Radio Resource Management and Metric Estimation for Multicarrier CDMA systems

Moti M. Tabulo

THE UNIVERSITY OF EDINBURGH

A thesis submitted for the degree of Doctor of Philosophy.
The University of Edinburgh.
April 2005

Abstract

It is envisaged that the 4th generation (4G) of wireless networks will need to carry a variety of heterogeneous traffic types with different Quality of Service (QoS) requirements. A number of Medium Access Control (MAC) and Physical (PHY) layer technologies have been proposed for 4G networks. These schemes are typically based on multicarrier transmission and require a large amount of scarce radio frequency spectrum relative to current systems. As a consequence, there has been increased interest in dynamic radio resource allocation (RRA) algorithms that aim to make decisions on the optimal usage of resources to provide the QoS required by 4G networks.

The majority of RRA algorithms manage resources on a per layer basis and either make assumptions about the nature of traffic and propagation conditions, or assume that they have perfect, real-time estimates of these conditions. In practice a per layer management of resources may offer poorer performance benefits when compared to a cross layer approach. In addition, the assumptions made about the traffic and propagation conditions mean that it is hard to take full advantage of the variability in these conditions. These issues are further compounded for multicarrier systems as such systems have an additional degree of freedom by virtue of their frequency component.

This thesis investigates the management of radio resources in the PHY and MAC layers of multicarrier CDMA (MC-CDMA) systems and how the estimation of metrics in the various layers may be used in performing a cross layer management of resources to provide increased QoS whilst making optimal usage of the radio resource.

At the PHY layer, the grouping and subcarrier allocation problem for a grouped MC-CDMA system is formulated as an integer linear programming problem. Two algorithms are proposed to solve this problem, namely a Branch and Bound based algorithm and a mixed greedy-probabilistic Local Search algorithm. The Local Search algorithm is found to offer increased QoS (in terms of BER) for more users at a lower complexity than any of the other algorithms. At the MAC layer, a new multi-rate model – multi-group MC-CDMA (MG-MC-CDMA) – is introduced and the performance of power control and multi-group allocation algorithms in the MG-MC-CDMA system examined. A generalised processor sharing scheduler that takes advantage of the particular features of the MG-MC-CDMA system is proposed. The performance of some of these MAC layer algorithms is found to be limited by their assumptions as to how much capacity is available for use, underscoring the desirability of accurate capacity estimates to enable the management of resources.

A capacity model, incorporating an interference analysis and that takes into account the nature of the traffic types carried in the system, is outlined. In addition to MAC layer metrics characterising the traffic in the system, the capacity model has, as some of its required metrics, PHY layer parameters such as the ratio of inter-cell interference to total received power and information of whether or not a mobile is in a cell's edge region. New techniques are proposed to estimate these metrics. It is shown that the resulting dynamic capacity estimation framework can accurately and dynamically measure the capacity.

The final contribution of the thesis is the use of the proposed dynamic capacity estimation framework to develop new radio resource management algorithms that work across the PHY and MAC layers to deliver enhanced QoS.

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 Institute for Digital Communications at the University of Edinburgh.

The software programs used to perform the simulations were written by myself with the following exceptions:

- The routines used to generate Gaussian distributed noise and uniformly distributed samples were obtained from *Numerical Recipes in C* [1].
- The routines used to perform fast Fourier transforms were obtained from the *"Fastest Fourier Transform in the West (FFTW)"* project [2].
- The source code for the maximum likelihood multi-user detector used in Chapter 3 was provided by Dr. Emad Al-Susa.
- The routines for the simplex algorithm utilised in Chapter 3 were obtained from the *"LP Solve"* project [3].

Moti M. Tabulo

Acknowledgements

This work has been the culmination of a long process that has been assisted by many people to whom I would like to express my appreciation.

First and foremost to my two supervisors; Dr. Dave Laurenson and Prof. Steve McLaughlin. Their guidance throughout my Ph.D. studies has contributed greatly to my research and in the development of research skills. I would particularly like to thank them for their feedback in the last few months. Dave for his speed in doing so and Steve for finding time in a busy schedule to read through my chapters. Sadly, that was not reciprocated on my part by an even faster rate of chapter production! I wish to thank Dr. Yeonwoo Lee for many fruitful discussions and for introducing me to many interesting topics. On a similar note, I would also like to thank Dr. Emad Al-Susa for discussions in the early part of my research.

I would not have been able to commence my studies in Edinburgh were it not for the financial aid provided by the Mobile Virtual Centre of Excellence (Mobile VCE) and the Institute of Digital Communications; their support is gratefully acknowledged.

It has been a pleasure working with present and former colleagues in the Institute of Digital Communications; I've had a great time both socially and on the job. To my friends both in, and outside of Edinburgh, thanks for reminding me that there is a life away from TLC and, in the past year and a half, room 2.12!

Someone who has not directly affected this thesis, but still had an impact, is Dr. Paul, my former tutor from UMIST, who encouraged me to think about applying to do a Ph.D.

Getting this far has been made much easier with the continuous support of my family; my brothers Mwanze and Ruwa, and my sister Khadija. Most importantly, I would like to express my deepest gratitude to my parents for the sacrifices they have made that have allowed me to reach this point. *Mama na Baba, asanteni sana.*

Kujikwaa si kuanguka, bali ni kuendelea mbele

"To stumble is not to fall, instead it is to continue going forward"

– Old Swahili proverb.

Contents

List of figures

List of tables

Acronyms and abbreviations

3G	3rd generation
4G	4th generation
ADR	Aggregated data rate
ARQ	Automatic repeat request
AWGN	Additive white Gaussian noise
BER	Bit error rate
BPSK	Binary phase shift keying
BS	Base station
CAC	Call admission control
CDF	Cumulative Distribution Function
CDMA	Code division multiple access
CIR	Channel impulse response
cRTP	Compressed real-time transport protocol
DCA	Dynamic channel allocation
DCE	Dynamic capacity estimator
DCL	Degree of confidence level
DCPC	Distributed constrained power control
DPC	Distributed power control
DFT	Discrete Fourier transform
DS	Direct sequence
ETSI	European Telecommunications Standards Institute
FaR	Fair resource
FDMA	Frequency division multiple access
FH	Frequency hopping
FR-RR	Fixed resource round robin
GMC-CDMA	Grouped multicarrier code division multiple access
GoS	Grade of service
GPS	Generalised processor sharing
HARQ	Hybrid automatic repeat request

IDFT	Inverse discrete Fourier transform
IP	Internet protocol
ISO	International Standards Organisation
ITU	International telecommunications union
LAN	Local area network
LP	Linear Program(ming)
MAC	Medium access control
MAI	Multiple access interference
MC-CDMA	Multicarrier code division multiple access
MGA	Multi-group allocation
MG-GPS	Multi-group generalised processor sharing
MG-MC-CDMA	Multi-group multicarrier code division multiple access
MIMO	Multiple input multiple output
ML-MUD	Maximum likelihood multi-user detector
MMSE	Minimum mean square error
MRC	Maximum ratio combining
MS	Mobile station
MSE	Mean square error
MUD	Multi-user detector
NMSE	Normalised mean square error
OFDM	Orthogonal frequency division multiplexing
OFDMA	Orthogonal frequency division multiple access
ORC	Orthogonality restoring combining
OSI	Open systems interconnect
PHY	Physical
PSK	Phase shift keying
QAM	Quartenary amplitude modulation
QoS	Quality of service
QPSK	Quadrature phase shift keying
RDP	Relative downlink power
RE	Resource estimator
RME	Resource metric estimator
RMS	Root mean square

SINR	Signal to interference plus noise ratio
SIR	Signal to interference ratio
SS	Spread spectrum
TCP	Transmit control protocol
TDD	Time division duplex
TDMA	Time division multiple access
UDP	User datagram protocol
VoIP	Voice over Internet protocol
WFQ	Weighted fair queueing
WMSA	Weighted multi-slot averaging
WWW	World-wide web

Nomenclature

α	Orthogonality factor, or Pareto shape parameter
$\alpha_{0,u,g}$	Orthogonality factor for the 0-th BS's u-th MS's g-th group
$\hat{\alpha}_{0,u,g}$	Inverse orthogonality factor for the 0-th BS's u-th MS's g-th group
β_{max}	Total allocated subcarrier amplitude
β_{new}	Total subcarrier amplitude for a new allocation
$\beta_{u,k}$	Amplitude of the u-th user's k-th subcarrier
γ	Energy per bit to interference spectral density
γ^t	Target energy per bit to interference spectral density
γ_{t_m}	Target energy per bit to interference spectral density for traffic type t_m
$\gamma_{u,g}$	Energy per bit to interference spectral density for the u-th MS in the g-th group
$\hat{\gamma}_{u,g}$	Cross-layer estimate of the energy per bit to interference spectral density
Δx	Distance between two points
ε	Other-cell interference to total in-cell received power ratio
η	Downlink power factor
$\theta_u^{t_m}$	Activity factor of the traffic type t_m assigned to the u-th user
ϑ	Forgetting factor for exponential moving average
κ	Interference plus noise enhancement factor
λ_{t_m}	Arrival rate for the traffic type t_m
$\mu_{\mathrm{D_d}}$	Mean number of packets in a packet call
$\mu_{\mathrm{D_{pc}}}$	Mean reading time between packet calls
μ_{t_m}	Departure rate for the traffic type t_m
$\rho_{u,g,k}$	Sharing factor for the u-th user in the g-th group on the k-th subcarrier
$\rho_{u,k}$	Sharing factor for the u-th user on the k-th subcarrier
ϱ_k	Per k-th subcarrier equalised channel coefficient
σ_y	Mean of the other-cell downlink relative interference plus orthogonality factor
ϕ_u	GPS weight for the u-th session
φ	Diversity gain
Ψ	Proportion of total BS power allocated to traffic channels
ψ_i	Fraction of total BS downlink power allocated to the i-th MS

B_{pkt}	Packet size
B_{pc}	Amount of data in a packet call
B_u	Amount of backlogged data for the u-th MS
b_k	Linear programming slack variable
$C_{DCE,g}$	DCE per group estimate of the maximum capacity
C_g	Amount of free resource in the g-th group
C_{max}	Estimated maximum capacity
C_{norm}	Load at which capacity was estimated
c_k	Code chip for the k-th subcarrier
D_{d}	Mean time between two consecutive packets in a packet call
D_{pc}	Reading time between two consecutive packet calls
D_{VoIP}	Mean length of a VoIP session
$D_{\text{on,VoIP}}$	Mean ON duration of a VoIP session
$D_{\text{off,VoIP}}$	Mean OFF duration of a VoIP session
$\frac{E_b}{I_0}$	Energy per bit to interference spectral density
e_j^2	Euclidean distance between received and all possible sequences
f_k	Frequency for the k-th subcarrier
$f_{subc_{g,k}}$	Subcarrier frequency for the g-th group's k-th subcarrier
d_u	Transmitted data symbol for the u-th user
$\widehat{d}_u$	Data estimate for the u-th user
G	Number of groups
GRP_g	User-to-groups vector
$\tilde{g}$	Source group
$\hat{g}$	Target group
$grp_{k,g}$	User index at k-th position of the g-th group
H_g	Rayleigh fading matrix for the g-th group
$h_{u,k}$	Complex channel gain for the u-th user's k-th subcarrier
I_{in}	In-cell interference
I_{out}	Out-cell interference
K	Length of spreading code, number of subcarriers in each group
$L_{b,u}$	Path-loss between the b-th BS and the u-th MS
$\mathcal{N}(s)$	Neighbourhood space of the current solution
m_u	Number of codes allocated to the u-th user

m_y	Mean of the other-cell downlink relative interference plus orthogonality factor
N	Noise
N_c	Number of OFDM subcarriers
N_{cp}	Length of OFDM cyclic prefix
N_g	Vector of noise components
N_{pc}	Number of packets in a packet call
n_k	Noise at the k-th subcarrier
$n(t)$	Time domain noise sample
$\mathbf{P}$	Optimum point in a search space
P	Power of the desired receive signal
P_{best}	Probability of greedy search
P_{thr}	Maximum power per group threshold
P_{out}	Outage probability
$P_{out,max}$	Maximum allowable outage probability
P_u	Desired receive signal power for the u-th user
$p_{g,max}$	Maximum total transmit power per group
$p_{tot,g}$	Total transmit power for the g-th group
$p_{u,g,max}$	Maximum transmit power for the u-th MS in the g-th group
$p_{u,g,min}$	Minimum transmit power for the u-th MS in the g-th group
$p_{u,g}$	Transmit power for the u-th MS in the g-th group
$\mathbf{Q}$	Optimum point in a reduced search space
R	Data rate
$R(\Delta x)$	Normalised auto-correlation function, with distance, of shadow fading
R_g	Vector of the g-th group's received components
R_{t_m}	Data rate for the traffic type t_m
$R_u(\tau, t)$	Amount of service the u-th session receives in a time interval $(\tau, t]$
R_{res}	Amount of data that can be transmitted by one group/code resource unit in a time slot
r_k	Complex baseband received signal at the k-th subcarrier
r_u	Amount of resources allocated to the u-th MS
S_0	In-cell total received power
S_g	Synchronous addition of chip streams for all users in the g-th group
$\widehat{S_g}$	Estimate of the received sequence of the g-th group
$s_{g,k}$	k-th element of the synchronous chip stream S_g

$SUBC_g$	Subcarrier-to-groups vector for g-th group
$subc_{k,g}$	subcarrier index for the k-th subcarrier in the g-th group
T_s	Data symbol duration
T_{os}	OFDM symbol duration
T_{pc}	Duration a packet call is active
t_m	m-th traffic type
U	Number of users
U_b	Number of users attached to the b-th BS
$\tilde{u}$	Source user
$\hat{u}$	Target user
V_j	Set of all transmitted sequences
W	System bandwidth
W_{grp}	Bandwidth of one group
$x(t)$	Time domain IDFT output
y_u	Other-cell relative downlink interference plus orthogonality factor for the u-th user
Z_{t_m}	SIR and activity factor weighted sum of other-cell relative downlink interference
z_k	Combining gain for the k-th subcarrier
$E\{\ \}$	Expectation operator
ln	Natural logarithm
$\log_2$	Logarithm to base 2
$\log_{10}$	Logarithm to base 10
$\bigwedge_{n=1}^{N}$	Logical AND of N terms
$\bigvee_{n=1}^{N}$	Logical OR of N terms
$P(n = N)$	Occupancy distribution
$Pr(\ )$	Probability of an event
$\lfloor\ \rfloor$	Round down to the nearest integer
$\lceil\ \rceil$	Round up to the nearest integer
$\oplus$	Exclusive OR

Chapter 1
Introduction

1.1 Motivation

The last four decades have seen the explosion of the Internet from a research program at the US Army's Advanced Research Projects Agency (ARPA) to a network that pervades everyday life providing access to a wide variety of information and a plethora of applications and services such as downloadable music, video, video telephony etc.

Forecasts on mobile usage suggest that as the penetration of cellular services saturates, the revenues from traditional circuit switched speech will stagnate. If the wireless communications industry is to sustain its business growth and continue to satisfy customer demand, then the provisioning of data services over wireless networks will be key. The wireless communications industry has coped with such changes by introducing progressive evolutions of wireless systems beginning with the first generation of circuit-switched analogue systems through to the current third generation (3G) of digital mixed circuit and packet switched systems.

The wireless industry is not resting on its 3G laurels and has already begun to focus on the next generation of wireless systems, otherwise known as beyond-3G or 4G systems. The vision of what a next generation system will look like is still under much discussion. There are, however, some common themes: next generation systems will be entirely packet switched, will require large amounts of throughput and will support an even larger variety of traffic types with heterogeneous quality of service (QoS) requirements relative to current 3G systems. Furthermore, to maximise their revenue, operators would need to provide these services whilst making optimal usage of the limited and expensive radio resource.

A number of high throughput Media Access Control (MAC) and physical layer (PHY) schemes have been proposed for next generation systems. The majority of these schemes have the common characteristic that they are based on multicarrier transmission. These schemes are the likes of Orthogonal Frequency Division Multiple Access (OFDMA, [4]) and spread multicarrier systems such as Multi-Carrier Code Division Multiple Access (MC-CDMA, [5, 6]), Orthogonal

Frequency and Code Division Multiplexing (OFCDM, [7]), etc. Of these, spread multicarrier systems offer the most potential for next generation services due to their combining the advantages of both CDMA and multicarrier systems. OFCDM allows for two dimensional spreading in both the frequency and time domains, whereas MC-CDMA only spreads information in the frequency domain. MC-CDMA may therefore be thought of as a special case of OFCDM. To better examine the key component of multicarrier systems, i.e. the frequency domain, this thesis focuses only on MC-CDMA.

The usage of MC-CDMA to meet QoS requirements whilst providing resource efficient services raises a number of interesting questions:

- The frequency component of MC-CDMA systems offers an additional degree of freedom compared to conventional CDMA systems. How should radio resource management algorithms take advantage of this degree of freedom?
- The capacity of MC-CDMA systems is interference limited, optimal usage of the capacity will aid in the efficient usage of resources. What metrics characterise the capacity and the state of the wireless resource in MC-CDMA systems, how may these be measured and how may they be used to provide better quality of service?
- The state of the PHY layer impacts on the MAC layer and vice versa, how can the metrics that characterise the state of the wireless resource be leveraged to enable the cross-layer management of resources to provide better performance?

The aims of this thesis are thus to investigate the management of resources in MC-CDMA systems, how measurements that characterise the radio resource are taken and used to aid in the management of resources to provide increased QoS.

1.2 Thesis structure

The structure of the remained of this thesis is as follows:

- **Chapter 2:** In this chapter multicarrier systems are introduced in some depth, CDMA is discussed and the advantages gained by combining multicarrier transmission with CDMA based multiple access presented. QoS and methods to provide QoS in wireless systems are highlighted.

- **Chapter 3:** A study of PHY level subcarrier allocation for a grouped MC-CDMA system is performed in this chapter. Several criteria for grouping and subcarrier allocation are presented, the received per subcarrier amplitude is chosen as the metric upon which to perform subcarrier allocation. It is shown that MC-CDMA subcarrier and grouping allocation with this metric can be formulated as an integer linear program. Two techniques to solve this linear program are proposed; a branch and bound based technique and a low complexity local search based technique. It is shown that both these techniques offer improved QoS for more users than other approaches in the literature.

- **Chapter 4:** Here the performance of various resource management algorithms are evaluated for MC-CDMA. Three power control variants are evaluated for the grouped MC-CDMA system and the best is found to be one which includes both per MS and per group constraints. A multi-rate allocation model termed multi-group allocation, that takes into account the frequency dimension of MC-CDMA systems is proposed. The performance of three multi-group allocation schemes is evaluated for real-time traffic and it is found that in the case of traffic requiring absolute bandwidth guarantees a multi-group allocation algorithm that performs allocations on a super-group basis is preferred to flexible schemes that perform per-group allocation. A new scheduler, termed the multi-group generalised processor sharing (MG-GPS) scheduler, that takes into account the specific nature of the grouped MC-CDMA system is proposed. This scheduler is shown to offer better quality of service for non-real time traffic than the other schedulers evaluated.

- **Chapter 5:** To maximise the use of the capacity in CDMA systems it is advantageous to be able to measure it. In this chapter two schemes for MC-CDMA capacity estimation are proposed. The first is for single cell MC-CDMA and does not take into account the specific nature of the carried traffic. For the second scheme an in depth capacity analysis is highlighted that allows the inclusion of traffic information in obtaining capacity estimates for the multi-cell downlink. The capacity analysis is then used as a basis to define what metrics are required. The chapter then goes on to propose various techniques in which the requisite metrics may be measured. The resulting capacity estimation framework termed the Dynamic Capacity Estimator (DCE) is evaluated for various PHY layer imperfections in the estimates.

- **Chapter 6:** The focus of this chapter is to combine the techniques developed in the previous chapters to allow for the cross-layer management of resources. The management of resources is is divided into two approaches depending on the traffic type. For real-time traffic the resource management issue is taken to be whether or not a call should be admitted into the network. A joint call admission and multi-group allocation algorithm utilising the DCE is proposed. This algorithm is compared to not using any admission control and to using a joint multi-group and power threshold based admission control algorithm. The DCE based admission control algorithm is shown to have equivalent performance to the power threshold algorithm without the need to pre-set an admission threshold. Both admission control algorithms perform better than not using any admission control. For non-real time traffic, scheduling of data and resources is taken to be the key resource management issue. A scheduling algorithm utilising a DCE based MG-GPS scheduler in conjunction with a cross-layer aware metric calculator is proposed to allow for the cross-layer management of resources. This scheduling algorithm is found to offer QoS performance gains at moderate to high loads over other scheduling algorithms.

- **Chapter 7** The objective of the final chapter is to summarise the work presented, present some conclusions, highlight limitations of the work performed and provide pointers to potential future work.

Chapter 2
Wireless Communications and Quality of Service

In order to establish the contributions of this thesis, it is necessary to review the background behind the concepts to be investigated. This chapter begins with a description of orthogonal frequency division multiplexing (OFDM) in section 2.1, section 2.2 describes multiple access techniques with an emphasis on CDMA. The combination of multicarrier transmission with CDMA multiple access is explained in section 2.3. Quality of service and its provisioning is described in section 2.4. The concept of resource metric estimation first broached by Jorguseski *et al.* [8] is explained in section 2.5. The chapter ends with a summary in section 2.6.

2.1 OFDM

The history of OFDM dates back to the work by Chang [9] in which he set out the principle of transmitting simultaneous messages through a linear band limited channel without either inter-symbol interference (ISI) or inter-channel interference (ICI). This was followed by work on orthogonal multiplexing [10]. Weinstein and Ebert [11] made a contribution that was to aid in the implementation of OFDM in practical systems with their introduction of the use of the discrete Fourier transform (DFT) in generating the multiple subchannels. The problem of keeping the subcarriers orthogonal to achieve zero ICI even after the signal has passed through a time-dispersive channel was addressed by Peled and Ruiz [12]. They solved this problem with the use of a cyclic prefix or cyclic extension in which a guard space between symbols is filled with a cyclic extension of the OFDM symbols.

OFDM is based on frequency division multiplexing in which data is transmitted over several frequency sub-bands each separated by a sufficiently large guard band to prevent adjacent bands interfering with each other. This implementation results in an inefficient use of spectrum, which can be overcome if orthogonal sub-bands are used that allow the sub-bands to overlap. A set of transmitter receivers pairs can be used to implement such a modem, which is inefficient from a receiver complexity point of view. A more efficient implementation is possible through the use of the DFT to perform baseband OFDM modulation and demodulation.

Figure 2.1 shows the structure of an OFDM system.

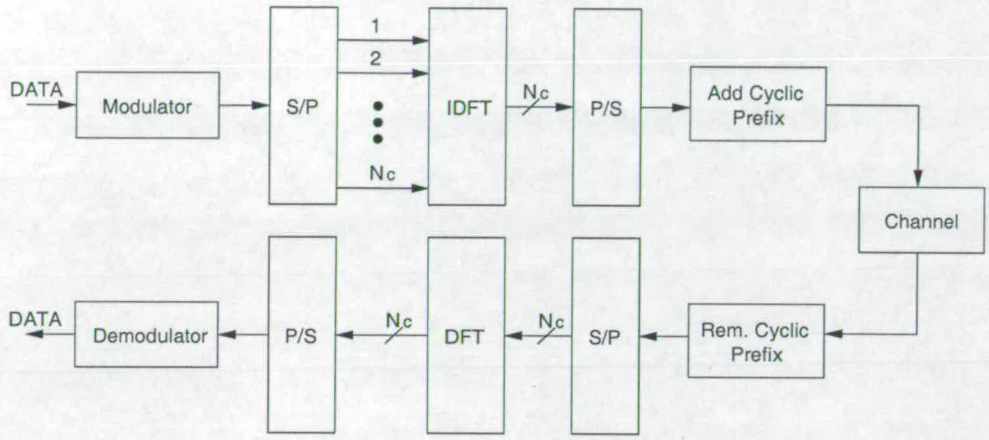


Figure 2.1: *OFDM system*

The serial, binary data for transmission is mapped to data symbols with a symbol rate of $1/T_s$ using some phase and amplitude modulation technique e.g. m-ary phase shift keying (m-PSK), quadrature amplitude modulation (QAM) etc. Conversion of the modulated data into time domain samples is performed using an N_c-point IDFT, via a serial to parallel conversion. The resulting time domain samples at the output of the parallel to serial converter form an OFDM symbol whose symbol rate is $1/(N_c T_s)$ i.e. the symbol rate has been reduced by a factor of N_c.

In order to counteract the inter-symbol interference introduced when the symbols are passed through a time dispersive channel, a cyclic prefix or guard interval is appended to the OFDM symbol. The cyclic prefix is composed of samples from the end of the OFDM symbol as shown in Figure 2.2. The length of the cyclic prefix, N_{cp} is chosen to be longer than the multipath delay spread such that components of one symbol do not interfere with the next symbol [13]. The resulting length of the OFDM symbol plus cyclic prefix is $N_c + N_{cp}$. In the receiver, the

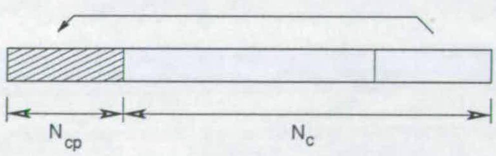


Figure 2.2: *Appendage of the cyclic prefix*

channel corrupted OFDM symbol has the N_{cp} samples of the cyclic extension discarded to leave the N_c samples of the original symbol. A spectral decomposition of the OFDM symbol into its N_c constituent parts is performed by an N_c-point DFT via a serial to parallel conversion. The N_c modulated data symbols are restored to serial order before being passed to a demodulator for conversion back to binary data.

2.1.1 Advantages of OFDM over single-carrier transmission

The channel equaliser is typically one of, if not the most computationally expensive part of a wireless communication receiver. For wireless channels, the bandwidth-delay spread product is a measure of the relative amount of frequency selectivity [13]. A channel is generally said to be frequency selective if the bandwidth delay spread product exceeds 0.1 [14]. The complexity of a time domain equaliser is governed by the number of taps it requires, a large bandwidth delay spread product translates into a large number of taps. The large throughput demands of next generation systems will require large amounts of bandwidth meaning a correspondingly large bandwidth delay spread product.

OFDM offers a way to manage this complexity; the increase in the symbol duration by a factor of N_c results in a symbol duration that is larger relative to a given delay spread. An operation which is equivalent to increasing the channel's coherence bandwidth by a factor of N_c. Therefore, with judicious choice of N_c, the OFDM system can be designed such that flat-fading narrowband propagation conditions apply to each sub-carrier which can be easily corrected by a complex multiplication [15]. Coupled with the use of the cyclic prefix to combat ISI, OFDM is able to deal with large delay spreads without the implementation complexity associated with having a long equaliser. It should be noted that the complexity difference between OFDM and single-carrier transmission is reduced if a frequency domain equaliser is utilised, however the equalisation process will still be twice as complex as that of OFDM due to the frequency domain equaliser needing to perform both a DFT and an IDFT [16].

There are some disadvantages to using OFDM; the use of a cyclic prefix reduces the data rate by a factor of $N_c/(N_c+N_{cp})$, OFDM systems are more sensitive to synchronisation errors and their signal output has a higher peak to average power ratio (PAPR) than single-carrier systems. More information on these degradations can be found in the work by Hanzo *et al.* [15] and the references therein.

2.2 Multiple Access Techniques

Multiple access techniques are used to allow the simultaneous sharing of a finite radio resource by multiple users without causing a significant degradation in the performance each user experiences [17].

There are several ways in which multiple users may access the channel depending upon how the radio resource is partitioned. The three major techniques are treated briefly below:

- In frequency division multiple access (FDMA), the radio bandwidth is divided into smaller units (or channels) and each user given one of these channels for the duration of its call.
- Division of the radio resource into small time units (or time slots) is the basis for time division multiple access (TDMA). Each user in a TDMA system occupies the entire bandwidth, but is only periodically allowed to use this bandwidth for a duration corresponding to the length (or a multiple of the length) of a time slot.
- Spread spectrum multiple access (SSMA) uses a pseudo noise sequence (also known as a spreading code) to increase (or spread) the bandwidth of a signal. Users in an SSMA system are separated in the code domain and may exist in the same bandwidth at the same time. There are two main types of SSMA techniques, frequency hopped spread spectrum (FH-SS) and direct sequence spread spectrum (DS-SS). In FH-SS the carrier frequencies of individual users are varied in a pseudo-random fashion within a wideband channel [17]. DS-SS systems spread the information signal by multiplying it with a larger bandwidth spreading code [18]. DS-SS is also known as direct sequence code division multiple access (DS-CDMA). Of the two aforementioned CDMA techniques, DS-CDMA is more prevalent in wireless standards and is the spread spectrum multiple access technique considered in this thesis. Henceforth the term CDMA will be taken to mean DS-CDMA.

It is also possible to have hybrids of the above techniques, for example FDMA/TDMA such as is used in the GSM standard, or CDMA/TDMA as used in some 3G standards.

2.2.1 Cellular CDMA

In a CDMA receiver the despreading operation used to detect a CDMA signal results in an increase in the power of the desired signal by a factor proportional to the amount by which the bandwidth of the original signal was expanded (i.e. the spreading gain). Spreading gain is what gives a CDMA system its robustness against interference. Theoretically, in a single-cell environment affected only by additive white Gaussian noise (AWGN), and assuming synchronous orthogonal codes (or synchronous pseudo-noise codes with an optimal multi-user detector) there is no self-interference and the spectral efficiencies for FDMA, TDMA and CDMA are equal [18].

However in a multi-cell environment the situation is different. A finite radio resource means that the same resource must be reused in other cells of a multi-cell system, this process is known as frequency reuse. The reuse of resources introduces inter-cell interference. To cope with the degradation caused by the inter-cell interference, TDMA and FDMA systems define a certain minimum distance, D, at which resources in another cell may be reused; the larger the separation, the smaller the inter-cell interference. The separation does come at a price though; the spectral efficiency of the multiple access technique is reduced. Herein lies the main advantage of CDMA, the robustness to interference afforded by the spreading gain allows the CDMA system to operate at much higher interference levels than a non-spread system with the very beneficial effect that resources can be reused in every adjacent cell (i.e. a frequency reuse factor of 1). The reuse of resources in every cell means that the spectral efficiency of CDMA in a cellular environment is higher than that of either TDMA or FDMA which are not interference robust. It is for this reason that CDMA techniques are used for 3G systems, and are being considered for next generation wireless networks. Further spectral efficiency improvements are possible with the use of sectorised antennae, discontinuous transmission to exploit the inactivity of users and exploitation of the wider system bandwidth to resolve multiple propagation paths with higher accuracy to take advantage of multipath diversity.

There are some provisos to obtaining this spectral efficiency in CDMA [19, 20]:

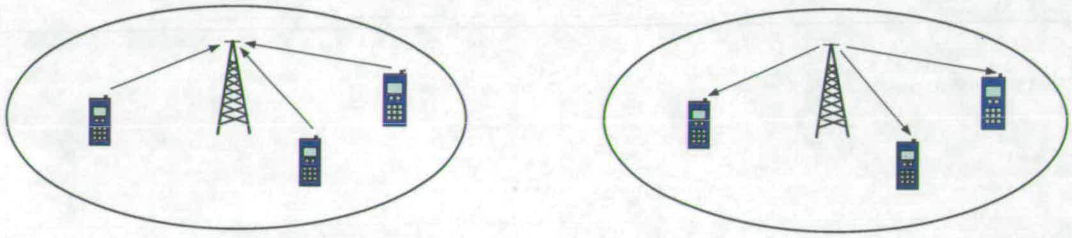
1. Perfect power control is required without which the spectral efficiency can be reduced significantly.

2. MSs are assigned to communicate with the BS to which they experience the lowest path loss, rather than just being assigned to the nearest BS. This means that for MSs with

mobility special attention must be paid to hand them off from unsuitable BSs when this criteria is not met. Handoff must also be performed for FDMA and TDMA systems, however due to the large separation between cell's using the same resource, handoff is not as critical an issue as it is for CDMA systems.

2.2.1.1 Simple capacity analysis for downlink CDMA

The downlink refers to the direction of data flow from a BS to an MS(s). It is also known as the forward link. The information flow in the opposite direction is known as the uplink, or reverse link. The two link directions are illustrated in Figure 2.3. The asymmetry of throughput



(a) The uplink **(b)** The downlink

Figure 2.3: *Cellular link directions*

requirements (in favour of the downlink) for some of the services envisaged for next generation networks suggests that the downlink will be the limiting link in these networks [21]. For example, internet surfing generally requires large amounts of throughput in the downlink to transmit downloaded content such as images, media files etc., whereas the uplink carries relatively low throughput page requests and protocol acknowledgement messages etc. Thus this thesis focuses on the downlink.

What follows is a simple analysis of the downlink capacity to reveal some of its characteristics. Before proceeding to do so, it is important to clarify the term "capacity" as used in this thesis. The capacity in wireless communications systems is normally defined as the sum of information bits that can be transmitted by all users in a cell within a certain time period. This may be referred to as the "transmission capacity". The transmission capacity is governed by the interference conditions, the PHY layer techniques used such as error correction, advanced modulation etc. An alternative measure of capacity to the transmission capacity is the maximum number of users that can be simultaneously accommodated in the cell also with a given quality [22]. The capacity in this sense may be referred to as the "connection capacity". The connection capacity assumes that users are given sufficient capabilities to meet the transmission capacity requirements. The objective thereafter is to maximise the number of users that can simultaneously connect and achieve this transmission capacity. Say for example each user can achieve a transmission capacity of 100 kbps at a certain bit energy to interference spectral density ratio, E_b/I_0, the system design aim with regards to the connection capacity will then be to maximise the number of users that can simultaneously achieve this E_b/I_0 and thus transmission capacity. The connection capacity also allows for an accounting of the statistical multiplexing gain to be obtained from the silence periods in user's transmissions (e.g. during silence periods in speech or when a WWW traffic user is reading a downloaded page).

In the remainder of the thesis whenever the term capacity is mentioned, it will be taken to mean the achievable maximum number users that can connect simultaneously with a given quality, i.e. the connection capacity.

The E_b/I_0 for a CDMA user can be expressed as follows:

$$(\frac{E_b}{I_0})_u = \gamma_u = \frac{W}{R} \cdot \frac{P_u}{I_{in} + I_{out} + N} \tag{2.1}$$

in which W is the spreading bandwidth, R is the data rate, P_u is the power of the u-th user's desired signal, I_{in} is the in-cell interference power, I_{out} is the other-cell interference power and N is the noise power.

