effect of delays, and performance enhancements," *EURASIP Journal on Wireless Communications and Networking*, no. 3, pp. 298–307, 2005.

[100] R. Fischer, C. Stierstorfer, and J. Huber, "Precoding for point-to-multipoint transmission over MIMO ISI channels," *International Zurich Seminar on Communications*, pp. 208–211, 2004.

[101] S. Catreux, V. Erceg, D. Gesbert, R. Heath Jr, I. Inc, and C. San Jose, "Adaptive modulation and MIMO coding for broadband wireless datanetworks," *IEEE Communications Magazine*, vol. 40, no. 6, pp. 108–115, 2002.

[102] A. Goldsmith and S. Chua, "Variable-rate variable-power MQAM for fading channels," *IEEE Transactions on Communications*, vol. 45, no. 10, pp. 1218–1230, 1997.

[103] R. Berry and E. Yeh, "Cross-layer wireless resource allocation," *IEEE Signal Processing Magazine*, vol. 21, no. 5, pp. 59–68, 2004.

[104] T. View, "Cross-layer wireless multimedia transmission: challenges, principles, and new paradigms," *IEEE Wireless Communications*, vol. 12, no. 4, pp. 50–58, 2005.

[105] S. Shakkottai and A. Stolyar, "Scheduling algorithms for a mixture of real-time and non-real-time data in HDR," *Proceedings of International Teletraffic Congress (ITC)*, pp. 793–804, 2001.

[106] D. Haccoun, "An Overview of Scheduling Algorithms in MIMO-Based Fourth-Generation Wireless Systems," *IEEE Network*, pp. 43–48, 2005.

[107] G. Ginis and J. Cioffi, "Vectored transmission for digital subscriber line systems," *IEEE Journal on Selected Areas in Communications*, vol. 20, no. 5, pp. 1085–1104, 2002.

[108] R. Jain, D. Chiu, and W. Hawe, "A quantitative measure of fairness and discrimination for resource allocation in shared computer systems," *Digital Equipment Corp. Eastern Research Lab, DEC-TR-301*, pp. 1–37, 1984.

[109] G. Berger-Sabbatel, A. Duda, M. Heusse, and F. Rousseau, "Short-term fairness of 802.11 networks with several hosts," *Proceedings of the 6th IFIP/IEEE International Conference on Mobile and Wireless Communication Networks.*

[110] R. Blahut, *Principles and practice of information theory.* Addison-Wesley Longman Publishing Co., Inc. Boston, MA, USA, 1987.

[111] S. Chung, A. Lozano, H. Huang, A. Sutivong, and J. Cioffi, "Approaching the MIMO Capacity with a Low-Rate Feedback Channel in V-BLAST," *EURASIP Journal on Applied Signal Processing*, vol. 2004, no. 5, pp. 762–771, 2004.

[112] L. Rabiner and B. Juang, "An introduction to hidden Markov models," *IEEE Signal Processing Magazine*, vol. 3, no. 1 Part 1, pp. 4–16, 1986.

[113] H. Heffes and D. Lucantoni, "A Markov Modulated Characterization of Packetized Voice and Data Traffic and Related Statistical Multiplexer Performance," *IEEE Journal on Selected Areas in Communications*, vol. 4, no. 6, pp. 856–868, 1986.

[114] A. Adas, "Traffic models in broadband networks," *IEEE Communications Magazine*, vol. 35, no. 7, pp. 82–89, 1997.

[115] W. Leland, M. Taqqu, W. Willinger, and D. Wilson, "On the self-similar nature of Ethernet traffic (extended version)," *IEEE/ACM Transactions on Networking*, vol. 2, no. 1, pp. 1–15, 1994.

[116] W. Willinger, M. Taqqu, R. Sherman, and D. Wilson, "Self-similarity through high-variability: statistical analysis of Ethernet LAN traffic at the source level," *IEEE/ACM Transactions on Networking (TON)*, vol. 5, no. 1, pp. 71–86, 1997.

[117] O. Cappe, E. Moulines, J. Pesquet, A. Petropulu, and X. Yang, "Long-range dependence and heavy-tail modeling for teletraffic data," *IEEE Signal Processing Magazine*, vol. 19, no. 3, pp. 14–27, 2002.

[118] T. Karagiannis, M. Molle, and M. Faloutsos, "Long-range dependence ten years of Internet traffic modeling," *IEEE Internet Computing*, vol. 8, no. 5, pp. 57–64, 2004.

[119] R. Addie, M. Zukerman, and T. Neame, "Broadband traffic modeling: simple solutions to hard problems," *IEEE Communications Magazine*, vol. 36, no. 8, pp. 88–95, 1998.

[120] J. Yu, A. Petropulu, and H. Sethu, "Rate-limited EAFRP-a new improved model for high-speed network traffic," *IEEE Transactions on Signal Processing*.

[121] M. Fischer, D. Masi, D. Gross, and J. Shortle, "One-Parameter Pareto, Two-Parameter Pareto, Three-Parameter Pareto: Is There a Modeling Difference?," *The Telecommunications Review*, pp. 79–92, 2005.

[122] D. Gross, J. Shortle, M. Fischer, and D. Masi, "Simulation input analysis: difficulties in simulating queues with Pareto service," *Proceedings of the 34th conference on Winter simulation: exploring new frontiers*, pp. 407–415, 2002.

[123] J. Gentle, *Random Number Generation and Monte Carlo Methods*. Springer, 2003.

Comparison of MIMO Transceiver Techniques with Practical Considerations

S. H. Lee and J. S. Thompson

University of Edinburgh, Institute for Digital Communications, Edinburgh, UK
{S.Lee|John.Thompson}@ed.ac.uk

Abstract – **Recently, multiple-input multiple-output (MIMO) techniques have received a lot of attention because of their high spectral efficiency. Recent studies have been focused on the practical implementation of these MIMO techniques. In this paper, a comparison of MIMO transceiver techniques such as Tomlinson-Harashima precoding (THP), zero forcing dirty paper coding (ZF-DPC) and vertical Bell Laboratories Layered Space-Time (V-BLAST) is performed. We consider the effect of practical impairments such as imperfect CSI (channel state information) and a transmit power constraint. If the total power is restricted at the transmitter for fairness, ZF-DPC performs worst among those techniques although it does not suffer modulo loss and error propagation. Due to the pre-subtraction of the spatial interference, precoding techniques like ZF-DPC and THP are more sensitive to the effect of imperfect CSI than V-BLAST.**

1. Introduction

In recent years, the multiple-input multiple-output (MIMO) technique has emerged as one of the most promising approaches for next-generation wireless systems to achieve very high data rates as well as high spectral efficiencies. Different space-time coding techniques have been proposed to achieve the capacity advantages of MIMO techniques. Vertical BLAST [2], a simplified version of Diagonal BLAST (D-BLAST) [1], has been proposed due to the complexities of D-BLAST implementation. While the transmit data streams of D-BLAST are dispersed diagonally across antennas and time, each data stream in V-BLAST is sent to a different transmit antenna. This configuration can reduce the computational complexity at the receiver compared with D-BLAST. However, in V-BLAST, error propagation may occurr because received symbol streams are decoded and cancelled successively during the detection procedure.

Error propagation can be eliminated with precoding techniques because the successive interference cancellation is performed at the transmitter, where all data streams are known. Practical precoding techniques have been developed by using the result of dirty paper coding. In [3], Caire and Shamai proposed a zero-forcing dirty paper coding which forces to zero the interference caused by users $j > i$ on each user i. This is done by using a QR-type decomposition of the channel combined with the pre-subtraction of the noncausal interference at the transmitter. Although the SNR of each spatial channel of ZF-DPC is unchanged, the pre-subtraction increases the total transmit power. Hence, Tomlinson-Harashima precoding, which employs a modulo operation, is an effective means to minimize the total transmit power. For an overview of THP in MIMO channels, see [4]. Although the modulo operation of THP prevents a possibly large increase in the total transmit power, it still suffers from three kinds of losses [5]. Power loss is generated by the increase in power due to the extension of the original constellation at the transmitter. Modulo loss is due to the existence of more neighbors at the edge of the original constellation at the receiver. Because the transmit signal constellation of THP is a uniform cubic shape, there also exists a shaping loss of 1.53dB relative to the Shannon capacity.

In general, MIMO transceiver techniques use channel state information (CSI) at the transmitter to achieve the best performance by adapting the transmission rate and power to channel state variations. However, the amount of channel state information needed at the transmitter is different according to the type of MIMO transceiver techniques. MIMO precoding techniques like THP and ZF-DPC require full CSI, which includes the magnitude and phase of each spatial path, for pre-subtracting spatial interference at the transmitter. For receive BLAST technique, only channel covariance matrix is sufficient for determining the adaptive modulation order of each spatial channel at the transmitter. If channel state information is not perfect, as is usual in most practical cases, the performance is degraded. The performance criterion to be optimized can be either to maximize the capacity or to minimize the average error rate. In [4], the average error rates of THP and V-BLAST were investigated. But it considers neither the effect of imperfect CSI nor the transmit power constraint. Indeed, it did not include the average error rate of ZF-DPC.

In this paper, we compare the average symbol error rates of different MIMO transceiver techniques such as THP, ZF-DPC and V-BLAST considering the effect of imperfect CSI with the transmit power constraint. For a fair comparison, we make the total power of each transceiver technique identical at the transmitter.

The remainder of this paper is organized as follows. Section 2 describes the channel model. In section 3, we discuss different MIMO transceiver techniques. Section 4 is for loading, which adjusts the transmission rate and power of the MIMO transceiver technique to channel variations. Numerical results for the average error rates of different MIMO transceiver techniques are shown in section 5. Conclusions are presented in section 6.

2. Channel Model

We use boldface to denote matrices and vectors. $|A|$ denotes the determinant. For any general matrix $\boldsymbol{B}$, $\boldsymbol{B}^H$ denotes the conjugate transpose and $\mathrm{Tr}(\boldsymbol{B})$ denotes the trace. $\boldsymbol{I}$ denotes the identity matrix and $\mathrm{diag}\{\lambda_i\}$ denotes a diagonal matrix with the (i, i) entry equal to λ_i and $E[\bullet]$ denotes expectation.

We focus on a full rank square MIMO channel with an equal number of K transmit and receive antennas. The $K \times 1$ transmit signal vector is $\boldsymbol{x}$, where each element of $\boldsymbol{x}$ has transmit power P_i and its covariance matrix $\boldsymbol{R}_{xx} = E[\boldsymbol{x}\boldsymbol{x}^H]$. The $K \times K$ matrix $\boldsymbol{H} = [h_{rt}]$ consists of gain factors h_{rt} between transmit antenna t and receive antenna r which are independent and identically distributed (i.i.d.) zero mean circular symmetric complex Gaussian (ZMCSCG) random variables with unit variance. The $K \times 1$ vector $\boldsymbol{n}$ denotes additive white Gaussian noise (AWGN) whose covariance matrix is equal to the identity matrix scaled by σ_n^2. The total transmit power P per symbol period is constrained to the average power $\overline{P}$ as

$$E[P] = E\left[\mathrm{Tr}\left(\boldsymbol{R}_{xx}\right)\right] \le \overline{P}. \quad (1)$$

Then the $K \times 1$ received signal $\boldsymbol{y}$ can be expressed as

$$\boldsymbol{y} = \boldsymbol{H}\boldsymbol{x} + \boldsymbol{n}. \quad (2)$$

We assume a simple additive measurement noise model for imperfect CSI as

$$\hat{\boldsymbol{H}} = \boldsymbol{H} + \Delta\boldsymbol{H}, \quad (3)$$

where the estimated channel $\hat{\boldsymbol{H}}$ consists of an exact channel matrix $\boldsymbol{H}$ and a measurement noise $\Delta\boldsymbol{H}$ of $K \times K$ matrix whose entries are i.i.d. ZMCSCG variables with variance σ_e^2. Unlike the perfect CSI case, the received signal includes the disturbance component $\Delta\boldsymbol{H}\boldsymbol{x}$, which is proportional to both the measurement noise and the transmit signal power.

3. MIMO Transceiver Techniques

Figure 1 illustrates a general MIMO transceiver in the fading MIMO channel.

MIMO transceiver techniques such as THP, ZF-DPC and V-BLAST are considered within this generalized structure. The data vector $\boldsymbol{a}$ consists of complex symbols taken from M-QAM constellation, i.e. each entry a_i (i = 1, 2, ..., K) belongs to the set $\mathfrak{A} = \{\zeta(a_I' + ja_Q') \mid a_I', a_Q' \in \{\pm 1, \pm 3, ..., \pm(\sqrt{M}-1)\}\}$, where ζ is a normalization factor which makes the average power (σ_a^2) identical regardless of the modulation order M.

For each MIMO transceiver technique, all blocks in Figure 1 have their configurations according to Table 1. For example, ZF-DPC has no modulo operation at either the transmitter or the receiver. V-BLAST needs no transmit gain adjustment because the successive interference cancellation is performed at the receiver. Although the modulo operation in THP increases the average transmit power according to the modulation order M, the increase can be neglected at high SNR.

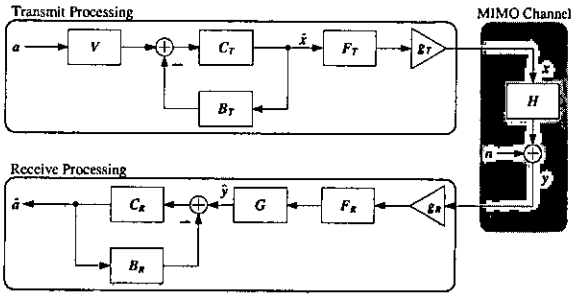


Figure 1 : A block diagram of the generalized MIMO transceiver

	Block	ZF-DPC	THP	V-BLAST
TX	C_T	I	MOD (•)	I
	B_T	$B - I$	$B - I$	0
	F_T	F	F	I
	g_T	$\sqrt{\frac{K}{\mathrm{Tr}\left(B^{-1}B^{-H}R_W\right)}}$	$\sqrt{\frac{K}{\left(\sum_{i=1}^{K}\frac{M_i}{M_i-1}v_{ii}^2\right)}}$	1
RX	C_R	I	MOD (•)	DEC(•)
	B_R	0	0	$B - I$
	F_R	I	I	F^H
	g_R	$1/g_T$	$1/g_T$	1

Table 1 : Configuration of MIMO tranceivers (MOD : modulo, DEC : decision)

Given the channel matrix $\boldsymbol{H}$, the QR-type decomposition of the estimated channel matrix $\hat{\boldsymbol{H}}$ is performed. Since the feedforward matrix $\boldsymbol{F}$ at the transmitter side is indispensable in the MIMO BC, the LQ decomposition should be performed for the precoding technique. The LQ decomposition of $\hat{\boldsymbol{H}}$ reads

$$\hat{\boldsymbol{H}} = \boldsymbol{S}\boldsymbol{F}^H, \quad (4)$$

where $\boldsymbol{F}$ is a unitary matrix and $\boldsymbol{S}$ is a lower triangular matrix. The diagonal scaling matrix is then given as $\boldsymbol{G} = \mathrm{diag}\{s_{11}^{-1}, ..., s_{KK}^{-1}\}$ and the feedback matrix reads $\boldsymbol{B} = \boldsymbol{G}\boldsymbol{F}\hat{\boldsymbol{H}} = \boldsymbol{G}\boldsymbol{S}$. $\boldsymbol{B} = [b_{ij}]$ has unit diagonal, $b_{ii}=1$, and $b_{ij} = 0$ for $j > i$. However, for the V-BLAST technique, the QL decomposition is performed, i.e.

$$\hat{\boldsymbol{H}} = \boldsymbol{F}\boldsymbol{S}^H. \quad (5)$$

While the QL decomposition can be seen as performing Gram-Schmidt orthogonalization on the columns of the channel, the LQ decomposition corresponds to the orthogonalization along the rows of the channel [4].

In the remainder of this section, we discuss implementation issues relating to these algorithms.

3.1. ZF-DPC

Because of the triangular structure of the feedback matrix $\boldsymbol{B}$, the output symbols of the feedback filter $\hat{x}_k$ are successively generated from the input data symbols $a_k \in \mathfrak{A}$

$$\hat{x}_k = v_{kk}a_k - \sum_{l=1}^{k-1} b_{kl}\hat{x}_l, (k = 1, ..., K). \quad (6)$$

The transmit power of ZF-DPC is increased due to the pre-subtraction of the interference. Since the matrix $\boldsymbol{F}$ does not change the total power, the total transmit power of ZF-DPC with loading is given as

$$\mathrm{Tr}\left(E\left[\boldsymbol{x}\boldsymbol{x}^H\right]\right)=\mathrm{Tr}\left(\boldsymbol{B}^{-1}\boldsymbol{B}^{-H}\boldsymbol{R}_{VV}\right)\sigma_a^2, \qquad (7)$$

where σ_a^2 is the per-component variance of the input data vector $\boldsymbol{a}$ whose covariance matrix is $E[\boldsymbol{a}\boldsymbol{a}^H]=\sigma_a^2\boldsymbol{I}$ and $\boldsymbol{R}_{VV}$ is the covariance of the power allocation matrix $\boldsymbol{V}$. Then the normalization factor of ZF-DPC to keep the transmit power $P=K\sigma_a^2$ is

$$g_T=\sqrt{\frac{K}{\mathrm{Tr}\left(\boldsymbol{B}^{-1}\boldsymbol{B}^{-H}\boldsymbol{R}_{VV}\right)}}. \qquad (8)$$

However, ZF-DPC needs full CSI at the transmitter for the pre-subtraction of spatial interference.

In the next section, we discuss THP and derive the normalization factor of THP to satisfy the power constraint at the transmitter.

3.2. THP

Similar to ZF-DPC, the output symbols of the feedback filter $\hat{x}_k$ are successively generated from the input data symbols $a_k \in \mathfrak{A}$. But THP uses the modulo operation to avoid a possibly large increase in transmit power as

$$\hat{x}_k=\mathrm{MOD}\left(v_{kk}a_k-\sum_{l=1}^{k-1}b_{kl}\hat{x}_l, 2\zeta\sqrt{M}\right), (k=1,\ldots,K), \qquad (9)$$

where the modulo operation reduces $\hat{x}_k$ symbol to the interval $[-\zeta\sqrt{M},\zeta\sqrt{M})$. Then the constellation is bounded by the square region of width $2\zeta\sqrt{M}$. However, since transmit symbols of THP are uniformly distributed over the square region, the total average transmit power of THP with loading is approximated as

$$\mathrm{Tr}\left(E\left[\boldsymbol{x}\boldsymbol{x}^H\right]\right)\cong\left(\sum_{i=1}^{K}\frac{M_i}{M_i-1}v_{ii}^2\right)\sigma_a^2, \qquad (10)$$

where M_i is the modulation order of ith spatial channel. Hence, the normalization factor of THP becomes

$$g_T=\sqrt{\frac{K}{\left(\sum_{i=1}^{K}\frac{M_i}{M_i-1}v_{ii}^2\right)}}. \qquad (11)$$

Note that THP also requires full CSI at the transmitter like ZF-DPC for pre-processing the transmit signal. Unlike precoding techniques, the receive V-BLAST technique needs no power normalization at the transmitter because it does not perform any pre-processing which could increase the transmit power. We will discuss it further in the next section.

3.3. V-BLAST

Since the successive detection of V-BLAST is equivalent to the operation of the decision feedback equalizer (DFE), the V-BLAST receiver can be interpreted as a special case of the generalized DFE [7]. The conventional V-BLAST algorithm finds the optimal ordering of the detection to maximize the signal to interference plus noise ratio (SINR). This can be performed in the DFE structure by introducing the permutation matrix $\boldsymbol{P}$ as equation (13) to find the optimal ordering. The transmit signal vector $\boldsymbol{x}$ equals to the input data vector $\boldsymbol{a}$ at the V-BLAST transmitter. It means there is no increase in transmit power with the V-BLAST technique. Instead, for each stage of successive cancellation at the decision feedback equalizer; the stream with the highest SINR is decoded and stripped away from the received signal vector, which causes the propagation of decoded errors. Unlike precoding techniques like ZF-DPC and THP, V-BLAST only requires channel covariance matrix at the transmitter for the adaptive modulation of each spatial channel according to its channel gain.

4. Loading

Loading adjusts the transmission rate and power to make the best use of spatial channels. Several algorithms are known to find the optimal rate and power distribution for loading. The algorithm [11] determines the rate distribution according to the capacity of the spatial channels. In this paper, we follow the approach at [10], where loading is performed to minimize the average error probability. Denoting the gain factors of the K spatial channels as s_{kk}, the diagonal components of the lower triangular matrix by the QR-type decomposition of the estimated channel $\hat{\boldsymbol{H}}$, the total fixed-sum rate R is distributed according to (12) such that the overall symbol error rate is minimized [4]. Note that the signal to noise ratio of each spatial channel is determined by the gain factor s_{kk} because the noise variance of each spatial channel is identical for single user multi antenna scenarios.

$$R_k=\frac{R}{K}+\frac{1}{K}\cdot\log_2\left(\frac{|s_{kk}|^{2K}}{\prod_{l=1}^{K}|s_{ll}|^2}\right). \qquad (12)$$

These rates are quantized to integers in an allowed range. In order to compensate for rate quantization errors and to make the same error rates for all spatial channels, the transmit signals are multiplied by a residual transmit power allocation matrix $\boldsymbol{V}$, where $\boldsymbol{V}=\mathrm{diag}\{v_{11},\ldots,v_{KK}\}$ and $\mathrm{Tr}(\boldsymbol{V})=1$, for fine adjustment as in [10].

Since the encoding/decoding order affects the amount of the interference from previously encoded/decoded signals, the performance can be varied according to the ordering. For the fixed sum rate loading, we use the permutation matrix (13) to maximize the minimum spatial channel SNR as [12] because the worst spatial channel dominates the overall error rate.

$$\boldsymbol{P}_{opt}=\arg\max_{\boldsymbol{P}}\min\left\{|s_{11}|^2,\ldots,|s_{KK}|^2\right\}. \qquad (13)$$

5. Numerical Results

Throughout the simulations, we assume $K = 4$, i.e. the channel matrix $\boldsymbol{H}$ is a square 4×4 matrix. We use M-QAM signal sets with cardinality $M = 2^2,\ldots,2^8$, which correspond to integer rates ranging from 2 to 8 bits per symbol interval. We also assume that the total rate is fixed to 16 bits per symbol interval. Although there is some improvement by adapting the minimum mean square error (MMSE) criterion at low SNR, we only focus on the ZF solution.

For the fair comparison, we use an effective SNR (σ_x^2/σ_n^2), the signal variance with the transmit power constraint applied, divided by the noise variance, instead of SNR. This is because the transmit power and received SNR is different according to the normalization factor of the MIMO transceiver technique. For the receive V-BLAST technique, the effective SNR is identical to SNR at the receiver. But the transmit power of precoding techniques consists of the transmit signal and the pre-subtracted interference, which reduces the overall SNR at the receiver.

5.1. Effect of the Transmit Power Constraint

Figure 2 compares the average symbol error rates of different MIMO transceiver techniques with and without the transmit power constraint. We use the fixed sum rate loading (12). In the case of no transmit power constraint, ZF-DPC performs best because it has neither modulo loss nor error propagation. Since the effect of the modulo loss in THP and that of the error propagation in V-BLAST is almost the same, the performance difference between THP and V-BLAST is negligible without the transmit power constraint.