The downlink is synchronous and as such orthogonal spreading codes may be used, however multipath channels degrade the orthogonality of these codes resulting in in-cell interference. An orthogonality factor α describes the loss in orthogonality with a '0' indicating perfect orthogonality between codes and a '1' completely non-orthogonal codes. The downlink in-cell interference at the u-th user's location may then be expressed as

$$I_{in} = \alpha S_{0,u} \tag{2.2}$$

in which $S_{0,u}$ is the total received power from the in-cell BS (the 0-th BS) at the u-th user.

The other-cell interference at each user location can be described in relation to the total in-cell received power via an other-cell interference power to total in-cell received power ratio:

$$\epsilon_u = \frac{I_{out}}{S_{0,u}} \tag{2.3}$$

the other-cell interference is then simply $I_{out} = \epsilon_u S_{0,u}$.

If all U users in the desired cell are assigned the same transmit power, then the total received power from the in-cell BS will be:

$$S_{0,u} = UP_u \tag{2.4}$$

The in-cell and other-cell interference can then be respectively written as:

$$I_{in} = \alpha UP_u \tag{2.5}$$

$$I_{out} = \epsilon_u UP_u \tag{2.6}$$

and the E_b/I_0 for the u-th user expressed as:

$$\gamma_u = \frac{W}{R} \cdot \frac{P_u}{\alpha UP_u + \epsilon_u UP_u + N} \tag{2.7}$$

For the purpose of this analysis N is assumed to be negligible compared to the interference. Further, it is assumed that all users are positioned in the cell in such a way that their gain to the BS is the same and they therefore experience the same received power and interference conditions making their E_b/I_0, other-cell interference to total received power ratio and received power values the same:

$$\gamma_1 = \gamma_2 = \ldots = \gamma_u = \gamma$$

$$\epsilon_1 = \epsilon_2 = \ldots = \epsilon_u = \epsilon$$

$$P_1 = P_2 = \ldots = P_u = P$$

to give

$$\gamma = \frac{W}{R} \cdot \frac{P}{\alpha UP + \epsilon UP} \tag{2.8}$$

Equation 2.8 is easily manipulated to give the capacity of the system:

$$U = \frac{W}{R} \cdot \frac{1}{\gamma(\alpha + \epsilon)} \tag{2.9}$$

Despite the limitations of this capacity analysis, it illustrates some important points about CDMA in general and the downlink in particular. Firstly, the capacity depends upon the processing gain of the system, as the bandwidth W is increased, the ability to reject interference increases correspondingly, however an increase in spreading bandwidth is limited by the amount of expensive radio resource spectrum available. Secondly, the capacity can be improved by having a lower required E_b/I_0. Finally, the amount of capacity is dependent on the interference experienced from inside and outside the cell. For instance a small orthogonality factor means less in-cell interference and higher capacity, a corresponding increase in capacity is possible with smaller other-cell interference. The CDMA capacity is therefore rather flexible depending on the prevailing interference conditions and is said to be interference limited.

In deriving equation 2.9, some assumptions were made namely that the gain between each user and the BSs in the system was the same and that the in-cell BS assigned all its users the same transmit power level. In actual fact each user will experience different gains to the BSs in the system, resulting in different desired receive power levels and interference powers at each user location. Variations caused by different traffic loads in cells will also result in variation in the transmit power from each BS. Therefore accounting for all these factors to calculate and estimate the capacity of a CDMA system turns out to be a non-trivial task.

2.3 Spread OFDM Techniques

Combined OFDM and DS-CDMA transmission systems have recently received significant interest as potential multiple access methods for the next generation of high data rate wireless systems. OFDM modulation provides robustness against multipath fading with minimal receiver complexity, while the CDMA part is used for its multiple access flexibility and ability to provide higher capacity over other multiple access techniques [23].

There are two main types of combined OFDM/CDMA schemes. One converts the original data stream from serial-to-parallel, spreads the converted data using a spreading code and then modulates a different subcarrier with the resulting spread data stream [24] (this is similar to a normal DS-CDMA system where the spreading operation is performed in the time domain). In the second type, the original data stream is spread using a spreading code and then each chip of the resulting spread data stream is used to modulate a different subcarrier [5,6](this is equivalent to the spreading operation being performed in the frequency domain). These two schemes are commonly known as MC-DS-CDMA and MC-CDMA respectively. A third scheme is orthogonal frequency and code division multiplexing (OFCDM) which combines MC-CDMA with MC-DS-CDMA to perform spreading in both the frequency and time domains.

Abeta *et al.* [25] performed a study of the performance of MC-CDMA and MC-DS-CDMA. They concluded that MC-CDMA was superior to MC-DS-CDMA in the downlink due to MC-CDMA's ability to better take advantage of the channel's frequency diversity. Accordingly, in this thesis only MC-CDMA is considered, the combination of MC-DS-CDMA with MC-CDMA is not examined due to the focus of the thesis being on how the frequency dimension can be better utilised.

The following sub-section treats MC-CDMA in greater depth.

2.3.1 MC-CDMA

The structure of a single-user MC-CDMA transmitter is shown in figure 2.4. The data to be transmitted is modulated, copied N_c times (where N_c is the number of subcarriers and the length of the spreading code) and then multiplied by a spreading code C. The resulting spread data is then OFDM modulated by performing an Inverse Discrete Fourier Transform (IDFT). This whole operation is equivalent to transmitting each chip of the spreading code through a different subcarrier.

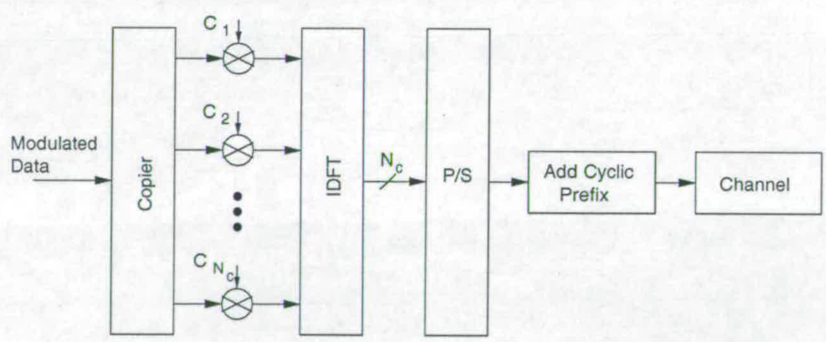


Figure 2.4: *Single-user MC-CDMA transmitter*

A conventional single-user MC-CDMA receiver is shown in Figure 2.5. Information received from the channel is serial-to-parallel converted, demodulated by passing it through a Discrete Fourier Transform (DFT) and despread by multiplying each subcarrier by its corresponding code chip. The N_c subcarriers are then combined into one received data estimate using a combining technique.

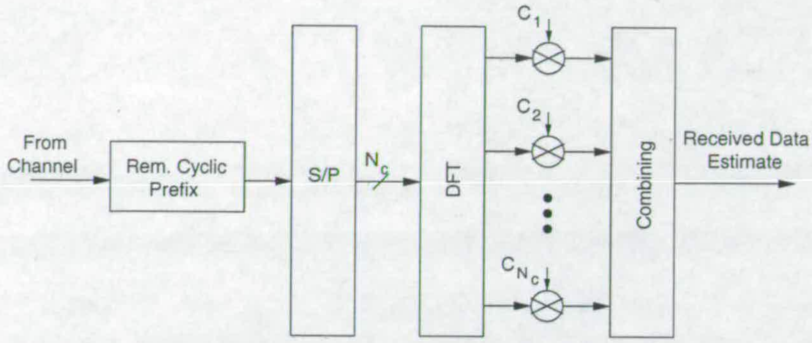


Figure 2.5: *MC-CDMA receiver*

Figure 2.6 shows the internal structure of the combiner, which is so called due to its combining the frequency domain information into one received data estimate. After the k-th subcarrier

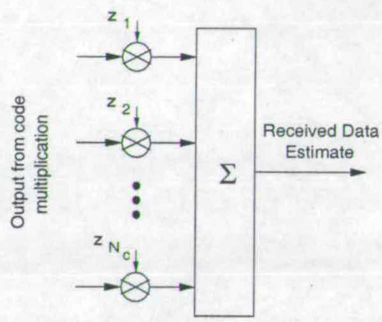


Figure 2.6: *MC-CDMA combiner*

is multiplied by the corresponding chip of the spreading code, it is multiplied by a combining gain, z_k, before the individual chips are summed to produce the decision variable for the u-th user:

$$s^u = \sum_{k=1}^{N_c} \varrho_k \tag{2.10}$$

where ϱ_k is the per subcarrier equalised channel coefficient given by

$$\varrho_k = z_k r_k \tag{2.11}$$

here z_k is the combining gain and r_k is the complex baseband component of the received signal after down conversion [23]:

$$r_k = \sum_{u=1}^{U} d^u h_k^u c_k^u + n_k \tag{2.12}$$

in which d^u, h_k^u and n_k are the transmitted symbol, complex channel gain and noise, respectively, on the k-th subcarrier for the u-th user. U is the number of users belonging to the same BS. For the downlink channel h_k^u can be assumed to be h_k for all users belonging to the same BS.

There are various ways to select the combining gain z_k, some of these are briefly covered below. More information can be found in the works by Hara [23], Kaiser [26, 27] and the references therein.

- Orthogonality Restoring Combining (ORC) or Zero Forcing: the combining gain is chosen as

$$z_k = \frac{h_k^\star}{|h_k|^2} \tag{2.13}$$

  ORC can completely eliminate the multi-user interference, however this comes at the price of an amplification of noise when low-level subcarriers are multiplied by high gains.

- Maximum Ratio Combining (MRC): the combining gain is given by

$$z_k = h_k^\star \tag{2.14}$$

  In the case of one user, MRC is able to minimise the BER. However with multiple users MRC tends to heighten the loss of orthogonality, rendering it practically unusable.

- Minimum Mean Square Error (MMSE): sets the combining gain as

$$z_k = \frac{h_k^\star}{|h_k|^2 + 1/(\gamma_b U)} \tag{2.15}$$

  in which γ_b is the present signal to noise ratio and U the number of users. The MMSE criterion limits the size of the combining gain for weak subcarriers, reducing the amount of noise amplification. For strong subcarriers the combining gain approaches that of ORC, restoring the orthogonality amongst users. For these reasons, the MMSE criterion is the preferred combining technique from the class of single-user detectors.

2.4 Quality of Service

2.4.1 Introduction

The quality of service (QoS) is a measure of how well a network does its job of transferring various kinds of information from a source to its destination. The QoS a network offers refers to a conglomeration of several performance metrics, the most important of these are [28, 29]:

1. Throughput: is the effective data transfer measured in bits per second. Sharing a network lowers the per user realisable throughput.
2. Latency: or delay, is the time taken by data to travel from source to destination.
3. Residual Bit Error Ratio: is the undetected bit error ratio in the delivered data units.
4. Packet loss due to congestion: is the number of packets lost when the source's finite length data queue overflows due to the link being congested.
5. Availability: is the time a network is available for use, ideally it should be available 100% of the time.

In this thesis BSs are assumed to be available 100% of the time and to have infinite length queues; therefore the QoS parameters of most relevance are the throughput, latency and residual bit error ratio.

Different services have different requirements for these parameters. Services may be broadly classified as falling into one of two classes [29]:

- Real-Time (RT) Traffic. Real-time services are those considered to place strict latency requirements to transfer data within a certain maximum delay. Due to these strict latency requirements, error correction and data retransmission possibilities are limited. Therefore the RT application must be able to cope with errors introduced in transmission.
- Non-Real Time (NRT) Traffic. These have less stringent latency requirements than RT traffic, but typically demand error free transmission. The lenient latency requirements allow for more robust error recovery mechanisms such as retransmissions.

2.4.2 QoS provisioning

For wireless systems there are a number of ways in which QoS over the air interface may be provided, these approaches have been divided into those that are generic to all wireless systems, those particular to multicarrier systems and those particular to CDMA.

2.4.2.1 Generic QoS provisioning

Some ways, common to most wireless systems, of providing improved air interface QoS over and above that provided by basic receiver algorithms (e.g. modulation, equalisation, channel coding, etc.) are:

- Link Adaptation. This involves adapting some parameter of the wireless system to obtain optimum performance in a given temporal channel condition. Parameters that may be adapted include the choice of modulation scheme, coding rate and transmit power [30, 31].

- Dynamic Channel Allocation. In dynamic channel allocation all resources to be used are placed in a common pool and dynamically assigned to users depending on the prevailing interference and traffic conditions [31].

- Automatic Repeat reQuest (ARQ). Errors in transmission may be coped with by repeating the transmission which is done using some ARQ protocol. The retransmission of data reduces the error rate at the expense of an increase latency and reduction in overall throughput.

- Scheduling. Wireless scheduling algorithms provide a means to consider the state of the current interference and traffic conditions to allocate bandwidth and provide multiplexing at the packet level [32].

2.4.2.2 QoS provisioning particular to multicarrier systems

In multicarrier systems some subcarriers may experience significantly deep fades, making communication on these subcarriers very difficult, conversely some subcarriers may experience large gains and be able to support larger bit rates. The following techniques for QoS provisioning in multicarrier systems take advantage of the different subcarrier gains:

- Bit and power loading: here each subcarrier may be allocated a different power and number of bits [15, 33].

- Subcarrier allocation. Rather than waste bandwidth transmitting data on weak subcarriers, in a multi-user scenario it is much better to take advantage of the multi-user diversity by dynamically allocating subcarriers to users with the aim of giving each user its best subcarriers. Since a subcarrier may not be equally faded for all users, then performance may be improved [34].

There are also algorithms that combine the two above steps, i.e. subcarrier allocation followed by bit and power loading [35, 36].

2.4.2.3 QoS provisioning particular to CDMA systems

In CDMA systems the QoS provisioning centres on managing the interference, some techniques for doing so are:

- Multi-user detection seeks to minimise or eliminate the interference coming from the same cell [37]. As was shown in the simple capacity analysis presented earlier, this will have the effect of increasing the number of users that can access the network with a given QoS.

- Call admission control. If the load is allowed to increase excessively, the QoS of existing users may not be guaranteed. Before admitting a new user, admission control must check whether the admission of that user will cause the interference situation to deteriorate to the extent that the quality of existing calls suffer as a result [38].

- Handoff. Chebaro and Godlewski [19] showed that allocating MSs to the nearest BSs rather than to those to which they experienced the lowest path loss could result in a factor of 4 increase in the interference. As users are mobile, careful design of algorithms that handoff from one BS to another can avoid exacerbation of the interference situation.

2.4.3 QoS provisioning in a layered structure

It is useful to highlight where the air interface QoS provisioning techniques outlined in section 2.4.2 fall in relation to the ISO/OSI layer model. This model describes the different layers in a network operating system and is shown in Figure 2.7. In terms of a wireless communication

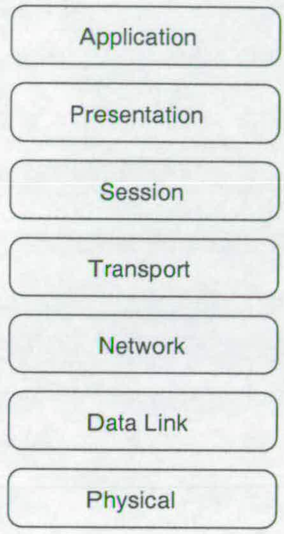


Figure 2.7: *The ISO/OSI layer model*

network, the most important parts of this model are those that deal with how the mobile gets access to the fixed portion of the network. In wireless communications circles these parts of the model are also referred to as the air interface; it is comprised of the physical (PHY), data link and network layers.

The air interface may be redrawn as in Figure 2.8 [39] in which the data link layer has been broken down into two further sub-layers; the link access layer and medium access control layer. The PHY layer performs transmission of data between MS and BS over the wireless channel.

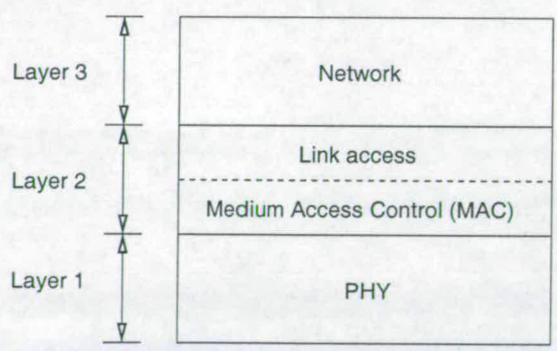


Figure 2.8: *The air interface*

The medium access control (MAC) layer coordinates and controls access to the PHY layer. The link access control performs functions necessary to set up, maintain and release a logical connections. The network layer contains functions to control information flow between nodes

(MSs and BSs) in the network.

The QoS provisioning techniques may now be placed in the air interface structure:

- Bit and power loading, subcarrier allocation and multi-user detection exist in the PHY layer. Typically link adaptation exists in the PHY layer, however in the case of power control it may exist in the MAC layer[1].
- Dynamic channel allocation, ARQ and scheduling exist in the MAC layer[1]. Call admission control may exist in the MAC or the data link layer depending on what type of information it takes in to make its admission decision and the granularity with which an admission decision is made e.g. on a call level, portion of call level etc.

The above illustrates that to properly provide air interface QoS, the different layers must be considered.

2.5 Resource Metric Estimation

From the previous section it is seen that QoS is provided across a variety of layers. This leads to a question as to whether a framework can be defined to provide QoS across layers as the layers in the air interface are not completely uncoupled, e.g. a large residual bit error rate at the PHY layer means more retransmissions at the MAC layer which may lead to congestion at the network layer.

In particular, a framework is desired that allows the exploitation of the interference limited capacity to ensure that the diverse QoS requirements of next generation systems are met. Jorguseski *et al.* [8] broached such a new approach, in the form of the Resource Estimator (RE), which takes in inputs such as interference conditions, current system load, traffic requirements etc., and uses these inputs along with built in capacity models to assist the RRA algorithms. Espinosa *et al.* [40] then outlined the use of a Resource Estimator and Control mechanism to improve resource utilisation by adaptively controlling the real time and non-real time traffic.

[1] It should be noted that some of the radio resource functions attributed here to the MAC layer were originally placed in the network layer in 3G and previous wireless networks. This corresponded to the functionality being carried out by radio network controllers (RNC) i.e. units that control several BSs. In high speed downlink packet access (HSDPA) and next generation networks it has been recognised that the best place for some of this functionality is at the BS to allow for faster, more adaptive resource allocation decisions [38].

However, they focused on pure CDMA systems and did not consider how resources could be managed across layers in a multicarrier CDMA system. Therefore in this thesis the resource estimator is developed into a Resource Metric Estimator (RME) suitable for multicarrier systems. Figure 2.9 depicts the multicarrier RME. Measurements of the interference and radio

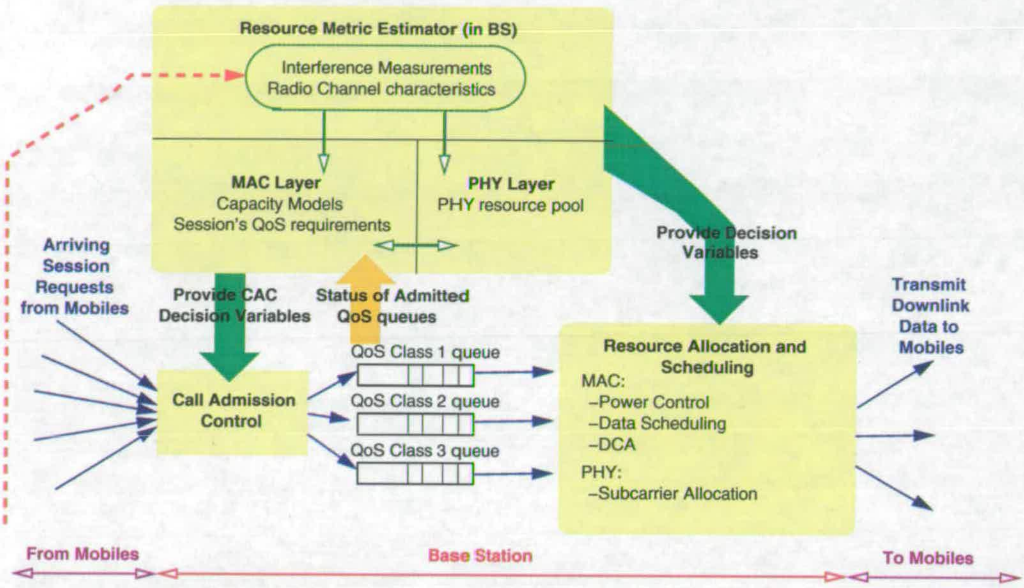


Figure 2.9: *The MC-CDMA resource metric estimator*

channel conditions are taken from MSs, these are passed onto a capacity model that uses them to estimate the capacity in conjunction with information on PHY layer capabilities and the QoS requirements of any admitted flows. The capacity model's information is used to aid in call admission control and in making resource management decisions such as the allocation of powers, scheduling of data, dynamic channel allocation and subcarrier assignment.

2.6 Summary

This chapter contained a brief review of the principles behind the concepts to be explored in the remainder of this thesis. Specifically, the advantages of multicarrier systems over single carrier systems were outlined, cellular CDMA and its interference limited capacity was presented and a description of quality of service and how to provide it given.

Multicarrier systems are preferred over single carrier systems for the ease with which they deal with multipath with relatively smaller implementation complexity. There are distinct advantages to be had by combining multicarrier transmission with CDMA due to the latter's resource efficiency in its ability to reuse the radio resource in every cell. This however requires some thought to be put into how quality of service can be maintained whilst maximising the capacity. Definitions for quality of service were provided and several ways to provide quality of service outlined. It was seen that QoS provisioning must by necessity consider the various layers in the air interface. An approach from the literature termed the Resource Estimator to provide QoS was described, and subsequently an enhanced version to provide QoS in MC-CDMA presented.

The approach taken in the coming chapters will be to first consider QoS provisioning in the different layers, namely the PHY and MAC. Thereafter a description of a multicarrier capacity model and estimation framework will be presented, the final technical chapter will cover the use of the RME with the newly developed QoS provisioning techniques to achieve a cross-layer allocation of resources for better QoS performance.

Chapter 3
PHY-layer Allocation for MC-CDMA

The objective of this thesis is to explore the use of radio resource metric estimation techniques in the management of resource across the PHY and MAC layers. As such, it is necessary to investigate methods of using and allocating resources in the different layers. Therefore, in this chapter the PHY level allocation of resources – primarily focusing on the frequency component – in MC-CDMA systems is investigated. Chapter 4 will deal with the issue of allocating resources at the MAC/system level.

Orthogonal spreading codes can be used in an MC-CDMA system, but in a frequency selective channel, the attenuations in subcarriers destroy the orthogonality of the codes [5]. Hence the capacity of an MC-CDMA system is limited by multiple access interference (MAI). A multi-user detector (MUD) can be used to mitigate the effects of MAI. There are number of MUD algorithms available, of these the best performance is obtained by using a maximum likelihood multi-user detector (ML-MUD) [6]. The disadvantage of the ML-MUD is that its complexity grows exponentially with the number of users. To reduce the aforementioned complexity, a grouped MC-CDMA scheme was proposed by Kaiser [26] whereby the transmitted signals of all the users were sub-divided into a set of smaller groups. Each group was then transmitted on a subset of all the subcarriers[1].

In this chapter a linear programming approach is formulated to solve the grouping and subcarrier allocation problem. In section 3.1, grouped MC-CDMA is described in more detail and the system model presented. Section 3.2 explains how the grouping and subcarrier allocation problem is formulated as a linear program (and that it in fact belongs to the subset of linear programs known as integer linear programs). Two techniques – based on local search and branch and bound – to solve the integer linear program are described in section 3.3. In section 3.4, simulation results are presented showing the performance of the integer linear programming approach with the two solution approaches, these results are compared to those obtained by

[1] A fact that was not mentioned in the paper by Kaiser is that the presence of distinct groups also allows for an additional multi-rate mechanism; allocation of multiple groups to one user. The issue of multi-group allocation will be dealt with in further detail in subsequent chapters.

two approaches from the literature. The penultimate section presents a complexity comparison of these approaches. Conclusions are drawn in the final section.

3.1 Grouped MC-CDMA and System Model

3.1.1 Multi-User MC-CDMA Transceivers

In a frequency selective fading channel, generally all of the subcarriers have different amplitude levels and different phase shifts; degrading the orthogonality of the spreading codes. In a single-user system, this is of little consequence if the system is designed properly. However in a multi-user system, any degradation in the orthogonality of the spreading codes results in multiple access interference (MAI), which is where users interfere with each other.

To mitigate the effects of MAI, a MUD can be used. A MUD uses the channel, timing and code information for multiple users to jointly detect each individual user. As mentioned in the preamble to this chapter, when it comes to the choice of MUD algorithm, an ML-MUD would be ideal in most applications. Unfortunately, because the ML-MUD searches through a set of all the possible transmitted sequences, its complexity grows exponentially with the number of users. For example if the system had 8 users then the ML-MUD would have to search through $2^8 = 256$ sequences, if there were 64 users, then there would be $2^{64} = 1.84\text{x}10^{19}$ sequences to search through! Clearly the use of the ML-MUD in its conventional form is impractical for a large number of users.

3.1.2 Grouped MC-CDMA

To reduce the complexity of the ML-MUD[2], Kaiser [26] suggested a grouped MC-CDMA scheme where the transmitted signals of U users are clustered into G groups, resulting in each group having a maximum capacity of U/G users. The complexity of the ML-MUD is thus reduced from having to search through 2^U sequences to searching through only $G2^{\frac{U}{G}}$ sequences. In his paper, the choice of user grouping and allocation was fixed i.e. the channel/resource information was ignored and the same users always went into the same groups which were then allocated to the same subcarriers.

[2]There are other approaches to multi-user detection, which while being suboptimal, have low complexity and provide good performance relative to the ML-MUD. Examples of these algorithms are the likes of successive and parallel interference cancellation [41–44].

In an alternative approach, Al-Susa and Cruickshank [45] developed an adaptive algorithm that performed grouping and subcarrier allocation algorithm using a combination of ranking and swapping. Henceforth this adaptive algorithm is referred to as the *swapping algorithm*. This algorithm performs the grouping and subcarrier allocation by dividing the procedure into three stages:

1. *Cluster Users:* Users are clustered into groups based on how similar their Channel Impulse Responses (CIR) are.

2. *Ranking:* All the subcarriers for each user are ranked according to their frequency domain powers (the subcarrier with the highest power for a particular user is ranked first).

3. *Allocate Subcarriers to Groups:* Using the ranking information, allocate subcarriers to groups; for each group select the subcarriers that share the highest number of best ranks. In case a subcarrier is selected by more than one group, allocate that subcarrier to the group with smallest power and assign the other group(s) a subcarrier from those subcarriers which have not been selected by any of the other groups. And then, in order to further improve the average received power of the overall weakest user, the weakest users from the different groups are swapped around, if this produces a better overall weakest user, then the new user grouping is kept. The weakest user swapping is terminated when no further improvement is achieved.

Here an adaptive grouped MC-CDMA scheme, for the downlink, is described that takes advantage of grouping to allow the use of an ML-MUD, but performs grouping and allocation adaptively according to some predefined criteria. Figure 3.1 is a diagram of the downlink adaptive Grouped MC-CDMA system with ML-MUD.

At a particular time instance, using criteria and a technique that will be explained later in section 3.2, the grouping and allocation algorithm allocates subcarriers to groups resulting in a subcarrier grouping vector

$$SUBC_g = [subc_{0,g},\ subc_{1,g},\ \ldots,\ subc_{K-1,g}], \qquad g = 0, 1, \ldots, G-1 \tag{3.1}$$

consisting of K components (where K is the length of each group's spreading code). This vector shows which K subcarriers are assigned to the g^{th} group. Thus the total number of carriers used in the system is $N_c = KG$. The individual elements of $SUBC_g$;

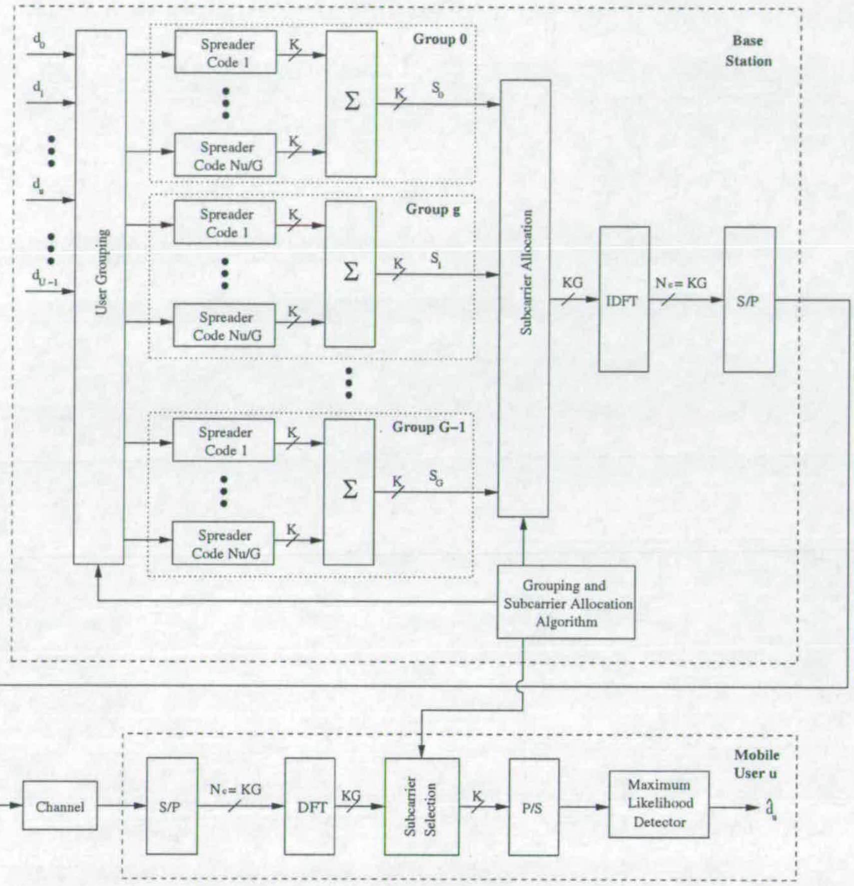


Figure 3.1: *Downlink adaptive grouped MC-CDMA receiver*

$subc_{0,g}$, $subc_{1,g}$, $\ldots$, $subc_{K-1,g}$, are subcarrier indices chosen, by the grouping and allocation algorithm, from the set of all possible subcarriers $subc_{k,g} \subset [0, 1, \ldots, N_c - 1]$. The grouping and allocation algorithm also assigns users to groups, giving a user-to-groups vector

$$GRP_g = \left[grp_{0,g},\ grp_{1,g},\ \ldots,\ grp_{\lceil \frac{U}{G} \rceil - 1,g}\right], \qquad g = 0, 1, \ldots, G-1 \tag{3.2}$$

which is composed of $\lceil \frac{U}{G} \rceil$ components (the notation $\lceil x \rceil$ indicates that x is rounded up to the nearest integer) and whose elements $grp_{0,g}, grp_{1,g}, \ldots, grp_{\lceil \frac{U}{G} \rceil - 1,g}$ correspond to user indices taken from the set of all possible users, $grp_{i,g} \subset [0, 1, \ldots, U-1]$.

Following the grouping and allocation, the data bits d_u of each user u in the g^{th} group – as indicated by GRP_g – are spread by a spreading code vector C_u of length K. The chip streams of all users in the g^{th} group are then added synchronously in the base station to yield the vector

$S_g,\ g = 0, 1, \ldots, G-1$, with

$$S_g = \sum_{u \in GRP_g} d_u C_u = [s_{g,0},\ s_{g,1},\ \ldots,\ s_{g,K-1}] \tag{3.3}$$

consisting of K components resulting in a diversity of K at the receiver. The output of each group is passed to a subcarrier assignment block that simply assigns the elements of S_g to its chosen subcarriers (indicated by $SUBC_g$). After the subcarrier assignment block, the K elements for each group are input to an IDFT for OFDM modulation onto K subcarriers. Thus, the total number of subcarriers used in the system is $N_c = KG$.

The IDFT output, $x(t)$, can then be expressed as

$$x(t) = \sum_{g=0}^{G-1} \sum_{k=0}^{K-1} s_{g,k} \exp(j2\pi f_{subc_{g,k}} t), \qquad t \in [0, T] \tag{3.4}$$

where $f_{subc_{g,k}}$ is an element taken from the gth subcarrier to groups assignment vector $SUBC_g$ and defines the subcarrier frequencies allocated to all the users in the gth group. If T_{os} and T_{cp} are defined as the duration of one basic OFDM symbol and the duration of the cyclic prefix respectively, then the total OFDM symbol duration (i.e. basic OFDM symbol with cyclic prefix attached), T, is $T = T_{os} + T_{cp}$. (The cyclic prefix is inserted to prevent Inter Symbol Interference – ISI – and Inter Carrier Interference – ICI.) The subcarrier frequencies are taken from a set of all the subcarriers allowed in the system

$$f_{subc_{g,k}} \subset [f_0,\ f_1, \ldots,\ f_{KG-1}] \tag{3.5}$$

where f_0 is the lowest subcarrier frequency, f_{KG-1} is the highest subcarrier frequency, and the distance between subcarriers is $1/T_{os}$; therefore $f_{KG-1} = f_0 + (KG-1)/T_{os}$.

It is assumed that the signal is transmitted through a frequency selective Rayleigh fading channel. In addition, the following reasonable assumptions are made: that the system is designed such that the duration of the cyclic prefix is greater than the multi-path spread of the channel and that the channel varies slowly compared to the symbol duration. With the latter two assumptions, the received signal has each of its subcarriers being affected by frequency non-selective, slow Rayleigh fading. It is also assumed that the noise is additive, white and Gaussian

(AWGN). The received signal, in analytical form, is therefore given by

$$r(t) = \sum_{g=0}^{G-1}\sum_{k=0}^{K-1} h_{u,subc_{g,k}} s_{g,k} \exp(j2\pi f_{subc_{g,k}} t) \ + \ n(t), \qquad t \in [0, T] \tag{3.6}$$

In equation (3.6), $h_{u,subc_{g,k}} = \beta_{u,subc_{g,k}} e^{j\phi_{u,subc_{g,k}}}$ accounts for the overall effects of fading ($\beta_{u,subc_{g,k}}$) and phase shift ($e^{j\phi_{u,subc_{g,k}}}$) for the kth subcarrier, of the desired user u that is in the gth group. As the focus is on the downlink, at the desired user all signals from other users using the same BS are received over the same channel.

At the receiver, the signal is spectrally decomposed, using a DFT, into a set of KG subcarriers. For analysis purposes only, it is assumed that the receiver has, via some feedback channel, perfect knowledge of which group it belongs to and which subcarriers are allocated to that group. With this knowledge, the receiver can extract the desired subcarriers from the set of all received subcarriers. Therefore, the received signal, R_g, can be written as a vector of K components with

$$R_g = H_g S_g^T + N_g \tag{3.7}$$

where H_g is a diagonal matrix[3] that describes the flat Rayleigh fading on the subcarriers assigned to group g. S_g^T represents the transposed transmitted sequence and N_g is the noise vector.

As mentioned in subsection 3.1.1, an ML-MUD detector is used in the receiver. Denoting $V_j, j = 1, \ldots, 2^L$ as the set of all possible transmitted sequences, then the estimate of the received signal $\widehat{S}_g = V_j$ if the Euclidean distance e_j^2 between the received and all possible transmitted sequences is minimised. i.e.

$$e_j^2 = min|R_g - H_g V_j^T|^2 \tag{3.8}$$

The estimate of the data bit of the desired user, $\widehat{d}_u$, can be obtained from $\widehat{S}_g$.

[3]Inter-subcarrier interference is caused by the channel characteristics changing during a symbol [46]. The lack of inter-subcarrier interference in equation 3.7 follows from the earlier assumption that the channel varies slowly in relation to the MC-CDMA symbol duration.

3.2 Grouping and Subcarrier Allocation

3.2.1 Criteria for Grouping and Subcarrier Allocation

There are several criteria that may be used to execute the grouping and allocation of subcarriers. Namely these include

1. Global received amplitude.
2. Bit Error Rate (BER).
3. Quality of Service (QoS) requirements.
4. Any combination of the above three.

Using global received amplitude as a criterion means optimising the grouping and allocation such that the received amplitude for all the users in a single cell is as high as possible. The idea behind this criterion is that having a high amplitude on the subcarriers will aid in diversity reception of the different received data chips, correspondingly, a low amplitude will hinder diversity reception of the transmitted chips. The disadvantage of this criterion is that if not explicitly taken into account, those users whose subcarriers are very weak will have little influence on the allocation and may consequently be allocated the poorest subcarriers. An advantage of the global received amplitude as a criterion is that this information is readily available in the form of the channel estimates that are used in detecting received user data at the base station (this is assuming that the system uses time division duplexing - TDD).

With BER as a criterion, the grouping and allocation could be performed to minimise the BER seen by all the users in a cell. The advantage of BER as a criterion is that, unlike global received amplitude, BER is a direct system performance measure. In contrast the relationship between BER (and any other system performance measure such as throughput) and global received amplitude is not direct and depends on other factors such as the noise level, rate of fading etc. The major disadvantage of BER as a criterion is the latency involved in estimating the BER per subcarrier; a large amount of time is required to accumulate enough error events to enable the BER per subcarrier to be estimated with reasonable accuracy. In fact, to the author's knowledge there are as yet no real-time algorithms that provide direct estimates of the BER per subcarrier.

QoS requirements may be used on their own, or in conjunction with the above two criteria in deciding which users get priority in the grouping and allocation procedure. For example, users that are transmitting real-time data (e.g. video) with higher BER/latency requirements would get priority over users transmitting data with lower BER/latency requirements (e.g. file transfer). Similarly, operators could decide to give users who pay a price premium priority over users that opt only for a basic service.

In this chapter, due to the ready availability of estimates, only the global received amplitude is used as a criterion for grouping and allocation. The combination of QoS requirements and global received amplitude is left as an item for subsequent chapters.

3.2.2 Maximisation of Global Received Amplitude

As has been stated in the previous section, the aim in performing the grouping and allocation is to maximise the global received amplitude in a single cell. With reference to equation (3.6), this is equivalent to

$$maximise \sum_{g=0}^{G-1} \sum_{u=0}^{U-1} \sum_{k=0}^{GK-1} \beta_{u,subc_{g,k}} \tag{3.9}$$

Note that in equation (3.9) the bounds of the last two summations have been modified to indicate that in performing the maximisation users are to be selected from the whole set of available users, whilst subcarriers are to be selected from the whole set of available subcarriers. Herein lies the problem, in its present form it is difficult to apply any optimisation technique to equation (3.9) as it does not allow the specification of items like how many users are allowed in a group and how many subcarriers are to be allocated to a group.

To assist in converting equation (3.9) into a more optimisable format, a sharing factor is introduced

$$\rho_{u,g,k} \in [0,1] \tag{3.10}$$

If $\rho_{u,g,k} = 1$, this indicates that the uth user in the gth group is assigned to the kth subcarrier. Conversely, if $\rho_{u,g,k} = 0$, this indicates that the uth user in the gth group is **not** assigned to the kth subcarrier.

If the sharing factor is added to the maximisation problem of equation (3.9), the following modified maximisation problem is obtained

$$maximise \sum_{g=0}^{G-1} \sum_{u=0}^{U-1} \sum_{k=0}^{GK-1} \beta_{u,subc_{g,k}} \rho_{u,g,k} \tag{3.11}$$

If equation (3.11) is to be used as the objective function in performing a maximisation, then examination of this equation reveals that such an objective function will have UG^2K elements. (Recall that U = Total Number of Users in the system, G = Total Number of Groups and K = Number of Subcarriers per Group.) In addition to equation (3.11), constraints will also be needed to

1. Ensure only a certain number of subcarriers are allocated to each group
2. Ensure only a certain number of users can be allocated to each group
3. Restrict the allocation of subcarriers to one group only. Otherwise, it is possible that a potential allocation would seek to assign subcarriers to more than one group, meaning that users in those groups would cease to be orthogonal and hence interfere heavily with each other.

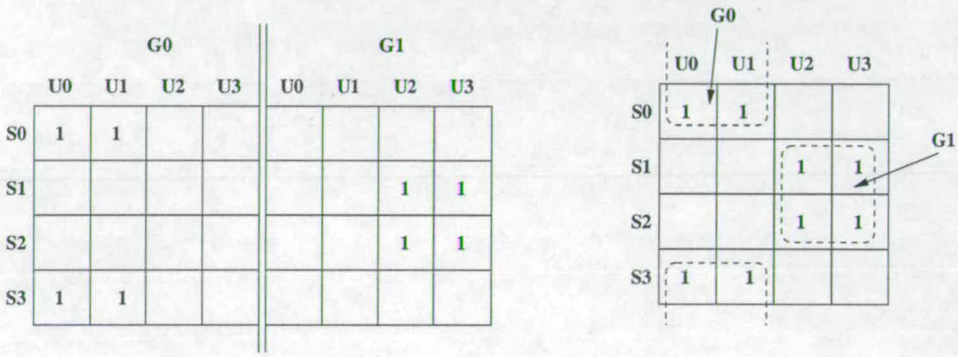
It is possible to reduce the number of elements in the maximisation problem by cognisance of the fact that it is not necessary to explicitly take the groups into account. Instead the problem can be redefined such that if a user shares more than one subcarrier with another user, then it must share all its other subcarriers with that user. In this way the grouping will be inherent in the result i.e. users assigned the same subcarriers must be in the same group. Thus, if the groups information is dropped from equation (3.11) and the new notation $\beta_{u,k}$ is adopted to indicate the amplitude of the kth subcarrier belonging to the uth user, and $\rho_{u,k}$ to indicate the allocation of the kth subcarrier to the uth user, then the following smaller maximisation problem results

$$maximise \sum_{u=0}^{U-1} \sum_{k=0}^{GK-1} \beta_{u,k}\rho_{u,k} \tag{3.12}$$

The objective function from the maximisation problem in equation (3.12) has a reduced number of elements; UGK compared to UG^2K in the original equation (3.11).

An example is introduced to illustrate this principle. In this example a system with $U = 4$, $G = 2$ and $K = 2$ is assumed. Figure 3.2(a) shows a conceptual matrix diagram of the objective function in equation (3.11). The rows represent subcarriers, the columns users, the left half of the matrix represents those users assigned to $Group$ 0 and the right half those users assigned to $Group$ 1. A '1' in any position of the matrix indicates that the user on that column is assigned to the subcarrier on that row. An empty position in the matrix is equivalent to a '0' and indicates no assignment; the actual zeroes are omitted in the interests of clarity. Figure

3.2(a) shows that users $u0$ and $u1$ are grouped into $Group$ 0 and allocated subcarriers $s0$ and $s3$. Correspondingly, users $u2$ and $u3$ are grouped into $Group$ 1 and allocated subcarriers $s1$ and $s2$. If the objective function is now modified to that in equation (3.12), then this is equivalent to merging both halves of the matrix diagram of Figure 3.2(a), to result in the smaller matrix diagram of Figure 3.2(b). As can be seen from Figure 3.2(b), unlike in Figure 3.2(a), the grouping of users is not explicitly defined in the matrix structure, but can still be extracted by noting that those users that share the same subcarriers are in the same group.



(a) Allocation with explicit grouping

(b) Allocation with no explicit groups

Figure 3.2: *An example of the two grouping and allocation methods*

With equation (3.12), the following constraints can now be added to the maximisation problem

$$\sum_{k=0}^{GK-1} \rho_{u,k} = Number\ of\ Subcarriers\ Allowed\ per\ User \qquad \forall\, u \in [0, 1, \ldots, U-1] \tag{3.13}$$

$$\sum_{u=0}^{U-1} \rho_{u,k} \leq Number\ of\ Users\ Allowed\ per\ Group \qquad \forall\, k \in [0, 1, \ldots, GK-1] \tag{3.14}$$

$$\bigvee_{u=0}^{U-1} \rho_{u,k} \oplus \rho_{u,k'} \rightarrow \bigwedge_{u=0}^{U-1} \overline{(\rho_{u,k} \cdot \rho_{u,k'})} \qquad \forall\, k \in [0, 1, \ldots, GK-1] \quad \forall\, k' \in [0, 1, \ldots, GK-1],\ k' \neq k \tag{3.15}$$

Equation (3.13) limits the number of subcarriers allocated to a user, while equation (3.14) limits the total number of users allocated to a group. Equation (3.15) results in a set of logical constraints that serve to ensure that subcarriers can only be allocated to one group, the notation

$$\bigvee_{n=0}^{N} x_n$$

is used to indicate the logical OR-ing of a particular sequence of terms x_n, $\forall\, n \in [0, 1, \ldots, N]$:

$$\bigvee_{n=0}^{N} x_n \equiv x_0 \vee x_1 \vee x_2 \vee \ldots \vee x_N$$

Similarly,

$$\bigwedge_{n=0}^{N} x_n$$

indicates the logical AND-ing of a sequence of terms x_n, $\forall\, n \in [0, 1, \ldots, N]$:

$$\bigwedge_{n=0}^{N} x_n \equiv x_0 \wedge x_1 \wedge x_2 \wedge \ldots \wedge x_N$$

Equation (3.15) takes special advantage of the structure of the grouping and allocation matrix introduced in Figure 3.2(b). Thus, it is perhaps not intuitive to see how this equation limits the allocation of subcarriers to one group. This is explained with the aid of the example in Figure 3.3 which shows an infeasible grouping. To understand why it is infeasible recall that a group is defined as a collection of subcarriers. In the figure there are 2 groups to which a maximum of 2 subcarriers may be allocated. From the grouping shown it can be seen that user $U2$ by virtue of its subcarrier allocation belongs to both group $G0$ and $G1$. The constraint of equation (3.15) should be able to identify such an infeasible allocation. To begin with an examination is made of how the constraint deals with subcarriers $S0$ and $S1$. The first half of the constraint is used to identify whether or not the allocation of two subcarriers is different, and in the case of $S0$ and $S1$ would evaluate to

$$(\rho_{0,0} \oplus \rho_{0,1}) \vee (\rho_{1,0} \oplus \rho_{1,1}) \vee (\rho_{2,0} \oplus \rho_{2,1}) \tag{3.16}$$

The XOR condition is used to test if the two input values to the XOR are different, if they are then the result will be TRUE. It is easy to see that if the required values for $\rho_{u,k}$ (from Figure 3.3) are put in, then the whole expression in equation (3.16) will evaluate as TRUE i.e. some of the elements in the allocation of $S0$ and $S1$ differ. As they differ, the next part of the constraint is to ensure that in the allocation of $S0$ and $S1$, the subcarriers are not allocated to the same users. The NAND condition's two inputs can take on any logic values apart from both being '1', i.e. it makes sure that both subcarriers are not assigned to the same users. In the case of $S0$ and $S1$ this would mean that $\rho_{1,0}$ and $\rho_{1,1}$ cannot both be '1'. The current solution would have to be discarded and another solution found. (A feasible allocation would for example be $U2$ being allocated to $S2$ instead of $S1$.)

To summarise the two parts of the constraint:

> if (a user is assigned to subcarrier *k* and not subcarrier *k'*)
>
> then,
>
> make sure that none of the users assigned to subcarrier *k* are at the same time assigned to subcarrier *k'*

Applying the constraint to the case of subcarriers $S1$ and $S2$, it can be seen that the first part of the constraint will evaluate to being FALSE as all the users assigned to subcarrier $S1$ are also assigned to subcarrier $S2$.

	U0	U1	U2	
S0	$\rho_{0,0}$ 0	$\rho_{1,0}$ 1	$\rho_{2,0}$ 0	G0
S1	$\rho_{0,1}$ 0	$\rho_{1,1}$ 1	$\rho_{2,1}$ 1	
S2	$\rho_{0,2}$ 1	$\rho_{1,2}$ 0	$\rho_{2,2}$ 0	G1
S3	$\rho_{0,3}$ 1	$\rho_{1,3}$ 0	$\rho_{2,3}$ 1	

Figure 3.3: *Example of bad grouping*

The maximisation problem may now be stated as shown in equation (3.17). It is possible to convert the last set of logical constraints of this equation into a linear form. If this is done, then the optimisation problem of equation (3.17) is recognisable as a multi-dimensional (i.e. it is seeking to maximise both user grouping and subcarrier allocation) linear combinatorial optimisation problem that may be solved using integer linear programming techniques. Unfortunately, when the logical constraints are expanded, they result in such a large number of linear constraints that the maximisation becomes too complex to solve in real time.

$$
\begin{aligned}
maximise \quad & \sum_{u=0}^{U-1}\sum_{k=0}^{GK-1} \beta_{u,k}\rho_{u,k} \\
subject\ to: \quad \sum_{k=0}^{GK-1} \rho_{u,k} &= Number\ of\ Subcarriers\ Allowed\ per\ User \\
& \forall\, u \in [0,1,\ldots,U-1] \\
\sum_{u=0}^{U-1} \rho_{u,k} &\leq Number\ of\ Users\ Allowed\ per\ Group \\
& \forall\, k \in [0,1,\ldots,GK-1] \\
\bigvee_{u=0}^{U-1} \rho_{u,k} \oplus \rho_{u,k'} &\rightarrow \bigwedge_{u=0}^{U-1} \overline{(\rho_{u,k} \cdot \rho_{u,k'})} \\
& \forall\, k \in [0,1,\ldots,GK-1] \\
& \forall\, k' \in [0,1,\ldots,GK-1],\ k' \neq k \\
\rho_{u,k} &\in [0,1] \qquad (3.17)
\end{aligned}
$$

3.2.3 Sub-optimal Approach to the Maximisation

To reduce the complexity of the maximisation algorithm, a sub-optimal approach is investigated. In particular, the goal is to change the maximisation problem from being a multi-dimensional one to a one-dimensional problem. This is done by dividing the problem into two steps.

3.2.3.1 Group subcarriers

In the first step, subcarriers are grouped together resulting in the Subcarrier to Groups vector $SUBC_g$ that was introduced in equation (3.1). A simple random assignment of subcarriers to groups is performed.

3.2.3.2 Assign users to groups

With the subcarriers grouped, the second step is to allocate the users to these subcarriers (and hence to groups) by solving the one-dimensional optimisation problem in equation 3.18.

The purpose of the third constraint in equation (3.18) is to ensure that the subcarriers from step 1 of the algorithm are always grouped together. Essentially this constraint says that if a user u is to be assigned a subcarrier that is in a group $SUBC_g$, then that assignment can only proceed if the user is assigned to all the subcarriers in $SUBC_g$. This constraint makes use of slack variables b_{k+uGK} that can only take the integer values 0 or 1. In this form, the optimisation problem in equation (3.18) results in a much smaller number of constraints than that for the optimisation problem in equation (3.17).

$$
\begin{aligned}
& maximise && \sum_{u=0}^{U-1}\sum_{k=0}^{GK-1} \beta_{u,k}\rho_{u,k} \\
subject\ to: & \sum_{k=0}^{GK-1} \rho_{u,k} &=& \ Number\ of\ Subcarriers\ Allowed\ per\ User \\
& && \forall\, u \in [0,1,\ldots,U-1] \\
& \sum_{u=0}^{U-1} \rho_{u,k} &\leq& \ Max.\ Number\ of\ Users\ Allowed\ per\ Group \\
& && \forall\, k \in [0,1,\ldots,GK-1] \\
& \sum_{k\in SUBC_g} \rho_{u,k} &=& \ (1-b_{k+uGK})\ Max.\ Number\ of\ Users\ Allowed\ per\ Group \\
& && \forall\, u \in [0,1,\ldots,U-1] \\
& && \forall\, g \in [0,1,\ldots,G-1] \\
& \rho_{u,k} &\in& \ [0,1] \\
& b_{k+uGK} &\in& \ [0,1]
\end{aligned} \tag{3.18}
$$

The optimisation in equation (3.18) is a linear program. More specifically, because the variables are limited to take on the integer values of 0 or 1, it is of the subset of linear programs known as integer linear programs. Some algorithms to solve this integer linear program are explained in the next section.

3.3 Solving the Integer Linear Program

Before explaining the methods used to solve integer linear programs, it is useful to highlight the difference between a linear and an integer linear program. Linear programming, or equivalently linear optimisation, consists in trying to find the optimal value of a linear function with a certain number of variables, given a set of linear constraints on these variables. Integer linear programming consists of finding the optimal value of a linear function with a certain number of variables, given integer and linear constraints on these variables.

Integer linear programs belong to the class of $\mathcal{NP}$-complete problems [47]. An $\mathcal{NP}$-complete problem is one for which no polynomial-time algorithm exists; that is any algorithm that correctly solves an $\mathcal{NP}$-complete problem will, in the worst case, require an exponential amount of time. This often leads to computation times that are too high to be useful [47, 48]. The practical significance of a problem being $\mathcal{NP}$-complete is that attention can be focused away from developing algorithms that always find an exact solution, to alternative algorithms that give good solutions in a practical time frame. This section focuses on such alternative algorithms.

There are a variety of techniques that may be used to solve integer linear programs, these techniques may be divided into two broad categories of being either general purpose algorithms, or special purpose algorithms.

General purpose algorithms are the likes of [47, 49]

1. total enumeration
2. and, branch and bound.

The main advantage of general purpose algorithms is that they are fairly generic in nature and thus applicable to a wide set of problems with little, or no modification. Their main disadvantage lies in their generality in that for many types of problems they are computationally intensive.

Some examples of special purpose algorithms are [47, 49]

1. dual ascent
2. and, local search.

Special purpose algorithms usually outperform general purpose algorithms in terms of computational time. Their disadvantage is that they are not generic, and in order to make savings in computational time, careful design needs to be done to fit each special purpose algorithm to a particular problem.

One algorithm from each of the aforementioned categories is selected to solve the integer program of equation (3.18). From the category of general purpose algorithms, the branch and bound algorithm is selected. From the category of special purpose algorithms the local search algorithm is selected. The key features of these two algorithms are highlighted in the following sub-sections.

3.3.1 Branch and Bound

3.3.1.1 Introduction to Branch and Bound

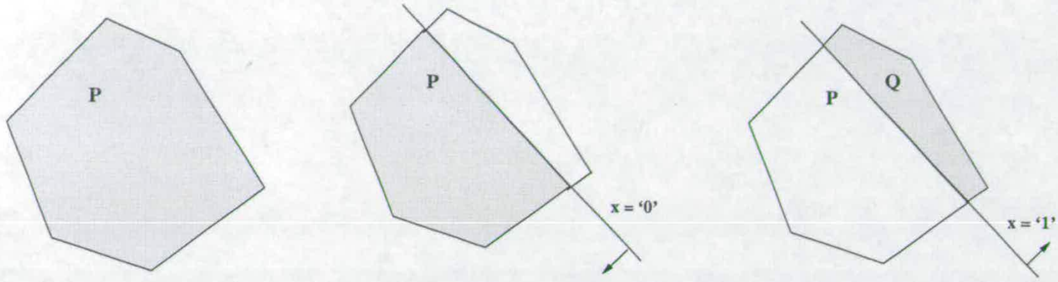
In the integer program of equation (3.18), each variable can take on only one of two values, either a '0' or a '1'. Hence one technique to solve this integer program would be simply to enumerate all these possibilities, working out the objective function for each and then selecting the solution with the highest value. This approach is feasible for problems with a small number of variables, for example if there are 5 variables there needs to be an enumeration of $2^5 = 32$ possibilities, however if the problem has anything over approximately 25 variables then the enumeration technique becomes computationally impractical, for example with 100 variables there are in the region of $2^{100} = 10^{30}$ possibilities.

Branch and Bound is a general purpose linear programming based tree search that systematically enumerates feasible solutions such that the optimal integer solution is found without having to enumerate all the possible solutions [49]. In the branch and bound approach, nodes of a search tree are examined, if at any node the algorithm can show that the optimal solution cannot occur at any of its descendants, then the tree can be pruned at that node.

How does the Branch and Bound technique determine whether or not the optimal solution can exist in the descendants of a node? Starting from the root node, the integrality requirement on all the variables is relaxed and they are allowed to take on values in the range of 0 to 1. With the integrality requirements relaxed, the integer program now becomes a straightforward linear program. The linear program may then be solved using any of the conventional linear programming algorithms such as the simplex algorithm. Solving the linear program will result in a set of values that the variables should have so as to give an optimal solution to the relaxed problem. The values that these variables have will often be non-integer and therefore the branch and bound algorithm now seeks to add tighter bounds to the problem to force those variables with non-integer values to take on integer values. Tighter bounds are achieved by solving the problem with minimal relaxation. (The tightest bound is achieved when the problem is solved with no relaxation at all i.e. all the integer constraints are restored.) To tighten the bound at a node, a variable with a non-integer value is selected and forced to take on integer parts. As the branch and bound algorithm is unable to tell which integer value the variable must have to give an optimal solution, it examines all the possible integer values that variable can take. In the case of an integer program with '0' or '1' values, this will result in the root node branching into two descendant nodes, the left descendant with the constraint that the variable takes a '0' value and the right descendant with the constraint that the variable takes a '1' value. Without any loss of generality from this point onwards the explanation of the Branch and Bound algorithm is restricted to the case where the variables can only take on '0' or '1' values.

Using a linear programming algorithm such as the simplex algorithm [50], the relaxed integer program at each of the two descendant nodes is solved and two solutions obtained (which may, or may not, have variables that are all integers). The descendant node with the lowest solution is pruned i.e. it is not considered any more in the branch and bound process. The rationale behind the pruning of descendant nodes with the smallest solution is as follows. The solution at the parent node had a larger feasible region; by adding integer constraints (of either '0' or '1') at the descendant nodes, the branch and bound algorithm is in effect decreasing the feasible region. The feasible region may be equated to a space in which solutions are searched for. A smaller feasible region means there is less space from which a solution may be obtained. The linear programming algorithm employed to solve the relaxed integer program at each node guarantees to find the optimal solution (if that optimal solution exists) within a certain search space. If the solution at the parent node was optimal in that search space, then making the search space smaller means the resulting optimal value can only – at best – be equal to the value before the search space was reduced. In the worst case, by making the search space smaller, the original solution is excluded leaving behind a solution that has a smaller optimal value than which was obtained before the search space was reduced. This concept is illustrated in Figure 3.4.

The original search space is the grey shaded area shown in Figure 3.4(a), the optimum point in this search space is found – using a linear programming technique – to be located at **P**. If a



(a) Original search space

(b) Search space with constraint x = '0'

(c) Search space with constraint x = '1'

Figure 3.4: *Adding constraints to a search space*

constraint *x='0'*, is added, then this is tantamount to reducing the size of the search space as shown in Figure 3.4(b). Adding *x='0'* however does not affect the best solution; it remains at **P**. If instead of *x='0'* a constraint *x='1'* is added as shown in Figure 3.4(c), then this means that the optimum point **P** would be excluded from the new search space. The optimum point of the reduced search space of Figure 3.4(c) is found to be at **Q**. Looking back to Figure 3.4(a), points **P** and **Q** were both in the original search space, but **P** was found to be the optimum point i.e. **P** was better than **Q**. Therefore, regardless of how the original search space is partitioned (by adding constraints), no point will be better than that found at **P**.

To sum up the above, the solutions (with tightened bounds) obtained from the descendant nodes can never be better than that of the parent node because additional constraints (in the form of integer constraints) are imposed on the problem reducing the problem's feasible region. As such nodes cannot give rise to better integer solutions, there is no need to consider them; they are pruned.

This branching process can be carried out recursively; each of the descendant nodes results in two additional nodes. Eventually after enough bounds are placed on the variables, a solution to the relaxed integer program whose variables are all integers with '0' or '1' values is obtained. The value of the best integer solution found so far is retained. If there are no other nodes to examine then this is the optimal solution to the integer program.

3.3.1.2 Specialised Branch and Bound solution to the Integer Program

Some specialised heuristics are introduced to the generic Branch and Bound technique presented in the preceding section. These specialisations aim to minimise the time required by the technique to converge to a good solution, and are as follows:

1. *Active node selection:* To select which node is selected for processing next, the node with the highest bound is always selected; this search heuristic is also known as a 'best-first' search. Another example of an active node selection rule – that is not used here – is a 'depth-first' search. In a 'depth-first' search one path of the tree is descended as far as possible until a solution is found, this solution is then used as a bound. The algorithm then continues descending the remaining paths and storing their solutions. When all the paths have been evaluated, the path with the best solution is selected.

2. *Branching variable selection:* To choose branching variables, the variable whose value in the solution is furthest from being integral is always selected. The direction of processing is always the left branch (which – as explained in the introduction to this section – is that branch with the value '0').

3. *Branch pruning rule:* A GMC-CDMA feasible solution is defined as one in which subcarriers exist in only one group and all the constraints about users/subcarriers per group are met. Then, as an addition to pruning branches based on their bounds, those branches that result in a GMC-CDMA feasible solution are pruned regardless of whether or not all the slack variables are integers. (For a GMC-CDMA feasible solution it is not necessary for the slack variables, b_{k+uGK}, to be integers so long as the main constraints are met and the main variables, $\rho_{u,k}$, take on integer values).

In Figure 3.5, an example is provided to illustrate the operation of the specialised branch and bound algorithm. At the root of the tree all the integer constraints are relaxed. The resulting linear program is solved, using the simplex algorithm (this is an algorithm that is used to solve linear programs [50]), to give a solution with a value of 71 that is not GMC-CDMA feasible. The first branching variable $\rho_{0,0}$ is selected and the relaxed linear program at the left node, $Node$ 1, is solved (using the simplex algorithm) with the constriction that $\rho_{0,0} = 0$. This results in a non-GMC-CDMA feasible solution with a value of 50. The linear program at the right branch, $Node$ 2, is solved with the constriction that $\rho_{0,0} = 1$, this results in a solution with a value of 60 that is also not GMC-CDMA feasible. Due to $Node$ 2 having a value greater than that of $Node$ 1, the branch $\rho_{0,0} = 0$ is pruned i.e. it is not necessary to examine the descendants of $Node$ 1. The next branching variable, $\rho_{2,1}$, is selected as that which is furthest from being integral. A similar procedure to that outlined for $Node$ 1 and $Node$ 2 is followed resulting in $Node$ 3 that has a non-GMC-CDMA solution with a value of 54 and $Node$ 4 with a GMC-CDMA feasible solution with a value of 55. $Node$ 4 has a higher bound than $Node$ 3, therefore $Node$ 4 is taken as the final solution.

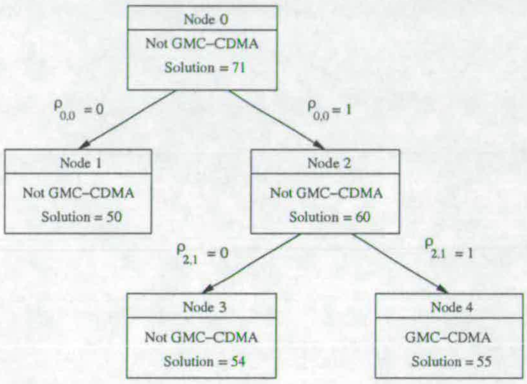


Figure 3.5: *Example of specialised Branch and Bound*

3.3.2 Local Search

3.3.2.1 Introduction to Local Search

Local search is based on what is probably the oldest optimisation technique – that of trial and error. Local search starts from an initial solution and iteratively tries to replace the current solution, s, with a better solution from an appropriately defined neighbourhood, $\mathcal{N}(s)$ of the current solution [47, 48]. So long as an improved solution exists, it is adopted and the neighbourhood search is repeated from the new solution. When a local optimum is reached, the algorithm stops.

The general local search algorithm is as shown in algorithm 1.

Algorithm 1 The general local search algorithm

$s \leftarrow$ Generate_Initial_Solution()
while (improvement still occurs) **do**
  $s \leftarrow$ Improve($\mathcal{N}(s)$)
end while

To apply this algorithm to a problem a number of choices have to be made. Firstly, how is an initial decision made? Secondly, what neighbourhood is to be used? A small neighbourhood might be quick to search, but this may come at the expense of providing poorer local optima than if a larger neighbourhood were used. Thirdly, how is the neighbourhood to be explored? One option is to take the first solution that results in an improvement, another is to exhaustively search the neighbourhood and take the solution with the best value. Very little theory is available as a guide, so these choices are usually made on the basis of empirical evidence [47]. The performance of local search on combinatorial optimisation problems depends highly on these choices and can be quite poor due to the tendency of the algorithm to get stuck in local optima of the neighbourhood space [48]. Modifications of the basic local search therefore primarily focus on the addition of mechanisms, such as annealing, indexing etc., that allow the search algorithm to escape from local optima. Because the stopping condition is then more complex than reaching a local optima, an additional termination condition is usually required.

Within the wireless communications arena, local search algorithms and their derivatives have mainly been applied to the channel assignment problem in radio network planning [51–54], they have also been applied to perform access point placement for wireless LANs [55] and to perform dynamic channel allocation for satellite systems [56].

3.3.2.2 Mixed Probabilistic-Greedy Local Search

In this section, a local search algorithm is presented that uses a mixed probabilistic-greedy scheme to navigate the neighbourhood space and hence avoid getting trapped in a local optimum. The objective function of this algorithm is as highlighted in equation 3.18, that is to maximise the total subcarrier amplitude subject to constraints on how many users can be in one group. Greedy algorithms have found applicability as solution approaches for a wide range of problems e.g. multi-user detection [57], contention-resolution protocols [58], image processing [59] etc. With regards to their use in multicarrier systems, Kivanc *et al.* [36] proposed a greedy descent algorithm for an orthogonal frequency division multiple access system (i.e. OFDMA - a combination of OFDM and FDMA) that used each user's rate requirements and average channel gains to decide the number of subcarriers to assign to a user. Kim *et al.* [60] proposed a greedy algorithm for the OFDMA uplink that used marginal rate functions to allocate subcarriers.

The neighbourhood space for the mixed probabilistic-greedy local search algorithm is defined as follows, firstly a group $\hat{g}$ is selected at random, the neighbourhood space is then that set of users not allocated to this group. Using the notation of the user-to-groups vector from equation (3.2), the neighbourhood space can be expressed as

$$\begin{aligned} \mathcal{N}(s) &= GRP_g \\ & \quad \forall \ \{g \in [0, 1, \ldots, G-1] \mid g \neq \hat{g}\} \end{aligned} \tag{3.19}$$

This is potentially a large neighbourhood space, but in a real wireless system the actual number of users that are active at any one time, for any one base station, is normally in the order of tens of users.

The structure for the mixed probabilistic-greedy algorithm is shown as algorithm 2; the same notation that was used in earlier portions of this chapter is adopted. As the algorithm functions by exchanging users between groups, the concept of a source group and a target group is applied. The target group, $\hat{g}$, is the group on which the algorithm attempts to perform a neighbourhood search at any one point. The source group, $\tilde{g}$, is a group from the neighbourhood space of the target group and is the group from which the algorithm has decided a (source) user, $\tilde{u}$, should be drawn from and placed into the target group in place of (or in addition to) the target user, $\hat{u}$.

The algorithm begins with an initialisation phase whereby the subcarrier-to-groups and user-to-groups vectors, $SUBC_g$ and GRP_g respectively, are randomly initialised. This initialisation is done in such a manner so as to ensure that the constraints of equation (3.2) are observed to give a feasible initial solution. The total allocated subcarrier amplitude, β_{max}, is then calculated with this initial solution.

The target group is then selected randomly from the set of all groups. With a target group the neighbourhood space, $\mathcal{N}(s)$, can then be explored. Two strategies are used to explore the neighbourhood space, the first is an exhaustive greedy search that finds the user whose subcarriers have the largest total amplitude in the current group. The second exploration strategy is one where the user is selected from the neighbourhood space randomly. This random strategy copes with local minima by introducing perturbations in the neighbourhood space. The decision as to which of the two exploration strategies is used is decided probabilistically. The probability of the exhaustive greedy search being used is given by P_{best} and that of the random perturbation is $1 - P_{best}$.