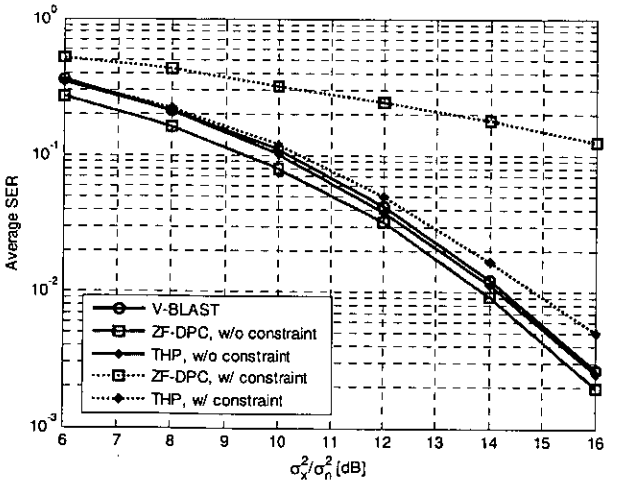


Figure 2 : Average SER of different MIMO transceiver techniques with fixed sum rate loading and perfect CSI ($K = 4$, R=16)

However, with the total power constraint, the performance of the precoding technique is degraded because the effective SNRs of all spatial channels at the receiver are decreased due to the constraint compared to the receive V-BLAST technique. Since the power increase of ZF-DPC is much larger than that of THP, the performance of ZF-DPC is much worse than that of THP with the transmit power constraint. Although the modulo operation of THP prevents a possibly large increase in power, THP still has the modulo loss, which causes the decrease of effective SNR. Hence, the performance of THP is slightly worse than that of V-BLAST with the transmit power constraint.

5.2. Effect of Imperfect CSI

In Figure 3 and Figure 4, we compare the average symbol error rates of MIMO transceiver techniques with measurement noise variance σ_e^2=-20dB and σ_e^2=-15dB relative to the transmit power respectively.

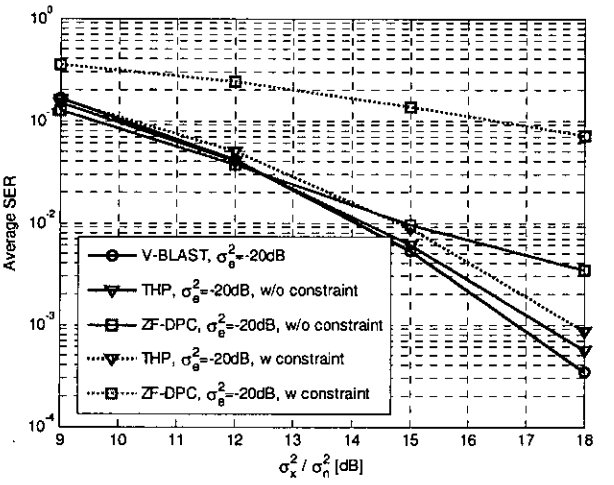


Figure 3 : Average SER of different MIMO transceiver techniques with fixed sum rate loading and imperfect CSI (σ_e^2=-20dB, $K = 4$, R=16)

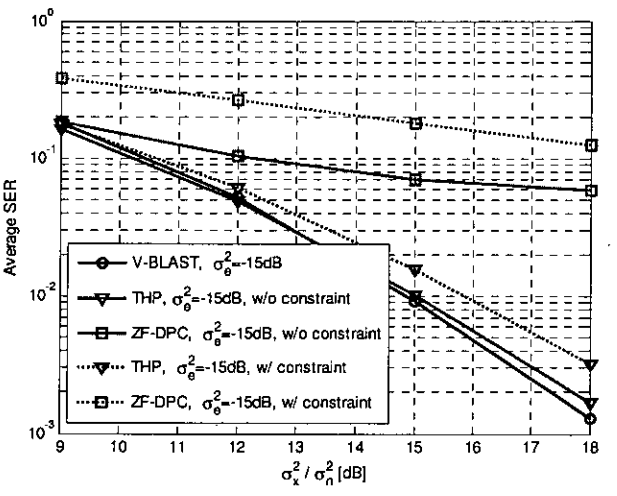


Figure 4 : Average SER of different MIMO transceiver techniques with fixed sum rate loading and imperfect CSI (σ_e^2=-15dB, $K = 4$, R=16)

Because the pre-subtraction boosts the effect of channel measurement noise, the precoding technique is more sensitive to the effect of imperfect CSI than the V-BLAST technique regardless of the transmit power constraint. In contrast to the case of perfect CSI, THP performs slightly worse than V-BLAST with imperfect CSI without the transmit power constraint at high SNR.

The amount of performance degradation of the precoding technique is proportional to both SNR and measurement noise variance. Between two precoding techniques, ZF-DPC is more sensitive to the effect of imperfect CSI than THP because the total amount of spatial interference is proportional to transmit power in the additive measurement noise MIMO channel. The performance degradation of ZF-DPC with imperfect CSI is so severe that ZF-DPC is not appropriate for practical dirty paper coding.

6. Conclusions

In this paper, we compared two transmit precoding techniques, namely THP and ZF-DPC, and the receive V-BLAST technique taking into account practical considerations.

Simulation results showed that ZF-DPC had the worst average symbol error rate performance among those MIMO transceiver techniques, if we limited the transmit power for the fair comparison. This is despite the fact that it had neither the modulo loss of THP nor the error propagation of V-BLAST. The transmit power constraint reduced the effective SNR of the precoding technique, which caused performance degradation for both THP and ZF-DPC compared to the receive V-BLAST technique. Precoding techniques were more sensitive to the effect of CSI than V-BLAST because the pre-subtraction operation at the transmitter amplifies the effect of imperfect CSI. The performance degradation of THP is negligible compared to V-BLAST.

Practically, THP may be particularly useful for the downlink or broadcast channel, where a multiple antenna base station sends data packets to a group of users, each possibly only having a single receive antenna. In that scenario, V-BLAST is not normally feasible because it needs receive antenna cooperation for successive interference cancellation at the receivers.

However, the presented model in this paper does not include all practical considerations. Hence, further work should be performed to compare the performance of various MIMO transceiver techniques in a more realistic MIMO channel model.

ACKNOWLEDGEMENT

This work was supported by Toshiba Research Europe Laboratory in UK.

REFERENCES

[1] G. Foschini, *"Layered space-time architecture for wireless communication in a fading environment when using multi-element antennas,"* Bell Labs Tech. Journal, pp. 41-59, 1996

[2] P. W. Wolniansky, G. J. Foschini, G. D. Golden, and R. A. Valenzuela, *"V-BLAST: an architecture for realizing very high data rates over the rich scattering wireless channel,"* URSI International Symposium on Signals, Systems and Electronics, pp. 295-300, 1998.

[3] G. Caire and S. Shamai, *"On the achievable throughput of a multiantenna Gaussian broadcast channel,"* IEEE Trans. Information Theory, vol. 49, no. 7, pp. 1691-1706, Jul. 2003.

[4] C. Windpassinger, R. F. Fischer, T. Vencel, and J. B. Huber, *"Precoding in multiantenna and multiuser communications,"* IEEE Trans. Commun., vol. 3, no. 4, pp.1305-1316, July 2004.

[5] W. Yu, *"Trellis and convolutional precoding for transmitter-based interference pre-subtraction,"* accepted in IEEE Trans. Commun., 2005, *URL: http://www.comm.toronto.edu/~weiyu/publications.html.*

[6] C. Windpassinger, T. Vencel and R. F. H. Fischer, *"Precoding and loading for BLAST-like systems,"* Proceedings of IEEE International Conference on Communications (ICC) 2003, Anchorage, Alaska, USA, May 2003.

[7] G. Ginis and J. M. Cioffi, *"On the relation between V-BLAST and the GDFE,"* IEEE Commun. Letters, vol. 5, no. 9, pp. 364-366, Sep. 2001.

[8] S. Chung, A. Lozano, H. C. Huang, A. Sutivong, and J. M. Cioffi, *"Approaching the MIMO capacity with a low-rate feedback channel in VBLAST,"* EURASIP Journal on Applied Signal Processing, pp.762-771, May 2004.

[9] T. Yoo and A. Goldsmith, *"Capacity of fading MIMO channels with channel estimation error,"* IEEE ICC, Paris, France, pp. 808-813, June 2004.

[10] R. F. H Fischer and J. B. Huber, *"A new loading algorithm for discrete multitone transmission,"* IEEE Global Telecommun. Conference, pp. 724-728, Nov. 1996.

[11] P. S. Chow, J. M. Cioffi, and J. A. C. Bingham, *"A practical discrete multitone transceiver loading algorithm for data transmission over spectrally shaped channels,"* IEEE Trans. Commun. Vol. 43, no. 2/3/4, pp. 773-775, Feb./Mar./Apr. 1995.

[12] T. Vencel, C. Windpassinger, and R. H. Fischer, *"Sorting in the V-BLAST Algorithm and Loading,"* In Proceedings of Communication Systems, pp. 304-309, Malaga, Spain, Sep. 2002.

QoS-guaranteed Sequential User Selection in Multiuser MIMO Downlink Channels

S. Lee and J. S. Thompson,
University of Edinburgh, Institute for Digital Communications, Edinburgh, UK
{S.Lee, John.Thompson}@ed.ac.uk

Abstract— In recent years, multiuser applications of MIMO techniques have received a lot of attention because of their large capacity gains for next generation multiuser wireless systems. In this paper, we propose a new sequential user selection algorithm and analyze its performance in multiuser QoS-guaranteed MIMO downlink channels. Multiple QoS-guaranteed users can be supported simultaneously while non-QoS users can receive the available residual throughput without sacrificing QoS-guarantees by flexibly allocating spatial gains to different QoS classes. The proposed algorithm can also reduce computational complexities of user selection procedures by dividing the whole user group into several subgroups according to their QoS priorities. We assume throughput-constrained constant rate users, delay-constrained real time users and no-constraint best effort users for different QoS classes. Numerical results demonstrate that our model can provide QoS-guaranteed services for constant rate users and real time users simultaneously, at the cost of some reduction in the total throughput for the MIMO downlink channels.

I. Introduction

Recently, multiple-input multiple-output (MIMO) techniques have received considerable attention because they can provide not only diversity gain which increases the reliability of transmission, but multiplexing gain which boosts data rate by exploiting the space dimension. MIMO techniques were first studied in single user scenarios. It is well known that the single user capacity of a MIMO system increases linearly with the minimum number of transmit and receive antennas. In multiuser MIMO broadcast channels (BCs or downlink, i.e., channels from the base station to mobile users), dirty-paper coding [1] (DPC) can achieve the sum-rate capacity [2]. However, not only is it difficult to implement in practice, DPC also considers throughput and neglects various quality-of-service (QoS) parameters such as data rate and packet delay. In most real scenarios, each user requests different types of service such as video, file transfer protocol (FTP) and hypertext transfer protocol (HTTP), each with different QoS requirements. Hence, the overall system performance should be considered to satisfy these QoS requirements of all users, not only the achievable capacity. This makes a system-level design for multiuser MIMO broadcast channels more complicated than mere implementation of throughput-optimal DPC techniques. Indeed, due to the time-varying nature of wireless channels, providing different users with different QoS requirements in wireless channels is a more challenging task than in wired networks. Various 'channel-aware' scheduling algorithms have been proposed to achieve optimal performance with different QoS requirements [3]-[5]. Most previous works limit their scope to opportunistic scheduling where only the best user is selected at any time for transmission. However, in the MIMO BC, multiple users can be served simultaneously, which necessitates finding the optimal user subset to satisfy the different QoS requirements of users as well as maximizing the total throughput of the system. Due to the prohibitive complexity of finding an optimal subset of users in the MIMO BC, several complexity-reduced suboptimal scheduling algorithms have been proposed. In [6] and [7], zero-forcing (ZF) precoding instead of DPC was used to reduce the complexity because it provides spatial separation among users and thereby support multiple users simultaneously. It has been shown that the sum-rate of ZF schemes approaches that of DPC as the number of users goes to infinity. However, these papers did not consider different QoS requirements. QoS-based user scheduling with transmit beamforming for multiuser MIMO systems has been considered in [8]. The beamforming strategy used in [8] made each transmit antenna pattern take independent scheduling decisions for user allocation with conventional opportunistic scheduling, which transforms the multiuser scheduling problem into the spatial multiplexing of several single user schedulers. In [9], users were classified into QoS-guaranteed real-time (RT) users and no QoS-guaranteed best-effort (BE) users. It showed that the combination of zero forcing beamforming and existing QoS algorithms like modified largest weighted delay first (M-LWDF) could provide statistical QoS guarantees to RT users. However, it did not provide any scheduling method to support BE users while simultaneously guaranteeing the QoS of RT users. Most existing models applied zero-forcing beamforming techniques to the physical layer for reducing the complexity at the cost of reduced sum-rate capacity compared to DPC.

In this paper, we propose a new sequential user selection algorithm, which can be combined with practical DPC techniques for different QoS-guaranteed services in the MIMO BC. The proposed algorithm can reduce the computational complexity of the selection of users by dividing the whole user

This work was supported by Toshiba Research Europe Limited (TREL) Telecommunications Research Laboratory (TRL) in Bristol, UK.

group into several subgroups according to their QoS priorities. Users are sequentially selected from the higher priority subgroups first and the QoS criterion of each subgroup determines the trade-off of spatial gain between multiplexing for lower priority users and for higher priority users. However, the QoS user selection scheme results in the decreased capacity compared to DPC. This is because even the exhaustive optimal selection becomes suboptimal due to the reduction of dimension for multiuser diversity compared to the selection of users from the whole user group. Note that this result arises because there is a trade-off between maximizing the total throughput and guaranteeing certain QoS requirements. In other words, QoS is guaranteed at the cost of reducing the total throughput.

The rest of this paper is organized as follows. In section II, the system model is introduced. Section III explains the proposed sequential user selection algorithm. Numerical results are presented in Section IV. Conclusions are drawn in Section V.

II. SYSTEM MODEL

We use boldface to denote matrices and vectors. For any general matrix $\mathbf{A}$, $\mathbf{A}^T$ denotes the transpose and $\mathbf{A}^H$ denotes the conjugate transpose, and diag$\{\lambda_i\}$ denotes a diagonal matrix with the (i, i) entry equal to λ_i.

Let a MIMO broadcast channel has M_T transmit antennas at the base station and K users each with a single receive antenna and $\mathbf{h}_k \in \square^{M_T \times 1}$ denote the channel between the transmit antenna arrays and the receive antenna for user k. Then the downlink channel can be represented as

$$y_k = \mathbf{h}_k^T \mathbf{x} + z_k, \qquad k = 1, \cdots, K, \tag{1}$$

where $\mathbf{x} \in \square^{M_T \times 1}$ is the transmit signal vector, y_k is the received signal for user k and z_k is the complex additive white Gaussian noise (AWGN) for user k with zero mean and unit variance. The transmit signals are assumed to experience path loss, log-normal shadow fading and block multipath fading. The channel is assumed to be fixed during a time slot and to vary independently over time slots. This means recalculating path loss, shadowing and fading coefficients for each user. Then the channel $\mathbf{h}_k$ at any time slot can be expressed as [12]

$$\mathbf{h}_k = \sqrt{SNR_0 \cdot (r_k/R)^{-\alpha} \cdot s_k} \cdot \mathbf{g}_k, \tag{2}$$

where SNR_0 denotes the average signal-to noise-ratio (SNR) at the boundary of cell radius R, r_k is the base-mobile distance of user k, α denotes the path loss exponent, s_k is a log-normal shadow fading variable with standard deviation σ_s. The column vector $\mathbf{g}_k = [g_{k1}\ g_{k2} \cdots g_{kM}]^T$ represents the channel response between the transmit antenna arrays and user k, whose components are independent and identically distributed (i.i.d.) zero mean circular symmetric complex Gaussian (ZMCSCG) random variables with unit variance.

III. SEQUENTIAL USER SELECTION ALGORITHM

In this section, we explain the proposed sequential user selection algorithm for QoS-guaranteed scheduling in the MIMO BC. The motivation of this algorithm is that multiple QoS-guaranteed users should be supported simultaneously while non-QoS users may receive the maximum possible throughput without sacrificing the performance of QoS-guaranteed users.

For simplicity, we assume that there are three kinds of QoS users: throughput-constrained constant rate (CR) users, delay-constrained real time (RT) users and best effort (BE) users with no constraints. Each QoS class has its priority for sequential selection: CR users have the highest priority, RT the second highest and BE the lowest priority. Initially, each class has its own number of transmit antenna patterns but there is priority for the actual transmission using these antenna patterns. For example, when the throughput performance of constant rate users does not satisfy a certain throughput threshold, real time users and best-effort users have no chance of using their pre-assigned transmit antennas for transmission because their priorities are lower than constant rate users.

Since the channel hk is random, the capacity associated with the MIMO BC is also a random variable. It means that no scheduling algorithm can always guarantee 'deterministic' QoS in wireless channels. Fortunately, in most real wireless systems, these hard QoS conditions are not usually required. They can tolerate a small amount of instantaneous QoS violation [9] (i.e., through buffering). For this reason, guaranteed-throughput means the average throughput during certain time slots is above throughput thresholds, and guaranteed-delay means the average delay is within delay thresholds with a certain outage probability.

Let $U := \{u_k \mid k = 1, \cdots, K\}$ denote the total user set and $U^{CR} := \{u_l \mid l = 1, \cdots, K_{CR}\}$, $U^{RT} := \{u_l \mid l = 1, \cdots, K_{RT}\}$ and $U^{BE} := \{u_l \mid l = 1, \cdots, K_{BE}\}$ denote the set of constant rate, real time and best effort users respectively. U^{CR}, U^{RT} and U^{BE} are disjoint where $K = K_{CR} + K_{RT} + K_{BE}$. In addition, let S^{CR} ($S^{CR} \subset U^{CR}$), S^{RT} ($S^{RT} \subset U^{RT}$) and S^{BE} ($S^{BE} \subset U^{BE}$) denote the set of selected constant rate, selected real time and selected best effort users respectively. M_s^{CR}, M_s^{RT} and M_s^{BE} are the number of pre-assigned transmit antennas for constant rate, real time and best effort users respectively. We assume that zero-forcing DPC [10] is used for the MIMO precoding technique. ZF-DPC exploits the DPC principle [1]; the capacity of a system with known interference at the transmitter is the same as if there were no interference present. It decomposes the MIMO channel into transmit beamforming matrix **F** and lower triangular matrix **B** using QR decomposition. Because of the lower triangular matrix B, the first encoded user has no interference, the second encoded user experiences interference only from the first encoded user and so on. In other words, any interference caused by users j > i on each user i is forced to zero by the pre-subtraction at the transmitter. For the sum-rate capacity, the number of selected users for transmission is assumed to be the same as that of transmit antennas. Then the sum-rate capacity of ZF-DPC is given by [10]

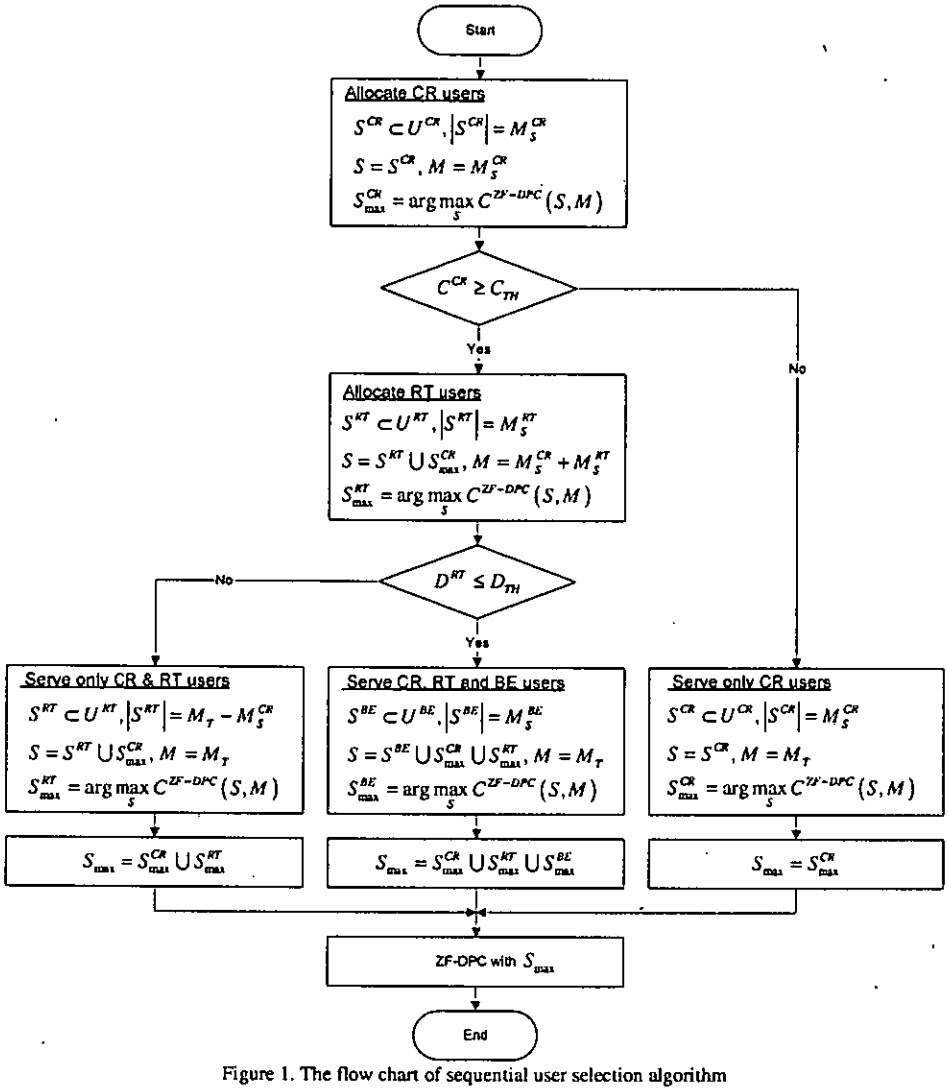


Figure 1. The flow chart of sequential user selection algorithm

$$C^{ZF-DPC} = \sum_{k=1}^{M_T} \log_2\left(1 + b_{kk}^2 P_k\right), \quad (3)$$

where b_{kk} is the kth diagonal entry of **B**, b_{kk}^2 is the effective channel gain to user k and P_k is the input signal power allocated to user k with a power constraint of P by $\sum_{k=1}^{M_T} P_k \leq P$. We also assume that the selection criterion is to maximize the sum-rate capacity of ZF-DPC, which is given by

$$S_{max} = \arg \max_{S \subset U, |S| = M_s} C^{ZF-DPC}(S), \quad (4)$$

where M_s is the number of selected transmit antenna beam patterns ($M_s \leq M_T$) and S is every possible user group which selects M_s users from all users. We assume that the number of selected users for transmission is the same as that of transmit antennas.

Figure 1 shows the flow chart of the proposed algorithm. At first, the algorithm selects S_{max}^{CR} by the maximum sum-rate rule among all possible combinations of CR user set (S^{CR}) with M_s^{CR} transmit antennas. There are $\binom{U^{CR}}{S^{CR}}$ possible ways of choosing S^{CR} from U^{CR} for the selection of CR users. If the sum-rate capacity of S_{max}^{CR} (C^{CR}) does not exceed the throughput threshold (C_{TH}), all transmit antennas are assigned to users in S_{max}^{CR} to improve the probability of throughput-guaranteed transmission. If the number of CR users is less than that of transmit antennas, we do use 'diversity gain' for CR users. Otherwise, S_{max}^{RT} is chosen from all RT users and previously selected S_{max}^{CR}. In this case, there are $\binom{U^{RT}}{S^{RT}}$ possible ways of choosing RT users because already-selected S_{max}^{CR} does not increase the combinations of possible user sets of S^{RT}. It only affects the sum-rate capacity of S_{max}^{RT}. There are some restrictions for the selection of RT users with the maximum

sum-rate rule if $S_{\max}^{CR}$ is included. Any RT user who has high cross correlation with users in $S_{\max}^{CR}$ might be excluded even if they have high individual channel gain. In other words, to find a set of the maximum sum-rate in U^{RT} is to select the 'most orthogonal' RT users with respect to previously selected CR users as well as RT users themselves, which results in the reduction of multiuser diversity gain. The selection of RT users does not affect the performance of the selected CR users because of ZF-DPC, which pre-subtracts interference from other users, after the selection procedure. To decide whether BE users can be transmitted to, the delay of RT users (D^{RT}) is compared with the delay threshold (D_{TH}). If the delay of real time users is less than DTH, BE users will be selected in the same way as RT users. In this case, both $S_{\max}^{CR}$ and $S_{\max}^{RT}$ are considered to find the maximum sum-rate capacity selection including BE users. Otherwise, all transmit antennas are allocated to CR and RT users.

IV. Numerical Results

In this section, numerical results for the performance of the proposed sequential user selection algorithm are presented. Throughout the simulations, we assume $M_T = 4$, $SNR_0 = 6$ dB, $\alpha = 4$, R = 1 km and shadowing standard deviation of 8 dB for the MIMO downlink channel.

Figure 2 shows the sum-rate capacity of the proposed algorithm and Figure 3 shows the computational complexity, the number of combinations for the maximum sum-rate rule, as a function of the number of users. Due to the decrease in the number of users for multiuser diversity effect, the capacity becomes worse as the number of subgroups increases. However, it is obvious that as K approaches to infinity, the throughput of the proposed algorithm goes to the throughput without division of user group regardless of the number of subgroups provided the number of subgroups is finite and does not increase with the number of users. This is because we can always find an orthogonal set of M_T users among the finite subgroups with infinite number of users.

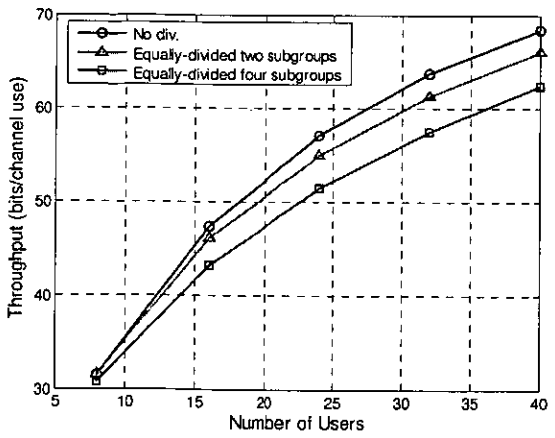


Figure 2. Sum-rate throughput comparison of the proposed algorithm with different division of the user group for $M_T = 4$ and ZF-DPC

In contrast to the capacity, the computational complexities are rapidly decreased as the number of subgroups increases. The whole group search, with no division of users, rapidly becomes prohibitive even for practical values of K, e.g. K > 30. Thus the whole group search will not be feasible for real MIMO BC scenarios, where there may be more than 4 transmit antennas and over 30 users.

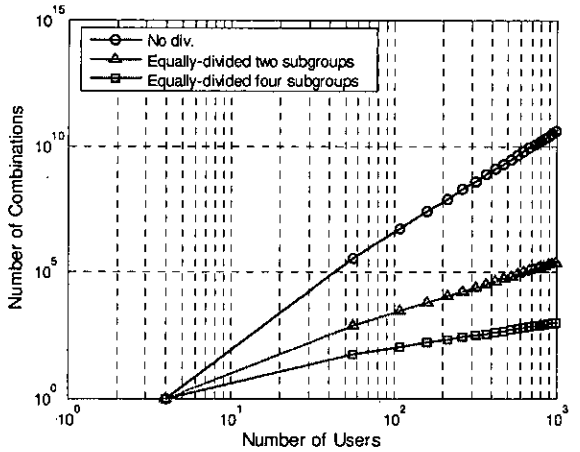


Figure 3. Number of combinations for the maximum sum-rate rule of the proposed algorithm for $M_T = 4$ and ZF-DPC

In Figure 4, we plot the throughput performance of the proposed algorithm versus the number of non-CR (i.e., RT and BE) users for different throughput thresholds (C_{TH}). These plots are derived assuming $U^{CR} = 2$, $S^{CR} = 2$ (i.e., no CR user selection and the number of CR users is fixed), $M_s^{CR} = 2$, 25% of RT users among non-CR users, $M_s^{RT} = 1$, $M_s^{BE} = 2$ and D_{TH} = 1024 slots. We assume 90% outage delay for the delay performance where the delay is guaranteed for 90% of the channel realizations so that 10% of delay values can exceed the delay threshold. We also assume that the signal power P = 1 and $P_1 = P_2 = \cdots = P_K = 1/M_T$. Data packets for all users arrive in queues at any time slot with mean arrival rate $\lambda = 3$ bits/channel use of a Poisson process independently and wait for transmission. If there is no trade-off of spatial gain, where all RT and BE users as well as CR users always have transmission opportunities through their pre-assigned transmit antenna patterns, the total throughput is maximized due to the absence of QoS constraints, which can maximize the multiuser diversity effect. We notice that the total throughput decreases QoS constraints are present. When C_{TH} = 4 bits/ channel use, the throughput of BE users begins to decrease over 15 non-CR users. This is because few transmission opportunities are given to the lowest priority BE users in order to support higher priority users. When C_{TH} is 12 bits/channel use, the throughput of non-CR users is very small because the algorithm usually assigns all transmit antennas to CR users to satisfy the guaranteed throughput. For throughput guaranteed CR users, the throughput within the MIMO capacity region is always satisfied regardless of the number of RT and BE users, who

have lower priorities than CR users. In this case, the throughput performance of RT users, not to mention of BE users, is worse than the case of no trade-off, and the throughput of CR users becomes the total system throughput.

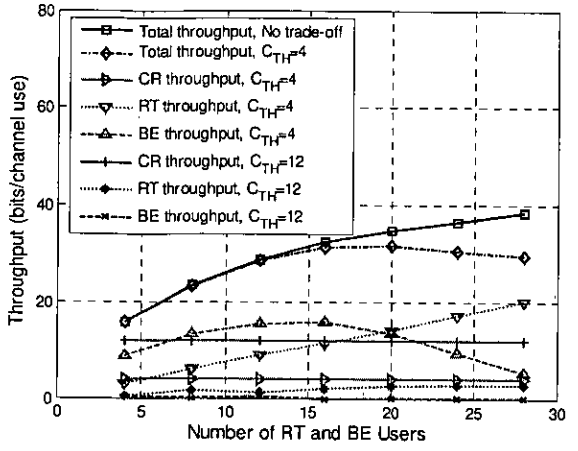


Figure 4. Throughput Performance of CR, RT and all users with the proposed algorithm for different QoS constraints

Figure 5 shows the delay performance of RT users for different throughput thresholds of CR users. For the delay performance, we evaluate the head-of-line (HOL) delays of RT users. All parameters are the same as those in Figure 4. We notice that when there is a trade-off and C_{TH} = 4 bits/channel use, the delay threshold D_{TH} = 1024 slots is guaranteed up to about 32 non-CR users. However, when C_{TH} =16 bits/channel use, the QoS of RT users can not be guaranteed regardless of the number of non-CR users because the highest priority CR users take most of the transmission opportunities to attain the throughput threshold.

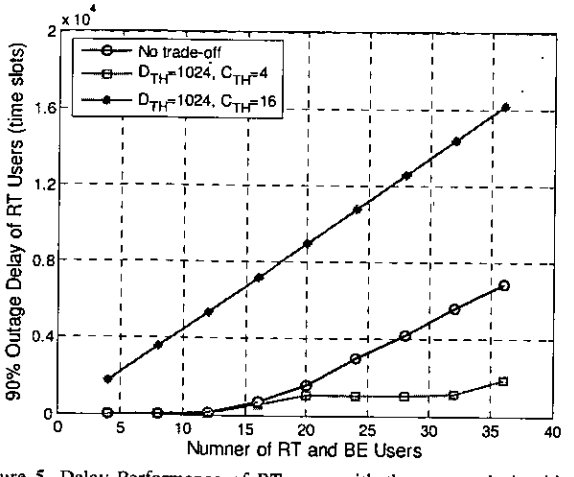


Figure 5. Delay Performance of RT users with the proposed algorithm for different QoS constraints

V. Conclusion

In this paper, we proposed a sequential user selection algorithm which can provide QoS-guaranteed services in MIMO broadcast channels. The proposed algorithm exploits the spatial gain of multiple transmit antennas in the MIMO BC to maximize the possible throughput of all users by the maximum sum-rate rule while simultaneously guaranteeing the throughput of constant rate users and the delay of real time users within certain thresholds. The proposed algorithm performs the trade-off of spatial multiplexing not only to maximize the total throughput but to improve the performance of QoS-guaranteed users. Users are sequentially selected from the higher priority subgroups first and transmission opportunities for lower priority subgroups are determined by performance criteria of higher priority subgroups. Any interference from other users is pre-subtracted at the transmitter by ZF-DPC, which assures no performance degradation of already-selected higher priority users.

Numerical results showed that the proposed algorithm could provide throughput-guaranteed service to constant rate users and delay-guaranteed service to real time users simultaneously. It could also reduce the computational complexity.

However, for QoS-guaranteed services, the degradation of the total throughput is inevitable. Hence, maximizing the total capacity of the MIMO BC is necessary if the system wants to admit more QoS-guaranteed services.

References

[1] M. Costa, "Writing on dirty paper," IEEE Trans. Info. Theory, vol. 29, pp.439-441, May 1983.

[2] S. Vishwanath, N. Jindal, and A. Goldsmith, "Duality, achievable rates, and sum-rate capacity of Gaussian MIMO broadcast channels," IEEE Trans. Info. Theory, vol. 49, no. 10, pp. 2658-2668, Oct. 2003.

[3] A. L. Stolyar, K. Ramanan, "Largest weighted delay first scheduling: large deviations and optimality," The Annals of Applied Probability, vol. 11, pp.1-48, 2001.

[4] M. Andrews, K. Kumaran and K. Ramanan, "CDMA data QoS scheduling on the forward link with variable channel conditions," Bell Labs. Preprint, 2000.

[5] S. Shakkottai and A. L. Stolyar, "Scheduling algorithms for a mixture of real-time and non-real-time data in HDR," in Teletraffic Engineering in the Internet Era, Proc. ITC-17, pp. 793-804, 2001.

[6] T. Yoo and A. Goldsmith, "On the optimality of multiantenna broadcast scheduling using zero-forcing beamforming," IEEE Journal on Sel. Commun. Vol. 24, no. 3, pp. 528-541, March 2006.

[7] C. Swannack, E. Uysal-Biyikoglu and G. Wornell, "Low complexity multiuser scheduling for maximizing throughput in the MIMO broadcast channel," in Proc. 42nd Annual Allerton Conf. on Communications, Control, and Computing, Monticello, Illinois, 2004.

[8] M. Kountouris, A. Pandharipande, H. Kim and D. Gesbert, "QoS-based user scheduling for multiuser MIMO systems," VTC- 2005 Spring, Stockholm, Sweden, vol. 1, pp. 211-215, 2005.

[9] W. Ajib, D. Haccoun and J. Frigon, "An efficient QoS based scheduling algorithm for MIMO wireless systems," VTC-2005 Fall, Dallas, U.S., vol. 3, pp. 1579-1583, 2005.

[10] G. Caire and S. Shamai, "On the achievable throughput of a multiantenna Gaussian broadcast channel," IEEE Trans. Info. Theory, vol. 49, no. 7, pp. 1691-1706, July 2003.

[11] S. Shakkottai, T. S. Rappaport and P. C. Karlsson, "Cross-layer design for wireless networks," IEEE Commun. Mag. pp. 74-80, Oct. 2003.

[12] O. Shin and K. B. Lee, "Antenna-assisted round-robin scheduling for MIMO cellular systems," IEEE Commun. Letters, vol. 7, no. 3, pp. 109-111, March 2003.

Trade-offs of Spatial Gain for QoS-guaranteed services in the MIMO Broadcast Channels

S. Lee and J. S. Thompson,
University of Edinburgh, Institute for Digital Communications, Edinburgh, UK
{S.Lee, John.Thompson}@ed.ac.uk

Abstract— **Although the capacity of multiple-input multiple output (MIMO) broadcast channels (BCs) can be achieved by dirty-paper coding (DPC), it is difficult to apply the results directly to practical quality-of-service (QoS)-guaranteed multiuser wireless systems. In this paper, we propose a new multiuser scheduling algorithm which exploits spatial gain of MIMO techniques for QoS-guaranteed services and analyze its performance with several MIMO precoding techniques. We assume that there are real-time (RT) users and best effort (BE) users in the MIMO BC. Numerical results demonstrate that the average delay of RT users can be guaranteed within certain delay thresholds by trading off spatial multiplexing for different user classes. Nonlinear Tomlinson-Harashima precoding (THP) is more robust to different types of user selection algorithms and shows better performance than linear zero-forcing beamforming (ZFBF) both in delay and throughput performance.**

I. Introduction

Over recent years, multiple-input multiple-output (MIMO) techniques have received a lot of attention because they can provide not only diversity gain which increases the reliability of transmission but multiplexing gain which boosts data rate by exploiting the space dimension. MIMO techniques were first studied in single user scenarios. It is well known that the single user capacity of a MIMO system increases linearly with the minimum number of transmit and receive antennas. In multiuser MIMO broadcast channels (BCs or downlink, i.e., channels from the base station to mobile users), dirty-paper coding (DPC) [1] can achieve the sum-rate capacity [2]. However, because DPC is difficult to implement in real systems, practical precoding techniques have been developed by using the result of DPC. Tomlinson-Harashima precoding (THP) [3] is one of practical implementation techniques of DPC. It pre-subtracts interference from other antennas successively at the transmitter and employs a nonlinear modulo operation to prevent a possible increase in the transmit power due to the pre-subtraction. A linear beamforming (BF) strategy is the other approach for MIMO broadcast channels. It has been considered as a practical MIMO precoding technique, despite not using the DPC principle, by several authors [4]-[6] because of its reduced complexity and spatial separation among transmit antennas. In addition, the gap between the capacity of DPC and that of linear beamforming is reduced if multiuser diversity can be exploited.

This work was supported by Toshiba Research Europe Limited (TREL) Telecommunications Research Laboratory (TRL) in Bristol, UK.

Multiuser diversity is a form of selection diversity among users in the MIMO BC. If there are sufficiently large numbers of users, the transmitter can select the 'most orthogonal' users with favorable channel conditions for transmission. This means that nonlinear precoding techniques like THP are no longer required to achieve the capacity of DPC, because there is no interference among orthogonal users. It has been shown that the sum-rate of beamforming approaches that of DPC as the number of users goes to infinity [5].

The other implementation issue of MIMO systems for downlink channels is how to provide quality-of-service (QoS) to different users. In real MIMO BC scenarios, each user may request different types of service such as video, file transfer protocol (FTP) and hypertext transfer protocol (HTTP), each with different QoS requirements. Several QoS-based multiuser scheduling algorithms have been suggested for the MIMO BC. In [7], joint linear beamforming and conventional opportunistic scheduling algorithms was proposed. Because of spatial separation due to the beamforming technique, each transmit beamformer can take an independent scheduling decision, which transforms the multiuser scheduling problem into the spatial multiplexing of several single user schedulers. In [8], users were classified into QoS-guaranteed real-time (RT) users and no QoS-guaranteed best-effort (BE) users. It presented that the combination of zero-forcing beamforming (ZFBF) and existing delay-aware scheduling algorithms could provide statistical QoS guarantee to real-time users.

Although these existing algorithms exploit practical multiuser scheduling methods in the MIMO BC, they do not fully take advantage of MIMO techniques and apply only to one specific MIMO precoding technique. For example, in order to use an opportunistic scheduling algorithm in multiuser scenarios, the separation of spatial channels is essential to calculate the opportunistic scheduling rule with the individual rate of each separated channel. This means that nonlinear precoding techniques like THP can not be easily combined with the opportunistic scheduling rule because nonlinear precoding techniques usually require consideration of the capacity region, not the individual rate to get the optimal performance in multiuser scenarios.

In this paper, we propose a new QoS-based multiuser scheduling algorithm in the MIMO BC, which provides QoS-guaranteed services as well as maximizing the total throughput by trading off spatial multiplexing gain for different QoS classes. The object of this paper is that, by flexibly

allocating spatial gain to all users for maximizing the total throughput or to only QoS users for improving the reliability of them, multiple QoS users can be supported while simultaneously non-QoS users may receive the available residual throughput. The key idea of the trade-off is to support non-QoS users without sacrificing the performance of QoS users to maximize the total throughput.

The rest of this paper is organized as follows. In section II, the system model is introduced. In section III, we review several MIMO precoding techniques and their sum-rate capacity in brief. Section IV explains the proposed algorithm. Numerical results are presented in section V. Conclusions are drawn in section VI.

II. System Model

We use boldface to denote matrices and vectors. For any general matrix $\mathbf{A}$, $\mathbf{A}^T$ denotes the transpose and $\mathbf{A}^H$ denotes the conjugate transpose, and diag$\{\lambda_i\}$ denotes a diagonal matrix with the (i, i) entry equal to λ_i.

Let a MIMO broadcast channel has M_T transmit antennas at the base station and K users each with a single receive antenna and $\mathbf{h}_k \in \square^{M_T \times 1}$ denote the channel between the transmit antenna arrays and the receive antenna for user k. Then the downlink channel can be represented as

$$y_k = \mathbf{h}_k^T \mathbf{x} + z_k, \qquad k = 1, \cdots, K, \tag{1}$$

where $\mathbf{x} \in \square^{M_T \times 1}$ is the transmit signal vector, y_k is the received signal for user k and z_k is the complex additive white Gaussian noise (AWGN) for user k with zero mean and unit variance. The transmit signals are assumed to experience path loss, log-normal shadow fading and block multipath fading. The channel is assumed to be fixed during a time slot and to vary independently over time slots. This means recalculating path loss, shadowing and fading coefficients for each user. Then the channel $\mathbf{h}_k$ at any time slot can be expressed as [10]

$$\mathbf{h}_k = \sqrt{SNR_0 \cdot (r_k/R)^{-\alpha} \cdot s_k} \cdot \mathbf{g}_k, \tag{2}$$

where SNR_0 denotes the average signal-to noise-ratio (SNR) at the boundary of cell radius R, r_k is the base-mobile distance of user k, α denotes the path loss exponent, s_k is a log-normal shadow fading variable with standard deviation σ_s. The column vector $\mathbf{g}_k = [g_{k1}\ g_{k2} \cdots g_{kM}]^T$ represents the channel response between the transmit antenna arrays and user k, whose components are independent and identically distributed (*i.i.d.*) zero mean circular symmetric complex Gaussian (ZMCSCG) random variables with unit variance.

III. Review of MIMO Precoding Techniques

In this section, we briefly go over several MIMO precoding techniques (ZF-DPC, THP and ZFBF) and their sum-rate capacity. In this paper, for the sum-rate capacity of each precoding technique, the number of selected users for transmission is assumed to be the same as that of transmit antennas.

A. ZF-DPC

Zero-forcing dirty paper coding (ZF-DPC) [11] exploits the DPC principle [1]; the capacity of a system with known interference at the transmitter is the same as if there were no interference present. It decomposes the MIMO channel into transmit beamforming matrix $\mathbf{F}$ and lower triangular matrix $\mathbf{B}$ using QR decomposition. Because of the lower triangular matrix $\mathbf{B}$, the first encoded user has no interference, the second encoded user experiences interference only from the first encoded user and so on. Any interference caused by users $j > i$ on each user i is forced to zero by the pre-subtraction at the transmitter.

The sum-rate capacity of ZF-DPC is given by [11]

$$C^{ZF-DPC} = \sum_{k=1}^{M_T} \log_2\left(1 + b_{kk}^2 P_k\right), \tag{3}$$

where b_{kk} is the kth diagonal entry of $\mathbf{B}$, b_{kk}^2 is the effective channel gain to user k and P_k is the input signal power allocated to user k with a power constraint of P by $\sum_{k=1}^{M_T} P_k \le P$. The second term in the logarithm function, $b_{kk}^2 P_k$, is the effective SNR of user k at the receiver. It is useful to note that the total power of the MIMO precoding techniques are not constrained in this paper. There may be possible power increase in ZF-DPC due to the pre-subtraction regardless of the input signal power. However, the power increase at the transmitter does not affect the effective SNR at the receiver because it is cancelled through the MIMO channel.

B. THP

Due to the pre-subtraction of interference at the transmitter, the transmit power of ZF-DPC increases, which may not be feasible for the practical implementation. THP employs a modulo operation to prevent the possible large increase of the transmit power of ZF-DPC. However, the modulo operation incurs additional losses to THP [12], which degrades the capacity. Let the modulo loss be denoted as Γ_{THP} [12], then the capacity of THP can be represented as

$$C^{THP} = \sum_{k=1}^{M_T} \log_2\left(1 + \frac{b_{kk}^2 P_k}{\Gamma_{THP}^k}\right), \tag{4}$$

where Γ_{THP}^k is the modulo loss of user k. If the target bit-error rate (BER) of the system is very small (e.g. $BER \le 10^{-6}$), the modulo loss can be approximated as (5) because precoding loss [14] due to the power increase at the transmitter become the most evident among all the losses.

$$\Gamma_{THP}^k \square \frac{M^k}{M^k - 1}, \tag{5}$$

where M_k denotes the modulation order of M-QAM for user k. Note that the modulo loss becomes negligible as M_k increases.

C. ZFBF

Unlike the above two nonlinear precoding techniques, the beamforming vectors of linear ZFBF are selected to avoid

interference among users by inverting the MIMO channel matrix. This results in power inefficiency because any user in ZFBF can not transmit its signal in the direction of the other users. However, this gives complete spatial separation among users because of perfect nulling.

The capacity of ZFBF is given by [5]

$$C^{ZFBF} = \sum_{k=1}^{M_T} \log_2\left(1+\gamma_k P_k\right), \tag{6}$$

where γ_k is the effective channel gain of user k, which is represented as

$$\gamma_k = \frac{1}{\mathrm{diag}\left\{\left(\mathbf{HH}^H\right)_k^{-1}\right\}}. \tag{7}$$

IV. Trade-offs of Spatial Gain of MIMO Techniques

In this section, we explain the proposed scheduling algorithm. We assume that there are two kinds of QoS users: delay-constrained real-time (RT) users and best-effort (BE) users with no constraints. Each class has a number of pre-assigned transmit antennas but there are different priorities for the actual use of them for transmission. Real-time users have higher priority than best-effort users in using the transmit antennas. For example, when the performance of real-time users does not satisfy a certain delay threshold, best-effort users have no chance of using their pre-assigned antennas for transmission.

Since the channel $\mathbf{h}_k$ is random, the capacity associated with the MIMO BC is also a random variable. Hence, we use a statistical QoS for the analysis. It means that no scheduling algorithm can always guarantee 'deterministic' QoS in wireless channels. Fortunately, in most real wireless systems, these hard QoS conditions are not usually required [8]. They can tolerate a small amount of instantaneous QoS violation (i.e., by buffering). For this reason, a guaranteed-delay in wireless systems means that the average delay is within a target delay threshold with a certain outage probability.