Algorithm 2 The mixed probabilistic-greedy local search algorithm

% Initialisation
$SUBC_g \Leftarrow$ Random() *% Generate random subcarrier to groups allocation*
$GRP_g \Leftarrow$ Random() *% Perform initial random user to groups allocation*
$\beta_{max} \leftarrow$ CalculateTotalAmplitude($SUBC_g$, GRP_g, β)
$\hat{g} \leftarrow$ PickRandom($[0, 1, \ldots, G-1]$)
iter_no_improv $\leftarrow 0$

while (iter_no_improv < max_iter_no_improv) **do**
  % Explore the neighbourhood space
  if (UniformRandomNo.() $\leq P_{best}$) **then**
    for ($\mathcal{N}(s)$) **do**
      % Best exhaustive search (greedy)
      $\tilde{u}$, $\tilde{g} \Leftarrow$ IdentifyUserWithBestAmplitude($\hat{g}$)
    end for
  else
    % Random selection
    $\tilde{u}$, $\tilde{g} \Leftarrow$ PickRandom($\mathcal{N}(s)$)
  end if

  % Update the new solution and recalculate the objective function
  $\hat{u} \leftarrow$ PickRandom($GRP_{\hat{g}}$) *% Pick user from target group*
  UpdateUserGroupingVector($GRP_{g,new}$, $\hat{u}$, $\hat{g}$, $\tilde{u}$, $\tilde{g}$)
  $\beta_{new} \leftarrow$ CalculateTotalAmplitude($SUBC_g$, $GRP_{g,new}$, β)

  % Accept or reject the new solution, update iteration counter correspondingly
  if ($\beta_{new} > \beta_{max}$) **then**
    $GRP_g \leftarrow GRP_{g,new}$
    iter_no_improv $\leftarrow 0$
    $\beta_{max} \leftarrow \beta_{new}$
  else
    iter_no_improv $\leftarrow$ iter_no_improv + 1
  end if
end while

After a source user and group have been selected, the algorithm then updates the subcarrier-to-groups and user-to-groups allocation vectors, in effect temporarily exchanging the target user in the target group with the source user in the source group. The total new allocated subcarrier amplitude, β_{new}, for this allocation is calculated and if it is found to be higher than the previous value, β_{max}, then the new solution is accepted. Acceptance of a new solution means that a counter of the number of iterations without improvement is reset. Rejection of a solution implies this counter is incremented. The algorithm is deemed to have reached a good solution when the number of iterations without an improvement exceeds some predefined threshold.

From the above description there are two thresholds that must be set, the probability of a greedy search being used and the maximum number of iterations without an improvement. The choice of what value to use for these thresholds is a trade-off between computational time and the quality of the resulting solution. A large number of iterations means that a better solution may be found, but this is obviously at the expense of computational time. With regards to the P_{best} threshold, a high value may mean that only poor solutions are found in a given time as the algorithm may get trapped in a local optima. As mentioned in section 3.3.2.1, there is very little theory on how to select such parameters, therefore an empirical approach is followed whereby the computational time and resulting solutions for a number of channel instances are evaluated. The combination of the two thresholds is chosen such that a good compromise – between computational time and resultant solution – is achieved. For the channel types used in section 3.4 the values so chosen were $P_{best} = 0.25$ with a maximum number of iterations without a solution of 200.

3.4 Simulation Parameters and Results

3.4.1 Simulation Parameters

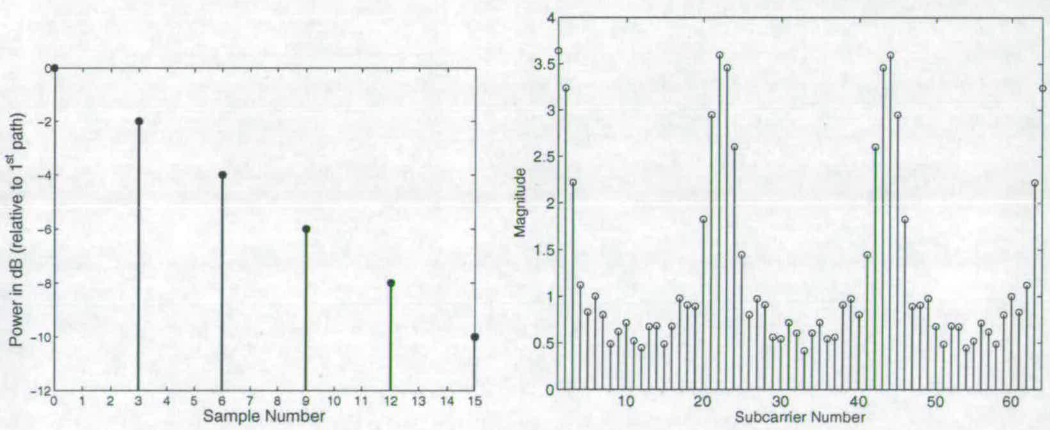
Table 3.1 outlines the parameters used in the simulation.

The two channel types used in the simulation are as shown in Figures 3.6 and 3.7. The two channels are modified versions (to take into account the smaller number of subcarriers used in the simulations) of the exponential models used by Maeda et al [61]. Channel 1 is a six tap exponential channel with a maximum delay spread of 15 samples and an RMS delay spread of 4.2903 samples. If a channel bandwidth of 20 MHz is used, this corresponds to a maximum delay spread of 0.75 μs and an RMS delay spread of 0.215 μs. Channel 2 is a twelve tap

Parameter	Value
Total Number of Users, U	64
Number of Groups, G	8
Number of Subcarriers per Group, K	8
Total Number of Subcarriers	64
Channel Models	6 and 12-tap exponential
Spreading Code Type	Walsh-Hadamard
Modulation Type	BPSK

Table 3.1: *Simulation Parameters*

exponential channel with a maximum delay spread of 11 samples and an RMS delay spread of 3.2944 samples. Again, with a channel bandwidth of 20 MHz, this corresponds to a maximum delay spread of 0.55 μs and an RMS delay spread of 0.165 μs. Such delay spreads are typical of microcells.

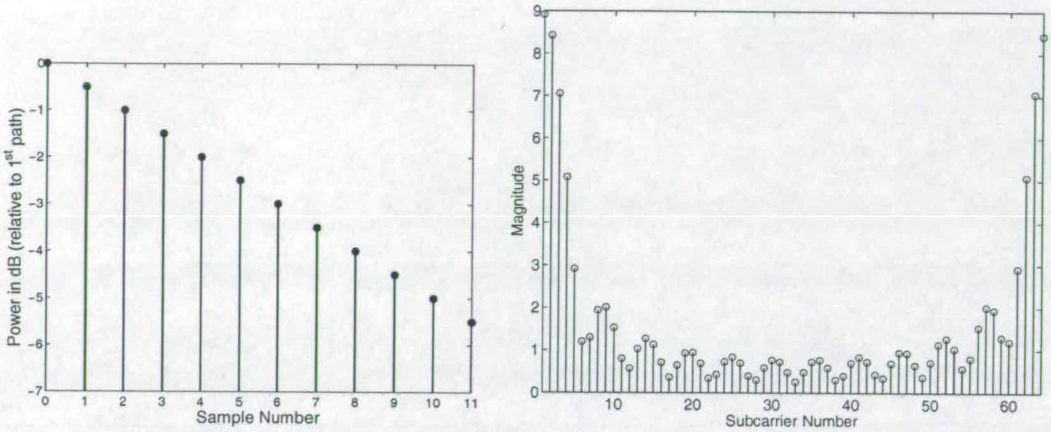


(a) Channel impulse response

(b) Non-faded channel frequency response

Figure 3.6: *Channel type 1*

It is assumed that the length of the cyclic prefix used in the system is long enough such that there is no inter-symbol interference (with the channel types simulated this condition is met with a cyclic prefix length of 16). This assumption allows the whole simulation to be performed in the frequency domain, thus saving simulation time.



(a) Channel impulse response

(b) Non-faded channel frequency response

Figure 3.7: *Channel type 2*

3.4.2 Simulated policies

Four grouping/allocation policies are simulated. Firstly, a fixed allocation as described in the preamble to this chapter.

The second grouping/allocation policy simulated is the swapping algorithm of [45] in which they aim to maximise global received power. The operation of this algorithm was explained in section 3.1.2.

The third policy simulated is the proposed linear programming algorithm with subcarriers assigned randomly to groups and the linear program solved by using the specialised Branch and Bound algorithm of section 3.3.1. This policy is referred to as the Branch and Bound algorithm (LP-BandB).

The final policy is the proposed linear programming algorithm with subcarriers assigned randomly to groups and the linear program solved by using the mixed probabilistic-greedy local search algorithm of section 3.3.2. This policy is referred to as the Local Search algorithm (LP-LSearch).

3.4.3 Effect of algorithms on subcarrier amplitudes

In this section an investigation is made of the effect of the different grouping/allocation policies on the user's subcarrier amplitudes. This investigation is performed for the two channels of section 3.4.1 and is done by executing the grouping and allocation algorithms on multiple realisations of the channel in question (500 realisations are used throughout this section).

3.4.3.1 Channel 1

Figure 3.8 shows a plot of the CDF of the subcarrier amplitudes. It can be seen that the grouping/allocation algorithm that results in the largest mean subcarrier amplitude is the Local Search algorithm, followed by the Branch and Bound algorithm, with the Swapping and Fixed allocation algorithms coming last with very similar means. The actual values are as shown in Table 3.2 below.

Grouping/Allocation algorithm	**Mean Subcarrier Amplitude**
Fixed Allocation	1.4111
Swapping Algorithm	1.4322
Proposed LP, Branch and Bound	1.5310
Proposed LP, Local Search	1.6816

Table 3.2: *Average subcarrier amplitudes for the different algorithms. Channel 1.*

The poor performance of the fixed allocation with respect to the other algorithms is hardly surprising as it makes no attempt to optimise the assignment of users to subcarriers.

What is surprising though is the relatively small gain from using the Swapping Algorithm. This discrepancy is due to an underlying unfairness inherent to the Swapping algorithm. As was explained in section 3.1.2, the first step of the Swapping algorithm is to cluster users into groups depending upon how similar their channel impulse responses are. Similar channel impulse responses translate into similar frequency domain transfer functions, thus this clustering is equivalent to users being grouped according to how similar their frequency domain transfer functions are. This means that the diversity in a group will be reduced with users sharing the same set of good subcarriers and equally the same set of bad subcarriers. This on its own does not necessarily have to lead to a poor allocation, however in the allocation of subcarriers to groups, the algorithm arbitrates between those users that select the same subcarrier by assigning such a subcarrier to the group with the lowest power. The problem then comes in that the other groups that selected this subcarrier are assigned subcarriers from the set of subcarriers that are not selected by other groups. By virtue of these subcarriers not being selected by other groups, they are likely to be of poor quality, furthermore, due to the lack of diversity in the groups, such subcarriers are likely to be poor for all users in a group. The algorithm then tries to improve the performance of the weakest users by swapping such users amongst the groups, however this is likely to result in little improvement for such users, and the system as a whole, as the basic allocation is so poor that no benefit can be gained.

Of the two linear programming based algorithms, the superior performance of the Local Search algorithm over the Branch and Bound is most likely due to the nature of the search heuristics introduced in the Branch and Bound search procedure. As it may be recalled, part of the purpose of these heuristics was to reduce the size of the search space to result in a quicker algorithm. However, in doing so the search heuristics may inadvertently cause the exclusion of better solutions with the result that the Branch and Bound algorithm finds local optima of the reduced search space. The Local Search algorithm avoids this with the use of randomised jumps to effectively increase the size of the search space.

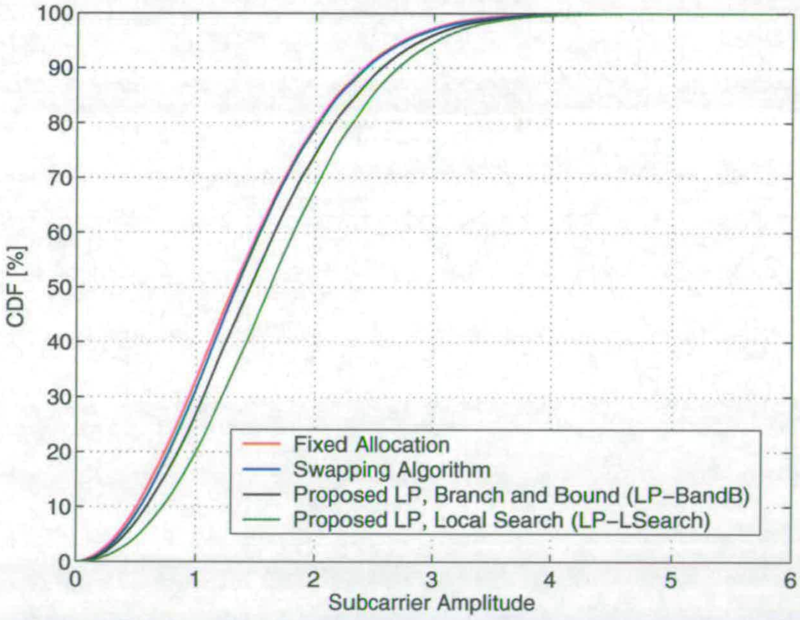


Figure 3.8: *CDF of user's allocated subcarrier amplitudes. Channel 1.*

To get a further insight into the effects of the allocation algorithms on the subcarrier amplitudes, an examination is made of how the different grouping/allocation algorithms affect the allocation of low value subcarriers. This is shown in Figure 3.9 in the form of a plot of the CDF of the the number of poor subcarriers per user. In this context a poor subcarrier is defined as one with an amplitude of less than 1. For example the value '0' on the x-axis indicates the percentage of users that had no poor subcarriers i.e. those with an amplitude of less than 1. It can be seen that the Local Search algorithm performs best in this regard, with the Branch and Bound, Swapping and Fixed allocation algorithms following. In addition, whilst the Swapping algorithm has a smaller percentage of users with 2 or fewer poor subcarriers than the Fixed allocation, this situation is reversed when a higher number of subcarriers is considered; the Swapping algorithm in fact has a larger percentage of users with 3 or more poor subcarriers, reinforcing the earlier notion that the Swapping algorithm results in users being allocated poor subcarriers.

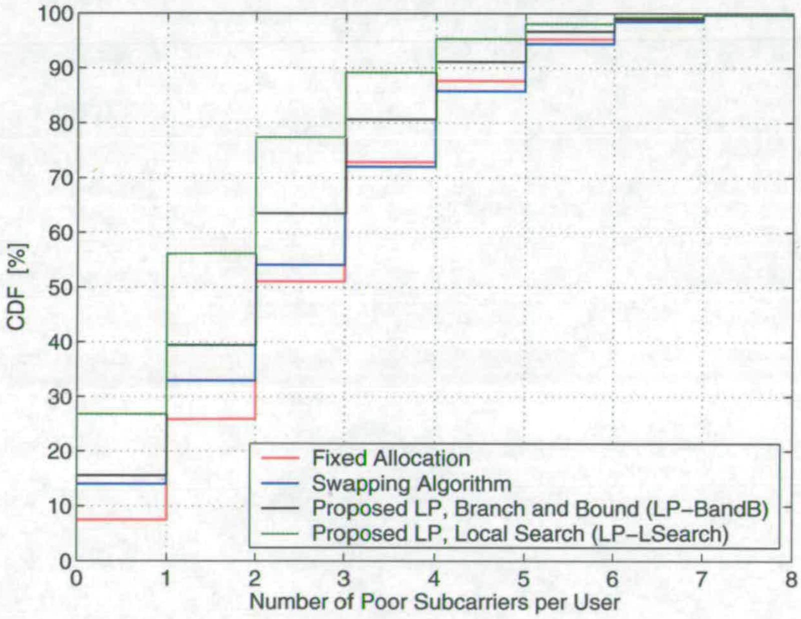


Figure 3.9: *CDF of the number of allocated subcarriers per user that have an amplitude of less than 1. Channel 1.*

3.4.3.2 Channel 2

Figures 3.10 and 3.11 show the CDF of the subcarrier amplitudes and the number of users with poor subcarriers, respectively, for channel 2. The mean of the subcarrier amplitudes is shown in Table 3.3. Similar to the data obtained for channel 1, the grouping and allocation algorithm with the highest subcarrier amplitude is the Local Search algorithm, with the Branch and Bound, Swapping and Fixed allocation algorithms following in that order. In terms of the number of poor subcarriers per user, the same order is preserved, with the Local Search algorithm allocating fewer users to poor subcarriers than any of the other algorithms.

Grouping/Allocation algorithm	Mean Subcarrier Amplitude
Fixed Allocation	2.3203
Swapping Algorithm	2.4225
Proposed LP, Branch and Bound	2.6181
Proposed LP, Local Search	2.8480

Table 3.3: *Average subcarrier amplitudes for the different algorithms. Channel 2.*

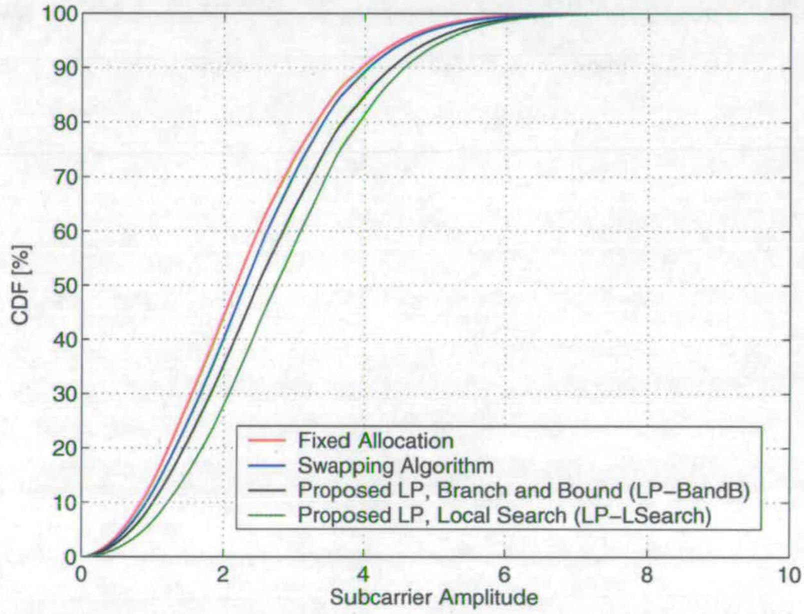


Figure 3.10: *CDF of user's allocated subcarrier amplitudes. Channel 2.*

3.4.4 QoS (BER) results

The focus is then shifted to the quality of service (QoS) – in terms of BER – experienced by users in the system. The majority of users will experience good QoS if the system has a low average user BER coupled with a low variation between individual user's BERs. The variation of BERs in the system is measured by calculating the standard deviation of user's BERs.

In the simulation, a static channel is used and the BER for one such channel realisation obtained. The final BER results are those after averaging the BER results from 100 channel realisations.

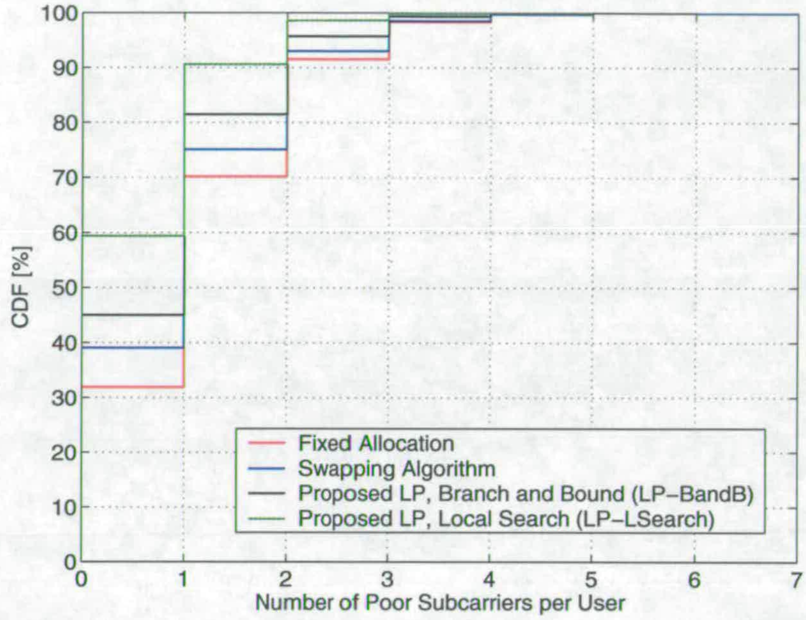


Figure 3.11: *CDF of the number of allocated subcarriers per user that have an amplitude of less than 1. Channel 2.*

3.4.4.1 Channel 1

Figure 3.12 shows the average user BER for the four grouping allocation policies when passed through channel 1. Also shown in this figure is the average user BER experienced if the only degradation is the addition of white Gaussian noise. As expected from the results of the grouping algorithms on the subcarrier amplitudes presented in section 3.4.3, the Local Search algorithm outperforms the other three algorithms, bringing the average user BER very close to that experienced in an AWGN only channel for an SNR of 0 to 8dB. The gain of the Local Search algorithm over the other grouping/allocation algorithms is tabulated in Table 3.4. As these results show, the performance of the Swapping algorithm is such that it is in fact better not to use any allocation algorithm than use the Swapping algorithm.

It must be noted that the performance of the Swapping algorithm relative to the Fixed allocation algorithm is at variance with what was presented in the original paper on the Swapping algorithm [45]. In particular, that paper showed that the Swapping algorithm performed better than the Fixed allocation algorithm. However, as the results in Figure 3.12 show, the Swapping algorithm does not in fact perform better than the Fixed allocation algorithm. Apart from differences in the channel used in the original paper [45], this is most likely due to the fact that the average BER was obtained by averaging over a small number of realisations of the static channel, making the performance of the Swapping algorithm seem more optimistic as compared to when the results are averaged across a much larger number of realisations.

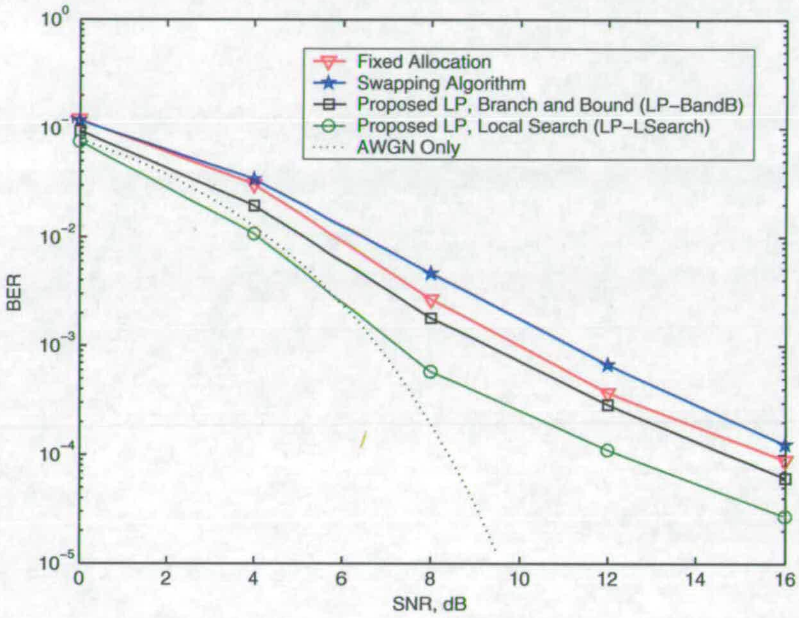


Figure 3.12: *Average user BER for various grouping/allocation policies (Channel 1)*

Grouping/Allocation algorithm	Relative gain of Local Search
Proposed LP, Branch and Bound	0.75dB
Fixed Allocation	1.75 dB
Swapping Algorithm	2.25 dB

Table 3.4: *Gain of the Local Search algorithm over the other grouping and allocation algorithms (taken at an SNR of 4 dB)*

Figure 3.13 shows the logarithm of the standard deviation of user's BERs. Once again the Local Search algorithm performs better than the other algorithms with the Branch and Bound algorithm's performance being at par with the fixed allocation, while the Swapping algorithm performs poorly in comparison to the rest.

3.4.4.2 Channel 2

Figures 3.14 and 3.15 show the average user BERs and the standard deviation of user's BERs respectively for channel 2. With this channel, the Local Search algorithm remains superior to the other algorithms, with the Branch and Bound coming in second and the Swapping and

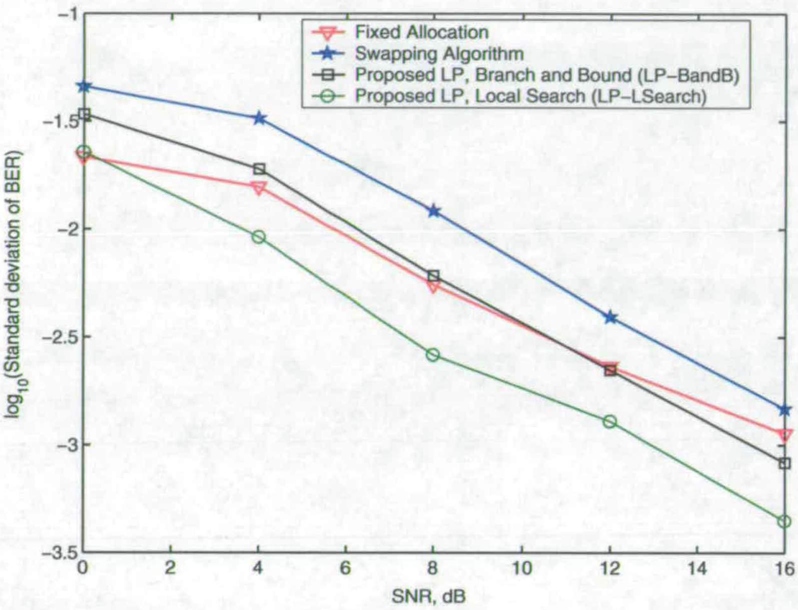


Figure 3.13: *Logarithm of the standard deviation of user's BER for various grouping/allocation policies (Channel 1)*

Fixed Allocation jointly last. The performance of the Swapping algorithm relative to the Fixed Allocation is slightly improved with the Swapping algorithm showing less of a performance loss than that observed with channel 1. However, it is still better to use no adaptive allocation than to use the Swapping algorithm.

3.5 Complexity Comparison

A superior algorithm is one that gives a performance improvement with little to no additional computational complexity over other comparable algorithms. An analysis of the complexity of the different grouping and allocation algorithms presented in this chapter is difficult due to the nature of the stopping conditions used by these algorithms. All the algorithms presented in this chapter (with the exception of the Fixed allocation which for all intents and purposes has zero complexity as it has no decisions to make) have no precisely defined stopping point. The algorithms essentially choose instead to stop computing when they decide no further improvement is possible/justified over the currently obtained solution.

Therefore an empirical approach to obtain the relative complexity of the algorithms is warranted. This is done by measuring the time it takes the algorithms to solve multiple, different

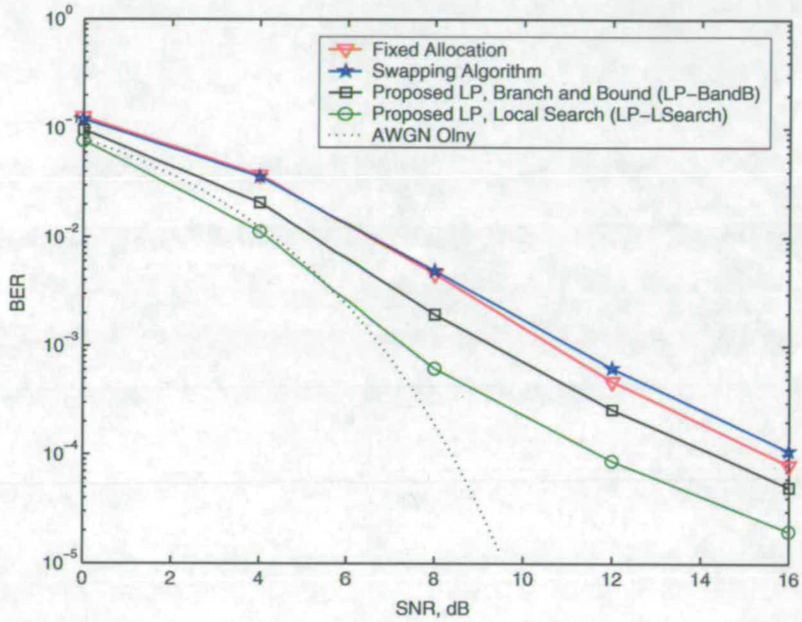


Figure 3.14: *Average user BER for various grouping/allocation policies (Channel 2)*

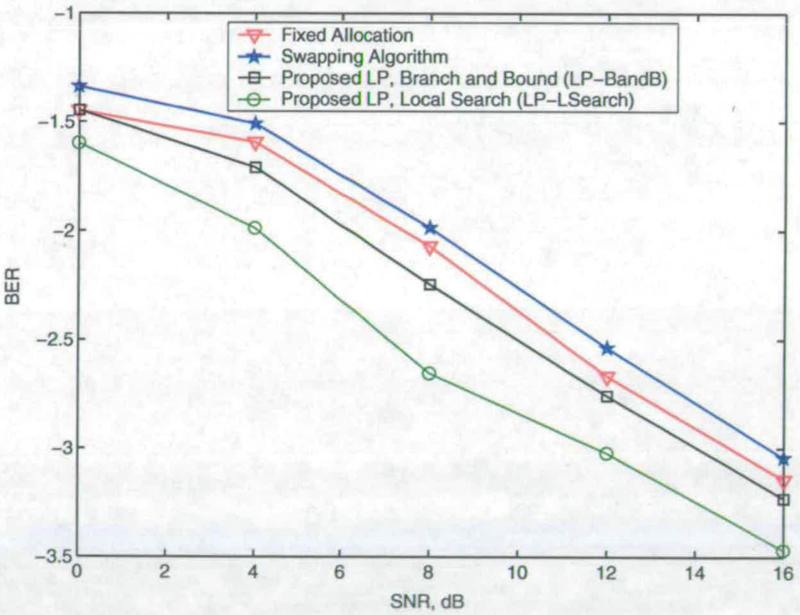


Figure 3.15: *Logarithm of the standard deviation of user's BER for various grouping/allocation policies (Channel 2)*

channel realisations and comparing the average solution times. Because the actual timing was done on a multi-tasking computer, the timing metric used was not the real world time (or wall clock time) taken to solve each channel instance, instead the number of CPU clock ticks is used to obtain a meaningful measure of the actual time spent by the CPU in performing each algorithm.

Table 3.5 shows the time in CPU clock ticks taken to perform each algorithm averaged over 1000 channel realisations. For the operating system and CPU on which the measurements were performed, there are 100 clock ticks per second, equivalently each clock tick corresponds to 10 ms. This figure is used to convert the CPU clock ticks into an appropriate number of seconds rounded to 2 significant figures.

Grouping/Allocation algorithm	**Average CPU Clock Ticks**	**Time in Seconds**
Swapping Algorithm	1.19	0.012
Proposed LP, Branch and Bound	18277.6	182.78
Proposed LP, Local Search	0.25	0.0025

Table 3.5: *Comparison of the algorithm complexity measured in CPU clock ticks*

The important thing to note from these results is not the absolute time taken to perform each algorithm, but that the Local Search algorithm is relatively less complex than either the Branch and Bound or Swapping algorithms. Despite the Local Search and Branch and Bound algorithms being based on the same foundations, there is a vast complexity reduction with the Local Search algorithm over the Branch and Bound algorithm. This may seem startling at first, however it may be explained by noting the differences in the way the two algorithms navigate the search space. In particular the Local Search algorithm has its search space constrained to only consist of the set of feasible solutions whereas the Branch and Bound algorithm's search space includes feasible as well as infeasible solutions. Furthermore, at each search node of the Branch and Bound algorithm it must perform a computationally expensive simplex algorithm to solve the relaxed linear program at the node.

3.6 Conclusion

This chapter investigated PHY layer grouping and subcarrier allocation methods for a grouped MC-CDMA system. The best solution to perform a multi-dimensional grouping and subcarrier allocation was found to be prohibitively complex. A new, sub-optimal approach was suggested whereby the subcarriers are first grouped and then users are assigned to groups by solving an integer linear program.

The best algorithm to solve the integer linear program was found to be a new mixed probabilistic-greedy local search algorithm. This algorithm when used resulted in larger mean subcarrier amplitudes than any of the other algorithms examined including a more computationally complex branch and bound based algorithm. The mixed probabilistic-greedy local search algorithm also meant fewer users were allocated poor subcarriers than with any other algorithm.

The improvement in subcarrier amplitudes obtained by the Local Search algorithm was translated into improved QoS performance, in terms of BER, for more users in the MC-CDMA downlink. Furthermore, this improvement was gained with a corresponding reduction in computational complexity as compared to the other algorithms examined.

Chapter 4
MAC-Layer Allocation for MC-CDMA Systems

One of the visions for the next generation of wireless systems is that they will have to carry a diverse mix of data services. A basic feature of such heterogeneous services is the requirement to support different bit rates. The support of different bit rates requires a multi-rate allocation model; in 3G systems, such as UMTS, multi-rate allocation is achieved by pooling multiple codes, multiple slots, or even, multiple codes and slots [16]. Unlike conventional CDMA systems, MC-CDMA systems – by virtue of their frequency component – allow for an additional degree of freedom in their resource management. To take advantage of this degree of freedom in the frequency domain, and utilising the grouped MC-CDMA structure of chapter 3, in this chapter an additional allocation model – termed multi-group allocation – is proposed. The resulting system is described as Multi-Group Multi-Carrier CDMA (MG-MC-CDMA). Whilst chapter 3 examined how allocations may be performed in the grouped MC-CDMA system when considering only PHY layer parameters, in this chapter the focus is extended to the allocation of MAC layer resources. In particular, an examination is made into how, in MG-MC-CDMA, power and groups may be allocated at the MAC layer to take into account of MS's slow fading and QoS requirements in the form of multiple data rates.

The transmitter power that an MS uses affects that MS's own link quality and also affects the total capacity of the wireless system. A terminal with low SIR may choose to increase its power, however any such power increase will come at a cost of increased interference to other MSs in the system which may then resort to increasing their own powers resulting in a power race [31]. Thus the choice of what transmission power to use is non-trivial. In section 4.1.1 of this chapter several power control variants are presented and their performance when applied to the MG-MC-CDMA system assessed. Three schemes for multi-group allocation are presented in section 4.1.2.