Let $U := \{u_k \mid k = 1, \cdots, K\}$ denote the total user set and $U^{RT} := \{u_l \mid l = 1, \cdots, K_{RT}\}$ and $U^{BE} := \{u_l \mid l = 1, \cdots, K_{BE}\}$ are the set of real-time and best-effort users respectively. U^{RT} and U^{BE} are disjoint, each has K_{RT} and K_{BE} elements respectively with $K_{RT} + K_{BE} = K$. Indeed, let S^{RT} $\left(S^{RT} \subset U^{RT}, |S^{RT}| = M^{RT}\right)$, S^{BE} $\left(S^{BE} \subset U^{BE}, |S^{BE}| = M_T - M^{RT}\right)$ denote the set of selected real-time and best-effort users respectively, where M^{RT} is the number of pre-assigned transmit antennas for real time users. In this case, the set of selected users for transmission by a certain selection criterion at any given time is represented as S $(S \subset U, |S| = M_T)$.

Figure 1 shows the concept of the proposed algorithm. The function $f(x, y, z)$ represents the user selection rule, where x is pre-allocated users, y is a user set for the selection and z is the number of transmit antennas left for the selection.

At first, the algorithm selects S^{RT} by a certain selection rule assuming M^{RT} pre-assigned transmit antenna beam patterns. The selection rule can be any performance criterion such as the sum-rate [4] selection which maximizes the sum-rate capacity of the selected users; the norm-based [9] selection which selects users with large channel norm $\|\mathbf{h}_k\|^2$; or the round-robin selection which gives equal opportunity for transmission to all users. For example, the sum-rate rule finds the real time user set $S^{RT}_{\max}$ by calculating the capacity formula (3), (4) or (6) in the form shown in (8) with every possible user group of M^{RT} users from all the real time users.

$$S^{RT}_{\max} = \arg \max_{S^{RT} \subset U^{RT}, |S^{RT}| = M^{RT}} C\left(S^{RT}\right). \tag{8}$$

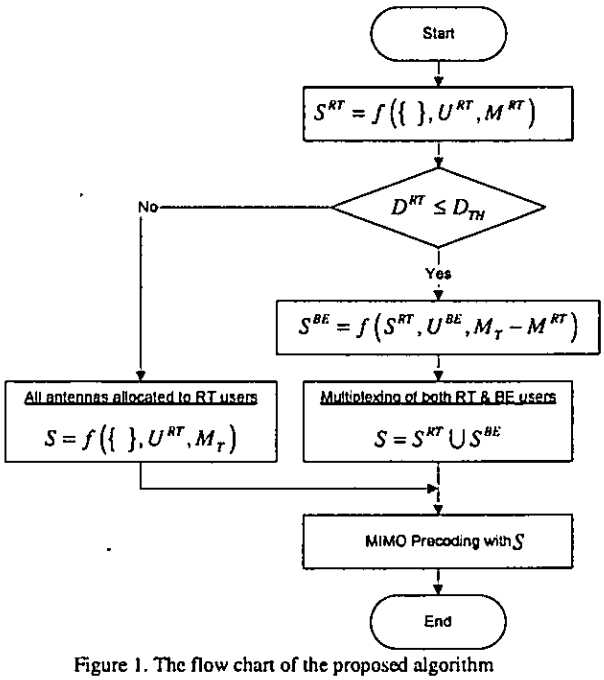


Figure 1. The flow chart of the proposed algorithm

There are $\binom{K_{RT}}{M^{RT}}$ possible ways of choosing S^{RT} from U^{RT} for the selection of real time users if a group-wise selection rule like the sum-rate rule is employed. However, the norm-based and the round-robin rules can select S^{RT} directly from U^{RT} without considering all possible combinations of user groups, which reduces the complexity of the selection of users compared to the sum-rate rule.

If the average delay of S^{RT} exceeds the delay threshold (D_{TH}), all transmit antennas are assigned to the real-time users to reduce the delay of the real-time users, and the set S_{BE} becomes the empty set. Otherwise, S^{BE} is chosen among best-effort users. The former case uses spatial multiplexing of real time users to guarantee the delay performance and the latter case uses spatial multiplexing of both real-time and best-effort users to increase

the throughput performance of all users. This trade-off of spatial gain is determined by the delay threshold for real-time users.

Note that any group-wise selection of best-effort users should include S^{RT}, the set of 'already-selected' real-time users. For example, if the sum-rate rule is employed, any best-effort user who has high cross correlation with users in S^{RT} might be excluded even though they have high individual channel gain. In other words, in order to find a set of the maximum sum-rate users from U^{BE}, one should select the most orthogonal users from previously selected real-time users as well as best-effort users themselves, which reduces the multiuser diversity effect. However, the selection of best-effort users does not affect the performance of the selected real-time users if nonlinear MIMO precoding techniques are used for these selected users. For example, if THP is used for the MIMO precoding technique, any possible interference to real-time users caused by best-effort users can be removed by the pre-subtraction of interference at the transmitter if the channel state information (CSI) is perfect.

V. Numerical Results

In this section, numerical results for the performance of the proposed algorithm are presented. Throughout the simulations, we assume SNR_0 = 6 dB, $\alpha = 4$, R = 1 km and shadowing standard deviation of 8 dB for the MIMO downlink channel. We also assume that the received SNR of each user experiences the saturation due to several practical impairments caused by antennas, radio frequency (RF) circuitry, analogue-to-digital converters (ADC), digital-to-analogue converters (DAC) and so on. We do not investigate the source of the saturation in this paper but assume the received SNR is saturated at 40 dB. The saturation function is given by

$$g(x) = \frac{x}{1+\beta^{-1}x}, \tag{9}$$

where β is a desired saturation value, and x is the SNR determined by the MIMO channel parameters except the practical impairments.

For MIMO precoding techniques, we use zero-forcing dirty-paper coding (ZF-DPC) [11], THP and ZFBF [5]. Each MIMO precoding technique uses M-QAM signal sets with cardinality $M = 2^2, \ldots, 2^8$, which corresponds to integer data rates ranging from 2 to 8 bits per symbol interval. This means that the capacity term of each user in (3), (4) and (6) is quantized to integer data rates from 2 bits/channel use to 8 bits/channel use by the floor function. Because we assume the received SNR is saturated at 40 dB, 256-QAM will be sufficient for any user with large channel gain. Although the encoding order of nonlinear MIMO precoding techniques affects the capacity of each individual user in the MIMO BC, the selected users are ordered according to their effective channel gain for simplicity. We also assume that the total input signal power of P = 1 and each user has the same input power as $P_1 = P_2 = \cdots = P_K = 1/M_T$. Data packets for all users arrive in queues at any time slot with mean arrival rate $\lambda = 2$ bits/channel use of a Poisson process independently and wait for transmission. For the delay performance, we evaluate the head-of-line (HOL) delays of real-time users.

Figure 2 shows the throughput performance versus the number of users for different combinations of MIMO precoding techniques and user selection criteria with $M_T = 4$. First, if all users are not divided into two subgroups ("No div."), the throughput of ZF-DPC is best among MIMO precoding techniques. THP performs worse than ZF-DPC due to modulo loss [12]. Although the capacity of linear beamforming goes to that of DPC as the number of users K goes to infinity [5], it performs worst with a practically finite number of users. Second, we notice that the division of users according to their QoS classes results in poorer performance due to the reduction of the multiuser selection diversity gain. However, we notice that the performance degradation of ZFBF is worse than that of THP with the division of users. Third, user selection algorithms affect the throughput performance. The group-wise sum-rate rule shows the best performance because it selects a user set which has the maximum capacity among all possible combinations but it requires much more computation than other selection rules. Because the norm-based rule does not consider interference among users, it has worse performance than the sum-rate rule. The round-robin rule shows the poorest capacity performance, which is unchanged regardless of the number of users. This is because the round-robin rule does not consider the channel conditions of all users when it chooses users for transmission. It is useful to note that the performance degradation of linear ZFBF with the norm-based rule is more severe than that of nonlinear THP with the norm-based rule, which shows that the performance of THP is more robust to the type of the user selection algorithm than ZFBF. If we compare the computational complexity of finding a maximum sum-rate user set with that of finding users with large norm of the channel, THP with the norm-based rule can be a reasonable engineering trade-off between complexity and performance for practical systems.

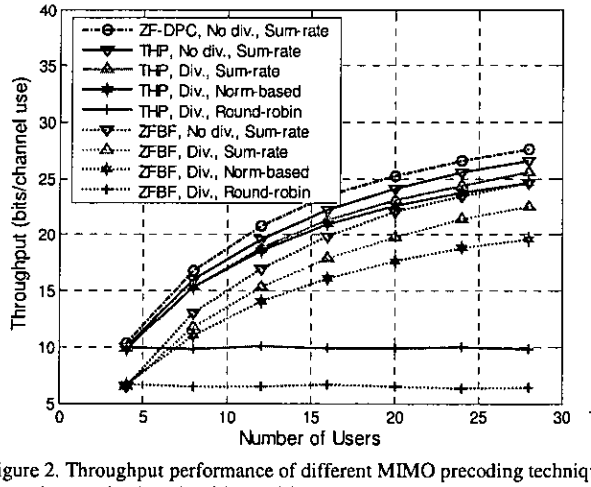


Figure 2. Throughput performance of different MIMO precoding techniques and user selection algorithms with the proposed algorithm for $M_T = 4$

Figure 3 shows the computational complexity of the proposed algorithm if the sum-rate rule is used for the selection of users. We notice that the computational complexities are rapidly reduced as the number of subgroups increases. The complete search with no division of users rapidly becomes prohibitive even for practical values of K, e.g., $K > 30$. However, it is obvious that as K approaches to infinity, the throughput of the proposed algorithm goes to the throughput without division of user group regardless of the number of subgroups provided the number of subgroups is finite and does not increase with the number of users. This is because we can always find an orthogonal user set of M_T users among the finite subgroups with infinite number of users.

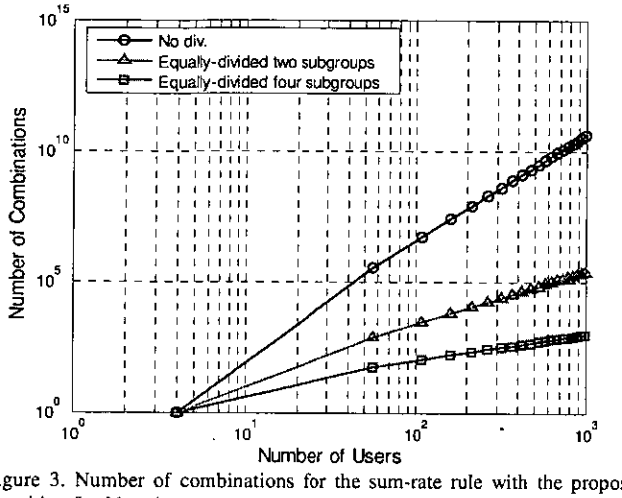


Figure 3. Number of combinations for the sum-rate rule with the proposed algorithm for $M_T = 4$

The delay performance of real-time users is presented in Figure 4. THP and the norm-based rule are used for the plot. It is assumed that there are 25 % of real-time users among all users and the delay threshold (D_{TH}) is the average HOL delay of real-time users. We assume 90% outage delay for the delay performance where the delay is guaranteed for 90% of the channel realizations so that 10% of delay values can exceed the delay threshold. We notice that the proposed algorithm can satisfy any delay threshold (D_{TH}) for a certain range of the number of users. While the trade-off of spatial gain occurs, the delay stays near the delay threshold but the delay grows rapidly when the number of active users is more than the theoretical users that the MIMO system can support. With $D_{TH} = 4096$ slots, the scheduler can guarantee the delay of the real time users up to 48 users when $M_T = 4$ is used. With $M_T = 8$, the scheduler can support about two times as many users as the case of $M_T = 4$ while maintaining the delay of the real time users at $D_{TH} = 4096$ slots. If there are more transmit antennas, the delay-guaranteed range is increased because the MIMO capacity is significantly increased.

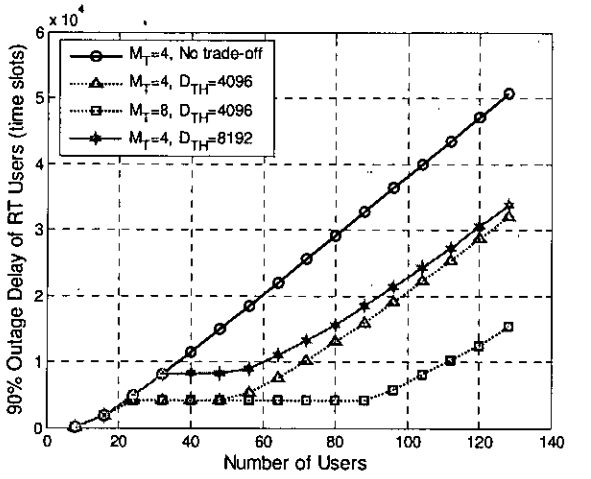


Figure 4. Delay Performance of real-time users with the proposed algorithm for different delay thresholds and antenna configurations for THP, norm-based rule and $M^{RT} = 1$

Figure 5 shows how the trade-off of spatial gain affects the throughput performance. We assume THP with the norm-based rule, $M_T = 4$ and $M^{RT} = 1$. If there is no trade-off, the total throughput increases as the number of users grows because spatial gain is always used as multiplexing for maximizing the throughput. However, with the trade-off, the total throughput is reduced to guarantee the delay of real time users. When spatial gain is used only for real-time users, the multiuser diversity gain is reduced. We notice that as the number of users increases, the throughput of best-effort users decreases because the real-time users take all the transmit antennas more frequently to guarantee their delay performance, which results in less chance of transmission for best-effort users.

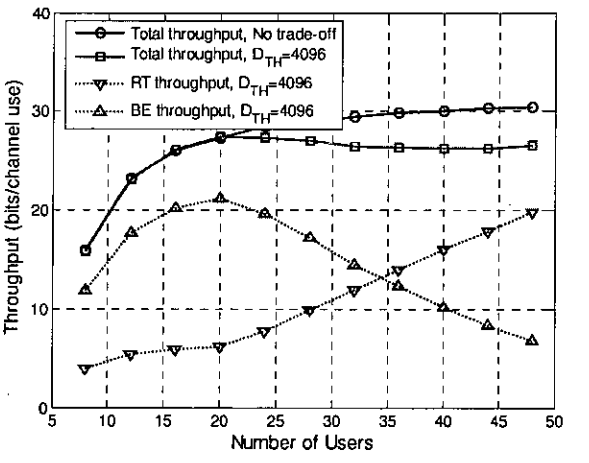


Figure 5. Throughput performance of real-time, best-effort and total users for the proposed algorithm for THP, norm-based rule, $M_T = 4$ and $M^{RT} = 1$

Figure 6 shows the delay performance for different MIMO precoding techniques and user selection algorithms for D_{TH} = 4096 slots, $M_T = 4$ and $M^{RT} = 1$.

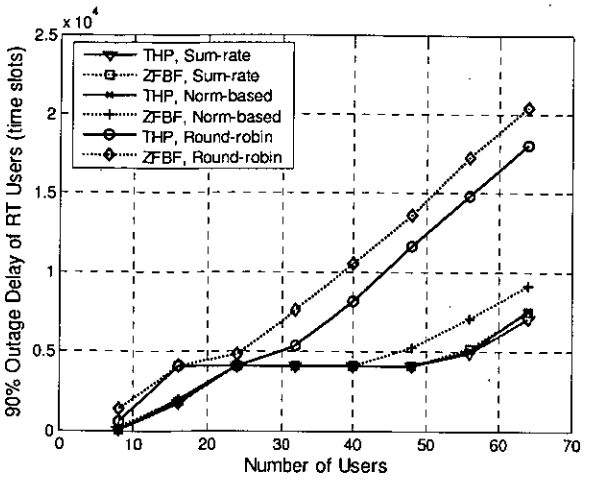


Figure 6. Delay Performance of real-time users with the proposed algorithm for different MIMO precoding and user selection algorithms for $D_{TH} = 4096$, $M_T = 4$ and $M^{RT} = 1$

We notice that the sum-rate rule has better performance than other selection rules regardless of the MIMO precoding technique used. In this case, the number of users guaranteeing the delay performance of the real time users ranges up to about 48 users. If the norm-based rule is used, the delay performance of THP is slightly worse than the case of the sum-rate rule. ZFBF with the norm-based rule shows worse delay performance than THP with the norm-based rule because THP can actively remove mutual interference by the pre-subtraction, which is similar to finding a set of orthogonal users. The round-robin rule has the poorset performance among selection rules because it does not consider channel conditions at all. Note that if the system should admit more than 48 users with THP and the norm-based rule with parameters in this paper, a system engineer may assign more transmit antennas, an extra hardware burden or use a more computationally complex user selection algorithm, an extra software burden.

VI. Conclusions

In this paper, we proposed a new multiuser scheduling algorithm which trades spatial gain of multiple transmit antennas between best-effort users and real-time users for QoS-guaranteed services in the MIMO broadcast channels. Because the trade-off of spatial gain is used for the delay performance of real time users, the selection rule can focus on how to maximize the total throughput of all users, not any other QoS parameters. In other words, the maximum throughput can be supported by the selection algorithm while guaranteeing the delay of real-time users by the flexible use of spatial gain.

In view of practical implementation, nonlinear Tomlinson-Harashima precoding has better performance both in delay and throughput and is more robust to different types of selection algorithms than linear zero-forcing beamforming. THP with the norm-based selection rule can be a reasonable combination, despite not being an optimal approach in the sense of maximizing the sum-rate capacity, for practical multiuser MIMO downlink systems with different QoS requirements.

However, the proposed algorithm in this paper does not consider changing the number of pre-assigned transmit antennas for real-time users when their average delay does not satisfy the delay threshold. It only switches between a real-time user transmission-only mode and a simultaneous real-time and best-effort user transmission mode. The QoS of real time users can be achieved at the cost of the total throughput. Hence, if there are several delay thresholds for real time users, the scheduler can make a gradual decision on trade-offs either by changing the number of pre-assigned transmit antennas or by switching the transmission mode. In this case, the system performance can be further improved.

References

[1] M. Costa, "Writing on dirty paper," *IEEE Trans. Info. Theory*, vol. 29, pp.439-441, May 1983.

[2] S. Vishwanath, N. Jindal, and A. Goldsmith, "Duality, achievable rates, and sum-rate capacity of Gaussian MIMO broadcast channels," *IEEE Trans. Info. Theory*, vol. 49, no. 10, pp. 2658-2668, Oct. 2003.

[3] C. Windpassinger, R. F. Fischer, T. Vencel, and J. B. Huber, "Precoding in multiantenna and multiuser communications," *IEEE Trans. Commun.*, vol. 3, no. 4, pp.1305-1316, July 2004.

[4] N. Jindal, "High SNR analysis of MIMO broadcast channels," *in Proc. International Symposium on Information Theory*, pp. 2310- 2314, 4-9 Sept. 2005.

[5] T. Yoo and A. Goldsmith, "On the optimality of multiantenna broadcast scheduling using zero-forcing beamforming," *IEEE Journal on Sel. Commun.*, vol. 24, no. 3, pp. 528-541, March 2006.

[6] C. Swannack, E. Uysal-Biyikoglu and G. Wornell, "Low complexity multiuser scheduling for maximizing throughput in the MIMO broadcast channel," *in Proc. 42nd Annual Allerton Conf. on Communications, Control, and Computing, Monticello, Illinois*, 2004.

[7] M. Kountouris, A. Pandharipande, H. Kim and D. Gesbert, "QoS-based user scheduling for multiuser MIMO systems," *VTC- 2005 Spring, Stockholm, Sweden*, vol. 1, pp. 211-215, 2005.

[8] W. Ajib, D. Haccoun and J. Frigon, "An efficient QoS based scheduling algorithm for MIMO wireless systems," *VTC-2005 Fall, Dallas, U.S.*, vol. 3, pp. 1579-1583, 2005.

[9] S. Sanayei and A. Nosratinia, "Antenna selection in MIMO systems," *IEEE Commun. Mag.* pp. 68-73, Oct. 2004.

[10] O. Shin and K. B. Lee, "Antenna-assisted round-robin scheduling for MIMO cellular systems," *IEEE Commun. Letters*, vol. 7, no. 3, pp. 109-111, March 2003.

[11] G. Caire and S. Shamai, "On the achievable throughput of a multiantenna Gaussian broadcast channel," *IEEE Trans. Info. Theory*, vol. 49, no. 7, pp. 1691-1706, July 2003.

[12] W. Yu, D. P. Varodayan and J. M. Cioffi, "Trellis and convolutional precoding for transmitter-based interference pre-subtraction," *IEEE Trans. Commun.*, vol. 53, no. 7, pp. 1220-1230, July 2005.

[13] C. F. Fung, W. Yu and T. J. Lim, "Precoding for the multi-antenna downlink: multi-user gap approximation and optimal user ordering," *accepted in IEEE Trans. Commun., 2006.* Available: www.comm.utoronto.ca/~weiyu/precoding.pdf.

[14] C. Windpassinger, R. F. Fischer, T. Vencel, and J. B. Huber, "Precoding in multiantenna and multiuser communications," *IEEE Trans. Commun.*, vol. 3, no. 4, pp.1305-1316, July 2004.

[15] S. Shakkottai, T. S. Rappaport and P. C. Karlsson, "Cross-layer design for wireless networks," *IEEE Commun. Mag.*, pp. 74-80, Oct. 2003.

[16] Q. H. Spencer, C. B. Peel, A. L. Swindlehurst and M. Haardt, "An introduction to the multi-user MIMO downlink," *IEEE Commun. Mag.*, pp. 60-67, Oct. 2004.

A Simple Channel Model for Investigation of Multiuser Scheduling in MIMO Broadcast Channels

S. Lee and J. S. Thompson
University of Edinburgh, Institute for Digital Communications, Edinburgh, UK
{S.Lee, John.Thompson}@ed.ac.uk

Abstract- **In this paper, a simple multiple input multiple output (MIMO) broadcast channel model is proposed and compared with measured MIMO channels. The proposed model uses a sum-of-sinusoids model for temporal correlation and the Kronecker model for spatial correlation. MIMO channel measurements were performed in the city of Bristol, UK as a part of the Mobile VCE programme. These were conducted in the 2GHz frequency band with two pairs of dual polarized transmit antennas (4 transmit antennas) at the base station and with two pairs of crossed dipoles (4 receive antennas) at the receiver for different locations in a dense urban cellular environment. For validation of the proposed MIMO model, the standard deviation of cross correlations and eigenvalues are compared between the measured data and the simulated data. Numerical results demonstrate that the temporal correlation is more meaningful than the spatial correlation in view of multiuser scheduling in MIMO broadcast channels. Indeed, the proposed model has similar characteristics to the measured MIMO channels taking into account temporal correlation properties.**

I. Introduction

For the original MIMO capacity analysis of [1] and [2], the entries of MIMO channel matrices were assumed to be independent and identically distributed (*i.i.d.*) complex Gaussian random variables. Afterwards, various MIMO channel models have been proposed to describe real channel environments. The Kronecker model is the most well-known simplified spatial model for MIMO channels. This concept was validated by comparison with measurement results [3][4]. The Kronecker structure of the channel covariance matrix is suitable for arrays with a moderate number of antenna elements. However, the elements of real MIMO channels are correlated both in time due to Doppler spread effects and space due to insufficient spacing between multiple antenna elements. Joint spatial temporal models were presented in [5], where spatial and temporal correlations are modeled based on angle of departure (AOD), Doppler spread, shadowing and path loss, in addition to conventional spatial model parameters such as antenna spacing, beamwidth and so on. As the number of parameters considered increases, the complexity of the model also grows, making it difficult to implement and run the MIMO channel efficiently despite its improved accuracy.

For a multiuser scheduler in a MIMO broadcast channel (BC, or downlink, i.e. channels from the base station to mobile terminals), the temporal correlation is one of the most crucial factors. The reason is as follows; if a MIMO precoding technique is used in the MIMO BC, channel state information (CSI) at the transmitter is essential for transmit processing. It is likely that there is a time difference between obtaining CSI and using that CSI at the transmitter in a temporally correlated channel. Because the channel is time-varying due to the Doppler spread, there always exist mismatches of the channel estimate caused by that difference. This degrades the performance of the MIMO precoding technique. However, if the multiuser scheduler exploits the statistical CSI (SCSI) such as mean and covariance from the past instantaneous CSI (ICSI), it can allocate users to different time slots more efficiently whilst considering the temporal correlation [6]. This can minimize the performance degradation due to imperfect CSI caused by temporal variations. Unlike the case of the temporal correlation, the effect of the spatial correlation caused by insufficient antenna spacing might not be overcome completely by the multiuser scheduler because it is determined by the physical antenna configuration and channel environments. In this case, the performance degradation can be minimized if the antenna spacing is sufficient to guarantee very small spatial correlation values or rich scattering, so that the MIMO channel matrix has *i.i.d.* elements. This means that parameters related to the temporal correlation are more important than those related to the spatial correlation to make a realistic simulation model for investigating the performance of the multiuser scheduler in the MIMO broadcast channel.

Therefore, the object of this paper is to provide a simple but realistic MIMO BC model for multiuser scheduling algorithms. There are a lot of factors that should be considered for the modeling of real MIMO BCs but it is not only difficult but too complex to include all these factors in a simulation model. Some parameters may have little effect despite the complexity to model them accurately. In order to validate our proposed model, we compare it with MIMO channel measurements followed by

This work was supported by Toshiba Research Europe Limited (TREL) Telecommunications Research Laboratory (TRL) in Bristol, UK.

an example in Section IV, which explains why the temporal correlation is more meaningful than the spatial correlation in view of multiuser scheduling.

The rest of this paper is organized as follows. In section II, the proposed MIMO broadcast channel model is introduced. Section III explains the MIMO channel measurements. Section IV compares the effect of spatial and temporal correlations on multiuser scheduling in terms of scheduling. Validation of the proposed model is presented in Section V. Conclusions are drawn in Section VI.

II. The Proposed MIMO Broadcast Channel Model

We use boldface to denote matrices and vectors. For any general matrix $\mathbf{A}$, $\mathbf{A}^T$ denotes the transpose, $\mathbf{A}^H$ denotes the conjugate transpose and Tr(A) denotes the trace.

Consider a MIMO broadcast channel with M_T transmit antennas at a base station and K users each with a single receive antenna. Let $\mathbf{h}_k(n) \in \mathbb{C}^{M_T \times 1}$ denote the channel at time n between the transmit antenna array and the receive antenna for user k. Note that we use a discrete time index n to represent the time variation. Then the downlink channel can be represented as

$$y_k(n) = \mathbf{h}_k^T(n)\mathbf{x}(n) + z_k(n), \qquad k = 1, \cdots, K \ , \tag{1}$$

where $\mathbf{x}(n) \in \mathbb{C}^{M_T \times 1}$ is the transmit signal vector with a power constraint $\mathrm{Tr}(\mathbf{R}_{xx}) \le P$, $y_k(n)$ is the received signal for user k and $z_k(n)$ is the complex additive white Gaussian noise (AWGN) with zero mean and unit variance for user k.

The channel is assumed to exhibit path loss, log-normal shadow fading and Rayleigh fading as in conventional radio channel models. Then the channel $\mathbf{h}_k(n)$ can be expressed as

$$\mathbf{h}_k(n) = \sqrt{SNR_k \left(\frac{r_k(n)}{R_c}\right)^{-\alpha_l}}\, s_k(n) \cdot \mathbf{g}_k(n) \ , \tag{2}$$

where SNR_k denotes the average signal-to noise-ratio (SNR) of user k, $r_k(n)$ denotes the distance between user k and the base station, R_c denotes the cell radius, α_l denotes the path loss exponent, $s_k(n)$ denotes the log-normal shadowing for user k. The vector $\mathbf{g}_k(n) = [g_{k1}(n)\, g_{k2}(n) \cdots g_{kM_T}(n)]^T$ represents the Rayleigh-distributed fading between the transmit antenna array and user k. Although path loss exists in real radio channels, it can be ignored if all users are located at the same distance from the base station, so that the proposed model does not consider path loss parameters. The effect of shadowing is modeled as a log-normal process with standard deviation, which can be expressed as [7]

$$s_k(n) = 10^{[\sigma_s \gamma_k(n)]/20} \ , \tag{3}$$

where $\gamma_k(n)$ is a sum-of-sinusoids (SOS) random process of user k at time instant n, which is given by

$$\gamma_k(n) = \sum_{i=1}^{N_s} c_{ki} \cos(2\pi f_{ki} n + \theta_{ki}) \ , \tag{4}$$

In (4), the gains c_{ki} and the frequencies f_{ki} are non-zero, real-valued constant quantities for user k, N_s is the number of sinusoids and the phases θ_{ki} are *i.i.d.* uniform random variables on the interval $(0, 2\pi]$ for user k. There are several methods to determine the parameters c_{ki} and f_{ki}. In this paper, we use the method of equal areas (MEA) [7] and the following equations are derived for those parameters

$$c_{ki} = \sqrt{\frac{2}{N_s}} \ , \ f_{ki} = \frac{v_k}{2\pi D} \tan\left[\frac{\pi(i-0.5)}{2N_s}\right], \tag{5}$$

where D denotes the decorrelation distance. For Rayleigh fading, the sum-of-sinusoids statistical model [8] is applied. Many conventional Rayleigh fading models are based on the Jakes model [9]. However, the Jakes model is a deterministic model, which has limitations when modeling *i.i.d.* random multiple antenna matrix entries. According to the Rayleigh fading model, the Rayleigh fading variable $g_{km}^{(n)}$ of the mth transmit antenna for user k is given by

$$\begin{aligned} g_{km}(n) = \sqrt{\frac{1}{N_s}} \Bigg\{ & \sum_{i=1}^{N_s} \cos(\psi_{i,km}) \cos\left[2\pi f_k^d n \cdot \cos\left(\frac{2\pi i - \pi + \theta_{km}}{4N_s}\right) + \phi_{km}\right] \\ & + j \sum_{i=1}^{N_s} \sin(\psi_{i,km}) \cos\left[2\pi f_k^d n \cdot \cos\left(\frac{2\pi i - \pi + \theta_{km}}{4N_s}\right) + \phi_{km}\right] \Bigg\}, \end{aligned} \tag{6}$$

where f_k^d denotes the Doppler frequency of user k, θ_{km}, ϕ_{km}, and $\psi_{i,km}$ are mutually *i.i.d.* uniform random variables over $[-\pi, \pi)$ for all i, m and k. With this model, the temporal correlation of any user with delay τ can be given by $J_0(2\pi f^d \tau)$, where J_0 is the zeroth order Bessel function of the first kind.

Assuming the base station selects M_T users and forms a channel matrix $\mathbf{H}\,(\mathbf{H} = [\mathbf{h}_1 \cdots \mathbf{h}_{M_T}]^T)$ for transmission, with the Kronecker spatial correlation model, the channel matrix $\mathbf{H}$ can be given by [4] and [10]

$$\mathbf{H} = \mathbf{R}_r^{1/2} \mathbf{H}_w \mathbf{R}_t^{1/2} \ , \tag{7}$$

where $\mathbf{R}_r$ is the $M_T \times M_T$ spatial correlation matrix at the receiver, $\mathbf{R}_t$ is the $M_T \times M_T$ spatial correlation matrix at the base station and $\mathbf{H}_w$ is a $M_T \times M_T$ matrix of which elements are *i.i.d.* complex Gaussian random variables with zero mean and unit variance. For the spatial correlation matrices, $\mathbf{R}_r$ can be assumed to be an identity matrix if users are separated enough to ignore the spatial correlation between them. Assuming a linear array of equally-spaced M_T parallel antenna elements at the base station, the spatial correlation matrix can be given by the following Toeplitz matrix [11]

$$\mathbf{R}_t = \begin{bmatrix} 1 & \rho_s & \cdots & \rho_s^{(M_T-1)^2} \\ \rho_s & 1 & \cdots & \cdots \\ \cdots & \cdots & \cdots & \rho_s \\ \rho_s^{(M_T-1)^2} & \cdots & \rho_s & 1 \end{bmatrix}, \tag{8}$$

where ρ_s represents the spatial correlation between two adjacent antenna elements, which can be approximated as [12]

$$\rho_s \approx \exp\left[-23\Lambda^2 \left(\frac{d}{\lambda}\right)^2\right]. \tag{9}$$

In (9), Λ denotes the angular spread, d denotes the distance between two adjacent transmit antenna elements and λ denotes the wavelength of the carrier frequency.

Table I summarizes the proposed channel model.

TABLE I

THE PARAMETERS FOR THE PROPOSED CHANNEL MODEL

Item	Parameters
Antenna configuration	- Transmitter: M_T - Receiver: K users each with a single antenna
Path loss exponent	α_l
Long term fading	Log-normal shadowing (σ_s, D)
Short term fading	Rayleigh
Doppler spread	f^d
Delay spread	None
Temporal correlation	$J_0(2\pi f^d \tau)$
Spatial correlation	Kronecker model (Tx)

III. THE MIMO CHANNEL MEASUREMENTS

The MIMO channel measurements were conducted in the city centre of Bristol, UK as part of a Mobile Virtual Centre of Excellence (M-VCE) elective programme [13]. This campaign used realistic antenna deployments, where two pairs of dual polarized transmit antennas at the base station and two pairs of crossed dipoles for reception were used at different locations in a dense urban cellular environment. The transmitting site was located on the roof of a building 33m above ground level and with a coverage area of one 65° sector. MIMO channel soundings were made with a user walking or sitting within the coverage area. Parameters for the MIMO channel soundings are given in Table II.

TABLE II

MEASUREMENT PARAMETERS FOR MIMO CHANNEL SOUNDINGS

Parameter	Setting
Carrier Frequency	2GHz
Measurement Bandwidth	20MHz
Number of TX antennas	4
Number of RX antennas	4
Time Grid for Sampling the Channel	6.144ms
Measurement Duration	6s

IV. EFFECT OF SPATIAL AND TEMPORAL CORRELATIONS ON MULTIUSER SCHEDULING

In this section, the performance of the SCSI-assisted multiuser scheduling algorithm [6] is analyzed to show why the temporal correlation is more meaningful than the spatial correlation. For this, Tomlinson-Harashima precoding (THP) [14] is used for the MIMO precoding technique.

With assumptions of an equal transmit power allocation and high SNR, the sum-rate capacity of THP with M_T selected user set can be given by

$$C^{THP} = \sum_{k=1}^{M_T} \log_2\left(1 + \frac{b_{kk}^2 P}{M_T}\right), \tag{7}$$

In order to investigate the effect of the temporal correlation, it is assumed that one downlink frame of the MIMO BC with time period T_f consists of N_{slot} ($N_{slot} > 1$) time slots each with time period T_s. Then the ICSI is updated once per every frame at the base station. Assuming the Jakes power spectrum as in Section II, the temporal correlation function when $n = mT_s$ after the initial accurate channel measurement at $n = 0$ can be written as

$$\rho_t(mT_s) = J_0\left(2\pi f_k^d m T_s\right), \quad m = 1, \cdots, N_{slot}. \tag{8}$$

In this case, any decrease in ICSI becomes a delay-induced channel estimation error due to the temporal correlation [14]. Then the sum-rate capacity considering the temporal variation can be written as

$$C^{THP}(n) = \sum_{k=1}^{M_T} \log_2\left(1 + \frac{b_{kk}^2 \rho_{t,k}^2(n) P}{M_T\left\{\left(1-\rho_{t,k}^2(n)\right)P + 1\right\}}\right), \tag{9}$$

where b_{kk}^2 is the effective channel gain.

Because a frame consists of multiple slots, with the SCSI-assisted multiuser scheduling algorithm, a user set for transmission is selected on a slot-by-slot basis using SCSI, so that the sum-rate capacity can be improved compared to a frame-by-frame multiuser scheduler without SCSI. Although the base station obtains ICSI once per every frame, it can utilize the previous ICSI for obtaining SCSI such as mean and covariance due to channel stationarity [15]. This SCSI can be used for estimating the temporal correlation of the MIMO BC when precise ICSI is not available. The SCSI-assisted sum-rate maximization rule finds a user set $S_{max}(n)$ at every slot index n according to

$$S_{\max}(n) = \arg \max_{S \subset \{1,\cdots,K\}, |S| = M_T} C^{THP}(n). \tag{10}$$

According to this selection rule, any user with a rapid decrease in the temporal correlation may be assigned to the first few slots in the frame. This may improve the sum-rate capacity in return for increased complexity. This is because the multiuser selection procedure has to be performed on a slot-by-slot basis if the SCSI-assisted algorithm is applied.

Figure 1 shows the performance of the SCSI-assisted multiuser scheduling algorithm for different spatio-temporal channel parameters. We assume the frame period $T_f = 2$ms, the number of transmit antennas $M_T = 4$, the number of users $K = 8$ and the number of time slots in a frame $N_{slot} = 5$. The relative Doppler spread values of all users $f^d T_f$, are assumed to be *i.i.d.* uniform random variables over [0.1, 0.5). The antenna spacing is assumed to be $d = 0.5\,\lambda$. When the relative Doppler spreads of all users have a uniform distribution over [0.1, 0.5), the SCSI-assisted multiuser scheduling algorithm (label 'SCSI') performs better than the normal multiuser scheduling algorithm (label 'No SCSI'), which does not use SCSI and applies a frame-basis multiuser selection. The performance difference between these two schemes is evident as the SNR increases because the delay-induced channel estimation error produces an irreducible error floor in the high SNR region. We notice that the performance improvement by using the SCSI-assisted multiuser scheduling algorithm can be achieved regardless of the angular spread (Λ), which is related with the spatial correlation. When the SNR_k = 20dB, the capacity increase is about 2 [bits/channel use] regardless of spatial channel

parameters. A smaller angular spread (Λ) means higher spatial correlation between antennas, which reduces the sum-rate capacity. This plot shows that the effect of the temporal correlation can be minimized by using an appropriate multiuser scheduling algorithm. However, the spatial correlation cannot be overcome by the multiuser scheduling algorithm used in this paper.

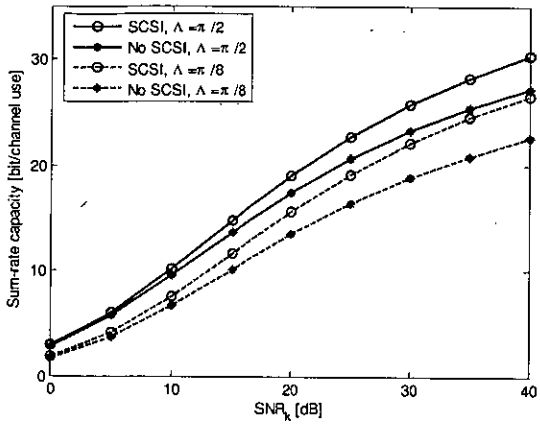


Figure 1. The performance of the SCSI-assisted multiuser scheduling algorithm in spatio-temporally correlated MIMO broadcast channels

V. Validation of the Proposed Channel Model

In this section, numerical results for the validation of the proposed model are presented. Throughout the simulations for the proposed MIMO broadcast channel model, we assume the number of transmit antennas $M_T = 4$, the number of users $K = 4$ and the average signal to noise ratio SNR_k = 0dB for all users and the number of sinusoids for Rayleigh fading and shadowing N_s=128, which is sufficiently large enough to have good agreement with desired properties [7][8].

In order to model the large scale environment of the MIMO channel measurements, the shadowing standard deviation of the measured data from 19 different locations was estimated. Assuming no significant path loss effect, the shadowing standard deviation is roughly estimated as $\sigma_s = 6.2$dB by mapping points with the received power and its distance from the base station. This result seems reasonable since, in [16], the measured values for standard deviation of shadowing are 7.5dB and 4.3 dB for suburban and urban environments respectively. The decorrelation distance for shadowing might be several tens of meters for the urban environment but the decorrelation distance D = 30m is assumed in this paper considering the geometrical characteristics of the city of Bristol. We also assume that the carrier frequency $f_c = 2$GHz and the Doppler spread $f^d = 8$Hz, which corresponds to walking pace for all users. The time period between discrete channel realizations is set to be 6.144 ms, which is the same as time grid used in the measurements. By normalizing each column vector of the channel matrix over the time axis separately, the effect of cross polarization discrimination (XPD), defined as a mean level difference between two orthogonal polarized components, can be ignored in the measured data.

Figure 2 shows a comparison between the simulated data and the measured data both in the time domain and in the Doppler domain. We notice that the Doppler spread of the walking reference is about 8Hz, which corresponds to a velocity of about 4.3km/h at the 2GHz center frequency. The Doppler spread of the simulated data is set to be the same as that of the measured data for comparison. In this case, we observe that the temporal variations of the measured data and the simulated data seem to match well.

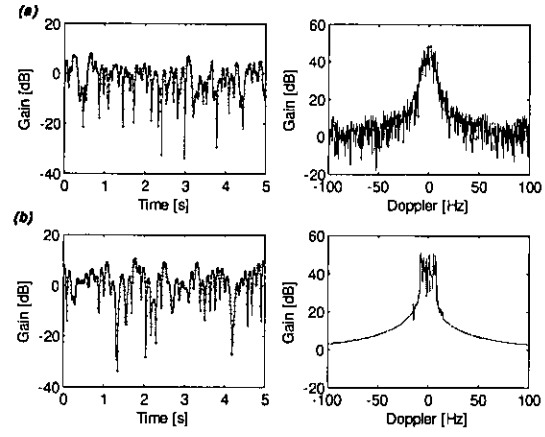


Figure 2. Time and frequency representation of (a) the measured data and (b) the simulated data

Figure 3 shows the cumulative distribution function (CDF) of the eigenvalues of the measured data and the simulated data. In this plot, it is assumed that there are four users, each with a single receive antenna and four transmit antennas at the base station, which forms a 4×4 MIMO broadcast channel and the angular spread $\Lambda = \pi/4$.

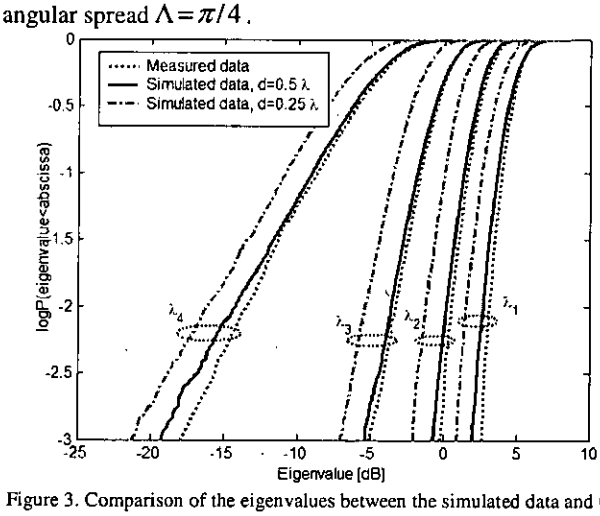


Figure 3. Comparison of the eigenvalues between the simulated data and the measured data with a 4×4 channel matrix ($M_T = 4$, $K = 4$, $M_R = 1$) with $\Lambda = \pi/4$

The measured data and the simulated data match well when the antenna spacing d is 0.5λ. When the antenna spacing d is 0.25λ, the eigenvalues decrease due to the spatial correlation, which leads to a decrease in capacity. This implies that a MIMO broadcast channel with multiple antennas at the base station and a single receive antenna for each mobile terminal can ignore the effect of the spatial correlation if the antenna spacing is sufficient to guarantee very small spatial correlation values.

Figure 4 shows the CDF of spatial correlation at the receiver. Denoting the 4×4 MIMO channel matrix as $\mathbf{H}$ every point in each curve is obtained by averaging all non-diagonal entries of the matrix $\mathbf{H}^H\mathbf{H}$. Then the CDF is constructed with these points from all channel realizations.

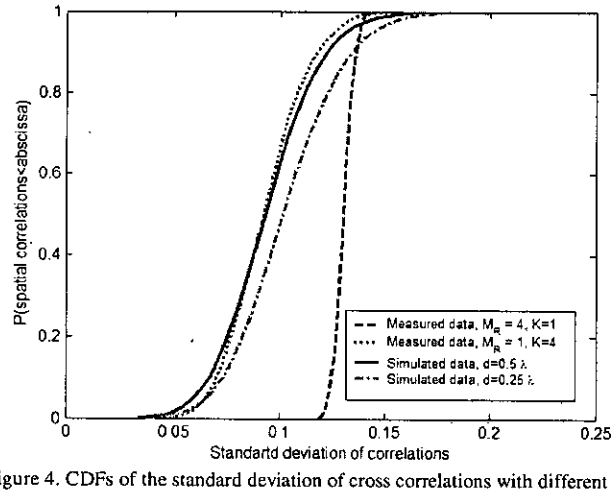


Figure 4. CDFs of the standard deviation of cross correlations with different antenna configurations with a full rank 4×4 channel matrix ($M_T = 4$) with $\Lambda = \pi/4$

We notice that the case of two pairs of crossed dipoles at the receiver (M_R = 4, K = 1) suffers from relatively high cross correlations (0.1 ~ 0.15) due to insufficient spacing of receive antennas. However, if the spatial correlation at the receiver is removed, namely, four different antennas are selected from four different users (M_R = 1, K = 4), to emulate the MIMO BC conditions, the cross correlation becomes small enough to be ignored compared to the case of four receive antennas from one user ($M_R = 4$), which has similar cross correlation characteristics to the simulated data with $d = 0.5\lambda$. However, as the antenna spacing decreases, the spatial correlation increases. When $d = 0.25\lambda$, the spatial correlation is larger than the case of $d = 0.5\lambda$.

V. Conclusion

In this paper, we proposed a simple MIMO broadcast channel model, which has reasonable complexity but can describe a realistic MIMO broadcast channel with multiple antennas at a base station and single receive antennas for each terminal. The proposed model was validated by comparing with MIMO channel measurements. Numerical results showed that the effect of temporal correlation can be minimized by multiuser scheduling, so that the temporal correlation is more meaningful than the spatial correlation when simulating multiuser scheduling algorithms. Indeed, the comparison between the measured data and the simulated data showed that the temporal correlation among channel parameters is sufficient to describe the realistic MIMO broadcast channel in terms of eigenvalues and cross correlations.

Acknowledgement

The authors would like to thank J. McGeehan, director of TRL in Bristol and Dr. McNamara for their support and technical discussions. M-VCE measurement data for this paper was used while the first author was on placement at TRL from August 2006 to October 2006.

References

[1] I. E. Telatar, "Capacity of multi-antenna Gaussian channels," *Eur. Trans. Telecommun.*, vol. 10, no. 6, pp. 585-595, 1999.

[2] G. J. Foschini, and M. J. Gans, "On limits of wireless communications in a fading environment," *Wireless Pers. Commun.*, vol. 6, no. 3, pp.311-335, 1998.

[3] J. P. Kermoal, L. Schumacher, K. I. Pederson, P. E. Mogenson, and F. Frederiksen, "A stochastic MIMO radio channel model with experimental validation," *IEEE Sel. Commun.*, vol. 20, no. 6, pp. 1211-1226, Aug. 2002.