When it comes to the choice of how many groups to assign to an individual MS, that choice primarily depends on the MS's traffic type. The focus is placed on two traffic types, namely a real-time service in the form of voice over Internet protocol (VoIP) traffic and a non real-time interactive service in the form of world-wide web (WWW) data. VoIP traffic has a fixed data rate, therefore the number of groups to assign is straightforward, whereas WWW data is often bursty with time varying amounts of data to be transmitted, thus the number of groups to allocate and how MSs are scheduled to those groups becomes an interesting issue. In section 4.2 a technique is described, for scheduling WWW data, that is based on generalised processor sharing and takes advantage of the structure of the MG-MC-CDMA system.

The simulation model used to assess the various radio resource allocation and management algorithms is presented in section 4.3. Simulation results showing the performance of the power control and multi-group allocation algorithms are shown in section 4.4. Results for the scheduling algorithm are shown in section 4.5. Finally, conclusions are presented in section 4.6.

4.1 Power Control and Multi-Group Allocation

4.1.1 Power Control

The power control models implemented are based on Distributed Constrained Power Control (DCPC [31]); DCPC aims to control the power to meet an SIR target. The three variants implemented are described in the next sub-sections.

4.1.1.1 DCPC with per MS power constraints for each group

Here the maximum power each MS is allowed to have per group is limited to $p_{u,g,max}$, the minimum power is $p_{u,g,min}$. The output powers for the u-th MS' g-th group, $p_{u,g}(n)$, are thus updated according to

$$p_{u,g}(n+1) = \min\left\{p_{u,g,max},\ \max\left(p_{u,g,min},\ \gamma^t.\frac{p_{u,g}(n)}{\gamma_{u,g}(n)}\right)\right\} \tag{4.1}$$

where $\gamma_{u,g}(n)$ is the current $\frac{E_b}{I_0}$ and γ^t is the target $\frac{E_b}{I_0}$ for the u-th MS' traffic type.

4.1.1.2 DCPC with per group total power constraints

The second DCPC variant does not limit the output power each MS is allowed to have. Instead, the output power is limited per group, so that each MS in a group can have any power as long as the total power allocated to that group, $p_{tot,g}(n)$, does not exceed the maximum per group power $p_{g,max}$, and does not go below the minimum per group power $p_{g,min}$:

$$
\begin{aligned}
p_{u,g}(n+1) &= \min\left\{\bar{p},\ \max\left(p^*,\ \gamma^t.\frac{p_{u,g}(n)}{\gamma_{u,g}(n)}\right)\right\} \\
\bar{p} &= p_{g,max} - p_{tot,g}(n) + p_{u,g}(n) \\
p^* &= p_{g,min} - p_{tot,g}(n) + p_{u,g}(n)
\end{aligned}
\tag{4.2}
$$

4.1.1.3 Hybrid DCPC with per MS and per group total power constraints

The third variant is a hybrid of the previous two. As well as limiting the total power per group, the power each MS may have is also limited. Thus each MS is only allowed powers that do not exceed its per MS maximum, $p_{u,g,max}$, with the proviso that the total power allocated to a group does not exceed the per group maximum, $p_{g,max}$, or fall below the per group minimum, $p_{g,min}$:

$$
\begin{aligned}
p_{u,g}(n+1) &= \min\left\{\hat{p},\ \max\left(p^*,\ \gamma^t.\frac{p_{u,g}(n)}{\gamma_{u,g}(n)}\right)\right\} \\
p^* &= p_{g,min} - p_{tot,g}(n) + p_{u,g}(n) \\
\hat{p} &= \min\left(p_{u,g,max,hybrid},\ \bar{p}\right) \\
\bar{p} &= p_{g,max} - p_{tot,g}(n) + p_{u,g}(n)
\end{aligned}
\tag{4.3}
$$

4.1.2 Multi-group Allocation

Three multi-group allocation algorithms are studied in this chapter. They are described as follows.

4.1.2.1 Random multi-group allocation

In the random multi-group allocation algorithm, when a traffic session arrives for an MS, the groups to which it is to be allocated are chosen randomly from the set of all groups.

4.1.2.2 Least Resource multi-group allocation

In the least resource multi-group allocation algorithm, each BS keeps a list of the groups ordered in terms of the amount of resources assigned to each group. MSs are allocated to the group(s) with the least allocated resource drawn from this list. This allocation algorithm is illustrated in Figure 4.1 which shows the allocation of 2 groups in a system with 8 groups.

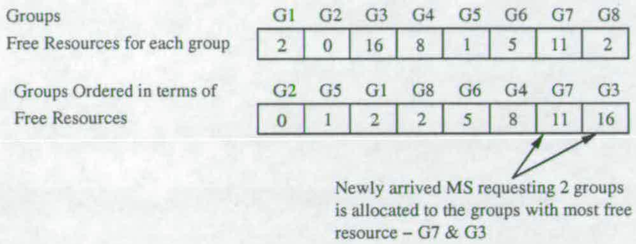


Figure 4.1: *Example of Least Resource multi-group allocation*

4.1.2.3 Least Interference multi-group allocation

In this allocation algorithm, at the start of a transmission the BS requests a measurement report at which point the MS measures the interference on the groups and communicates this to the BS. The MS is then allocated to the group(s) with the least amount of interference.

4.2 Data Scheduling

In a CDMA system, the radio resources are shared by all services and thus a decision on how and when these services access the resources must be made. This decision is governed by the requirements of the traffic class to which an MS's data belongs. In this thesis, for reasons of simplicity, only two traffic classes are considered; real-time conversational traffic and non-real time interactive traffic. (Further information on these traffic types can be found in section 4.3.2.) Real-time traffic typically has stringent latency requirements, but with low average bit rates, whilst non-real time interactive traffic has relatively lenient latency requirements with high peak and average bit rates. Due to the low latency requirements of real-time data, packets from this traffic class are given the highest priority, are not queued and are instead guaranteed to be transmitted on a first come first served basis. Resources for real-time traffic data packets are thus effectively reserved. The remainder of the radio resource is then left for utilisation by the non-real time interactive traffic. The amount of non-real time data to be transmitted using the remaining radio resource is decided by a packet scheduler that decides the transmission rate for each data connection based on factors such as the amount of data for transmission, the channel conditions, amount of radio resources available etc.

A scheduling discipline that is efficient in utilising radio resources and fair in scheduling users is the generalised processor sharing (GPS) discipline [62]; in its packet based emulation this is also known as weighted fair queueing (WFQ) [32, 63]. GPS uses the resources efficiently as it facilitates statistical multiplexing and it is fair as when resources become available it re-distributes those resources to active sessions. Furthermore, GPS guarantees that every session will receive its share of the network resources irrespective of the behaviour of other session and therefore offers perfect isolation [64].

GPS-based scheduling disciplines function by allocating each user a positive weight, the amount of traffic a user receives is then proportional to this weight. If a session u is assigned a weight ϕ_u and continuously has packets to transmit in a time interval $(\tau, t]$ then the amount of service $R_u(\tau, t)$ the session receives in that interval is

$$\frac{R_u(\tau, t)}{R_m(\tau, t)} \geq \frac{\phi_u}{\phi_m}$$

for all sessions m that have also received some service in this time interval [62].

There have been a number of implementations of GPS-based schedulers in wireless systems. Initially the focus of such schedulers was on the implementation of GPS based on a time-scheduling approach [65, 66]. This approach is suitable for use in hybrid TDMA/CDMA systems as shown by Arad and Leon-Garcia [67]. However in general CDMA systems the key issue is the variation in the interference in the network, schedulers for general CDMA systems that do not use time division multiplexing and account for this have also been proposed [64, 68–70].

4.2.1 Multi-Group Generalised Processor Sharing

In this section a simple multi-group generalised processor sharing (MG-GPS) scheduler is presented that adjusts the service rate that each MS flow receives by taking advantage of the multi-group MC-CDMA system to vary the number of groups allocated to an MS flow as well as the number of multiple codes assigned to such flows. The proposed MG-GPS scheduler schedules non-real time traffic flows in the following manner:

Firstly, the BS, using information from its non-real time traffic queues, obtains the amount of backlogged traffic, B_u, that is to be served for MS u ($u = 1, 2, \ldots, U$), during the current time slot.

Secondly, using the value B_u, the expected amount of service R_u which is to be received by the u-th MS is calculated:

- If $B_u = 0$, then $R_u = 0$
- If $B_u > 0$, then $R_u = r_u R_{res}$

  where R_{res} is the amount of data that can be transmitted by one group/code resource unit during a time slot of duration T and r_u is the resource allocated to MS u which is given by
  $$r_u = min\left(\left\lfloor \frac{\phi_u}{\sum_{i=1}^{M}\phi_i}\sum_{g=1}^{G} C_g \right\rfloor , \left\lfloor \frac{B_u}{R_{res}} \right\rfloor\right) \tag{4.4}$$
  in which the term C_g denotes the amount of free capacity – or, equivalently, free resource – that is available in the g-th group. (The term to the left of the comma is actually the minimum guaranteed resource for the u-th MS if that MS has enough backlogged data to utilise this amount of resource.)

- Each MS has its amount of backlogged data reduced by the total number of bits it is expected to transmit in the current scheduling interval (this is given by multiplying the number of resource units allocated to each MS by R_{res}). If the amount of total allocated resources, $\sum_{u=1}^{U} r_u$, is less than that of the maximum usable resources, $\sum_{g=1}^{G} C_g$, then the remaining resources are distributed to MSs that still have backlogged data to transmit. This distribution is done by repeating the MG-GPS procedure for those MSs requesting more resources.

4.2.2 Comparison schedulers

To evaluate the performance of the MG-GPS scheduler it is compared with the following two schedulers:

Fixed Resource - Round Robin (FR-RR) This scheduler assigns a fixed and predetermined amount of resource in each time slot to MSs that have data to transmit. The MSs to be scheduled for transmission are chosen in a random round robin fashion.

Fair Resource (FaR) The fair resource scheduler simply divides the resources amongst MSs that have data to transmit. It is akin to the MG-GPS scheduler with weights of 1, but does not consider how much backlogged data an MS has when deciding how much resource to allocate. Thus this scheduler is not efficient as resources that go unused by an MS with low amounts of backlogged data are not utilised by other MSs that desire to have more than their fair share of resources.

4.3 System-Level Simulation Model

4.3.1 Introduction to the Simulator

The pseudo-code for the simulator is as shown in algorithm 3.

Algorithm 3 The system-level simulation

```
for (Simulation Iterations) do
  Initialisation
  while (Simulation Time < Max Time) do
    for (All Mobiles) do
      Mobility Update
      Traffic Update
      Wireless Channel
      BS Ownership
      Multi-Group Allocation/Scheduling
      for (All groups assigned to this mobile) do
        Power Allocation
        SIR Calculation
      end for
    end for
    Identify blocked calls and calls to be dropped
    Identify successfully transmitted packets
    Increment Simulation Time
  end while
  Performance measures for 1 iteration
end for
Performance measures averaged across multiple iterations
```

The main thing to note from this pseudo-code is that there is an iteration that performs per group power control and SIR calculation.

Key features of the simulator are dealt with, in detail, in the following sub-sections.

4.3.1.1 Wireless Channel

The slow fading is modelled as a Gaussian random variable with an adjustable standard deviation. Fast, frequency selective fading is not considered at this stage.

In a cellular system, co-channel interference from neighbouring cells has a significant effect on the performance of a sample cell. To reduce the computing burden, it is normal to simulate only a finite, small set of cells that are taken to be representative of the cells in a real system. Taking a small set of cells means that cells at the boundary of the simulation area experience lower co-channel interference than cells at the centre of the simulation area as the boundary cells are not surrounded by cells on all sides. Taking results from such boundary cells would result in an over optimistic evaluation of the system's performance [31].

There are two approaches to eliminate such effects. The first – the 'sea of cells' technique – is to create a small set of core cells from which results are drawn and then surround these cells with more cells from which no results are drawn. The boundary effects are eliminated as each of the core cells is surrounded by other cells. The disadvantage of this approach is that even though no results are drawn from the surrounding cells, the same effort must be put into simulating the surrounding cells as the core cells. The result is that the simulation burden is very heavy. The second approach – the wraparound technique – has less of a simulation burden and involves making each one of the small set of core cells appear to be in the middle of the simulation area. This is done by making each boundary cell the neighbour of the boundary cell located directly opposite in the cell layout [71] and calculating the path gains in a corresponding manner. Conceptually this approach can be thought of as placing the cells into a parallelogram that is stitched together to create a torus shaped surface. A torus has no borders and thus all cells will have neighbours on all sides and there will be an infinity of straight paths between any two points of the torus [31]. To eliminate boundary effects in the system-level simulations in this thesis, the wraparound technique is used.

The path loss model used is:

$$L = 136.51 + 37.6\log_{10}(d) \tag{4.5}$$

where d is the MS-BS distance in km. This is based on the path loss model outlined in the ETSI selection procedures for UMTS [72], with a carrier frequency of 5 GHz.

The correlation of the shadow fading between the successive positions of a moving MS is implemented as defined by Gudmundson [73]. In this paper he characterises the normalised autocorrelation function, with distance, of the shadow fading as an exponential function

$$R(\Delta x) = \exp\left(\frac{\Delta x}{d_{corr}} \ln 2\right) \tag{4.6}$$

where Δx is the separation between two points and d_{corr} is the environment dependent decorrelation length.

4.3.1.2 SIR Calculation

Assuming that the u-th MS is attached to BS0, the downlink $\frac{E_b}{I_0}$ for this MS's g-th group is as given in equation (4.7). The numerator of this equation is the spreading gain multiplied by the received signal power (which is calculated as the power allocated to a group multiplied by the path loss - the assumption is made that the path loss is approximately constant across all groups). The denominator of the equation characterises the interference plus noise; the first term in the denominator is the inter-cell interference received in a user's group from other BSs. Despite orthogonal codes being used by the BS, the presence of multipath propagation may degrade the orthogonality of these codes. The second term is the intra-cell interference resulting from this degradation in orthogonality in the BS's own codes, with the orthogonality factor reflecting the degree of degradation. The final term is the noise power experienced in each group.

$$\left(\frac{E_b}{I_0}\right)_{u,g} = \frac{\frac{W_{grp}}{R_{t_m}} p_{0,u,g} L_{0,u}}{\sum_{b=1}^{B} P_{b,g} L_{b,u} + \alpha_{0,u,g} P_{0,g} L_{0,u} + N W_{grp}} \tag{4.7}$$

where

B	$=$	total number of BSs in the system
W_{grp}	$=$	spreading bandwidth for one group
R_{t_m}	$=$	data rate for the u-th MS' service type t_m
$p_{0,u,g}$	$=$	power transmitted by BS0 to the u-th MS's g-th group
$P_{b,g}$	$=$	total power transmitted by the b-th BS to the g-th group
$L_{b,u}$	$=$	path loss from the b-th BS to the u-th MS

$\alpha_{0,u,g} = 1 - \hat{\alpha}_{0,u,g}$

$\hat{\alpha}_{0,u,g}$ = downlink orthogonality factor for the u-th MS communicating on the g-th group to BS0, where
'0' = Completely non-orthogonal
'1' = Perfectly orthogonal

N = noise power density

4.3.1.3 Performance measures

The performance measures utilised depend on the traffic type. Because real-time services are guaranteed resources, the performance measure for this traffic class is the probability of calls being blocked from accessing the system and, once they are given access to the system, the probability that they are dropped from the system due to having poor quality resources i.e. the outage probability. Non-real time traffic sessions are generally always allowed into the system, thus the blocking probability is an inappropriate measure. In addition, in the case of an outage event, the user is not dropped, instead the data is re-transmitted, making the probability of outage an inappropriate measure of performance. More appropriate measures of performance for the non-real time traffic class are the packet delay and packet call throughput. All these measures are explained in further detail below.

Blocking and Outage Probability The blocking and outage probabilities are calculated from the number of new real-time traffic calls arriving, the admitted real-time traffic calls in the system and the real-time traffic calls in outage at the end of each simulation iteration. The number of blocked calls is the number of calls that were refused entry to the system during the traffic update. The number of calls in outage is based on the group outage probability. Whether or not an individual MS's group is in outage is decided based on that group's calculated $\frac{E_b}{I_0}$. If the $\frac{E_b}{I_0}$ is below the threshold for the service, then that group is deemed to be in outage.

As one MS can be communicating over multiple groups, the group outage needs to be related to the MS outage. The relationship depends on the MS's traffic type; if the traffic type is such that it is allowed to have a flexible data rate (e.g. instead of a fixed data rate it has a maximum and a minimum guaranteed data rate) then the MS goes into outage when the number of groups meeting the $\frac{E_b}{I_0}$ threshold drops below the required minimum number of groups. For example, if each group can transmit 12.5 kbps and the MS's traffic type requires a minimum data rate of

50 kbps and a maximum data rate of 100 kbps (corresponding to an allocation of 4 and 8 groups respectively). Such an MS will only go into outage if more than 4 of its groups do not meet the $\frac{E_b}{I_0}$ threshold. Conversely, if an MS's traffic type is such that it requires a fixed data rate, then that MS will be in outage if even one of its groups is unable to meet the $\frac{E_b}{I_0}$ threshold.

The actual $\frac{E_b}{I_0}$ threshold used will depend on the traffic type and the results of link level simulations.

Packet delay and throughput Two measures of performance for non-real time interactive traffic are used, these are the packet delay and packet call throughput (the definition of a packet call is explained in section 4.3.2.2). The packet delay is defined as the time a packet exists in the BS's queue i.e. the time taken from the moment it arrives in the queue to the moment it leaves the queue. A packet may only leave the queue if it has had all of its bits successfully transmitted, or optionally, if its time in the queue exceeds some threshold.

The MAC layer can only transmit a certain number of bits at each scheduling interval, these bits are stored in a MAC transport block. A packet, depending on its size, may be divided into one or more MAC transport blocks. A MAC transport block may also contain multiple packets of the same traffic class destined for the same MS. If the received E_b/I_0 of a MAC transport block equals, or exceeds, the E_b/I_0 threshold for that traffic class, then that transport block's bits are deemed to be received error free (thus too are the corresponding bits of the packet(s) being carried in the transport block). Should the transmission of a transport block be unsuccessful, then it is assumed that, via some ARQ protocol, the MS is able to inform the BS of the unsuccessful transmission and as a result all the transport block's bits are slotted for re-transmission in the next scheduling interval. In some ways this model for bit errors in a transport block is pessimistic as it behaves like a step function; all bits are received error free above a threshold and in error below that threshold. The situation in practice does not follow such a step function with the number of bit errors increasing as the E_b/I_0 decreases. However, it should be noted that without knowledge of what was transmitted, the only way a receiver can check for errors in a transport block is via cyclic redundancy checks (CRC). Part of the transport block is typically reserved for a CRC value calculated by the transmitter. The receiver calculates the CRC value of the data it has received and compares it with that sent by the transmitter. If the two values differ, then the transport block is erroneous and it is discarded. Consequently it is of no importance whether one bit, or an intermediate number of the transport block's bits were received in error; the end result is the same, that transport block is in error and all its bits must be re-transmitted. Therefore, the step function model while being inaccurate at the bit level, is sufficient to characterise the error behaviour at the transport block level.

The size of a MAC transport block in this thesis is defined to be the bit rate of one code/group combination multiplied by the simulation time interval. For example, if one code/group combination has a bit rate of 10.2 kbit/s and the simulation time step is 0.1s, then the size of a MAC transport block would be 1020 bits.

The packet call throughput is defined as the number of bits in a packet call, B_{pc}, divided by the time taken to successfully transmit that packet call's bits, T_{pc}:

$$\text{Packet call throughput} = \frac{B_{pc}}{T_{pc}} \tag{4.8}$$

T_{pc} is the duration between the first packet of a packet call arriving in the BS's queue, to the last packet of that packet call leaving the queue.

It should be noted from above that the performance of individual MSs is not explicitly measured.

4.3.1.4 Mobility Model

The mobility model is a pseudo random mobility model with semi-directed trajectories as used in the ETSI selection procedures for UMTS [72]. MSs are uniformly distributed across the cells and their direction is randomly chosen at initialisation.

The MS's positions are updated according to the decorrelation length of the long-term fading. Their direction can be changed at each position update according to a given probability, the maximum change in direction is limited.

An MS's speed is initialised at the beginning of each new traffic session; its speed is fixed for the duration of the traffic session. The speed is a random number drawn from a truncated (i.e. no negative speeds) normal distribution.

Table 4.1 summarises the parameters for the mobility model.

The first three parameters in the above Table are as defined in [72]. The last two parameters – the mean and standard deviation of the normal distribution used to generate the MS speed – are

Parameter	Value
Decorrelation length of long-term fading	20m
Probability of changing direction at position update	0.2
Maximum swing angle	45 degrees
Mean of normal distribution used to generate MS speed	5 km/h
Standard deviation of normal distribution used to generate MS speed	10 km/h

Table 4.1: *Mobility model parameters*

chosen to give speed values that are suitable for micro-cells. A histogram of the MS speed with these parameters is shown in Figure 4.2.

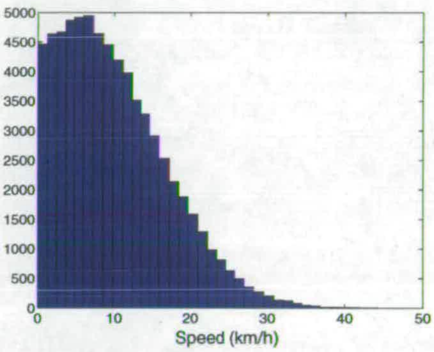


Figure 4.2: *Example histogram of MS speed*

4.3.2 Traffic Models

As mentioned earlier, only two traffic classes are considered, namely real-time conversational traffic and non-real time interactive traffic. The models used for these two traffic classes are as follows.

4.3.2.1 Real-Time Conversational Traffic

Under the real-time conversational traffic class the example is taken of Voice over IP (VoIP). The session arrival process is modelled by a Poisson process. A session lasts for a mean duration, D_{VoIP}, that is exponentially distributed. The previous parameters are the same as those for conventional circuit switched voice traffic.

Within a session, voice activity detection is implemented and no packets are transmitted in the silence periods. The durations of the talk and silence periods are exponentially distributed with mean ON duration $D_{\text{on,VoIP}}$ and mean OFF duration $D_{\text{off,VoIP}}$.

In each ON period, packets are generated by a voice codec. It is assumed that the ITU G.729A codec [74] is used. The codec output bit rate is 8 kbit/s. These output bits are produced as frames at 10ms intervals (or multiples thereof) forming the voice payload. After the codec, Internet Protocol (IP), User Datagram Protocol (UDP) and compressed Real-time Transport Protocol (cRTP) headers are added to the voice payload to make up the VoIP packet.

Table 4.2 lists the VoIP model's numerical parameters.

Parameter	Value
Mean length of a VoIP session, D_{VoIP}	120s
Mean ON Time, $D_{\text{on,VoIP}}$	1s
Mean OFF Time, $D_{\text{off,VoIP}}$	1.35s
Codec Bit rate	8 kbit/s
Codec voice payload interval	30 ms
Codec voice payload size	30 bytes
Number of packets per second	33.3
Size of IP/UDP/cRTP headers	32 bytes
VoIP packet size	62 bytes
Required data rate	16.53 kbit/s
Required BER	10^{-6}

Table 4.2: *VoIP parameters*

4.3.2.2 Non Real-Time Interactive Traffic

In this traffic class web browsing data is considered. The session arrival process is modelled by a Poisson process. The web traffic model is based on that used in the ETSI selection procedures [72], whereby a web browsing session is divided into packet calls in which a number of packets arrive. This is illustrated in Figure 4.3.

The number of sessions arriving per second is altered to give different loads. The number of packet calls per session, N_{pc}, is a geometrically distributed random variable. The reading time between two consecutive packet calls, D_{pc} is a geometrically distributed random variable

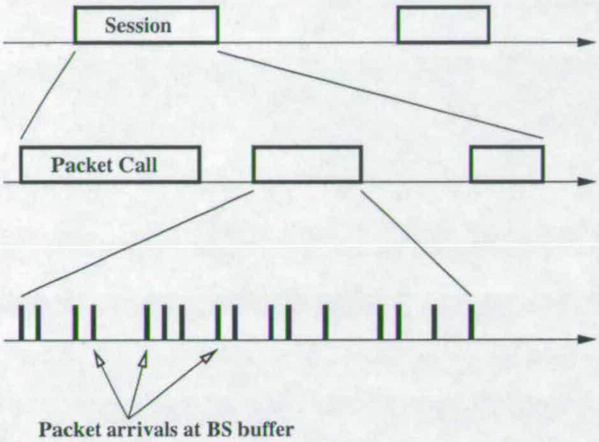


Figure 4.3: *WWW data model*

with mean $\mu_{\mathrm{D_{pc}}}$. The reading time starts when the last packet of the packet call is completely received; the reading time ends when the user makes a request for the next packet call. The number of packets in a packet call, N_d, is a geometrically distributed random variable with a mean $\mu_{\mathrm{N_d}}$.

The time interval between two consecutive packets inside a packet call, D_{d}, is a geometrically distributed random variable with mean $\mu_{\mathrm{D_d}}$. The packet size, B_{pkt}, follows a Pareto distribution with shape parameter α, tail parameter k, maximum allowed size m, minimum allowed size n and an average value μ_{n}.

Table 4.3 summarises the values these parameters are assigned. These are exactly the same as that used in the ETSI model, except for the reading time between packet calls and the mean time interval between packet calls. The reading time is changed to 5 seconds to facilitate shorter simulation times, while the mean time interval between packets in a packet call is altered to give a peak bit rate of 10 Mbps.

4.3.3 Simulation Parameters

Table 4.4 shows the general simulation parameters. The cell radius is chosen to be suitable for a microcell. The path loss coefficient is from the path loss model outlined in section 4.3.1.1. The standard deviation of lognormal shadowing is selected as a value suitable for the urban environment [17]. The number of BSs is set to either 7 or 19 corresponding to either 2 or 3 cell tiers respectively. 7 cells are used when the basic simulation burden is so large as to make the simulation of 19 cells prohibitive. The maximum power for a BS is selected as

Parameter	**Value**
Mean number of packet calls per session, N_{pc}	5
Mean number of packets in a packet call, N_d	25
Mean reading time between packet calls, D_{pc}	5s
Mean time between packets in a packet call, D_{d}	400 μs
Packet size; Pareto shape parameter, α	1.1
Packet size; Pareto tail parameter, k	81.5 bytes
Packet size; maximum size, m	66,666 bytes
Packet size; minimum size, n	81.5 bytes
Average packet size	480 bytes

Table 4.3: *WWW data parameters*

3W (34.77 dBm) and is a little above the 2W (33 dBm) maximum power recommended by the 3rd Generation Partnership Project (3GPP) for microcell BSs [75]. The system bandwidth and number of subcarriers are as chosen for a typical next generation multicarrier system by researchers at NTT DoCoMo in Japan [7]. The thermal noise density is calculated as kT where k is Boltzmann's constant and an ambient temperature, T, of 290 Kelvin is assumed [17].

The bit rate per group per user per time slot is calculated using the given number of subcarriers and system bandwidth in Table 4.4, along with the number of data symbols per slot as in Table 5.2 of section 5.4.1. Further, QPSK modulation is assumed with an error control coding rate of $2/3$. The bit rate is thus

$$\text{Bit rate} = \frac{\text{No. of Data symbols per group}}{\text{No. of Time Slots} \times \text{Time slot length}} \times \text{No. of bits per symbol} \times (1 - \text{Coding rate})$$

Where the time slot length is calculated as

$$\text{Total no. of MC-CDMA symbols} \times \frac{\text{No. of subcarriers} + \text{Cyclic prefix length}}{\text{System bandwidth}}$$

Putting in the appropriate values

$$\text{Time slot length} = 85 \times \frac{512 + 128}{80\text{MHz}} = 680\ \mu s$$

and

$$\text{Bit rate} = \frac{78}{15 \times 680\ \mu s} \times 2 \times \frac{2}{3} = 10.196\ \text{kbps}$$

Parameter	**Value**
Cell radius	300m
Maximum probability of outage	0.01
Path loss coefficient	3.76
Standard deviation of log-normal shadowing	8 dB
Number of cells	7 or 19
Max. total BS power	34.77 dBm
Min. total BS power	-10 dBm
User distribution	Uniform across network
System bandwidth	80 MHz
Number of subcarriers	512
Spreading factor	16
Total number of groups	32
Noise power spectral density	-174 dBm/Hz
Bit rate per group per user per time slot (after error control coding)	10.2 kbps
Target E_b/I_0, real-time traffic	7 dB
Target E_b/I_0, non-real time traffic	5 dB
Orthogonality factor	0.9
Simulation time step	0.1 s

Table 4.4: *System-level simulation parameters*

The target E_b/I_0 is the E_b/I_0 that gives the desired BER and is selected by running PHY simulations. In this case, the target E_b/I_0 for real-time traffic is set to 7 dB from the set of PHY simulations and field experiment results in the work by Kishiyama *et al.* [76].

Non-real time traffic's lenient latency requirements allows for re-transmissions in which a hybrid automatic repeat request (HARQ) scheme with packet combining may be used. HARQ differs from conventional ARQ in that erroneous packets are not discarded and are instead stored in the receiver for combining with a newly re-transmitted packet to form a more reliable packet [77, 78]. For non-real time traffic, the E_b/I_0 is set to 5 dB due to this traffic type allowing such HARQ re-transmissions.

The orthogonality factor is similarly dependent on PHY parameters such as the delay spread in the channel, the type of detector used in the receiver etc. A value of 0.94 has been recommended by the 3GPP for the evaluation of microcell systems [75], in this chapter's simulations it is set to the slightly lower value of 0.9. The simulation time step is chosen as the length of approximately 10 time slots.

4.4 Power Control and Multi-Group Simulation Results

4.4.1 Results for Multi-Group Allocation

Figure 4.4 shows results for the probability of outage (per MS's group and per MS) for the three multi-group allocation algorithms presented in section 4.1.2. For these results, VoIP traffic is assumed as presented in section 4.3.2.1. This service type requires a data rate of 16.5 kbps which can be provided by using two groups (each of which has a data rate of 10.2 kbps giving a total data rate of 20.4 kbps). As explained in section 4.3.1.3 if one of an MS's groups goes into outage (from which the per MS group outage value is calculated), then that MS also goes into outage (from which the per MS outage value is calculated) as the MS can no longer be provided with the required data rate. Outage is defined to be when the E_b/I_0 drops below the target E_b/I_0.

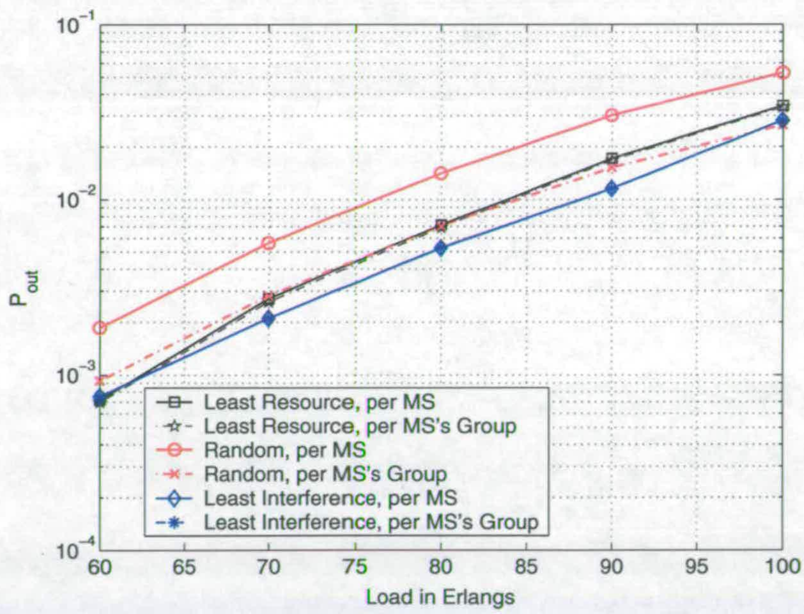


Figure 4.4: *Probability of outage for different multi-group allocation schemes*

Comparing the performance of the three multi-group allocation algorithms, it can be seen that the Least Interference algorithm outperforms the Least Resource and Random algorithms. This is due to the Least Interference algorithm's direct use of the measured interference values to preferentially assign users to those groups in which they experience the least amount of interference, i.e. where the users have the most chance of having low outage probabilities. Focusing on the performance of the Least Resource and Random algorithms, the Least Resource algorithm is seen to perform similarly to the Random allocation algorithm in terms of the per MS, per group outage. However, examining the per MS outage shows that the Least Resource algorithm outperforms the Random allocation algorithm. Allied to this is the fact the per MS group outage is virtually indistinguishable from the per MS outage for the Least Resource allocation whereas the two are very different for the Random Allocation algorithm. These phenomena can be explained by noting that the Least Resource algorithm tends to allocate groups to users in 'chunks', or groups of groups; making it very likely that each of an MS's groups will experience similar interference values. Hence, with the Least Resource algorithm, if one of an MS's groups is in outage, then it will be very likely that all of that MSs other groups will also be in outage. The Random algorithm by its very nature ensures that an MS's groups will face very different interference values, thus with this algorithm, even though one group is in outage (and thus the MS is in outage), the rest of the groups might still be able to support transmission giving rise to the difference between the two outage values.

4.4.2 Results for Power Control

Figure 4.5 shows results for the probability of outage (per MS's group and per MS) for the three power control variants presented in section 4.1.1. The outage assumptions are as presented in the preceding section. The Least Resource allocation algorithm is used to obtain all sets of results.

The maximum power limits for the two DCPC variants were calculated in the following manner:

- DCPC with per MS power constraints:

$$p_{u,g,max} = \frac{p_{bs,max}}{90} \tag{4.9}$$

  where $p_{bs,max}$ is the maximum power per BS. The number 90 is obtained experimentally from the results in Figure 4.4 and is the maximum number of MSs that can exist in the system for an outage probability of approximately 1%.

$$p_{u,g,min} = \frac{p_{bs,min}}{Number\ of\ Groups} \tag{4.10}$$

where $p_{bs,min}$ is the minimum power per BS.

- DCPC with per group total power constraints:

$$p_{g,max} = \frac{p_{bs,max}}{Number\ of\ Groups} \tag{4.11}$$

$$p_{g,min} = \frac{p_{bs,min}}{Number\ of\ Groups} \tag{4.12}$$

- Hybrid DCPC with per MS and per group total power constraints has $p_{g,max}$ and $p_{g,min}$ as in equations (4.11) and (4.12) with an added constraint on the maximum power per MS $p_{u,g,max,hybrid}$, which is calculated by performing a link budget analysis – using the general simulation parameters contained in Table 4.4 – to find out how much power an MS at the edge of a cell would need to meet the E_b/I_0 target.