[4] K. Yu, M. Bengtsson, B. Ottersten, D. McNamara, P. Karlsson, and M. Beach, "Modelling of wide-band MIMO radio channels based on NLoS indoor measurements," *IEEE Transactions on Vehicular Technology*, vol. 53, no. 3, pp. 655-665, 2004.

[5] G. J. Byers, and F. Takawira, "Spatially and temporally correlated MIMO channels: modeling and capacity analysis," *IEEE Trans. Vech. Tech.*, vol. 53, no, 3, pp. 634-643, May 2004.

[6] S. Lee and J. Thompson. "Statistical CSI-assisted multiuser scheduling in MIMO broadcast channels," *submitted to IEEE Communications Letters*, 2007.

[7] M. Patzold and V. D. Nguyen, "A spatial simulation model for shadow fading processes in mobile radio channels," *in Proc. PIMRC 2004*, Barcelona, Spain, vol. 3, pp. 1832-1838, Sep. 2004.

[8] Y. R. Zheng and C. Xiao, "Simulation model with correct statistical properties for Rayleigh fading channels," *IEEE Trans. Commun.*, vol. 51, no. 6, pp. 920-928, June 2003.

[9] W. C. Jakes, *Microwave Mobile Communications*, Piscataway, NJ, IEEE Press, 1994

[10] D. Gesbert, H. Bolcskei, D. Gore, and A. Paulraj, "Outdoor MIMO wireless channels: models and performance prediction," *IEEE Transactions on Communications*, vol. 50, no. 12, pp. 1926-1934, 2002.

[11] A. van Zelst and J. Hammerschmidt, "A single coefficient spatial correlation model for multiple-input multiple output (MIMO) radio channels," *Proc. 27th General Assembly of the International Union of Radio Science*, 2002.

[12] G. Durgin and T. Rappaport, "Effects of multipath angular spread on the spatial correlation of received voltage envelopes," *IEEE 49th Vehicular Technology Conference*, vol. 2, 1999.

[13] M. Hunukumbure and M. A. Beach, "MIMO channel measurements and analysis with prototype user devices in a 2GHz urban cell," *in Proc. PIMRC 2006*, Helsinki, Finland, Sep. 2006.

[14] C. Windpassinger, R. F. Fischer, T. Vencel, and J. B. Huber, "Precoding in multiantenna and multiuser communications," *IEEE Trans. Commun.*, vol. 3, no. 4, pp.1305-1316, July 2004.

[15] M. Vu and A. Paulraj, "A robust transmit CSI framework with applications in MIMO wireless precoding,@ *Proc. of 39th Asilomar Conference on Signals, Systems and Computers*, pp. 623~627, 2005.

[16] M. Gudmundson, "Correlation model for shadow fading in mobile radio systems," *Electronic letter*, vol. 27, no. 23, pp. 2145-2146, Nov. 1991.

Statistical CSI-assisted Multiuser Scheduling in MIMO Broadcast Channels

S. Lee and J. Thompson

***Abstract*—A new multiuser scheduling algorithm which can minimize the effect of scheduling delay in order to improve the sum-rate capacity in a temporally-correlated MIMO broadcast channel (BC) is presented. The scheduling delay induces an additive channel estimation error noise as well as decreasing the effective channel gain due to temporal variations. The proposed algorithm utilizing statistical channel state information (SCSI) allocates users with high Doppler spread in the first part of a frame so that the effect of scheduling delay can be minimized. Numerical results demonstrate that the proposed SCSI-assisted multiuser scheduling algorithm can improve the sum-rate capacity of the MIMO BC.**

***Index Terms*—Multiple input multiple output (MIMO), broadcast channel (BC), channel state information (CSI), multiuser scheduling, scheduling delay.**

I. Introduction

For multiuser MIMO broadcast channels (BCs, or downlink, i.e., channels from the base station to mobile users), many combined MIMO precoding and multiuser scheduling methods have been suggested [1] [2]. However, these multiuser scheduling algorithms do not consider temporal variations of the MIMO BC. The temporal variations cause mismatches of the channel estimate, which may affect the reliability of the channel state information (CSI) at the transmitter. This results in performance degradation because MIMO precoding techniques usually require CSI for transmit signal processing.

In this letter, we propose a new statistical channel state information (SCSI)-assisted multiuser scheduling algorithm for the temporally-correlated MIMO BC. The proposed algorithm utilizes SCSI for the selection of a user set, which can minimize the mismatch of channel estimates in order to improve the sum-rate capacity. Assuming a frame consists of multiple slots, with the proposed SCSI-assisted multiuser scheduling algorithm, a user set for transmission is selected on a slot-by-slot basis using SCSI, so that the sum-rate capacity can be improved compared to a frame-by-frame multiuser scheduler without SCSI.

The rest of this paper is organized as follows. In Section II, the system model is introduced. In Section III, the proposed SCSI-assisted multiuser scheduling algorithm is presented. Numerical results are presented in Section IV. Conclusions are drawn in Section V.

II. System Model

We use boldface to denote matrices and vectors. For any general matrix $\mathbf{A}$, $\mathbf{A}^T$ denotes the transpose, $\mathbf{A}^H$ denotes the conjugate transpose, $\mathrm{Tr}(\mathbf{A})$ denotes the trace and $\mathrm{diag}\{\lambda_i\}$ denotes a diagonal matrix with the (i,i) entry equal to λ_i. $\mathbf{I}$ denotes the identity matrix. $E[\cdot]$ denotes expectation and $\mathrm{MOD}[\cdot]$ denotes modulo operation.

Consider a narrowband temporally-correlated Rayleigh fading MIMO BC with M_T transmit antennas at a base station and $K(K \geq M_T)$ users each with a single receive antenna. Let $\mathbf{h}_k(t) \in \mathbb{C}^{M_T \times 1}$ denote the channel at time t between the transmit antenna array and the receive antenna for user k, whose elements are complex Gaussian random variables with zero mean and unit variance. Then the MIMO BC at time t can be represented as

$$y_k(t) = \mathbf{h}_k^T(t)\mathbf{s}(t) + z_k(t), \qquad k = 1, \cdots, K, \tag{1}$$

where $\mathbf{s}(t) \in \mathbb{C}^{M_T \times 1}$ is the transmit signal vector with a covariance matrix $R_{ss} = E[\mathbf{s}(t)\mathbf{s}(t)^H] = \sigma_s^2\mathbf{I}$ and a power constraint $\mathrm{Tr}(R_{ss}) \leq P$, the scalar $y_k(t)$ is the received signal for user k and $z_k(t)$ is the complex additive white Gaussian noise (AWGN) with zero mean and unit variance for user k at time instant t. Each user is assumed to experience independent frequency non-selective fading with different Doppler spread f_d^k. In this case, the temporal correlation of any user with delay τ can be given by $\rho_k(\tau) = J_0(2\pi f_d^k \tau)$, where J_0 is the zeroth order Bessel function of the first kind [3].

For the THP technique in the MIMO BC, a channel matrix $\mathbf{H}(S)$ is formed with a user set S ($|S| \leq M_T$) and decomposed into a unitary transmit beamforming matrix $\mathbf{F}$ and lower triangular matrix $\mathbf{B}$ by taking the QR decomposition. For simplicity of implementation, several assumptions are made. First, although the maximum sum-rate capacity of THP can be achieved by optimal ordering and transmit power allocation, these optimal strategies are ignored. Therefore, an equal power allocation over spatial channels is assumed. Secondly, the number of elements in a user set S is assumed to be the same as the number of transmit antennas ($|S| = M_T$). Lastly, if the target bit-error rate (BER) of the system is very small (*i.e.*, BER $\leq 10^{-6}$) and high signal-to-noise ratio (SNR) operation is assumed for all users, the modulo loss can be ignored [4]. Then, the sum-rate capacity of THP at time t can be approximated as

$$C_{THP}(S,t) = \sum_{k=1}^{M_T} \log_2\left(1 + \frac{b_{kk}^2(t)P}{M_T}\right). \tag{2}$$

With this, the sum-rate maximization rule with THP finds a user set S_{max} at time t according to

$$S_{\max}(t) = \arg \max_{S \subset \{1,\cdots,K\}, |S| = M_T} C_{THP}(S,t). \tag{3}$$

After the user set S_{max} is determined, THP is performed at the transmitter for transmission.

Fig. 1 shows the block diagram of THP with the selected user set S_{max} in the MIMO BC. The vector

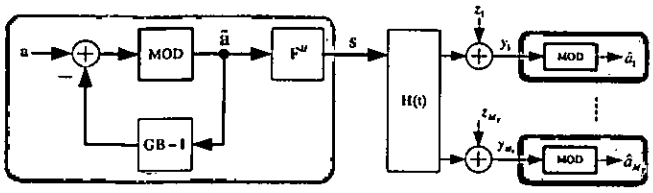


Fig. 1. The block diagram of THP in the MIMO BC

$\mathbf{a} \in \mathbb{C}^{M_T \times 1}$ denotes the input data vector with the covariance matrix $\mathbf{R}_{aa} = \sigma_a^2 \mathbf{I}$. The diagonal scaling matrix $\mathbf{G} = \text{diag}\left\{b_1^{-1}, \cdots b_{M_T}^{-1}\right\}$ determines the effective channel gain of all users at the receiver. For more detail on the operation of THP, see [5].

III. Statistical CSI-assisted Multiuser Scheduling

It is assumed that one downlink frame of the MIMO BC with time period T_f consists of N_S ($N_S > 1$) time slots with time period T_s. Then the ICSI is updated once per every frame at the base station. However, different users can be transmitted to in each time slot. Denoting the maximum Doppler spectrum as f_D, the frame period should be selected such that [6]

$$T_f < \frac{1}{2f_D}. \tag{4}$$

ICSI can be obtained by using either a feedback channel or the reciprocity principle in the case of time-division duplex (TDD) systems. There are several sources for imperfect ICSI but it is assumed that the ICSI is perfect at the beginning of every frame [3] for simplicity. In this case, the only imperfection is the mismatch of the channel estimate due to scheduling delay. This mismatch results in performance degradation, which will generally increase through the time slots of each frame. The amount of degradation for user k is determined by the temporal correlation, which is a function of the Doppler spread f_d^k and multiples of the slot period nT_s.

Based on the results in [3], an equivalent time-varying channel matrix reflecting scheduling delay within the frame at slot index n can be written as

$$\mathbf{H}(n) = \mathbf{P}(n)\mathbf{H}_0 + \sqrt{\mathbf{I} - \mathbf{P}(n)\mathbf{P}^H(n)}\mathbf{H}_m, \; n = 1, \ldots, N_S, \tag{5}$$

where $\mathbf{P}(n) = \text{diag}\{\rho_1(nT_s), \ldots, \rho_{M_T}(nT_s)\}$ denotes the autocorrelation matrix of the user set S_{max}, $\sqrt{\mathbf{I} - \mathbf{P}(n)\mathbf{P}^H(n)}$ represents the amplitude increase in the channel estimation error due to scheduling delay, $\mathbf{H}_0$ is the perfectly estimated channel matrix at the beginning of each frame and $\mathbf{H}_m$ is an uncorrelated estimation error matrix which has the same statistical characteristics as the estimated channel $\mathbf{H}_0$. Equation (5) implies that any decrease in ICSI becomes a delay-induced channel estimation error due to scheduling delay.

From (1), the received signal via the equivalent MIMO BC becomes

$$\mathbf{y}(n) = \mathbf{P}(n)\mathbf{H}_0\mathbf{s} + \sqrt{\mathbf{I} - \mathbf{P}(n)\mathbf{P}^H(n)}\mathbf{H}_m\mathbf{s} + \mathbf{z}(n), \tag{6}$$

where $\mathbf{y}(n) \in \mathbb{C}^{M_T \times 1}$ and $\mathbf{z}(n) \in \mathbb{C}^{M_T \times 1}$ are vector form of the received signal and complex AWGN respectively.

In (6), the first term denotes the desired signal and the second term denotes the additive measurement noise due to the delay-induced estimation error. After the modulo operation at the receiver, the output data vector becomes

$$\hat{\mathbf{a}} = \text{MOD}[\mathbf{P}(n)\mathbf{H}_0\mathbf{s}] = \mathbf{P}(n)\mathbf{G}^{-1}\mathbf{a}. \tag{7}$$

The variance of the estimate noise, $\sqrt{\mathbf{I} - \mathbf{P}(n)\mathbf{P}^H(n)}\mathbf{H}_m\mathbf{s}$, is given by

$$M_T\left(\mathbf{I} - \mathbf{P}(n)\mathbf{P}^H(n)\right)\mathbf{R}_{ss}, \tag{8}$$

because $E\left[\mathbf{H}_m\mathbf{H}_m^H\right] = M_T\mathbf{I}$. Since $\sigma_s^2 \simeq \sigma_a^2$ with THP [7] and $\sigma_a^2 = \frac{P}{M_T}$, the sum-rate capacity considering the temporal variation is written as

$$C_{THP}(S, \mathbf{H}(n)) = \sum_{k=1}^{M_T} \log_2\left(1 + \frac{b_{kk}^2 \rho_k^2(n) P}{M_T\left\{(1 - \rho_k^2(n))P + 1\right\}}\right). \tag{9}$$

Although the base station obtains ICSI once per every frame, it can utilize the previous ICSI for obtaining SCSI such as mean and covariance due to channel stationarity [3]. This SCSI can be used for estimating the temporal correlations of the MIMO BC when precise ICSI is not available. The SCSI-assisted sum-rate maximization rule at slot index $n (n = 1, \ldots, N_S)$ finds a user set $S_{max}(n)$ according to

$$S_{\max}(n) = \arg \max_{S \subset \{1, \cdots, K\}, |S| = M_T} C_{THP}\left(S, \mathbf{H}(n)\right). \tag{10}$$

The SCSI-assisted sum-rate maximization rule uses the equivalent estimated channel model (5) for the multiuser selection procedure. Any user with a rapid decrease in the temporal correlation may be assigned to the first few slots in the frame. This may improve the total throughput in return for the increased complexity. This is because the multiuser selection procedure has to be performed on a slot-by-slot basis if the proposed algorithm is applied. When the temporal variations of the MIMO BC are small compared to the coherence time, the proposed SCSI-assisted multiuser scheduling algorithm has little improvement from the normal sum-rate maximization algorithm, which is performed on a frame-by-frame basis. This would give the same choice in the multiuser selection procedure for all slots.

IV. Numerical Results

Throughout the simulations, we assume $M_T = 2$, $K = 32$ and $N_s = 5$. The relative Doppler spread values of all users, $f_d T_f$, are assumed to be independent and identically distributed (*i.i.d.*) uniform random variables over $[0.1, 0.5)$.

Fig. 2 shows the performance of the SCSI-assisted multiuser scheduling algorithm against SNR with different relative Doppler spread configurations. We assume $T_f = 2$ms in this plot.

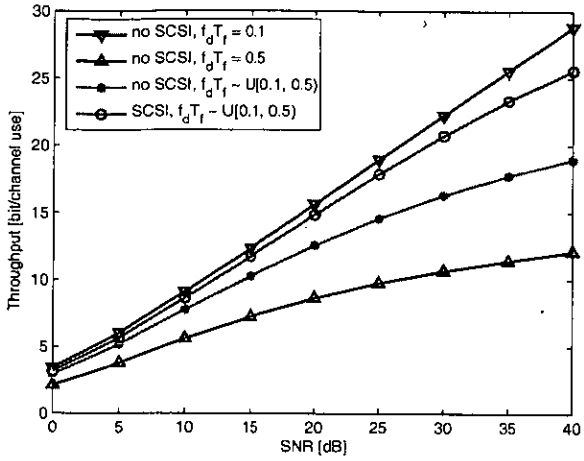


Fig. 2. The performance of SCSI-assisted multiuser scheduling algorithm against SNR with different relative Doppler spread configurations, $T_f = 2$ms, $M_T = 2$ and $K = 32$.

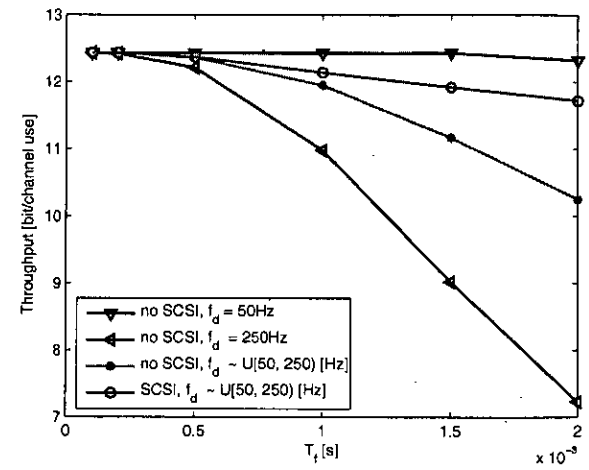


Fig. 3. The performance of SCSI-assisted multiuser scheduling algorithm against frame period with different relative Doppler spread configurations, $SNR = 15$dB, $M_T = 2$ and $K = 32$.

When the relative Doppler spreads of all users are identical, the throughput decreases as the relative Doppler spread increases. This is because the temporal channel correlation for all users decreases more rapidly as the relative Doppler spread increases. However, when the relative Doppler spreads of all users have a uniform distribution over $[0.1, 0.5)$, the proposed SCSI-assisted multiuser scheduling algorithm performs better than the multiuser scheduling algorithm which does not use SCSI and applies a frame-basis multiuser selection. The performance difference between these two schemes is evident as the SNR increases because the delay-induced channel estimation error produces an irreducible error floor in the high SNR region.

Fig. 3 shows the performance of the SCSI-assisted multiuser scheduling algorithm against the frame period T_f. We assume $SNR = 15$dB for all users in this plot.

When the frame period is small enough compared to the coherence time of the channel, the performance improvement of the proposed algorithm is negligible. However, as the frame period increases, the performance advantage of using SCSI becomes evident. The proposed SCSI-assisted multiuser scheduling algorithm performs better than the multiuser scheduling without SCSI considerations when $T_f > 0.2$ms. Note that the amount of performance improvement and complexity increase of the proposed algorithm depends on the channel conditions and the number of slots in the frame structure.

V. Conclusions

We proposed a SCSI-assisted multiuser scheduling algorithm for the temporally-correlated MIMO BC. Due to the temporal variations of the MIMO BC, mismatch in the channel estimate occurs, which may affect the reliability of CSI at the transmitter. Because the degree of mismatch is different among users if their Doppler spreads are unequal, the proposed algorithm can minimize the effect of such mismatches. This is achieved by allocating users with high Doppler spreads to the front part of a frame, so that it can improve the sum-rate capacity. The selected user set for each time slot may be different from one another if the SCSI-assisted multiuser scheduling algorithm is applied on a slot-by-slot basis. However, the normal multiuser selection algorithm without SCSI on a frame-by-frame basis would give the same choice to all slots. For this, the proposed algorithm makes use of SCSI as well as ICSI for the selection of a user set for transmission. By doing this, it can minimize the effect of scheduling delay whilst improving the sum-rate capacity of THP in the MIMO BC.

References

[1] T. Yoo and A. Goldsmith, "On the optimality of multiantenna broadcast scheduling using zero-forcing beamforming," *IEEE Journal on Selected Areas in Communications*, vol. 24, no. 3, pp. 528–541, 2006.

[2] C. Swannack, E. Uysal-Biyikoglu, and G. Wornell, "Low complexity multiuser scheduling for maximizing throughput in the MIMO broadcast channel," *Proc. Allerton Conf. Communications, Control and Computing*, 2004.

[3] M. Vu and A. Paulraj, "A robust transmit CSI framework with applications in MIMO wireless precoding," *Proc. 39th Asilomar Conf. Signals Systems, Computers*, pp. 623–627, 2005.

[4] R. Wesel and J. Cioffi, "Precoding and the MMSE-DFE," *Conference on Signals, Systems and Computers*, vol. 2, pp. 1144–1148, 1994.

[5] C. Windpassinger, R. Fischer, T. Vencel, and J. Huber, "Precoding in multiantenna and multiuser communications," *IEEE Transactions on Wireless Communications*, vol. 3, no. 4, pp. 1305–1316, 2004.

[6] S. Dragan, M. Narayan, and C. Dmitry, "Adaptive transmitter optimization in multiuser multiantenna systems: theoretical limits, effect of delays, and performance enhancements," *EURASIP Journal on Wireless Communications and Networking*, no. 3, pp. 298–307, 2005.

[7] R. Fischer, C. Stierstorfer, and J. Huber, "Precoding for point-to-multipoint transmission over MIMO ISI channels," *International Zurich Seminar on Communications*, pp. 208–211, 2004.

QoS-guaranteed Multiuser Scheduling in MIMO Broadcast Channels

Seung-Hwan Lee, *Student Member, IEEE*, and John Thompson , *Member, IEEE*

***Abstract*—This paper proposes a new multiuser scheduling algorithm, which can support a mixture of different quality-of-service(QoS) user groups simultaneously whilst satisfying fairness among users in the same QoS group in MIMO broadcast channels(BCs). For this, the proposed multiuser scheduling algorithm consists of two parts: a QoS-aware Fair (QF) scheduling in the same QoS group and an antenna trade-off scheme between different QoS groups. The proposed QF scheduling algorithm finds a user set which can satisfy fairness among users in terms of throughput or delay respectively. In addition, the proposed antenna trade-off scheme can minimize the QoS violations of a higher priority user group by trading off the number of transmit antennas allocated to different QoS groups. Numerical results demonstrate that the proposed QF scheduling method not only satisfies different types of fairness among users but can adjust the degree of fairness among users. The proposed antenna trade-off scheme can improve the probability of QoS-guaranteed transmission whilst maximizing the total throughput.**

***Index Terms*—Multiple input multiple output (MIMO), broadcast channel, dirty paper coding (DPC), weighted sum-rate capacity, quality of service (QoS), multiuser scheduling, fairness.**

I. Introduction

In multiuser MIMO broadcast channels (BCs), *i.e.* the channels from the base station to mobile users, dirty paper coding (DPC) [1] can achieve the sum-rate capacity [2]. However, because DPC is difficult to implement in real systems due to its excessively high complexity, practical precoding techniques such as Tomlinson-Harashima precoding (THP) and zero-forcing beamforming (ZF-BF) have been developed. Zero forcing dirty paper coding (ZF-DPC) [3] is a nonlinear suboptimal implementation of DPC based on interference pre-subtraction. THP applies a modulo operation to ZF-DPC to prevent a possibly large power increase due to the pre-subtraction of ZF-DPC. Linear ZF-BF is more simple method than THP because no user ordering for pre-subtraction is required at the transmitter. Throughput comparisons between these MIMO precoding techniques can be found in [4] and [5].

Alongside the choice of MIMO precoding technique, multiuser scheduling is one of the most important issues of the MIMO BC for satisfying different quality-of-service (QoS) requirements such as throughput, delay constraint and fairness among users. A lot of scheduling methods have been suggested [6] [7] [8]. Round robin (RR) is a simple algorithm which serves users in a cyclic fashion regardless of the channel conditions. Proportional fair (PF) scheduling [6] is designed to meet long-term throughput fairness among users by considering the wireless channel conditions and the amount of past throughput simultaneously to exploit the multiuser diversity effect. The Modified-largest weighted delay first (M-LWDF) [8] is a throughput optimal scheduling strategy which takes into account the channel conditions and the state of queues. With M-LWDF, different QoS requirements are satisfied in terms of the outage probability.

A number of studies focusing on combined MIMO precoding techniques and multiuser scheduling methods for the MIMO BC have been presented. In [9], the performance of ZF-BF with several scheduling algorithms with a multiuser selection algorithm of reduced-complexity has been analyzed. Joint ZF-BF and scheduling for the optimal throughput with reduced complexity has been considered in [10]. However, these studies do not consider different QoS requirements such as different data rates, delay constraints and different fairness requirements among users.

In this paper, a new QoS-guaranteed multiuser scheduling algorithm is proposed for supporting a mixture of different QoS users simultaneously whilst satisfying fairness among users in realistic MIMO BC scenarios, where a base station is providing a variety of services to different users with different QoS requirements and fairness considerations. For this, the proposed algorithm includes a new fairness metric for the achievement of different throughput or delay fairness among users in the same QoS group and a new antenna trade-off scheme for the QoS differentiation between different QoS groups. With the proposed antenna trade-off scheme, a higher priority group takes precedence in using multiple antennas for satisfying its QoS requirement. The number of transmit antennas allocated to each QoS group can be determined by the wireless channel conditions and QoS requirements. After determining the number of transmit antennas allocated to each QoS group using the antenna trade-off scheme, the proposed QoS-guaranteed multiuser scheduling algorithm finds user sets from the highest QoS group sequentially using the QF scheduling algorithm from each QoS group, so that the final user set for transmission consists of different QoS users who can maximize the total throughput as well as satisfying fairness requirements.

The main contributions of this paper are as follows:

- A new throughput fairness metric supporting different data rates among users as well as enabling a trade-off between throughput and degree of throughput fairness in the MIMO BC.
- A new delay fairness metric supporting different delay constraints among users as well as enabling a trade-off between delay and degree of delay fairness in the

S. Lee is with the Institute for Digital Communications, The University of Edinburgh, Edinburgh, UK. Phone: +44 131 650–5655, e-mail: s.lee@ed.ac.uk

J. Thompson is with the Institute for Digital Communications, The University of Edinburgh, Edinburgh, UK. Phone: +44 131 650–5585, e-mail: j.thompson@ed.ac.uk

MIMO BC.

- A new antenna trade-off scheme which changes the number of transmit antennas allocated to each QoS group according to the priority order for satisfying prioritized QoS between different QoS groups.

The rest of this paper is organized as follows. In Section II, the system model is introduced. In Section III, the proposed fairness scheduling algorithms are presented. Section IV explains the proposed antenna trade-off scheme. Numerical results are presented in Section V. Conclusions are drawn in Section VI.

II. System Model

We use boldface to denote matrices and vectors. For any general matrix $\mathbf{A}$, $\mathbf{A}^T$ denotes the transpose, $\mathbf{A}^H$ denotes conjugate transpose and $\mathrm{Tr}(\mathbf{A})$ denotes the trace. $E[\cdot]$ denotes expectation. For any general set B, $|B|$ denotes the cardinality of the set.

Consider a narrowband temporally-correlated fading MIMO BC with M_T transmit antennas at a base station and K ($K \geq M_T$) users each with a single receive antenna. Let $\mathbf{h}_k(t) \in \mathbb{C}^{M_T \times 1}$ denote the channel at time t between the transmit antenna array and the receive antenna for user k. Then the MIMO BC at time t can be represented as

$$y_k(t) = \mathbf{h}_k^T(t)\mathbf{s}(t) + z_k(t), \qquad k = 1, \cdots, K, \tag{1}$$

where $\mathbf{s}(t) \in \mathbb{C}^{M_T \times 1}$ is the transmit signal vector with a power constraint $\mathrm{Tr}\left(E[\mathbf{s}(t)\mathbf{s}(t)^H]\right) \leq P$, $y_k(t)$ is the received signal for user k and $z_k(t)$ is the complex additive white Gaussian noise (AWGN) with zero mean and unit variance for user k at time instant t. The transmit signals are assumed to experience path loss, log-normal shadowing and Rayleigh fading. In this case, the channel $\mathbf{h}_k(t)$ can be expressed as [11]

$$\mathbf{h}_k(t) = \sqrt{SNR_0 \left(\frac{r_k(t)}{R_c}\right)^{-\alpha_p}}\, \omega_k(t) \cdot \mathbf{g}_k(t), \tag{2}$$

where SNR_0 denotes the median of the mean signal-to noise-ratio (SNR) of all users averaged over fading at the cell boundary of distance R_c, $r_k(t)$ denotes the distance between user k and the base station, α_p denotes the path loss exponent and $\omega_k(t)$ denotes the shadowing for user k at time t given by [12]. The vector $\mathbf{g}_k(t) = [g_{k1}(t)\ g_{k2}(t)\ \cdots\ g_{kM_T}(t)]^T$ represents the Rayleigh-distributed fading between the transmit antenna array and user k at time t. The Rayleigh fading model [13] generates independent and identically distributed (*i.i.d.*) complex Gaussian random variables for the elements of $\mathbf{g}_k(t)$ with zero mean and unit variance. All users are assumed to be located at the same distance from the base station and experience frequency non-selective Rayleigh fading with Doppler spread f_d^k, which is determined by the carrier frequency f_c and the velocity v_k as $f_d^k = \frac{v_k f_c}{C}$, where C is the speed of light. In this case, the temporal correlation of any user with delay τ can be given by $\rho = m_s J_0(2\pi f_d \tau)$, where m_s is the area mean determined by shadowing and path loss, τ is the delay in time and J_0 is the zeroth order Bessel function of the first kind [14].

III. QoS-aware Fair (QF) scheduling algorithms

A. Fairness metrics for QF scheduling

In real MIMO BC scenarios, a base station is likely to provide a variety of services to different users, each with different quality of service (QoS) requirements. In this case, throughput fairness among users does not mean the allocation of the same amount of bandwidth to all users.

In order to support such heterogeneous user channels, define the scaled deviation as

$$\Delta x = \frac{x - \bar{x}}{x}, \tag{3}$$

where x is the observation and $\bar{x}$ is the required value of x. It can represent the relative degree of fairness regardless of the resource being considered. For example, there are two users with throughput requirements R_1 and R_2 respectively. It is assumed that the throughput requirement of the first user X_1 is two times as large as that of the second user X_2, namely, $R_2 = 2R_1$. If the data rates (observations) of X_1 and X_2 are $0.5R_1$ and R_1 respectively at a certain instant, the scaled deviation with respect to the throughput requirements (the required value) are $\Delta x_1 = (0.5R_1 - R_1)/(0.5R_1) = -1.0$ for X_1 and $\Delta x_2 = (R_1 - 2R_1)/R_1 = -1.0$ for X_2. In this case, throughput fairness is said to be satisfied in terms of the relative achievement in spite of the reduced amount of data delivered to each user.

In order to apply the concept of the scaled deviation to the QF criterion, an exponential function taking the argument Δx is used. According to the type of resource, either $\exp(\Delta x)$ or $\exp(-\Delta x)$ can be used as elements of the weight vector for a weighted sum-rate maximization rule. In the case of throughput fairness, $\exp(-\Delta x)$ is used so that any user with relatively smaller throughput than other users has a large weight value. On the contrary, in the case of delay fairness, $\exp(\Delta x)$ is used so that any user with relatively larger delay than other users has a large weight value.

Prior to applying the scaled deviation to fairness cases, let us introduce the other parameter m, which changes the slope of the exponential function with a form of $\exp(m\Delta x)$ or $\exp(-m\Delta x)$. Later, this slope parameter will be used for controlling the degree of fairness. Fig. 1 shows the characteristics of the scaled exponential function $\exp(m\Delta x)$ with $\bar{x} = 1$. When $m = 0.1$, the output of the scaled exponential function is saturated near one when $x > 1$, regardless of the scaled deviation. However, as m increases, the output of the scaled exponential function grows rapidly when $x > 1$. This makes a small change in x cause a significant increase in the output.

For throughput fairness, define the throughput fairness metric $\mu_k^t(t)$ for user k at time t as

$$\mu_k^t(t) = \exp\left(m_t \frac{\bar{R}(t) - a_k^t \bar{r}_k(t)}{\varepsilon + a_k^t \bar{r}_k(t)}\right), \tag{4}$$

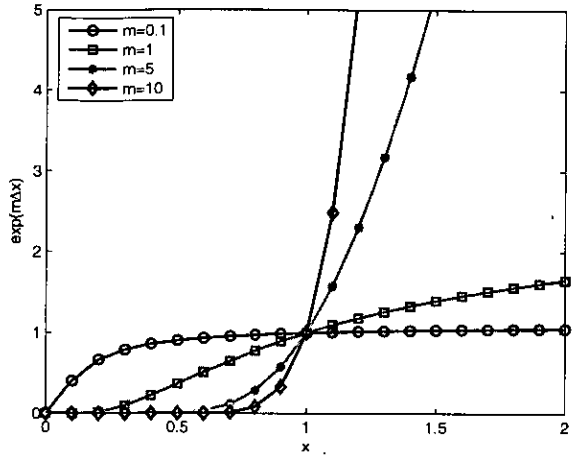


Fig. 1. The characteristics of the scaled exponential function $\exp(m\Delta x)$ with the slope parameter m, $\bar{x}=1$

where m_t ($m_t \geq 0$) is the throughput slope of the exponential function, which determines the sensitivity of (3) to throughput fairness. As might be expected from Fig. 1, smaller values of m_t mean less strict throughput fairness among users and typically gives only average throughput fairness among users. Indeed, when $m_t = 0$, the weighted sum-rate maximization problem reduces to the normal sum-rate maximization problem without any fairness consideration. The scalars a_k^t are constants to allow different throughput requirements among users. The variable ε is an appropriate small value for ensuring the denominator is nonzero. The value $\bar{r}_k(t)$ is the exponential moving average of the past throughput for user k, which is updated as

$$\bar{r}_k(t+1) = \left(1 - \frac{1}{\alpha_t}\right)\bar{r}_k(t) + \frac{1}{\alpha_t}R_k(t), \tag{5}$$

where $\alpha_t(\alpha_t > 0)$ is the smoothing factor and $R_k(t)$ is the data rate of user k at time t. The scalar $\bar{R}(t)$ in (4) is defined as $\bar{R}(t) = (1/K)\sum_{k=1}^{K}\bar{r}_k(t)$. According to (4), any user with $a_k^t\bar{r}_k(t) > \bar{R}(t)$ has a value less than one as its weight. In this case, that user might be excluded from a selected user set for transmission by the weighted sum-rate maximization rule to meet the throughput fairness constraint among users despite its high channel gain.

For delay fairness, define the delay fairness metric $\mu_k^d(t)$ for user k at time t as

$$\mu_k^d(t) = \exp\left(m_d\frac{a_k^d d_k(t) - \bar{D}(t)}{\varepsilon + a_k^d d_k(t)}\right), \tag{6}$$

where m_d ($m_d \geq 0$) is the delay slope of the exponential function, which determines the sensitivity of delay fairness as in the case of throughput fairness. Again, small values of m_d means less strict delay fairness among users. The scalars a_k^d are constants to allow different delay requirements among users. The scalar $d_k(t)$ denotes the head-of-line (HOL) delay in the queue of user k and $\bar{D}(t)$ is defined as $\bar{D}(t) = (1/K)\sum_{k=1}^{K}d_k(t)$.

B. Weighted sum-rate maximization with THP

For the THP technique in the MIMO BC, a channel matrix $\mathbf{H}(S_0)$ is formed with a user set S_0 ($|S_0| \leq M_T$) and decomposed into unitary transmit beamforming matrix $\mathbf{F}$ and lower triangular matrix $\mathbf{B}$ by taking the QR decomposition. Because of the lower triangular matrix $\mathbf{B}$, any interference caused by users $j > i$ on each user i is forced to zero by pre-subtraction at the transmitter. However, due to the pre-subtraction, the transmit power increases, so that THP employs the modulo operation to minimize this. Denoting the transmit power allocated to user k as P_k and b_{kk} as kth diagonal element of the matrix $\mathbf{B}$, the achievable sum-rate capacity of THP is given by

$$C(S_0) = \max\sum_{k\in S_0}\log_2\left(1 + \frac{b_{kk}^2 P_k}{\Gamma_{THP}^k}\right), \tag{7}$$
$$\text{subject to } P_k \geq 0,\ \sum\nolimits_{k\in S_0} P_k \leq P.$$

where Γ_{THP}^k denotes the modulo loss of user k [15]. For simplicity, several assumptions are made before applying the achievable sum-rate capacity to the weighted sum-rate maximization rule. First, although the maximum sum-rate capacity of THP can be achieved by the optimal transmit power allocation [16], these optimal strategies are ignored for simplicity. Therefore, an equal power allocation over spatial channels is assumed. Second, the number of elements in a user set S is assumed to be the same as that of transmit antennas ($|S| = M_T$). Lastly, if the target bit-error rate (BER) of the system is very small (*i.e.* BER $\leq 10^{-6}$) and high SNR is assumed for all users, the modulo loss can be ignored [17] except the shaping loss of 1.53dB, which can be achieved by using multidimensional lattice codes rather than M-QAM modulation. Then, the sum-rate capacity of THP at time t is approximated as

$$C_{THP}(S,t) = \sum_{k=1}^{M_T}\log_2\left(1 + \frac{b_{kk}^2(t)P}{M_T}\right). \tag{8}$$

From (4), (6) and (8), denoting $U := \{u_k | k = 1, \cdots, K\}$ as the total user set, the weighted sum-rate maximization rule considering throughput and delay fairness with THP at time t can be given as (9) and (10) respectively.

$$S_{\max}^t(t) = \arg\max_{S\subset U, |S|=M_T}\sum_{k=1}^{M_T}\mu_k^t(t)\log_2\left(1 + \frac{b_{kk}^2(t)P}{M_T}\right). \tag{9}$$

$$S_{\max}^d(t) = \arg\max_{S\subset U, |S|=M_T}\sum_{k=1}^{M_T}\mu_k^d(t)\log_2\left(1 + \frac{b_{kk}^2(t)P}{M_T}\right). \tag{10}$$

The weighting used for the existing PF algorithm for user k is the inverse of the exponential moving average of the past throughput, so that the weighted sum-rate maximization rule for throughput fairness with the PF algo-

rithm is given by

$$S^t_{\max}(t) = \arg \max_{S \subset U, |S| = M_T} \sum_{k=1}^{M_T} \frac{1}{\bar{r}_k(t)} \log_2 \left(1 + \frac{b^2_{kk}(t)P}{M_T}\right). \quad (11)$$

For delay fairness with the existing LWDF rule, the weighted sum-rate maximization rule is of the form

$$S^d_{\max}(t) = \arg \max_{S \subset U, |S| = M_T} \sum_{k=1}^{M_T} a^d_k d_k(t) \log_2 \left(1 + \frac{b^2_{kk}(t)P}{M_T}\right). \quad (12)$$

In [7], the constants a^d_k are defined by probabilistic QoS requirements. It is useful to note that the time index t will be omitted to simplify equations from the next section.

If there are two or more different QoS groups, the QoS-guaranteed multiuser scheduling algorithm determines the number of transmit antennas allocated to each QoS group. Then, the QF scheduling algorithm is applied at each stage to find a user set for that QoS group. In this case, the number of selected users from each QoS group is the number of transmit antennas allocated to each QoS group, which is determined by the antenna trade-off scheme introduced in the next section.

IV. Antenna Trade-off between Different QoS Group

In order to support different QoS users simultaneously in the MIMO BC, a new antenna trade-off scheme between different QoS groups is proposed in this section. We assume there are two QoS groups in the MIMO BC: a best effort (BE) user set $U_{BE} := \{u_k | k = 1, \cdots, K_{BE}\}$ and a delay-constrained real-time (RT) user set $U_{RT} := \{u_k | k = 1, \cdots, K_{RT}\}$, where K_{BE} and K_{RT} are the number of BE and RT users respectively and $K_{BE} + K_{RT} = K$. It is assumed that RT users have a higher priority than BE users in terms of QoS requirements. The objective of any multiuser scheduling algorithm in this configuration is to satisfy the delay constraint of RT users while maximizing the total throughput of all users. A simple multiuser scheduling algorithm may reserve a certain fixed number of transmit antennas to each group for supporting BE and RT users simultaneously. A more advanced way for serving different QoS groups is to change the number of transmit antennas assigned to each QoS group according to certain criteria. This can increase the degree of adaptability to fading MIMO channels. For this adaptive antenna trade-off scheme, each QoS group has its set of pre-assigned transmit antennas but there is a priority for their actual use in transmission. RT users are assumed to have higher priority than BE users in using multiple transmit antennas. For example, when the average HOL delay of RT users is within a certain delay threshold, RT and BE users use their pre-assigned groups of antennas for transmission. When the average HOL delay of RT users exceeds the delay threshold, RT users occupy a higher number of transmit antennas according to the antenna trade-off scheme, then BE users use any remaining transmit antennas for transmission. There are two methods for trading off transmit antennas: the hard trade-off and soft trade-off schemes. The hard trade-off scheme assigns all transmit antennas to RT users when the average delay of RT users exceeds the delay threshold. When this happens, BE users have no chance of using their pre-assigned antennas for transmission. The hard trade-off scheme determines the number of transmit antennas for RT users M_{RT} according to

$$M_{RT} = M_S + (M_T - M_S)\, \mathrm{u}\left[\bar{D}_{RT} - D_{TH}\right], \quad (13)$$

where $\bar{D}_{RT}$ denotes the average HOL delay of RT users, D_{TH} denotes the delay threshold of RT users, M_S ($M_S < M_T$) denotes the pre-assigned groups of antennas for RT users and $\mathrm{u}\{\cdot\}$ denotes unit step function. With the soft trade-off scheme, M_{RT} varies from the pre-assigned antennas M_S up to the total number of transmit antennas M_T determined by

$$M_{RT} = M_S + \sum_{i=0}^{M_T - M_S - 1} \mathrm{u}\left[\bar{D}_{RT} - (1 - \delta i)\, D_{TH}\right], \quad (14)$$

where δ determines the dynamic range of the soft trade-off scheme. When $\delta = 0$, equation (14) becomes the same as that for the hard trade-off scheme. This graceful adaptation by the soft trade-off scheme has better performance than the hard trade-off scheme. BE users have the remaining transmit antennas $M_{BE} = M_T - M_{RT}$ for their transmission. After the number of transmit antennas for each QoS group is determined, the multiuser scheduling algorithm selects a user set from each QoS group sequentially. It is assumed that the number of selected BE and RT users are the same as the number of determined transmit antennas for BE and RT users respectively. For the selection of RT users, delay fairness among RT users is employed. Similarly, for the selection of BE users, throughput fairness among BE users is considered. In other words, the proposed multiuser scheduling algorithm in this configuration not only considers throughput and delay fairness among BE and RT users respectively but includes the delay constraint of RT users. For this, the proposed multiuser scheduling algorithm uses the antenna trade-off scheme for determining the number of transmit antennas for each QoS group and the weighted sum-rate maximization rules for selecting an optimal user set with fairness considerations. The weighted sum-rate maximization rule with delay fairness finds a user set S^d_{RT} among all possible user set S_{RT} ($S_{RT} \subset U_{RT}, |S_{RT}| = M_{RT}$) according to

$$S^d_{RT} = \arg \max_{S_{RT}} \sum_{k=1}^{M_{RT}} \mu^d_k \log_2 \left(1 + \frac{b^2_{kk}P}{M_{RT}}\right). \quad (15)$$

For the selection of BE users, the weighted sum-rate maximization rule with throughput fairness finds a user set S^t_{BE} among all possible user sets S_{BE} ($S_{BE} \subset U_{BE}, |S_{BE}| = M_{BE}$) and the 'already selected' RT user set S^d_{RT} according to

$$S^t_{BE} = \arg \max_{S_{BE}} \left\{ \sum_{k=1}^{M_{BE}} \mu^t_k \log_2 \left(1 + \frac{b^2_{kk}P}{M_{BE}}\right) + C\left(S^d_{RT}\right) \right\}, \quad (16)$$

where $C\left(S_{RT}^{d}\right)$ is the weighted sum-rate capacity of the already selected RT users, which is of the form

$$C\left(S_{RT}^{d}\right)=\sum_{i\in S_{RT}^{d}}\mu_{i}^{d}\log_{2}\left(1+\frac{b_{ii}^{2}P}{M_{RT}}\right). \qquad (17)$$

After user sets are selected from each QoS group, THP is performed on the final user set $S_{\max}$ $\left(S_{\max}=S_{RT}^{d}\cup S_{BE}^{t}\right)$ for transmission.

V. Numerical Results

Throughout the simulations, we assume the number of transmit antennas $M_T=4$, the median SNR of all users $SNR_0=10$dB, the path loss exponent α_p=4 and the distance $r_k=R_c$. It is also assumed that the carrier frequency $f_c=2$GHz and the velocity $v_k=5.4$km/h for all users. For shadowing, standard deviation $\sigma_s=4.3$dB and decorrelation distance $D=8.3$m are assumed to emulate urban environments [18]. The data arrivals in queues for all users are assumed to be independent Poisson processes with arrival rate λ. If the queue dropping probability due to queue overflow and the error rate through the MIMO BC are negligible, throughput can be determined by the arrival rate and the sum-rate capacity. When the arrival rate is less than or equal to the sum-rate capacity, it defines the throughput. However, when the data arrival rate is greater than the sum-rate capacity, the sum-rate capacity determines the throughput.

Several fairness indices have been suggested for the measurement of fairness for different resource allocation schemes. In this paper, a new fairness index based on the Jain Fairness Index [19], [20] is proposed in order to consider different fairness requirements as

$$I_{QF}=\frac{1}{N_G}\sum_{i=1}^{N_G}\frac{\left(\sum_{j=1}^{K_i}\gamma_{ij}\right)^2}{K_i\sum_{j=1}^{K_i}\gamma_{ij}^2}, \qquad (18)$$

where N_G is the number of subgroups with different fairness requirements, K_i $\left(\sum_{i=1}^{N_G}K_i=K\right)$ is the number of users in the subgroup i and γ_{ij} is the fraction of transmission resource allocated to user j in the subgroup i, which satisfies $\sum_{j=1}^{K_i}\gamma_{ij}=1$. When perfect fairness is achieved, $I_{QF}=1$. I_{QF} decreases from one as degree of unfairness increases. When perfect unfairness (*i.e.* all resources are allocated to one user) is achieved, $I_{QF}=1/K$.

A. Throughput Fairness with QF Scheduling Algorithm

Fig. 2 and Fig. 3 compare the throughput and the fairness index of the PF algorithm and the proposed QF algorithm for different smoothing factors α_t against the throughput slope m_t respectively. The number of users $K=8$ and the constants a_k^t are assumed to be one. We notice that PF-THP and QF-THP have the same fairness index when the average throughout of each scheme is identical (cross-over point). For example, when $\alpha_t=10$, these two algorithms have the same average throughput of about 10.9 [bits/channel use] with fairness index of about 0.90.

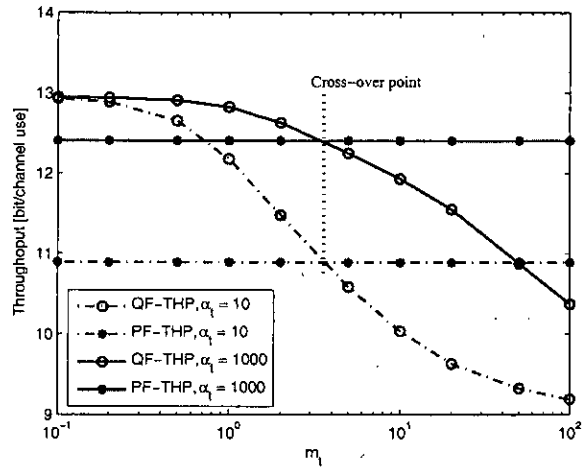


Fig. 2. The comparison of the PF algorithm and the proposed QF algorithm in terms of the average throughput for different smoothing factor against the throughput slope m_t, $M_T=4$, $SNR_0=10$dB, $f_c=2$GHz, $v_k=5.4$km/h, $\sigma_s=4.3$dB, $D=8.3$m, $K=8$ and $a_k^t=1$.