Table 4.5 gives numerical values for these limits calculated using the preceding formulae.

Parameter	Value
Maximum BS power	34.7 dBm
Minimum BS power	-10 dBm
Number of Groups	32
$p_{u,g,max}$	15.17 dBm
$p_{u,g,min}$	-25.05 dBm
$p_{g,max}$	19.65 dBm
$p_{g,min}$	-25.05 dBm
$p_{u,g,max,hybrid}$	15 dBm

Table 4.5: *Power limits*

Figure 4.5 shows the probability of outage results for the three power control schemes. It is seen that the algorithm with per MS power constraints performs poorly compared to the other two power control schemes at lower loads. This is because the per group and hybrid power control algorithms allow flexibility in allocating the downlink power resources, for example, if an MS is in a favourable propagation situation, the hybrid and per group algorithms may be able to allocate that MS appreciably less power than if the power was limited per MS.

In turn, the hybrid power control algorithm performs better than the per group algorithm. The hybrid power control trades off some of the flexibility of the per group algorithm for a limit on the power each MS may enjoy. The effect of this limit is that an MS that requests a power value beyond the relaxed per MS limit is deemed to be in an unsupportable location and thus is removed from the system before it can cause additional interference. The per group algorithm would continue ramping up such an MS's power causing yet more interference to other MSs.

At higher loads the performance of the per MS algorithm converges to that of the hybrid algorithm's as the power demanded by each MS with either algorithm is likely to be equally high. However, at these higher loads the probability of outage is greater than the conventionally accepted value of 1%. Thus for the main operating loads (i.e. those where the outage probability is less than 1%) the hybrid power control algorithm outperforms the other two algorithms. Furthermore the parameters for the hybrid algorithm are easier to set up than those of the per MS algorithm, as no assumption is made as to the maximum number of MSs that can exist in the system.

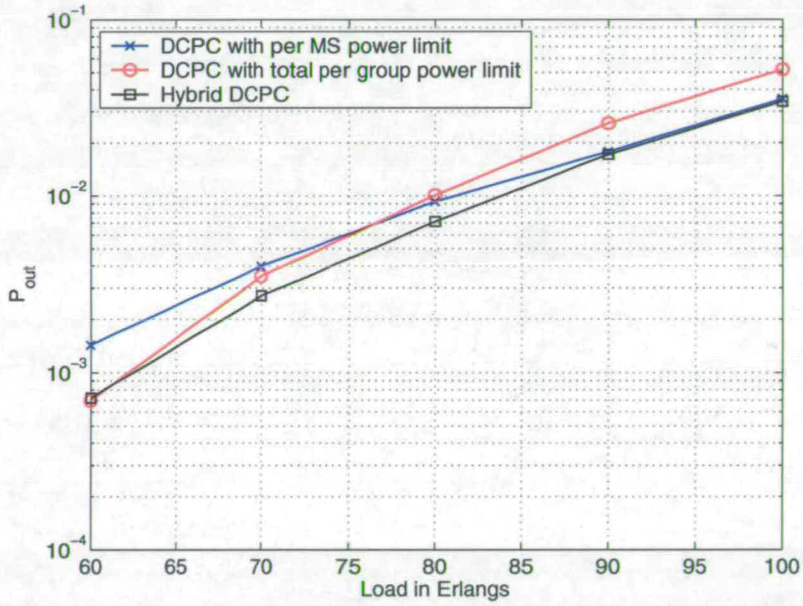


Figure 4.5: *Probability of outage for the three power control schemes.*

4.5 Data Scheduling Simulation Results

The simulations of the schedulers presented in section 4.2 were performed with the number of base stations set to seven to reduce the simulation duration. Only non-real time interactive WWW data was considered in these simulations. All other parameters were set as laid out in

Table 4.4.

The maximum number of usable resources per group was altered to examine the effect of changing this parameter on the performance of the different schedulers. A simple capacity estimate for single-cell CDMA (and thus requiring no assumptions about inter-cell interference) can be formed by manipulating the expression for E_b/I_0 to give [31]:

$$N = 1 + \frac{W/R}{E_b/I_0} \tag{4.13}$$

putting the values for $W/R = 16$ and $E_b/I_0 = 5$ dB $= 3.16228$, a simple estimate of the maximum usable number of resources per group can be calculated to be 6 codes. In the simulations, the maximum number of usable codes per group was set to either 3, 6, 10 and 16 codes. These correspond to pessimistic, conservative, optimistic and over-optimistic levels respectively for the number of free resources per group.

For the fixed resource round robin scheduler the amount of resources assigned to each user in a time slot is set to 64. The per MS weights for the MG-GPS scheduler are set to 1.

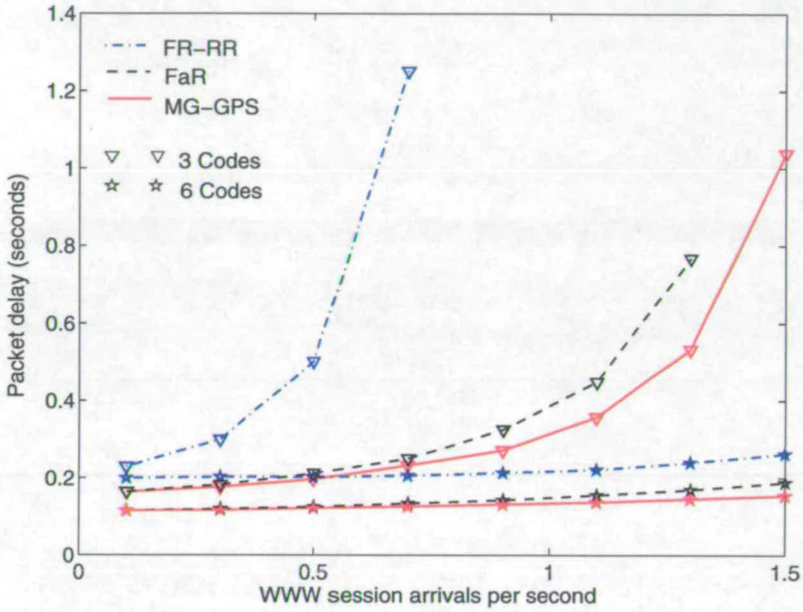
Figure 4.6(a) shows the average packet delay when 3 or 6 maximum usable codes per group are assumed, while Figure 4.6(b) shows the average packet delay when 10 or 16 maximum usable codes per group are assumed.

Examining the case of either 3 or 6 codes, it is seen that the FR-RR scheduler has poorer average packet delay performance than either the FaR or the MG-GPS scheduler. This is because the FR-RR scheduler can only schedule a certain fixed number of users at each time instance; the other MSs must await their turn to be scheduled, which means their data is held in the queue for longer time periods than is the case with the other schedulers. In addition, the FaR scheduler is more inefficient than the other schedulers as at the lower session arrival rates there is a large amount of unused resource as scheduled MSs may not have enough backlogged data to fully take advantage of their resource allocation. The FaR and MG-GPS allow for smaller delays as each MS is guaranteed to receive some resource in each time period, thus the time each packet spends in the queue is lower as each MS gets service in every scheduling instance. When a pessimistic number of maximum usable codes per group – i.e. 3 codes – is assumed, the delay is higher than that for the conservative number of codes of 6. This is due to the schedulers not having enough resources to fully meet the needs of the arriving traffic. Increasing the number of sessions arriving per second has the effect of increasing the amount of backlogged data in the packet queues and hence the average packet delay.

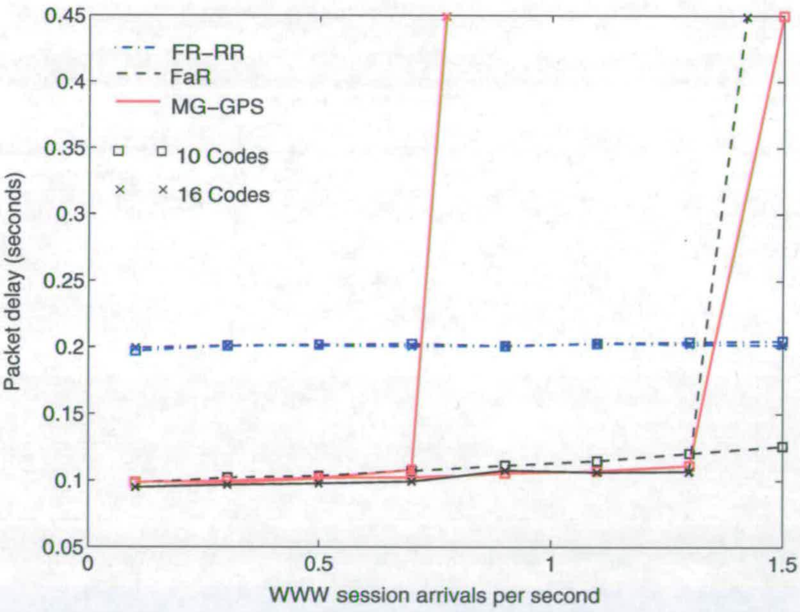
When the maximum number of usable codes per group is increased to either 10 or 16, the average packet delay begins to exhibit some tendencies peculiar to the nature of the schedulers. Initially, both the FaR and MG-GPS schedulers have lower packet delays than the FR-RR scheduler. However, this situation changes at higher session arrival rates; the exact point of change being dependent on the particular scheduler and the number of usable codes. When 10 codes are assumed, the MG-GPS scheduler's average packet delay increases abruptly after an arrival rate of 1.3 sessions per second (for reasons of clarity the Figure has been truncated to show only maximum average delays of up to 0.45 seconds). A similar effect is seen, when 16 codes are assumed, for the MG-GPS and the FaR schedulers for session arrival rates of 0.9 and 1.3 sessions per second respectively. This phenomena is due to the relative efficiencies of the different schedulers. The MG-GPS scheduler in particular is an efficient scheduler and allocates all unused resource to MSs that have backlogged traffic. Therefore, when the load – in the form of the session arrival rate – is high enough, the MG-GPS scheduler will attempt to make use of all the resources available in a group. Of course as the maximum number of codes per group is set at an optimistically high level, some of the groups may be unable to support use of all the codes causing packet transmissions on these groups to fail. These packets will then have to be retransmitted whereupon, because the amount of backlogged data has increased, the MG-GPS scheduler will once again attempt to make use of all resources. This explains why the loading point at which the MG-GPS scheduler exhibits failure is lower when 16 codes rather than when 10 codes are used. At the over-optimistic setting of 16 usable codes, the potential for large amounts of interference making most of these codes unusable is much higher. The inefficiency of the FaR scheduler relative to the MG-GPS scheduler is its strength in this situation. With the FaR scheduler, no attempt is made to allocate unused resources to MSs that request more than their fair share. Groups with such unused resource are given the benefit of lower interference levels meaning that the likelihood of such groups supporting the transmission is higher than if the more efficient MG-GPS scheduler was used.

Figures 4.7(a) and 4.7(b) show the average packet call throughput when the maximum usable amounts of codes per group is set to either 3 or 6 and 10 or 16 codes respectively.

When 3 or 6 codes are used, the MG-GPS scheduler performs better than either the FaR or FR-RR scheduler, this is due to the MG-GPS being more efficient in its use of resources. Similar to the case for the average packet delay, when 10 codes are used the MG-GPS scheduler

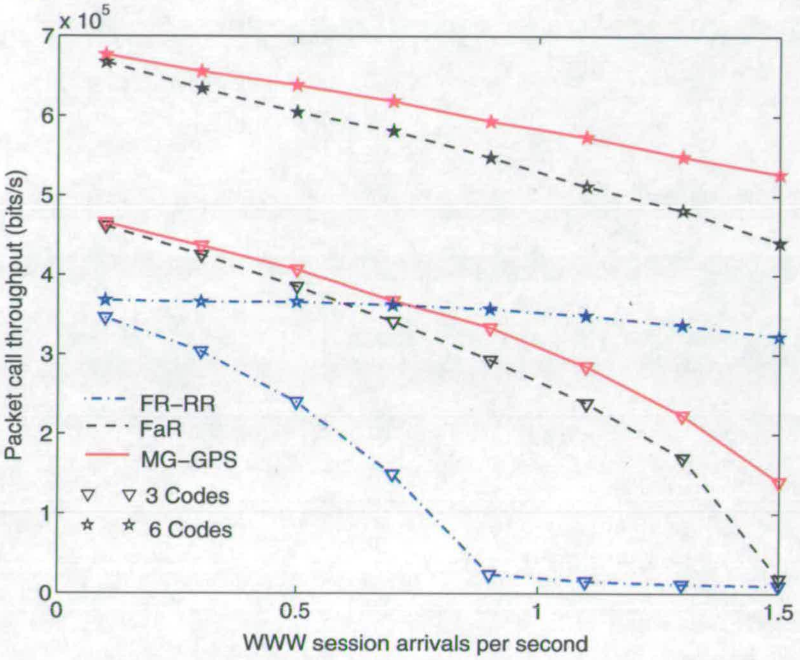


(a) Maximum usable codes per group = 3 and 6

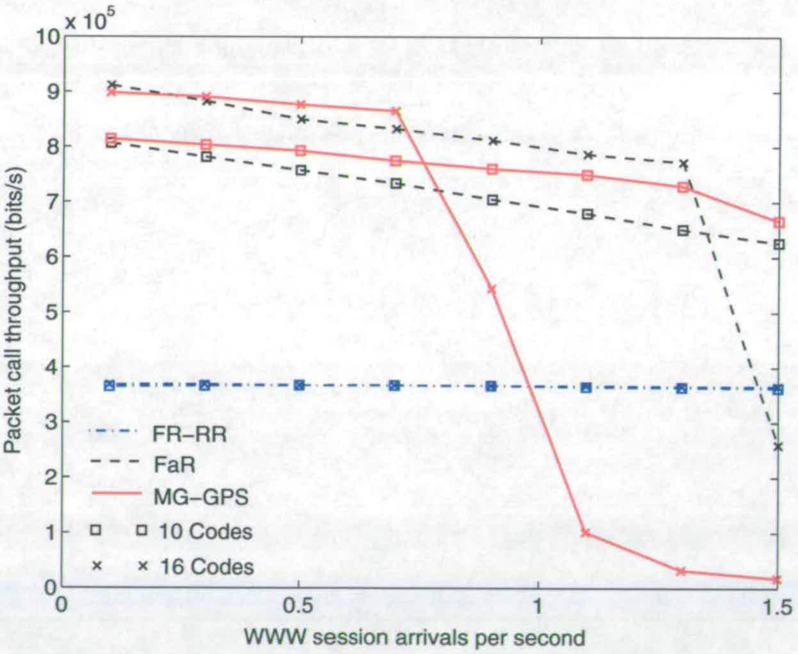


(b) Maximum usable codes per group = 10 and 16

Figure 4.6: *Average packet delay for different schedulers and maximum amounts of usable codes per group*



(a) Maximum usable codes per group = 3 and 6



(b) Maximum usable codes per group = 10 and 16

Figure 4.7: *Average packet call throughput for different schedulers and maximum amounts of usable codes per group*

exhibits better average throughput performance than the FR-RR scheduler due to the MG-GPS scheduler's efficient use of resources. When 16 codes are used the MG-GPS scheduler initially has better average throughputs for low loads, however as the load increases the MG-GPS, in its efficiency, aggressively attempts to use all 16 codes leading to large amounts of interference rendering these codes unusable and resulting in a large drop in the throughput.

A key pointer from this is that to obtain maximum performance out of the MG-GPS scheduler, a measure of the amount of usable resources per group should be obtained. This is even more important in a heterogeneous traffic scenario when the number of usable resources for the non-real time traffic becomes even more time varying due to some of these resources being used by other traffic types.

4.6 Conclusions

Work on multi-carrier systems has largely focused on the PHY layer aspects, in this chapter a performance evaluation of several resource management issues that impact on the grouped MC-CDMA system has been carried out.

It has been demonstrated that the best power control variant is a hybrid DCPC with per group and per MS constraints. With regards to multi-group allocation, a least interference multi-group allocation was found to perform best, however the additional complexity involved in using such an allocation relative to a least resource multi-group allocation means the least-resource allocation is in fact a better candidate for implementation. In addition, for traffic with a fixed data rate requirement, it was found to be better to allocate groups in super-groups as such groups experienced similar interference levels.

For non-real time data scheduling a simple multi-group generalised processor sharing scheme was proposed. This scheduler initially performed better than other schedulers, but then due to its efficiency in using resources, performed poorly to the other schedulers at high loads. This effect may be mitigated by the integration of estimates of the number of free resources in making the scheduling decision.

One major conclusion of this chapter is the need for accurate capacity estimation. For real-time traffic this will allow access to the system to be limited at high loads to protect ongoing calls. For non-real time data traffic, availability of accurate capacity estimates will aid in the scheduler deciding how much data can be transmitted at any one time avoiding the need for retransmissions which increase the packet delay and reduce the achievable throughput. The issue of dynamic capacity estimation is the topic addressed in the next chapter.

Chapter 5
Dynamic Capacity Estimation

The previous two chapters concentrated on techniques to manage radio resources at the PHY and MAC layers. It was shown that performance gains can be achieved by separately optimising the allocation decisions at these layers. However, whilst these decisions were treated separately, they are inherently linked and co-dependent. For example, in Chapter 3 the subcarrier allocation algorithm performed MS grouping and allocation with regard only to a PHY layer parameter; the amplitude of each subcarrier. Chapter 4 introduced the MAC layer allocation of resources which took into account the MS's slow fading and QoS requirements in the form of multiple data rates. In a real system both these separate elements will have to occur concurrently i.e. BSs will need to allocate both the PHY and MAC layer resources to MSs. The conventional approach is to do as was done in the previous chapters and allocate these resources separately. The Resource Metric Estimation framework diverges from this approach in that it aims to allocate resources across layers whilst taking cognisance of the advantages and limitations of each layer. With this approach additional performance gains are to be expected.

In a CDMA system the capacity is not fixed and is instead time-varying and dependent on the interference conditions experienced by MSs in the system. In this chapter capacity estimation is investigated in which algorithms and models are developed that deliver estimates of the maximum capacity and hence the free resource. These capacity estimation algorithms will then be used, in the subsequent chapter, to aid the Resource Metric Estimator in providing cross-layer resource allocation decisions by bridging the gap between the PHY/MAC and network layers. The definition of maximum capacity adopted is that number of users that can exist in a system below a certain, predefined, probability of outage.

In the first section of this chapter the issue of capacity estimation in the single-cell environment is examined, here some assumptions are also made relating to the traffic and interference characteristics; section 5.2 then presents some simulation results pertaining to this environment. Section 5.3 removes the assumptions of section 5.2 and looks at the issue of dynamic capacity estimation (DCE) in a multi-cell downlink environment, a capacity model is outlined and practical methods for estimating various parameters introduced. In section 5.4 the performance of the multi-cell DCE is presented. The chapter finishes with some conclusions in section 5.5.

5.1 Single-Cell Capacity Estimation

In this section an expression for the capacity of a single-cell grouped MC-CDMA system is developed. For the purposes of the analysis each user uses all the available groups and multiple data rates are provided by allocating more than one code to a user i.e. multi-code allocation.

Let m be the number of code allocations given to a user. With m code allocations the user will have a data rate of m times the basic rate R i.e. a rate of mR bit/s. Subrates may be achieved by discontinuous transmission of one of the assigned codes (such techniques are already utilised in current 3G systems). For a multi-code system, the received power for each of the m codes (of the same multi-code user) is the same. The multi-user interference experienced by each of the m codes is also the same [79]. With these insights an expression for the SIR in the multi-code case can be developed. The E_b/I_0 for any of the codes in the g-th group of the u-th user is given by

$$\left(\frac{E_b}{I_0}\right)_{u,g} = \gamma_{u,g} = \frac{KL_{u,g}P_{u,g}}{\sum_{j\neq u} m_j L_{j,g}P_{j,g} + (m_u - 1)L_{u,g}P_{u,g} + \eta_0 W} \tag{5.1}$$

Where m_u is the number of codes assigned to the u-th user, K is the spreading gain and $L_{u,g}$ is the path-loss between the BS and the u-th user on the g-th group. The first term in the denominator is the interference experienced from the m_j codes of the j-th user. The second term is the self-interference experienced from the u-th user's own codes where it is assumed the codes are completely non-orthogonal.

To obtain an expression for the overall E_b/I_0 for each user the multi-code E_b/I_0 is averaged across the groups

$$\gamma_u = \frac{1}{G}\sum_{g=1}^{G}\gamma_{u,g} = \frac{1}{G}\sum_{g=1}^{G}\left(\frac{KL_{u,g}P_{u,g}}{\sum_{j\neq i} m_j L_{j,g}P_{j,g} + (m_u - 1)L_{u,g}P_{u,g} + \eta_0 W}\right) \tag{5.2}$$

The long-term fading for each group is the same, $L_{u,g} = L_u$, furthermore if a user occupies all groups, then the interference and hence the power assigned to each group is the same. Thus equation (5.2) reduces to

$$\gamma_u = \frac{KL_uP_u}{\sum_{j\neq u} m_jL_jP_j + (m_u-1)L_uP_u + \eta_0 W} \tag{5.3}$$

In an interference limited system such as MC-CDMA, the power of an individual user must be controlled to limit the effect of interference on other users. Conversely, the feasibility of assigning an appropriate power level to a user may be used as a way to develop an expression for the capacity. Following the approach by Gürbüz [80], equation (5.3) can be rearranged and the following expression derived to indicate that a feasible power solution for each user can exist if and only if

$$\sum_{j=1}^{U} \frac{m_j\gamma_j}{\gamma_j + K} + \frac{\eta_0 W}{min_{1\leq u\leq N}\left[P_{max,u}L_u\frac{\gamma_j+K}{m_j\gamma_j}\right]} < 1 \tag{5.4}$$

Where $P_{max,u}$ is the maximum transmit power limit for the u-th MS.

The maximum capacity U is thus defined by the number of users that the system will support before the outage probability P_{out} exceeds a certain predefined threshold [81]. It is important to note that the maximum capacity developed above assumes that users are continually active, although an activity factor may be included to account for the effects of discontinuous transmission.

5.2 Single-Cell Capacity Estimation Simulation Model and Results

In this section some results are presented to indicate the maximum capacity values obtained. Table 5.1 shows the parameters used to obtain the theoretical capacity values. Three traffic types are investigated, voice, video and data. The video and data traffic types respectively require 8 and 16 times higher data rates relative to the voice traffic.

The impact of systematic errors in the power control is also investigated; Viterbi [82] showed that for CDMA systems with imperfect power control, the statistics of the E_b/I_0 followed a log-normal distribution with a mean corresponding to the desired E_b/I_0 level and a standard deviation on the order of 1.5dB to 2.5dB.

Figures 5.1 and 5.2 show the theoretical capacities with no systematic error and 2 dB of sys-

Parameter	Value
Number of groups	2
Spreading Gain	512
Total Number of Subcarriers	1024
Base Data Rate	8 Kbps
Data Rate per Serial-to-Parallel Channel	4 Kbps
System Bandwidth	1024*4 Kbps = 4.096 MHz
Noise Spectral Density	-174 dBm/Hz
Maximum Mobile Power Limit (voice, video and data)	5, 14.1, 17.1 dBm
Standard Deviation of Shadowing	10 dB
Mean of Required E_b/I_0 for voice, video and data	5, 5, 7 dB
Rate for voice, video and data	8, 64, 128 Kbps
Number of Code Channels required for voice, video and data	1, 8, 16
Outage Probability constraint	1%

Table 5.1: *Parameters used for single-cell capacity estimation*

tematic error respectively. As can be seen by comparing these two figures, the systematic error reduces the capacity in the system.

The data traffic has less stringent latency requirements relative to those for the voice and video traffic. Therefore, when there is real-time traffic present in the system the data traffic has to contend with lower priority and use the spare or residual capacity that is not being used by the real-time traffic. Figures 5.3 and 5.4 illustrate how the throughput of the real-time users affects the spare capacity for non-real time traffic with no systematic error and with a systematic error of 2 dB respectively. In these figures the Aggregated Data Rate (ADR) is the sum of data rates from users. The blue plots on these figures show the maximum ADR possible for different numbers of voice and video users K_{vid}. The red plots show the residual ADR which is the Maximum ADR minus the ADR from the voice and video users. The residual ADR is the data rate available for utilisation by non-real time data users. Similar to the observations for Figures 5.1 and 5.2, the presence of video users has a large effect on the achievable throughput for data users, with the presence of systematic error reducing the throughput data user can achieve.

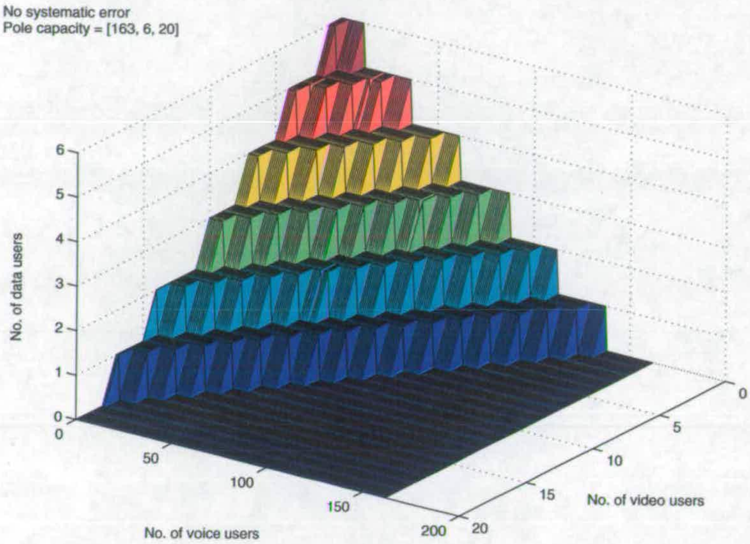


Figure 5.1: *Theoretical capacity of single cell multi-code MC-CDMA with no systematic error*

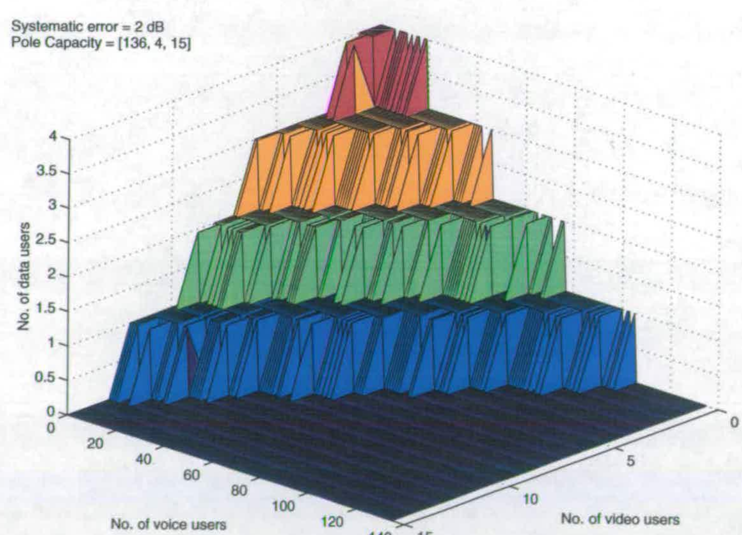


Figure 5.2: *Theoretical capacity of single cell multi-code MC-CDMA with systematic error = 2dB*

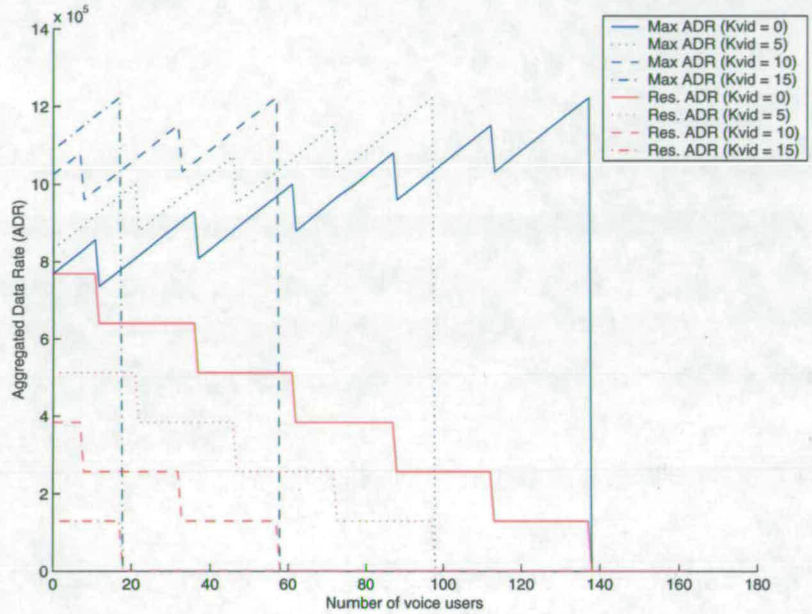


Figure 5.3: *Maximum and residual ADR with systematic error = 0dB*

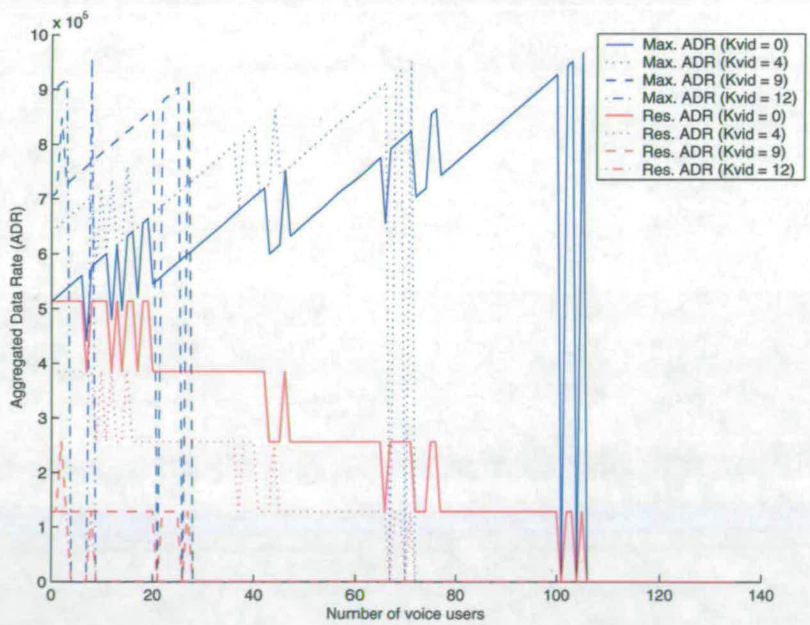


Figure 5.4: *Maximum and residual ADR with systematic error = 2dB*

5.3 Multi-Cell Dynamic Capacity Estimation

In the preceding section, an expression for the capacity of a single-cell MC-CDMA system was presented. However, there were some assumptions made which while simplifying the analysis, limit the usefulness of the resulting capacity model. Firstly, in the cellular environment, most CDMA networks will not have cells operating in isolation of other cells, multiple cells will be required. In a multi-cell CDMA network interference will be received from both intra and inter-cell interferers. The capacity model needs to explicitly account for these interferers for it to be of most use in the RME framework. Secondly, users were assumed to be constantly transmitting; to account for the characteristics of different traffic types a more sophisticated traffic model is required. Thirdly, the assumption was made that each group had similar interference characteristics. To take advantage of the groups as an extra dimension of resource allocation freedom, different groups may be assigned to different traffic types and hence such groups will have different interference characteristics. Finally, the capacity model assumed that users were completely non-orthogonal, in a real system there is bound to be some modicum of orthogonality between users. In this section a comprehensive capacity model is outlined that takes into account all the aforementioned factors. It largely follows an approach similar to the works by Viterbi [83] and Choi [84].

5.3.1 Analytical expression for capacity

To develop the capacity model a three stage process is followed. Firstly, the interference in the system is modelled, thereafter the traffic in the system is characterised. Finally, the two are combined to give an analytical expression for the capacity and hence the capacity model.

5.3.1.1 Interference Model

In the forward-link the power allocated to an individual mobile station (MS), u, can be considered as a fraction, ψ_u ($0 < \psi_u \leq 1$), of the total power allocated to traffic sources by a base station (BS) [83].

The received E_b/I_0 for the u-th MS communicating with BS0 using traffic type t_m (e.g. $t_1 =$

VoIP, t_2 = WWW data, t_3 = packetised video streaming data etc.) is as follows

$$\left(\frac{E_b}{I_0}\right)^u_{t_m} = \gamma^u_{t_m} = \frac{W}{R_{t_m}} \cdot \frac{\psi_u^{(t_m)} \Psi P_0 L_{0,u}}{\sum_{b=1}^{B} P_b L_b + \alpha P_0 L_{0,u} + N_0 W} \tag{5.5}$$

where

W = spreading bandwidth

R_{t_m} = data rate for traffic type t_m

P_b = total power transmitted by the b-th BS

Ψ = proportion of total BS power allocated to traffic channels

$L_{b,u}$ = path loss from the b-th BS to the u-th MS

α $= 1 - \hat{\alpha}$

$\hat{\alpha}$ = orthogonality factor, where,

'0' = Completely non-orthogonal

'1' = Perfectly orthogonal

N_0 = noise power

The following assumptions are now made

1. the background noise is negligible compared to the total signal power.
2. the same E_b/I_0 is required by MSs communicating with each respective service type (e.g. all voice users require the same E_b/I_0, γ_{t_1}, all data users γ_{t_2}, etc.).
3. no soft handoff, mobiles only implement hard handoff.

Further, the following notation is adopted; that the received signal power at the u-th MS from the 0-th BS is S_0 $(= P_0 L_{0,u})$ and that the out of cell interference received by the u-th MS is I_{out} $(= \sum_{b=1}^{B} P_b L_b)$ then

$$\gamma^u_{t_m} = \frac{W}{R_{t_m}} \cdot \frac{\psi_u^{(t_m)} \Psi S_0}{I_{out} + \alpha S_0} \tag{5.6}$$

If the user is required to meet a certain E_b/I_0, γ_{t_m}, for the traffic type t_m, then equation (5.6) can be rearranged to obtain the relative power allocation for the u-th user

$$\psi_u^{(t_m)} = \frac{\gamma_{t_m} R_{t_m}}{W\Psi} \left(\frac{I_{out}}{S_0} + \alpha\right) \tag{5.7}$$

The u-th user receives a fraction ψ_u of the transmitted power under the conditions

$$\sum_{u=1}^{U_b} \psi_u \leq 1 \tag{5.8}$$

where U_b is the total number of users attached to a BS.