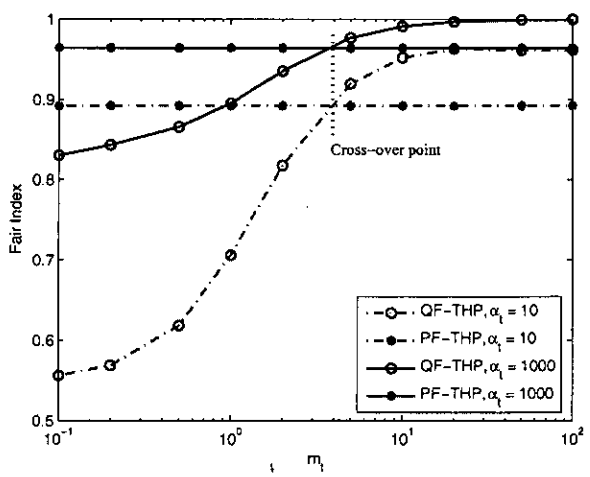


Fig. 3. The comparison of the PF algorithm and the proposed QF algorithm in terms of fairness index for different smoothing factor against the throughput slope m_t, $M_T=4$, $SNR_0=10$dB, $f_c=2$GHz, $v_k=5.4$km/h, $\sigma_s=4.3$dB, $D=8.3$m, $K=8$ and $a_k^t=1$.

This means that the performances of these two schemes are basically the same in terms of the average throughput and fairness. However, the proposed QF algorithm is able to control the degree of fairness by trading the average throughput for fairness among users. The larger m_t is, the higher the throughput fairness among users and the lower the throughput. The smoothing factor α_t determines the time window size for fairness. Larger time windows give more flexibility than smaller window sizes in selecting users due to the multiuser diversity. This increases the average throughput of all users. We notice that the case of $\alpha_t=1000$ has larger average throughput and fairness index than the case of $\alpha_t=10$. The percentage achievement

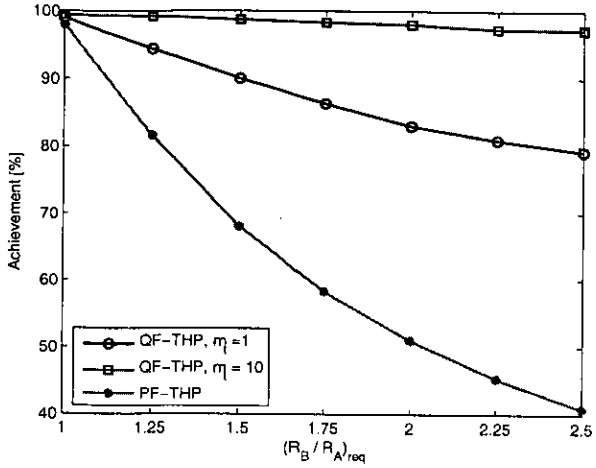


Fig. 4. The percentage achievement of different throughput requirements with the PF algorithm and the proposed QF algorithm, $M_T = 4$, $SNR_0 = 10$dB, $f_c = 2$GHz, $v_k = 5.4$km/h, $\sigma_s = 4.3$dB, $D = 8.3$m, $K = 8$ and $\alpha_t = 1000$.

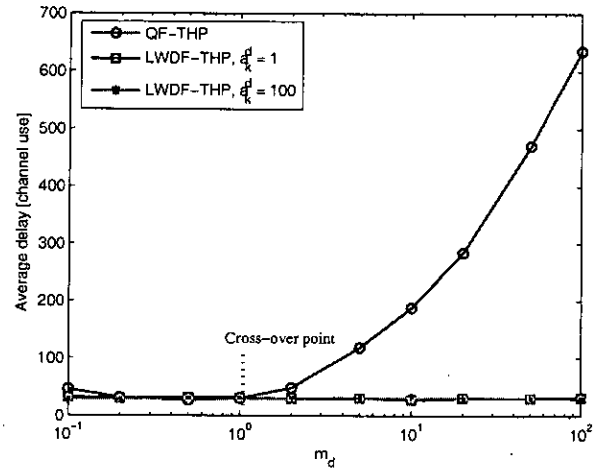


Fig. 5. The comparison of the the LWDF algorithm and the proposed QF scheduling algorithm in terms of the average delay against the delay slope m_d, $M_T = 4$, $SNR_0 = 10$dB, $f_c = 2$GHz, $v_k = 5.4$km/h, $\sigma_s = 4.3$dB, $D = 8.3$m, $\lambda = 1.25$ and $K = 8$.

is defined as

$$\frac{(R_B/R_A)_{out}}{(R_B/R_A)_{req}} \times 100\ [\%], \tag{19}$$

where $(R_B/R_A)_{out}$ is the ratio of the actual throughput of group B to group A and $(R_B/R_A)_{req}$ denotes the ratio of the required throughput of group B to group A. Fig. 4 shows the performance of the proposed QF algorithm for different throughput requirements. It is assumed that half of the users (group A) have a required throughput R_A and the others (group B) have a required throughput R_B. The smoothing factor $\alpha_t = 1000$ is also assumed. Because the PF algorithm does not consider different throughput requirements among users, it performs the worst in terms of the percentage achievement to the required throughput ratio. A large throughput slope m_t means strict fairness among users, so that the case of $m_t = 10$ with QF-THP performs better than the case of $m_t = 1$ in terms of the degree of fairness, which corresponds to the trend of the results in Fig. 3.

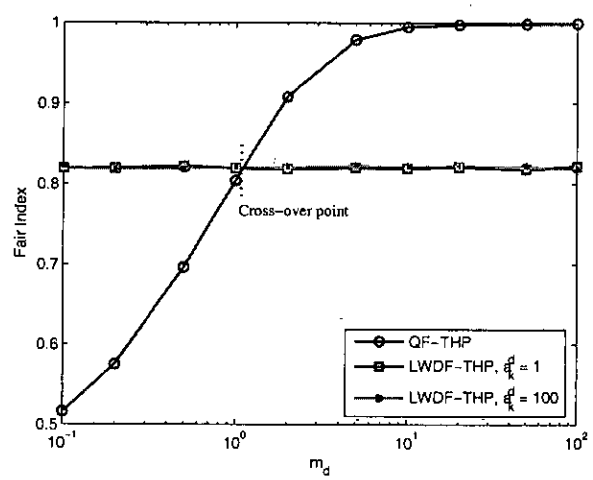


Fig. 6. The comparison of the LWDF algorithm and the proposed QF scheduling algorithm in terms of fairness index against the delay slope m_d, $M_T = 4$, $SNR_0 = 10$dB, $f_c = 2$GHz, $v_k = 5.4$km/h, $\sigma_s = 4.3$dB, $D = 8.3$m $\lambda = 1.25$ and $K = 8$.

B. Delay Fairness with QF Scheduling Algorithm

Fig. 5 and Fig. 6 compare the average delay and the fairness index of the LWDF algorithm and the proposed QF algorithm for different smoothing factors α_d against the delay slope m_d respectively. The number of users $K = 8$ and the constants a_k^d are assumed to be one. The arrival rate $\lambda = 1.25$ [bits/channel use] is assumed for all users. Unlike the LWDF algorithm, the proposed QF algorithm can adjust the degree of fairness by changing the delay slope m_d. The average delay and the fairness index of the LWDF algorithm is the same regardless of the constants a_k^d. Similar to the case of throughput fairness, the performance of these two algorithms in terms of the average delay and the fairness index is the same when $m_d \simeq 1$. The delay fairness metric increases the degree of fairness by increasing the average delay of all users. This results in the decrease of the average throughput of all users.

Fig. 7 and Fig. 8 show the effect of changing the degree of delay fairness on both the throughput and the delay performance respectively. As expected, the case of more strict delay fairness ($m_d = 10$) performs worst in terms of the throughput and delay performance. However, it achieves the highest degree of fairness in return for the performance degradation. The case of less strict delay fairness ($m_d = 0.1$) performs better than any other cases because it gives more flexibility than the case of $m_d = 10$ in selecting users for transmission.

Fig. 9 shows the performance of the proposed QF algorithm for different delay requirements. It is assumed that half of the users (group A) have a required delay D_A and

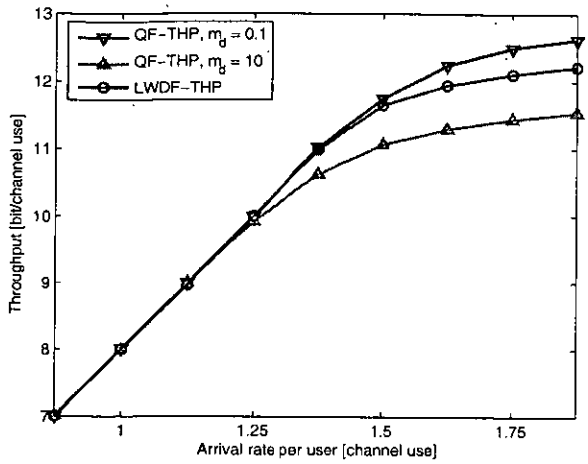


Fig. 7. Throughput performance of the LWDF algorithm and the proposed QF algorithm against different arrival rates per user , $M_T = 4$, $SNR_0 = 10$dB, $f_c = 2$GHz, $v_k = 5.4$km/h, $\sigma_s = 4.3$dB, $D = 8.3$m, $K = 8$ and $a_k^d = 1$.

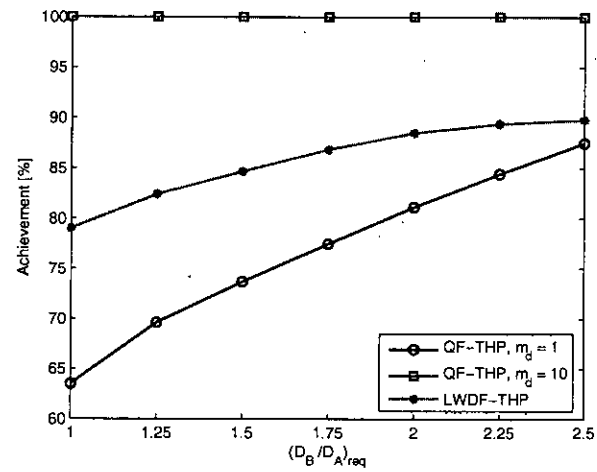


Fig. 9. The percentage achievement of different delay requirements with the LWDF algorithm and the proposed QF algorithm,, $M_T = 4$, $SNR_0 = 10$dB, $f_c = 2$GHz, $v_k = 5.4$km/h, $\sigma_s = 4.3$dB, $D = 8.3$m and $K = 8$.

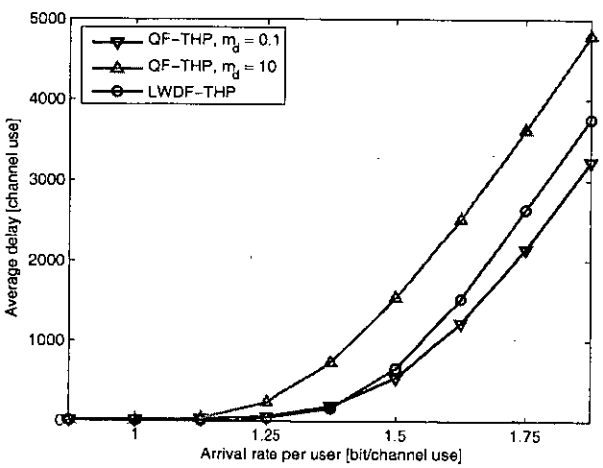


Fig. 8. Delay performance of the LWDF algorithm and the proposed QF algorithm against different arrival rates per user , $M_T = 4$, $SNR_0 = 10$dB, $f_c = 2$GHz, $v_k = 5.4$km/h, $\sigma_s = 4.3$dB, $D = 8.3$m, $K = 8$ and $a_k^d = 1$.

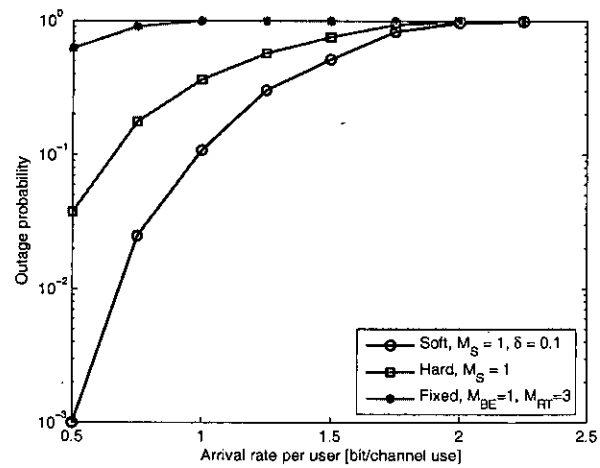


Fig. 10. The outage probability of the average delay of RT users exceeding the delay threshold D_{TH} for different antenna configurations, $M_T = 4$, $SNR_0 = 10$dB, $f_c = 2$GHz, $v_k = 5.4$km/h, $\sigma_s = 4.3$dB, $D = 8.3$m, $K_{RT} = 4$, $K_{BE} = 12$, $D_{TH} = 50$, $\alpha_t = 1000$, $a_k^t = 1$,$a_k^d = 1$ and $\delta = 0.1$.

the others (group B) have a required delay D_B. We notice that the proposed QF algorithm can satisfy the different delay requirements up to the ratio $(D_B/D_A)_{req} = 2.5$ when $m_d = 10$. When $m_d = 1$, the proposed QF algorithm performs worse than the LWDF algorithm. However, when the delay slope m_d becomes large, the average delay of all users also increases, which results in a decrease of the average throughput of all users.

C. Antenna Trade-off between Different QoS Groups

For the performance analysis of the proposed antenna trade-off scheme, it is assumed that there are 4 RT users ($K_{RT} = 4$) and 12 BE users ($K_{BE} = 12$) for each QoS group. The pre-assigned transmit antenna to RT users M_S is assumed to be one. For throughput fairness, $\alpha_t = 1000$ is also assumed. All users in BE and RT groups have the same throughput and delay requirements respectively. For the delay threshold for RT users, $D_{TH} = 50$ [channel use] is assumed. The parameter $\delta = 0.1$ is assumed for the dynamic range of the soft trade-off scheme.

Fig. 10 and Fig. 11 show the outage probability of the average delay of RT users exceeding the delay threshold D_{TH} and the average throughput with fairness considerations for different transmit antenna configurations respectively. We notice in Fig. 10 that the soft trade-off scheme performs best among all transmit antenna configurations in terms of the outage probability because it can change the number of transmit antennas assigned to RT users gracefully according to the channel conditions. This

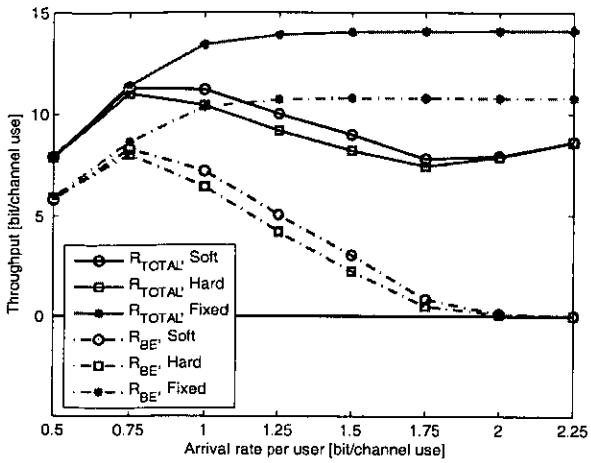


Fig. 11. The throughput performance for different antenna configurations, $M_T = 4$, $SNR_0 = 10$dB, $f_c = 2$GHz, $v_k = 5.4$km/h, $\sigma_s = 4.3$dB, $D = 8.3$m, $K_{RT} = 4$, $K_{BE} = 12$, $D_{TH} = 50$, $\alpha_t = 1000$, $a_k^t = 1$,$a_k^d = 1$ and $\delta = 0.1$.

enables the multiuser scheduling algorithm to minimize the outage probability of the average delay of RT users exceeding the delay threshold D_{TH}. The hard trade-off scheme performs worse than the soft trade-off scheme but better than the fixed scheme, which does not change the number of pre-allocated transmit antennas, in terms of the outage probability. However, the fixed scheme performs best in terms of the throughput of all users (R_{TOTAL}) and that of BE users (R_{BE}) because it can always include BE users regardless of the delay status of RT users. This maximizes the multiuser diversity gain. Nevertheless, the throughput results of the fixed scheme are misleading because it cannot satisfy the delay constraints of RT users unlike the other antenna trade-off schemes. Between antenna trade-off schemes, the soft trade-off scheme performs also better than the hard trade-off scheme in terms of the throughput. When the arrival rate per user is greater than about 2 [bits/channel use], the throughput of the soft trade-off scheme is almost the same as that of the hard trade-off scheme. This is because the number of transmit antennas selected by the soft trade-off scheme is likely to be the same as that by the hard trade-off scheme when the average delay of RT users is much larger than the delay threshold. Indeed, the throughput of BE users with antenna trade-off schemes is almost zero when the arrival rate is greater than 2 [bits/channel use]. This is because RT users have higher priority and they take more transmit antennas than the pre-assigned value M_S very frequently to satisfy the delay constraint. This results in less chance of transmission for BE users.

VI. Conclusions

We proposed and analyzed a new QoS-guaranteed multiuser scheduling algorithm, which consists of a QoS-aware Fair (QF) scheduling for selecting users in the same QoS group with throughput or delay fairness among users. We also discussed an antenna trade-off scheme for supporting different QoS users simultaneously with QoS differentiation. For the selection of a user set with the proposed QF scheduling algorithm, the weighted sum-rate maximization rule is exploited so that the selected user set satisfies the different fairness requirements whilst maximizing the sum-rate capacity. The exponential function with a scaled fairness deviation as its argument is used for fairness among users in terms of throughput and delay respectively. The proposed QF scheduling algorithm can also control the degree of fairness by adjusting the slope of the exponential fairness metric. In serving BE and RT users with delay constraints simultaneously, the antenna trade-off scheme performs better than the fixed antenna scheme in terms of both the delay performance of RT users and the throughput performance of all users. Between antenna trade-off schemes, the soft trade-off scheme shows better performance in terms of the outage probability and the throughput due to the graceful adaption to the time-varying channel conditions.

However, the computational complexity of weighted sum-rate maximization rule used by the proposed multiuser scheduling algorithm grows rapidly as the number of users increases. Hence, additional work should be done in order to decrease the complexity of the proposed algorithm for practical implementation.

References

[1] M. Costa, "Writing on dirty paper." *IEEE Transactions on Information Theory*, vol. 29, no. 3, pp. 439–441, 1983.

[2] S. Vishwanath, N. Jindal, and A. Goldsmith, "Duality, achievable rates, and sum-rate capacity of Gaussian MIMO broadcast channels," *IEEE Transactions on Information Theory*, vol. 49, no. 10, pp. 2658–2668, 2003.

[3] G. Caire and S. Shamai, "On the achievable throughput of a multiantenna Gaussian broadcast channel," *IEEE Transactions on Information Theory*, vol. 49, no. 7, pp. 1691–1706, 2003.

[4] R. Zhang, J. Cioffi, and Y. Liang, "Throughput comparison of wireless downlink transmission schemes with multiple antennas," *IEEE International Conference on Communications, ICC 2005*, vol. 4, 2005.

[5] T. Yoo and A. Goldsmith, "Sum-Rate Optimal Multi-Antenna Downlink Beamforming Strategy Based On Clique Search," *Global Telecommunications Conference, GLOBECOM 2005*, vol. 1, pp. 1510–1514.

[6] A. Jalali, R. Padovani, R. Pankaj, Q. Inc, and C. San Diego, "Data throughput of CDMA-HDR a high efficiency-high data ratepersonal communication wireless system," *51st IEEE Vehicular Technology Conference Proceedings, VTC2000-Spring Tokyo*, vol. 3, 2000.

[7] A. Stolyar and K. Ramanan, "Largest weighted delay first scheduling: large deviations and optimality," *The Annals of Applied Probability*, vol. 11, no. 1, pp. 1–48, 2001.

[8] M. Andrews, K. Kumaran, K. Ramanan, A. Stolyar, R. Vijayakumar, and P. Whiting, "CDMA data QoS scheduling on the forward link with variable channel conditions," *Bell Labs Technical Memorandum*, April, 2000.

[9] T. Yoo and A. Goldsmith, "On the optimality of multiantenna broadcast scheduling using zero-forcing beamforming," *IEEE Journal on Selected Areas in Communications*, vol. 24, no. 3, pp. 528–541, 2006.

[10] C. Swannack, E. Uysal-Biyikoglu, and G. Wornell, "Low complexity multiuser scheduling for maximizing throughput in the MIMO broadcast channel," *Proc. Allerton Conf. Communications, Control and Computing*, 2004.

[11] S. Catreux, P. Driessen, L. Greenstein, I. Wireless, and C. San Jose, "Data throughputs using multiple-input multiple-output (MIMO) techniques in a noise-limited cellular environ-

ment," *IEEE Transactions on Wireless Communications*, vol. 1, no. 2, pp. 226–235, 2002.

[12] M. Patzold and V. Nguyen, "A spatial simulation model for shadow fading processes in mobile radio channels," *15th IEEE International Symposium on Personal, Indoor and Mobile Radio Communications, PIMRC 2004*, vol. 3, 2004.

[13] Y. Zheng and C. Xiao, "Simulation models with correct statistical properties for Rayleigh fading channels," *IEEE Transactions on Communications*, vol. 51, no. 6, pp. 920–928, 2003.

[14] M. Vu and A. Paulraj, "A robust transmit CSI framework with applications in MIMO wireless precoding," *Proc. 39th Asilomar Conf. Signals Systems, Computers*, pp. 623–627, 2005.

[15] C. Fung, W. Yu, and T. Lim, "Precoding for the Multi-Antenna Downlink: Multiuser SNR Gap and Optimal User Ordering," *IEEE Transactions on Communications*, vol. 55, no. 1, pp. 188–197, 2007.

[16] N. Jindal, W. Rhee, S. Vishwanath, S. Jafar, and A. Goldsmith, "Sum power iterative water-filling for multi-antenna Gaussian broadcast channels," *IEEE Transactions on Information Theory*, vol. 51, no. 4, pp. 1570–1580, 2005.

[17] R. Wesel and J. Cioffi, "Precoding and the MMSE-DFE," *Conference on Signals, Systems and Computers*, vol. 2, pp. 1144–1148, 1994.

[18] M. Gudmundson, "Correlation model for shadow fading in mobile radio systems," *Electronics Letters*, vol. 27, no. 23, pp. 2145–2146, 1991.

[19] R. Jain, D. Chiu, and W. Hawe, "A quantitative measure of fairness and discrimination for resource allocation in shared computer systems," *Digital Equipment Corp. Eastern Research Lab, DEC-TR-301, September*, 1984.

[20] G. Berger-Sabbatel, A. Duda, M. Heusse, and F. Rousseau, "Short-term fairness of 802.11 networks with several hosts," *Proceedings of the 6th IFIP/IEEE International Conference on Mobile and Wireless Communication Networks.*