Let, y_u be

$$y_u = \left(\frac{I_{out}}{S_0} + \alpha \right) \tag{5.9}$$

where the $\frac{I_{out}}{S_0}$ term is the **other-cell relative downlink interference** and the α term is the orthogonality factor (see equation (5.5)).

Now a key, but temporary, constraint is made; that all mobiles are located at the boundary of the cell whose BS they are communicating with. Without this constraint, an analysis of the distribution of y_u, which depends on relative MS-BS distances and on log-normally distributed random variables, would prove to be analytically intractable [83]. Intuitively this also makes sense as mobiles that are located at the edge of the coverage area are likely to have a greater impact on the cell capacity than those at the centre. With this constraint however, y_u can be treated as a lognormally distributed random variable with mean dB value m_y and standard deviation of dB value σ_y. To account for the fact that MSs are in fact not all located at the cell's edge, but instead uniformly distributed across the cell, the constraint is relaxed by introducing a forward-link power factor η [84]. The forward link power factor is the ratio of the average I_{out}/S_0 to that of the worst case I_{out}/S_0 where all MSs are at the cell boundary. The outage probability can then be derived as

$$\begin{aligned} P_{out} &= Pr\left[\frac{\gamma_{t_1} R_{t_1}}{W} \sum_{u=1}^{U_{t_1}} \theta_u^{t_1} y_u + \cdots + \frac{\gamma_{t_m} R_{t_m}}{W} \sum_{u=1}^{U_{t_m}} \theta_u^{t_m} y_u > \frac{\Psi}{\eta} \right] && (5.10) \\ &= Pr\left[Z = Z_{t_1} + Z_{t_2} + \ldots + Z_{t_m} > \frac{\Psi}{\eta} \right] && (5.11) \end{aligned}$$

where $\theta_u^{t_m}$ is the activity factor of the traffic type t_m assigned to the u-th MS, U_{t_m} is the number of active sessions for traffic type t_m and $Z_{t_m} = \frac{\gamma_{t_m} R_{t_m}}{W} \sum_{u=1}^{U_{t_m}} \theta_u^{t_m} y_u$.

Since y_u are independent identically distributed lognormal random variables, Z_{t_m} can be approximated as lognormally distributed random variables conditioned on the number of active sessions [84]. The technique for calculating the mean and variance of Z (which is approximated by a lognormal random variable [85]) is outlined in the works by Choi and Sampath [84, 86].

The probability of outage, conditioned by the number of active sessions for each traffic type, is then

$$P_{out} = Pr\left[\tilde{Z} = \ln(Z) > \ln\left(\frac{\Psi}{\eta}\right) | U_{t_1}, U_{t_2}, \ldots, U_{t_m}\right] \tag{5.12}$$

5.3.1.2 Analytical descriptions of the traffic models

The number of active sessions for each traffic type is a random variable dependant upon the characteristics of that traffic type, the focus is on two traffic types; VoIP and WWW data.

VoIP As outlined in section 4.3.2.1, a VoIP session can be thought of as an ON-OFF source; during the ON-period the user is talking and fixed size packets are generated at regular time intervals by the voice codec. In the OFF-period the user is silent, no packets are transmitted and hence the wireless resource may be used by other users. The activity factor, θ_{t_1} (defined as the percentage of time the source transmits to the duration of the session) is

$$\theta_{t_1} = \frac{\text{ON-time}}{\text{OFF-time + ON-time}} \tag{5.13}$$

VoIP packets have strict latency requirements, thus the scheduling policy adopted is that VoIP packets are transmitted in the next appropriate time slot after each packet's arrival. By assuming the session arrival process and the session length process to be Markovian, with rates λ_{t_1} and μ_{t_1} respectively, then for m servers such a process can be modelled by an $M/M/m/m$ queue. This leads to the formula of equation (5.14) for the occupancy distribution of sessions [87, 88] in which the occupancy is decreased by the activity factor θ_{t_1}.

$$P(n = N) = \frac{(\theta_{t_1}\lambda_{t_1}/\mu_{t_1})^n/n!}{\sum_{j=0}^{m} \frac{(\theta_{t_1}\lambda_{t_1}/\mu_{t_1})^j}{j!}} \tag{5.14}$$

where n is a random variable representing the number of active sessions in the system at steady state and $P(n = N)$ – the occupancy distribution – indicates the probability of n being some value N.

WWW data With reference to the WWW data traffic model introduced in section 4.3.2.2, a WWW data session can be modelled as a series of packet calls within which a number of packets are transmitted. Each packet call can be thought of as the downloading of a page by a user. Molina *et al.* [89], show that for a large number of users the interarrival process of page requests at the BS is exponentially distributed. Thus, the page (packet call) arrival is a Poisson process. If the arrival rate of user's WWW sessions at the BS is λ_s and the mean number of packet calls per session is N_{pc} then the arrival rate of packet calls at the BS is

$$\lambda_{t_2} \approx \lambda_s N_{pc} \tag{5.15}$$

Within a packet call, a number of packets are transmitted and there is a time interval between two consecutive packet transmissions. This behaviour can be characterised by an activity factor θ_{t_2} [84]

$$\theta_{t_2} = \frac{\sum_P T_{ON}}{\sum_P T_{OFF} + \sum_P T_{ON}} \approx \frac{E[T_{ON}]}{E[T_{OFF}] + E[T_{ON}]} \tag{5.16}$$

where P is the average number of packets per packet call, T_{ON} is the duration of a packet transmission during a packet call and T_{OFF} is the duration of the time intervals between two successive packet transmissions in a packet call.

The serving time of a packet call does not follow a negative exponential distribution. As this is non-real time data with lenient latency requirements (relative to for example real-time services like VoIP), there is no blocking of unserved packet calls. The queue system may thus be modelled as an $M/G/\infty$ queue. The occupancy distribution of such a system is given by the following Poisson distribution [88]

$$P(n = N) = \frac{(\theta_{t_2}\lambda_{t_2}/\mu_{t_2})^n}{n!} e^{-\theta_{t_2}\lambda_{t_2}/\mu_{t_2}} \tag{5.17}$$

where μ_{t_2} is the serving rate of a packet call. The above equation describes the probability of n MSs (that are accessing WWW data) transmitting packets in the system at steady state.

5.3.1.3 Probability of outage and final capacity expression

Incorporating the information on the number of sessions developed in equations (5.14) and (5.17) into the expression for the probability of outage conditioned on the number of active sessions in equation (5.12) the following final expression for the probability of outage in the case of the two traffic types can be obtained

$$\begin{aligned} P_{out} &= \sum_{l=0}^{\infty} \frac{(\theta_{t_1}\lambda_{t_1}/\mu_{t_1})^l/l!}{\sum_{j=0}^{U_0^{t1}} \frac{(\theta_{t_1}\lambda_{t_1}/\mu_{t_1})^j}{j!}} \times \ldots \\ &\times \sum_{n=0}^{\infty} e^{-\theta_{t_2}\lambda_{t_2}/\mu_{t_2}} \frac{(\theta_{t_2}\lambda_{t_2}/\mu_{t_2})^n}{n!} \cdot Q\left(\frac{ln(\Psi/\eta) - E[\tilde{Z}]}{\sqrt{Var(\tilde{Z})}}\right) \end{aligned} \tag{5.18}$$

in which

$$U_0^{tm} = \lfloor \frac{W/R_{tm}}{\gamma_{tm}} \rfloor$$

The maximum capacity is then that load for which the outage probability is less than or equal to some pre-defined maximum value i.e. $P_{out} \leq P_{out,max}$.

5.3.1.4 Providing code capacity estimates

The DCE may also be used to provide estimates of the number of usable codes in each group – the code capacity. This is achieved by modifying the previous analysis to consider a code as an active user that is always on, the maximum number of usable codes can then be found as that number of always on users for which the outage probability is less than the pre-defined maximum value.

5.3.2 DCE Procedure

With the insight gained from the analytical expression for the capacity developed in section 5.3.1 the procedure for the Dynamic Capacity Estimator can be outlined. This is as follows:

1. **Statistics estimation** - Each individual MS estimates the mean other-cell downlink relative interference, y_u (see equation (5.9)) on some, or all, of its groups. This statistic is communicated to the BS at appropriate intervals. With information from the MSs in its cell, the BS can calculate the cell-wide mean, m_y, and standard deviation, σ_y, of the other-cell downlink relative interference.

2. **Cell edge detection** - Each MS determines whether it is at the edge of the coverage area of the BS it is communicating with. A single bit indicating this is passed to the BS along with the statistics of y_u.

3. **Maximum capacity per group** - With knowledge of m_y and σ_y the current outage probability per group can be calculated using equation (5.18). The maximum capacity per group is then obtained by iteratively placing higher (or lower) packet arrival and serving rates into the equation to find that load for which $P_{out} \simeq P_{out,max}$. To account for the fact that the statistics upon which this estimate is based are obtained from MSs, some of which are not right at the edge of the cell, a Degree of Confidence Level in the estimate of the current ($\mathrm{DCL_{curr}}$) capacity is required.

4. **Maximum total capacity for the whole system i.e. all groups** - The maximum total capacity in the system is a weighted sum of the maximum capacity per group. The exact nature of the summation depends on the scheme used to allocate users to groups.

Further details on most of the above steps can be found in the following sections.

5.3.3 Statistics Estimation

The key statistic to be estimated at the MS is y_u (equation (5.9)). y_u can be broken down into its constituent components; firstly, the orthogonality factor, α and, secondly, the other-cell downlink relative interference, I_{out}/S_0. (I_{out} is the out-cell interference, while S_0 is the in-cell received power.)

Channel estimates are required in the receiver to perform coherent detection, the orthogonality factor can be readily calculated from these estimates and thus estimation of this parameter is not treated further.

A scheme, termed the null pilot assignment plan, for separately measuring the out-cell interference and in-cell received power is explained in section 5.3.3.1.

5.3.3.1 Null pilot assignment plan

Conventionally, the BSs surrounding a cell and the in-cell BSs transmit at the same time, making it difficult to isolate and measure the out of cell interference I_{out}. The novel null pilot assignment plan is designed to overcome this problem.

The MC-CDMA frame format is as shown in figure 5.5. One frame is transmitted per time slot. There are two common pilot symbols time multiplexed at the beginning of the frame, these are mapped over all subcarriers. In the middle of the frame are 3 mid-amble symbols, these mid-ambles exist for each code and amongst other tasks are used to obtain SIR estimates, the mid-ambles are also mapped across all subcarriers. Within each frame there are a pair of what are termed 'null pilot symbols', these are similar to the mid-ambles with the exception that during the null pilot symbols the BS does NOT transmit any information. The out-cell interference is measured, by the MSs, during these null pilots. (The in-cell received power is measured outside these nulls, for example during the mid-amble.)

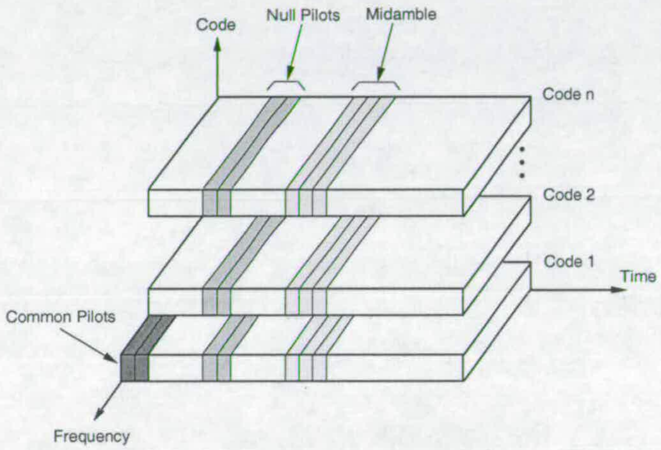


Figure 5.5: *MC-CDMA frame*

If – according to the PHY layer assumptions presented in section 5.4.1 – it is assumed that there are 85 symbols per frame, then the use of two null symbols per frame results in a small 2.5% reduction in the achievable throughput.

The exact position of the null pilots within the frame needs to be chosen such that in each frame every MS is permanently able to measure the out-cell interference. To achieve this, each cell in a cellular network is allocated a null pilot offset (this is simply the number of symbols from the beginning of the frame to the first null pilot). In a large cellular network it would obviously be impossible to have a unique offset for each and every cell; offsets have to be reused. With the reasonable assumption that the majority of the interference comes from the first two tiers of cells surrounding the desired cell, a technique from cell frequency planning can be borrowed to design the offset plan shown in figure 5.6. This figure shows a cellular network and the different offsets assigned to each cell. For example, cell 0 has null pilots at offset 'A' that is different from the offsets assigned to all the 18 neighbouring cells in the first two tiers. This is repeated across the network.

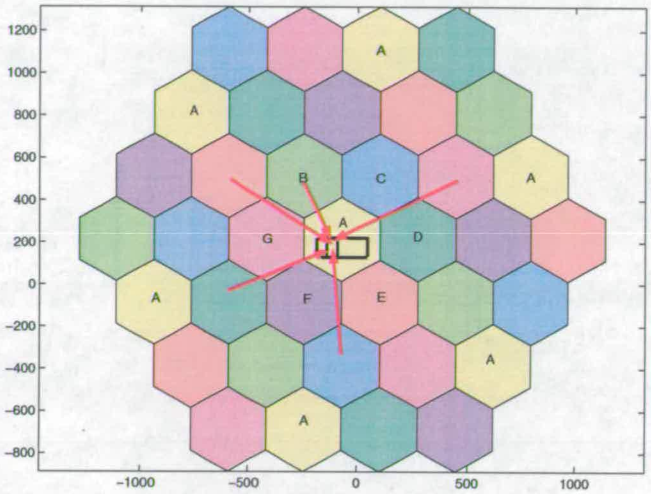


Figure 5.6: *Null pilot offset plan*

In selecting the offsets, care must be taken to ensure that the worst case propagation time does not result in the null pilot of a cell in the first two tiers clashing with that of the central cell. A result of this is that in cases where the cell radius is large, a larger offset reuse factor would be required.

5.3.3.2 Averaging techniques

The number of null pilot symbols is kept intentionally low to avoid adversely affecting the throughput per frame. This small number of pilot symbols means that it may be difficult for an estimator to form a reliable estimate of the out-cell interference, especially under heavily fluctuating traffic and propagation scenarios. Under such conditions, the choice of averaging scheme becomes important, in section 5.4 three averaging/filtering techniques are investigated;

1. Weighted Multi-Slot Averaging (WMSA) was proposed by Andoh *et al.* for pilot based channel estimation in CDMA systems [90], Lee *et al.* then applied it to the case of SIR estimation in TDD-CDMA [91]. Two different averaging lengths are used, the first (WMSA1) is 8 slots long and has coefficients *[0.3 0.8 1 1 0.8 0.3]*. The second (WMSA2) is 10 slots long and has coefficients *[0.15 0.3 0.8 1 1 0.8 0.3 0.15]*. The coefficients are chosen to give a tapering weight between the samples.
2. Moving average. The length of the moving average filter is chosen to correspond to that of the longest WMSA filter i.e. 10 slots.

3. Median filtering. The length of the median filter is also 10 slots.

A diagram of the full statistics estimation process is shown in figure 5.7. Firstly, the out-cell received power is measured by calculating $P_{out-cell} = E(|x_{in}|^2)$ during the null pilot symbols (where x_{in} is the input signal) and then, during the mid-amble and normal symbols, the total received power is measured using a similar procedure. These two values are manipulated resulting in an estimate of the other-cell downlink relative interference. This estimate can then be averaged across a suitable number of time slots before being passed to the BS.

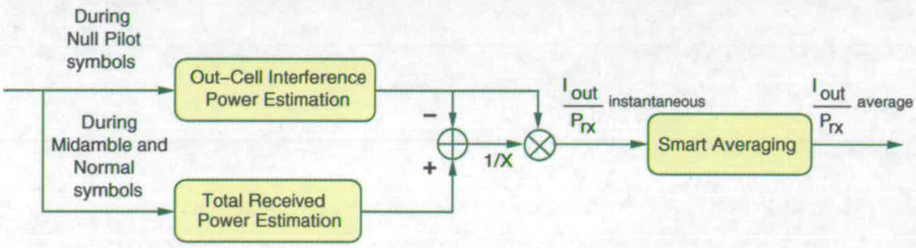


Figure 5.7: *Statistics estimation procedure*

5.3.4 Cell edge detection

MSs need to determine when they are in a position to effect a handoff from one BS to another. The measurements taken by an MS for the handoff decision are utilised to allow the determination of whether or not an MS is at the edge of its BS' coverage area. Thus the cell edge detection does not require any extra measurements/communication other than what is already provided for handoff detection.

The ideal scenario would be to take information only from MSs that are right at the edge of the cell. Unfortunately such a scenario is unrealistic for a number of reasons

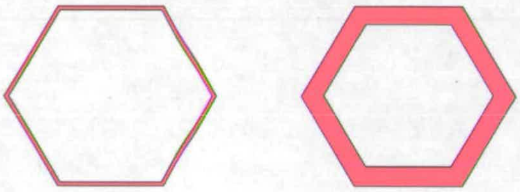
1. any MSs that were in such a position would be excellent candidates for handoff and would not be present at that position for any meaningful period of time

2. MSs are uniformly distributed over the cell area, thus the number of MSs right at the edge is likely to be extremely low which would mean that it would be harder to collect results that would form a statistically large enough sample to result in good estimation of the parameters of y_u

3. detection of the cell edge is not straightforward especially considering that the coverage

area for each cell is non-static and depends on a lot of parameters e.g. propagation conditions, traffic load, maximum power etc.

There is therefore a trade off between having a mobile at the edge of the coverage area and having enough mobiles present at the edge to give good data. To solve this, the requirement for MSs to be right at the cell edge is firstly relaxed (Figure 5.8(a)), instead this area is broadened into a cell edge region as seen in figure 5.8(b). As the cell edge region is larger than the cell edge, it provides a relatively large amount of MSs from which to collect the statistics of y_u, furthermore, these MSs are likely to exist in this region for a longer period of time than at the cell edge.



(a) Edge of cell **(b)** Cell edge region

Figure 5.8: *Expansion of the cell edge*

Handoffs are normally requested if the handoff metric crosses some threshold value. Cell edge region detection is then simply a matter of loosening that threshold, so that whenever the handoff metric crosses the looser threshold, the MS is deemed to be in the cell edge region.

The broadening of the area from which statistics of y_u are collected is bound to have an effect on the DCE algorithm. This is accounted for with a Degree of Confidence Level (DCL_{curr}) in the current maximum capacity. The DCL_{curr} is explained further in section 5.3.5.

5.3.5 Degree of Confidence Level in the current capacity

A key feature of the interference analysis presented in section 5.3.1.1 was that the other-cell downlink interference was taken from the edge of the cell. This constraint allowed the other-cell downlink relative interference to be modelled as a log-normally distributed random variable. In section 5.3.4 it was explained that it was more practical to take such measurements from a *cell-edge region* instead of the absolute cell-edge. However, a consequence of using a cell-edge region is that the other-cell downlink relative interference will no longer be strictly log-normal, but will also be affected by the varying MS-BS distances (i.e. distance between the MSs in the cell-edge region and their serving BSs). Therefore, this will result in the estimate of the mean of the other-cell relative downlink interference at the cell-edge being biased, with an attendant degradation in the capacity estimate calculated using this mean.

To account for these effects, a Degree of Confidence Level in the current (DCL_{curr}) estimate of the maximum capacity is introduced. The DCL_{curr} can be thought of as a quantifier of the distance between the maximum capacity estimate and the true capacity.

MSs are assumed to be uniformly distributed across the network, thus if a cell is viewed as a circle then due to the geometry of circles there is a higher probability of MSs existing at the cell-edge than at areas closer to the cell's centre. As a corollary to this, having a larger number of MSs equals a larger number of MSs at the cell-edge. Therefore, at high loads/high outage probabilities the estimate of the maximum capacity is likely to be more accurate as these situations equate to more mobiles being present in the cell/higher powers and thus more statistically meaningful results. The inputs to the DCL_{curr} will thus be the current load and the calculated outage probability. The DCL_{curr} is calculated via an empirical approach whereby MSs are distributed across a network and simulations run to determine the capacity of that network. Results are stored for various loads and current outage probabilities. These results are then compared to the maximum capacities calculated using the developed DCE framework, the result of this comparison is the DCL_{curr} i.e. the distance between the estimated and true capacities.

The procedure for calculating the DCL_{curr} is as shown in algorithm 4.

The DCL_{curr} may then be used to weight the maximum capacity estimate, C_{est}, as follows:

$$C_{est} = DCL_{curr}C_{max} \tag{5.19}$$

5.3.6 Maximum system capacity from per group capacity

The maximum capacity is first estimated per group, the total maximum capacity for the system (i.e. all groups) can then be obtained by appropriate summation of the per group capacities. In actual fact, the maximum system capacity depends on the model used to decide which user is allocated to which group. In the best case, the allocation model would be such that the

Algorithm 4 Calculating DCL_{curr}

if $(P_{out} \simeq P_{out,max})$ **then**
  $DCL_{curr} = 1$
else
  $DCL_{curr} = f(C_{max},\ C_{norm},\ P_{out},\ P_{out,max})$
end if

where,
C_{max} = Estimated maximum capacity
C_{norm} = Load at which the maximum capacity was estimated
P_{out} = Probability of outage for C_{norm}
$P_{out,max}$ = Maximum allowable probability of outage
$f(\ldots)$ = Function decided via an empirical approach which may for example involve running simulations where users are distributed across the cell and collecting data on capacity limits to decide the DCL_{curr}

maximum system capacity is the summation of the individual per group capacities, this can be considered an upper capacity limit. For some allocation models, the allocation of users to groups may result in undue interference on other users and hence the maximum system capacity being lower than the upper limit. The level at which the allocation model approaches the upper limit is what is termed, in this thesis, the 'group utilisation' of the model. If an allocation model has 100% group utilisation this means that its maximum system capacity is equal to the summation of individual per group capacities. An 80% group utilisation means that its maximum system capacity is equal to 80% of the summation of individual per group capacities, and so forth.

The simulation results presented in section 4.4.1 show that the least resource multi-group allocation has a 100% group utilisation, whilst the random multi-group allocation has an approximately 90% group utilisation.

5.3.7 Note on data transmission overhead

As outlined in the preceding sections, the multi-group multi-carrier DCE process is largely performed on a per group basis. The number of groups in the system depends on the length of the spreading code and the number of subcarriers; shorter spreading codes/more subcarriers equal more groups. Conceivably some objections could be raised regarding the data overhead (in terms of other-cell downlink relative interference measurements) that each MS has to transmit.

To assuage such fears, it may be useful to highlight that groups can be managed in groups of groups i.e. super-groups. If the allocation model now allocated users to groups on the coarser super-group boundaries, then each of the individual groups in that super-group would experience broadly similar long term fading and interference. The data transmission burden can then be reduced to the transmission of information for one group per super-group. There is unfortunately a trade-off in that the allocation model loses flexibility which could potentially impair its performance.

5.4 Multi-Cell DCE Simulation Models and Results

5.4.1 Estimation of the other-cell downlink relative interference

In this section methods for the estimation and averaging of the other-cell downlink relative interference are investigated using physical layer simulations of a MC-CDMA system.

Figure 5.9 shows the structure of the simulation. The out-cell interference is modelled as an MC-CDMA source with added AWGN noise.

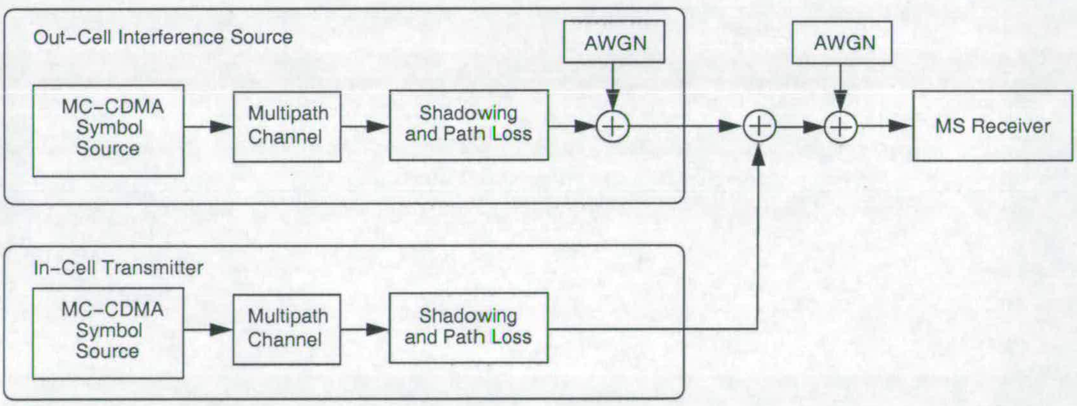


Figure 5.9: *MC-CDMA PHY layer simulation structure*

The PHY layer simulation parameters can be found in Table 5.2. The channel model used is a 24 tap exponential delay model as outlined by Maeda et al [61]. The maximum delay spread of this channel model is 1.15 μs, whilst its RMS delay spread is 0.2912 μs.

In evaluating the performance of the estimator and averaging techniques two performance measures will be used. The first is the normalised (to the true value of the ratio) mean square error (MSE) which is used to measure how close the results get to the true value. The second is the

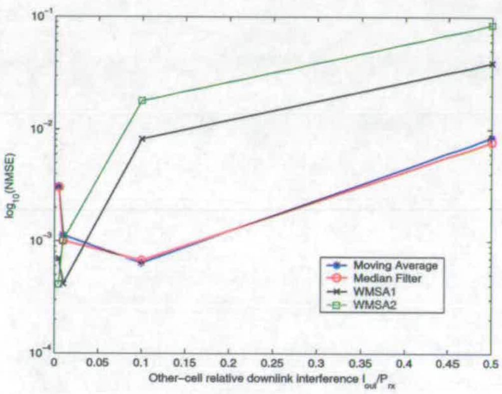
Parameter	Value
Total Number of Users	64
Total number of subcarriers	512
Spreading gain per group	16
Spreading code type	orthogonal Gold
Total number of groups	32
Number of groups per user	8
Group to subcarrier allocation	Fixed
Proportion of out-cell interference power that is AWGN	40%
Channel model	24-tap, exponential delay [61]
Channel model, maximum delay spread	1.15 us
Carrier frequency	5 GHz
Bandwidth	80 MHz
Number of midamble symbols	3
Total number of symbols per slot	85
Total number of slots simulated	1000

Table 5.2: *PHY layer simulation Parameters*

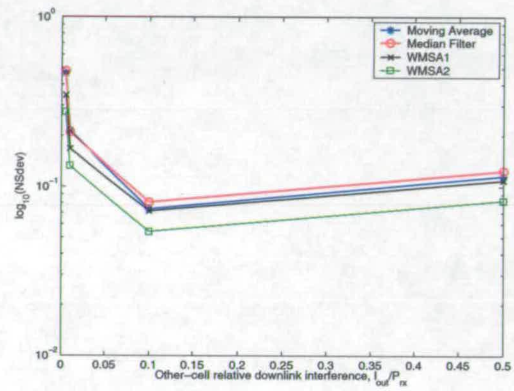
normalised (again to the true value) standard deviation which is used to measure the variability in the estimates.

Figure 5.10 shows the normalised MSE and the normalised standard deviation results of the estimation process for different other-cell downlink relative interference values at a Doppler frequency corresponding to a mobile velocity of 5 km/h and an SNR of 30 dB.

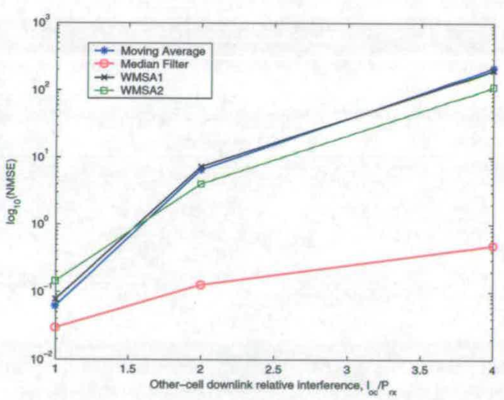
Simulation results by Viterbi [83] show that the majority of the other-cell downlink relative interference occurs in the range 0 to 0.5. The simulation results in Figure 5.10 show that the estimation process is most accurate at these values of interest. (N.B. Preliminary results showed that the trend seen in the figures was closely followed for intermediate values in the ranges 0.1 to 0.5 and 2 to 4. To reduce the overall amount of simulation time required, these results are not shown here.) Of the four averaging schemes, the WMSA variants initially perform better than the moving average and median filter for a small range of ratios between 0.005 and 0.025. Above these values, the moving average and median filter clearly outperform both variants of the WMSA filter (WMSA1 and WMSA2). The moving average and median filter at first exhibit broadly similar performance in the 0 to 0.5 range, however, in the 1 to 4 range the median filter is observably superior. This is due to the fact that at higher ratios there is a larger amount of power coming from outside of the cell; the power levels may have relatively large variations between symbols, because the number of the null pilots in such instances is not large enough to average out these variations, data outliers are produced. The median filter is more immune to such outliers than any of the other averaging techniques and is thus the best choice of averaging technique.



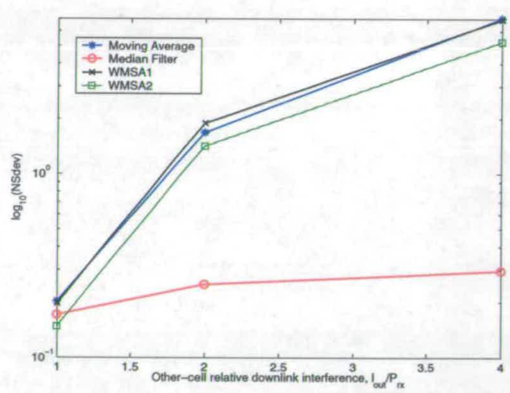
(a) Normalised MSE for ratios 0.005 to 0.5



(b) Normalised Std. dev. for ratios 0.005 to 0.5



(c) Normalised MSE for ratios 1 to 4

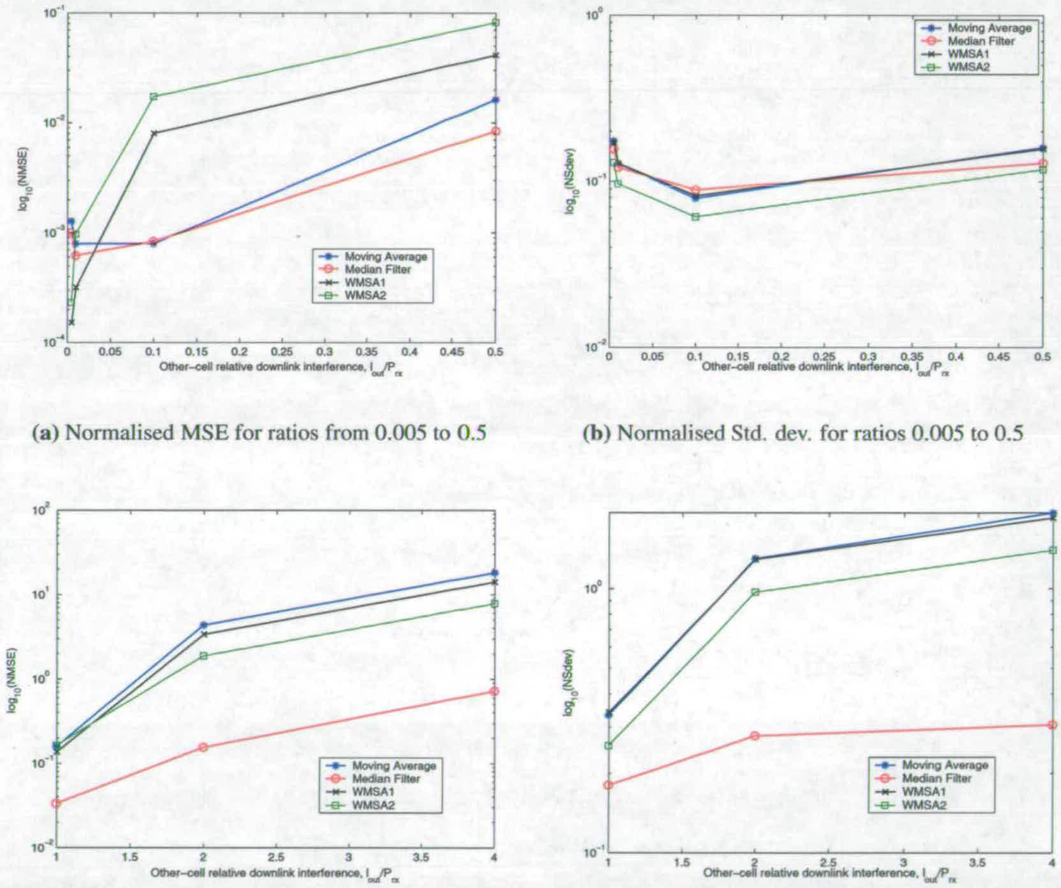


(d) Normalised Std. dev. for ratios 1 to 4

Figure 5.10: *Estimation of other-cell downlink relative interference, Doppler = 23 Hz (5 km/h), SNR = 30 dB*

Figure 5.11 shows results of when the estimation process is tested at a Doppler frequency corresponding to the higher mobile velocity of 30 km/h. These results do not differ appreciably from those at the lower Doppler frequency. Since the MC-CDMA system is assumed to be intended to be deployed in microcells where mobile velocities are not expected to be much higher than 30 km/h, the effect of even higher Doppler frequencies is not tested. Hence, for the system assumptions, the estimation process is robust to changes in Doppler frequency.

To check the robustness of the algorithm to changes in orthogonality factor, the channel was changed by reducing the maximum delay spread by 75% to 0.2875 μs (RMS delay spread of 0.0725 μs). Such a channel would have a higher orthogonality factor relative to the channel with the previous delay spread. There were, for all intents and purposes, no differences in the estimation results.



(a) Normalised MSE for ratios from 0.005 to 0.5

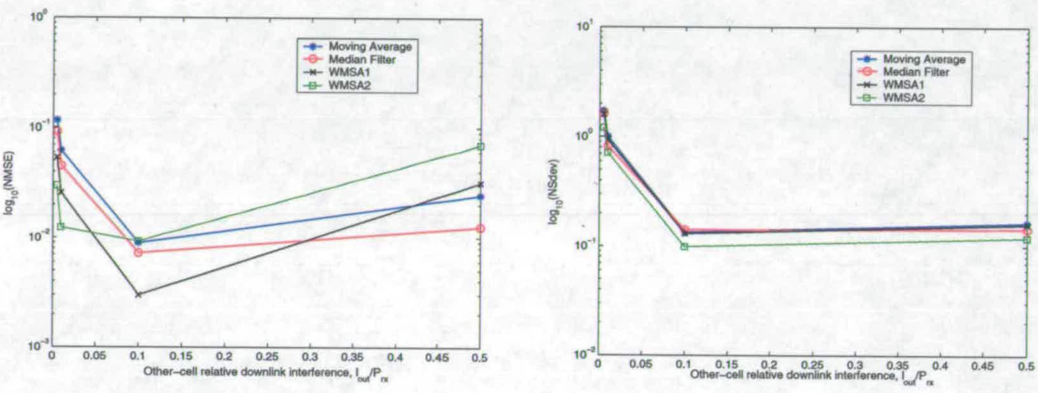
(b) Normalised Std. dev. for ratios 0.005 to 0.5

(c) Normalised MSE for ratios 1 to 4

(d) Normalised Std. dev. for ratios 1 to 4

Figure 5.11: *Estimation of other-cell downlink relative interference, Doppler = 139 Hz (30 km/h), SNR = 30 dB*

The effects of increased noise are then investigated. The results, for the range of major interest, can be seen in Figure 5.12 where the SNR is now 20 dB. Comparing Figures 5.11 and 5.12, it can be seen that in general the normalised MSEs for all the algorithms at the SNR of 20 dB are higher than those when the SNR is 30 dB. Thus, while the estimation process still works well at the lower SNR, the quality of the estimation process drops as the SNR is decreased. This is due to the increased noise power in the signal making it difficult to correctly estimate the power of the desired component.



(a) Normalised MSE for ratios from 0.005 to 0.5

(b) Normalised Std. dev. for ratios 0.005 to 0.5

Figure 5.12: *Estimation of other-cell downlink relative interference, Doppler = 139 Hz (30 km/h), SNR = 20 dB*

5.4.2 DCE

The simulation model for the system-level (i.e. cellular) simulation of the DCE process is the same as that used earlier in section 4.3. Table 4.4 in that section contains the main simulation parameters.

Simulation results are collected from all 19 cells, wraparound of the interference is performed to counteract edge effects. Power control is performed using the Hybrid DCPC scheme of section 4.1.1.3. Only one time slot is simulated. For the purposes of simplicity, only one traffic class, VoIP, is used to evaluate the performance of the DCE. Users are allocated to groups using the Least Resource multi-group allocation scheme of section 4.1.2.2.

The metric chosen for cell edge detection and handover is the MS-BS path gain, the MS is deemed to be at the cell-edge when the path gain between the serving BS and that of the next strongest BS differ by less than 6 dB. The MS-BS path gain is also the metric which the MSs use to determine to which BSs they belong to. To ensure that MSs do not fluctuate unnecessarily between BSs, there is a handoff margin; an MS is only allowed to handoff to a BS if the path gain to that BS is greater than the path gain to the current BS by an amount exceeding the handoff margin.

5.4.2.1 Degree of confidence level in the current capacity

An example of the degree of confidence level in the estimate of the current maximum capacity, DCL_{curr}, for the system assumptions is shown in figure 5.13. The figure shows the current load estimate, the calculated probability of outage and the DCL_{curr} in the estimate of the maximum capacity at these values. A DCL_{curr} of greater than 1 indicates that the estimate of the maximum capacity is overly pessimistic. Such situations tend to occur when the current load and/or consequent probability of outage values are high. A DCL_{curr} of less than 1 indicates that the maximum capacity estimate is overly optimistic. Such situations tend to occur when the current load and/or consequent probability of outage values are low.

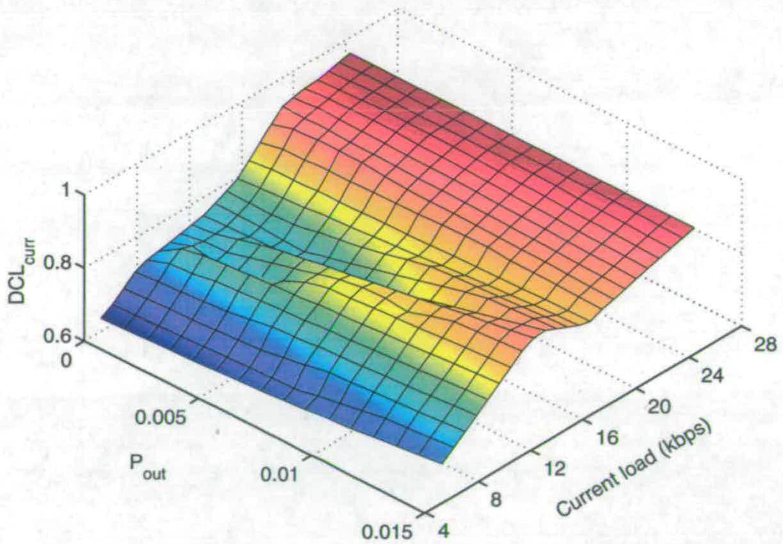


Figure 5.13: DCL_{curr}

5.4.2.2 Capacity estimation

Figures 5.14-5.16 show the estimated per group capacity (N.B. the capacity is after error control coding) performance of the DCE for different orthogonality factors, with and without the

DCE and with various levels of PHY estimation errors. These results are drawn by averaging the estimated per group capacity across all groups for a simulation time corresponding to 30 minutes of real world operation.

The length of the filter here describes the equivalent length, in seconds, of the moving average filter used to obtain averages of the other-cell downlink relative interference. The 'True Value' in the figures indicates the true capacity value after running the simulations and is that capacity value for which the outage probability is 1%.

The PHY estimation errors are added in the following manner. Firstly, simulations of the NPAP scheme – as explained in section 5.3.3.1 – are performed for various values of the other-cell relative downlink interference. Secondly, from these simulations, the mean and standard deviation of the estimated other-cell relative downlink interference after median filtering are calculated. These values are stored in a Look Up Table (LUT); the LUT's parameters are the desired other-cell relative downlink interference, the mean of its estimated value at the PHY layer and the standard deviation of the estimated value at the PHY layer. Finally, in the simulation a perfect calculation of the other-cell relative interference is made. To add the PHY layer estimation errors this value is fed to the LUT from which the mean and standard deviation of the other-cell relative interference after PHY estimation are drawn (after suitable interpolation). The resulting mean and standard deviation are used to generate a normally distributed number. This new number is the other-cell relative downlink interference after the PHY estimation errors.

The following three PHY estimation error scenarios are simulated:

- Perfect PHY estimate: No errors are made in the estimation of the other-cell relative downlink interference.
- PHY estimation errors (1): Small errors in the PHY estimation of the other-cell relative downlink interference, corresponding to the use of a Doppler frequency of 23 Hz (5 km/h) and an SNR of 30 dB.
- PHY estimation errors (2): Larger errors in the PHY estimation of the other-cell relative downlink interference, corresponding to the use of a Doppler frequency of 139 Hz (30 km/h) and an SNR of 20 dB.

A common trend in the results is that the use of the DCL_{curr} improves the performance of the capacity estimation process. With the DCL_{curr} using moving average filters with a length of 10

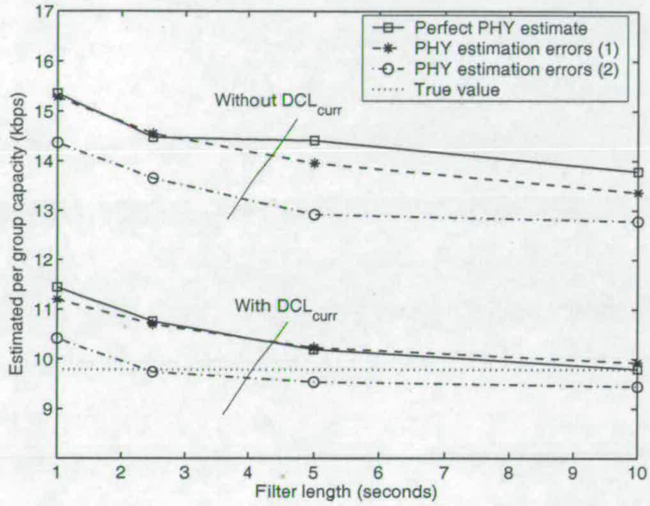


Figure 5.14: *Estimation of per group capacity with and without DCL_{curr} and with various levels of imperfect estimation (at the PHY layer) of the other-cell downlink relative interference. Orthogonality = 0.1.*

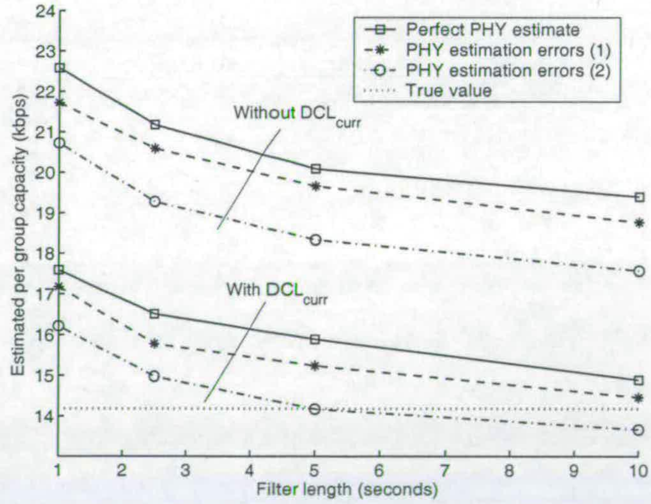


Figure 5.15: *Estimation of per group capacity with and without DCL_{curr} and with various levels of imperfect estimation (at the PHY layer) of the other-cell downlink relative interference. Orthogonality = 0.5.*

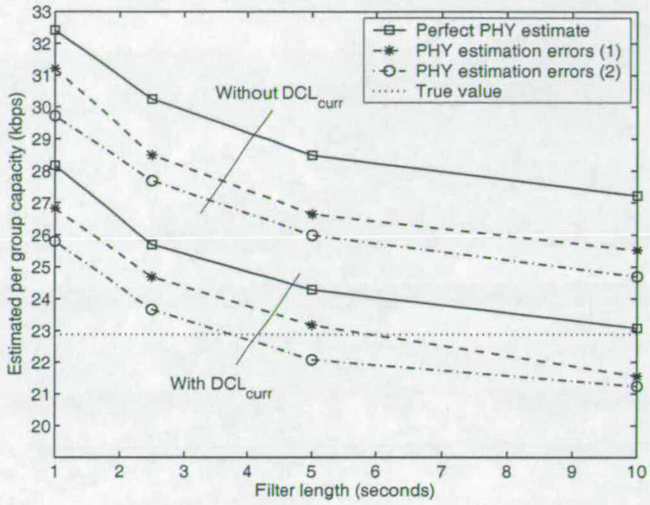


Figure 5.16: *Estimation of per group capacity with and without DCL_{curr} and with various levels of imperfect estimation (at the PHY layer) of the other-cell downlink relative interference. Orthogonality = 0.9.*

seconds results in an excellent estimate of the capacity for all orthogonality factors.

The addition of the PHY errors causes underestimation of the capacity. This effect is due to the PHY errors causing the overestimation of the other-cell downlink relative interference, resulting in a larger mean value than if no errors were present. Similarly, with PHY errors, the standard deviation of the other-cell downlink relative interference is larger than if no errors were present. Larger mean and standard deviation values correspond to larger P_{out} values for a given load and thus a lower estimated maximum capacity. The lower orthogonality factor is seen to be more immune to PHY error effects, this can be explained by examining equation 5.9 for the other-cell relative downlink interference. This equation is reproduced below for convenience.

$$y_u = \left(\frac{I_{out}}{S_0} + \alpha \right)$$

At lower orthogonality factors, the α term in this equation begins to dominate (recall that $\alpha = 1 -$ orthogonality factor) and errors in the $\frac{I_{out}}{S_0}$ term have less impact.

5.4.2.3 Results of the total capacity estimate

Figures 5.17-5.19 illustrate the performance of the DCE when estimating the total capacity. These results are drawn by averaging the estimated total capacity for a simulation time corresponding to 30 minutes of real world operation.

As in the previous subsection, the performance of the DCE improves with the use of the DCL_{curr} and is degraded when PHY estimation errors are added resulting in the DCE underestimating the total capacity.

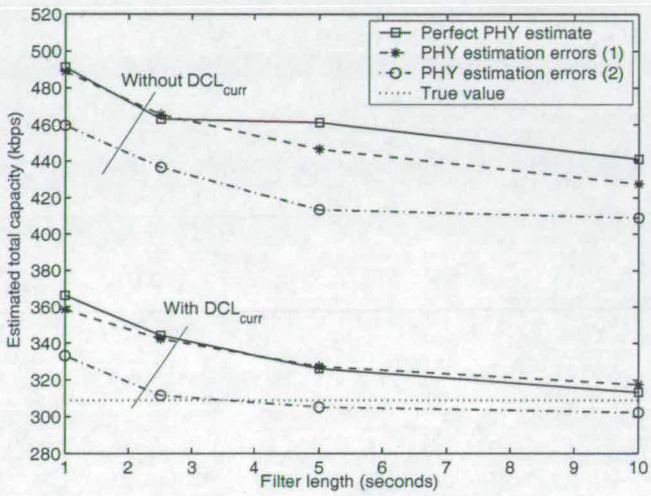


Figure 5.17: *Estimation of total (all groups) capacity with and without DCL_{curr} and with various levels of imperfect estimation (at the PHY layer) of the other-cell downlink relative interference. Orthogonality = 0.1.*

5.5 Conclusions

This chapter explored the concept of capacity estimation in multi-carrier CDMA. A simplified model for the single-cell capacity was presented. Thereafter a more comprehensive model for the multiple-cell capacity per group was performed. Using information gleaned from the comprehensive analysis, a Dynamic Capacity Estimation scheme and procedure has been proposed that is able to estimate the maximum capacity in one group and by extension the whole system. Further a degree of confidence level in the current capacity was utilised to account for errors produced during the estimation process. System level evaluation has shown that the DCE is able to capably deliver estimates of the maximum capacity aided by the DCL_{curr}.

Simulation results demonstrate that it is possible to estimate the main parameter (the other-cell downlink relative interference) required by the DCE with little modification to a conventional transmit frame, namely the introduction of null pilots. Although the use of some symbols as null pilots degrades the throughput achievable, under the system parameters such degradation

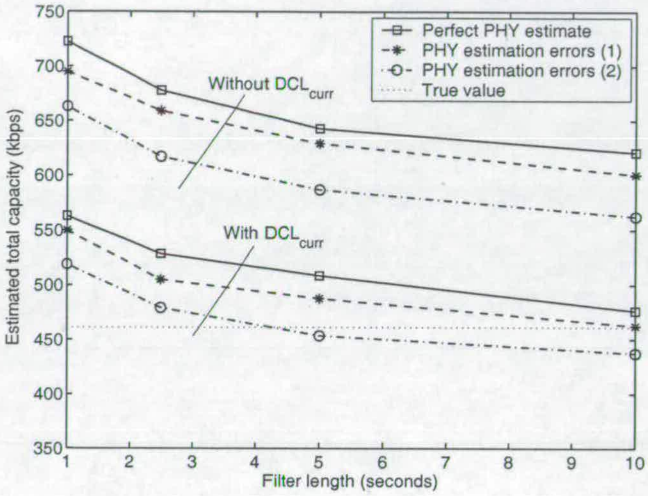


Figure 5.18: *Estimation of total (all groups) capacity with and without DCL_{curr} and with various levels of imperfect estimation (at the PHY layer) of the other-cell downlink relative interference. Orthogonality = 0.5.*

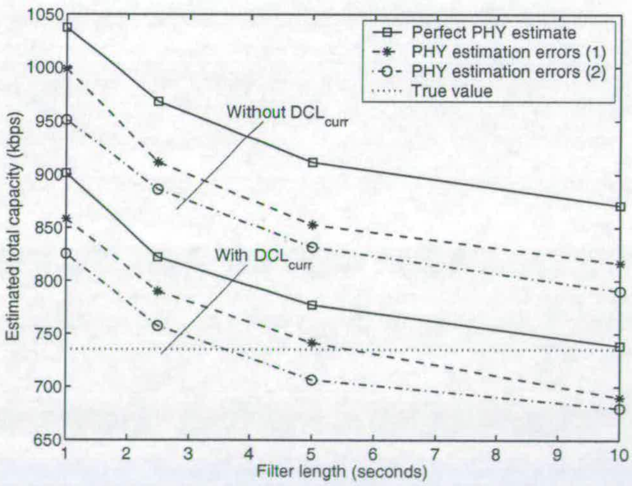


Figure 5.19: *Estimation of total (all groups) capacity with and without DCL_{curr} and with various levels of imperfect estimation (at the PHY layer) of the other-cell downlink relative interference. Orthogonality = 0.9.*

was limited to 2.5% of the throughput. To reduce this value, alternative schemes may be utilised for estimating the other-cell downlink relative interference. As the focus of this work is on the applications of RME to RRM such an investigation is left as an area for further research.

In chapter 6, algorithms that exploit the DCE to aid in the cross-layer scheduling of resources are presented.

Chapter 6
Resource Control and Scheduling

The preceding chapters have investigated the management of radio resources in both the PHY and MAC layers; in Chapter 5 a dynamic capacity estimation scheme was introduced that provided capacity estimates that could be used by the Resource Metric Estimator to bridge the gap between the PHY and MAC levels.

This chapter investigates the use of the DCE in aiding the provision of cross-layer resource allocation decisions. In making a decision on how to allocate the resources, the nature of each traffic type must be considered. For real-time traffic flows, their strict latency requirements means that once a real-time traffic flow is accepted into the network, then the quality of that flow should be guaranteed for the duration the flow is active. For non-real time flows, their less stringent latency requirements means that each flow can cope with temporarily poor transmission conditions by the resource controller choosing to allocate resources away from such a flow to other flows. Thus the resource allocation decisions for real-time and non-real time traffic may be reduced to decisions on call admission control and packet scheduling respectively. The admission control algorithm must decide if a real-time flow can be accepted into the network, whereas non-real time flows are always admitted into the network, meaning the scheduling algorithm must decide when packets may be given access to the network.

In this chapter the concepts from previous chapters are integrated via the Resource Metric Estimation framework with this unit taking in data from both layers and using this data to provide cross-layer allocation decisions. Section 6.1 focuses on admission control for real-time VoIP traffic. Section 6.2 deals with a cross-layer resource allocation and scheduling algorithm that uses the MG-GPS scheduler of chapter 4 and information from the DCE to make cross-layer aware decisions. The chapter finishes with conclusions in section 6.3.

6.1 Call Admission Control for Real-Time traffic

6.1.1 Introduction

A call admission control algorithm decides on whether or not to admit a call based on some criteria; the aim in making this decision is to admit as many calls as possible whilst ensuring that the quality of ongoing calls is preserved. In noise-limited systems based on orthogonal channels, the capacity is fixed at a value which normally corresponds to a multiple of the number of orthogonal channels available, the admission decision is thus based on whether or not an orthogonal channel is available for a newly arrived user. In interference-limited systems, such as CDMA, the capacity is time varying due to its dependence on other user's signal power, allowing a more flexible use of the wireless resources. However, to get the most out of this flexibility the question as to what metric the admission decision should be based on becomes crucial.

There have been a number of suggested metrics in the literature. Liu and Zarki [92] proposed an SIR based admission control that was shown to outperform a fixed call admission control. Kim [93] proposed an SIR based call admission algorithm whereby new calls are admitted provisionally into the system, power control conducted and the decision on whether or not to accept a call is made depending on whether or not the call's SIR meets a threshold value after the call has interacted with the system for some time. A drawback of this scheme is that it may take too long to come to a decision. Jeon and Jeong [94] proposed a scheme that uses the estimated average SIR and a CAC threshold SIR to make the admission decision. Gunnarson *et al.* [95] developed an admission control scheme that makes an admission decision based on the uplink relative load (a value that is related to the noise rise) in the target and surrounding cells. Admission control using a similar metric is mentioned in the book by Holma and Toskala [38]. Dimitriou *et al.* [96] proposed an admission control scheme that makes its decision based on whether or not the multiple access interference exceeds a certain threshold.

By far what appears to be the most popular class of admission control schemes is those that use the powers users are assigned/experience as an estimate of the interference condition in the cell and hence as a parameter upon which to base an admission decision. These schemes typically admit a user if the power metric (either received or assigned transmitted power for the uplink and downlink cases respectively) does not exceed some predefined threshold. Examples of such schemes are those proposed by Huang and Yates [97], Knutsson *et al.* [98], Kim and

Han [99], Perez-Romero *et al.* [100], Holma and Toskala [38], Aissa *et al.* [101], etc. The appeal of the transmit/received power as a metric is due to its ready availability, needing little to no measurement over and above that already being performed for other services in a wireless communication system.

6.1.2 DCE CAC algorithm

All the metrics and algorithms mentioned in section 6.1.1 perform well with the proviso that the threshold on which the admission decision is made is set correctly. This threshold will typically vary depending on the different classes of cells, propagation conditions etc., and in some cases may even need to be set on a cell by cell basis. Thus in this section a new downlink admission control scheme is investigated that utilises the earlier developed DCE framework to perform a joint admission control and multi-group allocation without the need for pre-set thresholds. This algorithm is termed the DCE CAC.

The performance of this algorithm is compared to that of two other algorithms; one without CAC using only dynamic multi-group allocation (termed No CAC) and the other a joint power threshold based CAC and dynamic multi-group allocation algorithm (termed Power threshold CAC).

Previous CAC algorithms that make a decision without a predefined threshold include works whereby the CDMA power control problem is set up as a constraint model, the feasible condition of which determines whether or not a call may be accepted [80, 102–104]. A drawback of these works is that they fail to consider the statistical multiplexing gains that may be obtained from the inactivity of users. Everitt [105] proposed a scheme, based on assumptions on the interference factors between a target cell and its interfering cells, that allowed the calculation of linear constraints that define the uplink capacity region in terms of the number of users. These inter-cell interference factors were assumed to be constant when in actual fact they are dynamic, exhibiting random behaviour dependent on the propagation conditions, MS movement etc. Ishikawa and Umeda [106] developed an analysis of the CDMA uplink that produced a calculation of an effective CAC threshold and Erlang capacity that could be used to make an admission decision; similar to the work by Everitt, they also assumed constant inter-cell interference factors. Evans [107] developed a traffic model for the uplink and introduced the concept of effective bandwidth in CDMA systems. He then proceeded to outline an effective bandwidth based call admission control algorithm.

6.1.2.1 Joint power threshold CAC and dynamic multi-group allocation

This algorithm combines a power threshold CAC algorithm (based on the downlink CAC algorithm proposed by Knutsson *et al.* [98]) with the Least Resource multi-group allocation algorithm of section 4.1.2.2. The algorithm is as follows:

1. A new session arrives and requests admission.
2. The Least Resource multi-group allocation algorithm selects a group(s) for this session.
3. Power threshold CAC is performed for the groups from step (ii). Admit the new session and assign it to the group(s) from step (ii) if

$$p_{tot,g} < P_{thr} \tag{6.1}$$

   where $p_{tot,g}$ is the total output power for the BS's g-th group before the new session is admitted and P_{thr} is a predefined power threshold. This power threshold is chosen by maximising the system's performance for a certain load.

   If the session is denied access to these group(s), then blacklist them and proceed back to step 2.
4. If the algorithm has iterated through all the groups, then the session is denied admission.

6.1.2.2 Joint DCE CAC and dynamic multi-group allocation

This algorithm combines the Least Resource multi-group allocation algorithm with a new DCE based CAC algorithm. The following steps describe the algorithm

1. A new session arrives and requests admission.
2. The Least Resource multi-group allocation algorithm selects a group(s) for this session.
3. DCE CAC is performed for the groups from step (ii), and admit the new session and assign it to the group(s) from step (ii) if

$$C_{norm,g} < C_{maxdcl,g} \tag{6.2}$$

Where $C_{norm,g}$ is the current load for the BS's g-th group and is estimated by multiplying the average per user throughput by the number of users currently in the system. $C_{maxdcl,g}$ is the maximum capacity for the g-th group as estimated by the DCE whilst taking into account the $\mathrm{DCL_{curr}}$.

If the session is denied access to these group(s), then blacklist them and proceed back to step 2.

4. If the algorithm has iterated through all the groups, then the session is denied admission.

The DCE CAC algorithm utilises information from the various layers to make a decision. PHY layer information is fed in to the DCE to develop a capacity estimate. From the upper layers information on the current carried traffic load is drawn.

6.1.3 Simulation results

In this section some simulation results illustrating the performance of the various CAC schemes outlined in the previous section are presented. The performance measures used to evaluate the CAC algorithms are the probability of outage, the call blocking probability (see section 4.3.1.3 for further explanation of these measures) and the Grade of Service (GoS). The GoS is a combination of the first two performance measures and is motivated by the well accepted notion that it is more annoying for a subscriber to have an ongoing session interrupted (due to it being dropped) than for that session to be blocked. The GoS can be stated thus

$$GoS = 10 * P_{out} + P_{block} \tag{6.3}$$

The system-level simulation parameters are the same as those of Table 4.4 in section 4.3.3. The orthogonality factor is fixed at 0.9. A number of power thresholds were tested for the Power Threshold CAC scheme, a value of $P_{thr} = -5dBm$ was found to be most suitable.

Figures 6.1, 6.2 and 6.3 show the probability of outage, blocking and the grade of service respectively for the different CAC algorithms.

It can be seen that the use of CAC reduces the probability of outage under high load conditions, allowing more users to enjoy a good QoS. The joint DCE CAC performs slightly better than the joint power threshold CAC. At low loads both the CAC algorithms perform slightly poorer

– but not appreciably so – than if no CAC was used.

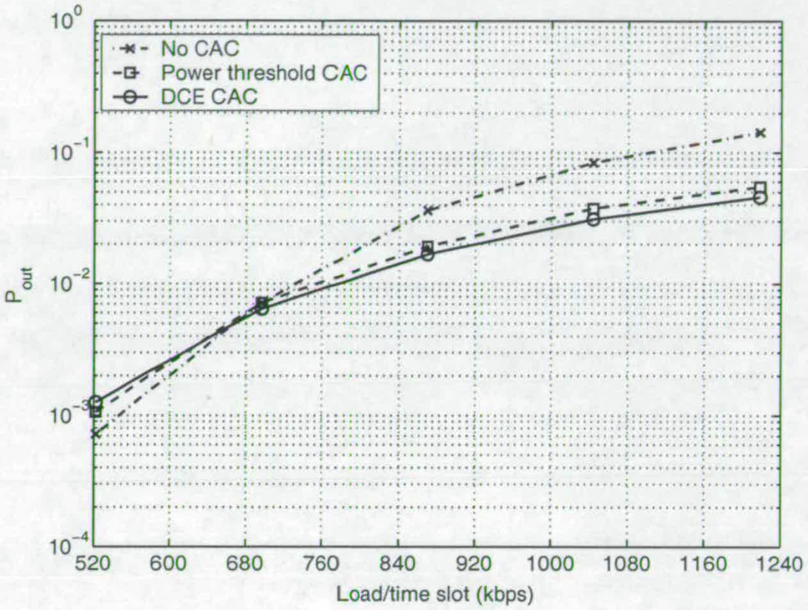


Figure 6.1: *Probability of outage for the different CAC schemes*

Of course the improvement in the outage probability comes at a price; an increase in the blocking probability. If no CAC is used, the blocking probability (for the loads in question) is zero. The CAC algorithms, on the other hand, will obviously have non-zero blocking probabilities. The joint Power threshold CAC has a lower blocking probability than the joint DCE CAC. Thus the DCE CAC trades blocking for outage.

The GoS results show that both the CAC algorithms deliver comparable GoS and are much better than using no CAC at higher traffic loads.

6.2 Resource Scheduling with MAC and Physical Layer Interactions/Criteria

6.2.1 Introduction

As researchers seek greater performance improvements in wireless systems, there has been a realisation that the previously held view that the performance should be optimised on a per layer basis is inadequate in the context of modern wireless systems. The issue of cross layer resource management has therefore recently become an area of increasing focus.

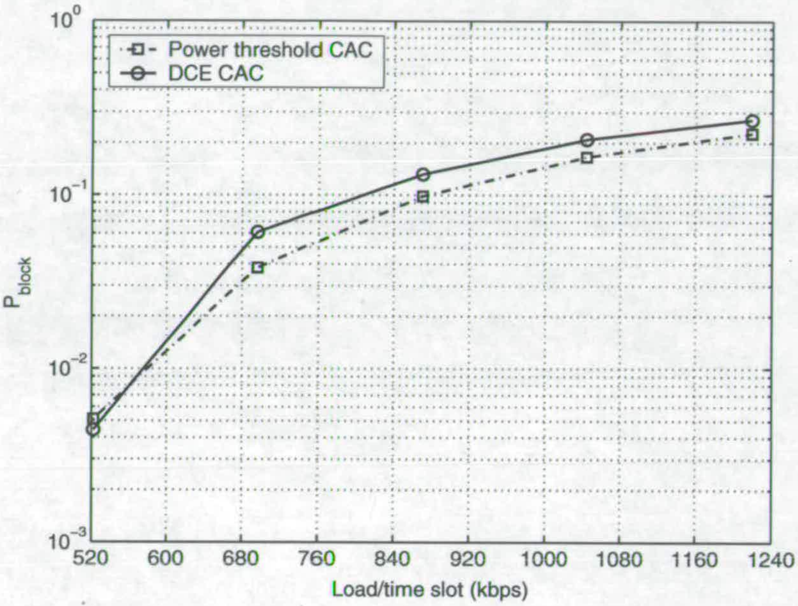


Figure 6.2: *Probability of blocking for the different CAC schemes*

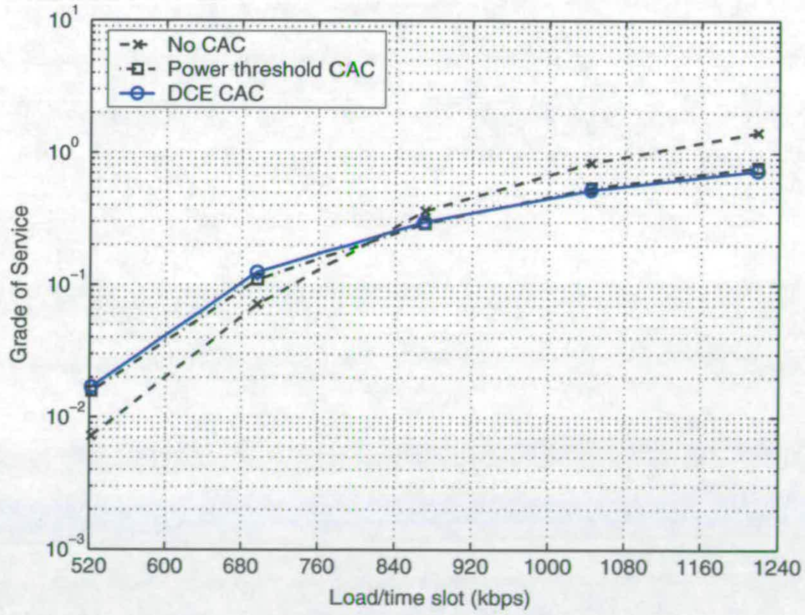


Figure 6.3: *Grade of Service for the different CAC schemes*

The approach taken in most cross-layer approaches is to marry together one or more physical and MAC layer algorithms and gain benefits from their interaction. Adaptive Modulation and Coding (AMC) at the PHY layer is combined with H-ARQ at the MAC layer by Liu *et al.* [108]. Their cross layer algorithm showed increased throughput under prescribed delay constraints. Friderikos *et al.* [109] proposed a cross layer algorithm that used information from the TCP state machine to perform power and rate allocation. A number of schemes have also been proposed that integrate information from smart antennas at the PHY layer with either H-ARQ or scheduling at the MAC layer to provide increased throughput [110–112].

With regards to resource management for multicarrier systems, the research traditionally focused on optimisation of QoS via power allocation, bit loading and subcarrier selection [33, 35, 113]. Currently there is little published work on cross-layer resource management for multicarrier systems, but it is highly unlikely that this situation will persist. Already there have been some publications integrating the PHY layer subcarrier, bit and power allocations of previous work with scheduling at the MAC layer. Cai *et al.* [114] proposed a scheme whereby power and subcarrier allocation were integrated with GPS scheduling via an optimisation condition that minimised the total allocated power. Their scheme showed better performance in terms of packet throughput and transmission delay than a packet GPS scheme and achieved similar fairness as a pure GPS scheme. A scheme that adopted a virtual clock scheduling algorithm integrated with an optimisation of the allocation of time slots, subcarriers and powers was proposed by Zhang *et al.* [115]. This scheme enhanced the system's spectral efficiency whilst improving the queueing performance. Li and Niu [116] considered the use of multiple-input, multiple-output (MIMO) antennas combined with OFDM. They found that the multiple transmit antennas in combination with the OFDM subcarriers could be considered as multiple parallel channels. They then proceeded to optimise the channel allocation using the received SINR as the optimisation metric whilst using a Weighted Fair Queueing scheduling algorithm. The optimisation of the subchannel allocation allowed the algorithm to simultaneously exploit the joint multi-user diversity and space frequency diversity whereas the WFQ scheduling provided throughput fairness. Their algorithm showed that it could provide minimum data rate guarantees and at the same time make efficient use of the wireless resource. A utility based framework for cross-layer optimisation was recently proposed by Song and Li [117, 118], they also proposed a set of algorithms for the cross-layer management of resources.

In contrast to the case of OFDM systems, there is, to the author's knowledge, currently no work that considers the cross-layer management of resources in an MC-CDMA system. Therefore, in this section a new scheme that takes into account both PHY and MAC layer information to manage the resources across these layers is proposed. The scheme utilises the techniques developed in the previous chapters, namely a modified version of the local search based algorithm of Chapter 3, the MG-GPS scheduler of Chapter 4 and the DCE of Chapter 5.

Figure 6.4 shows an overview of the new cross-layer resource management structure. The

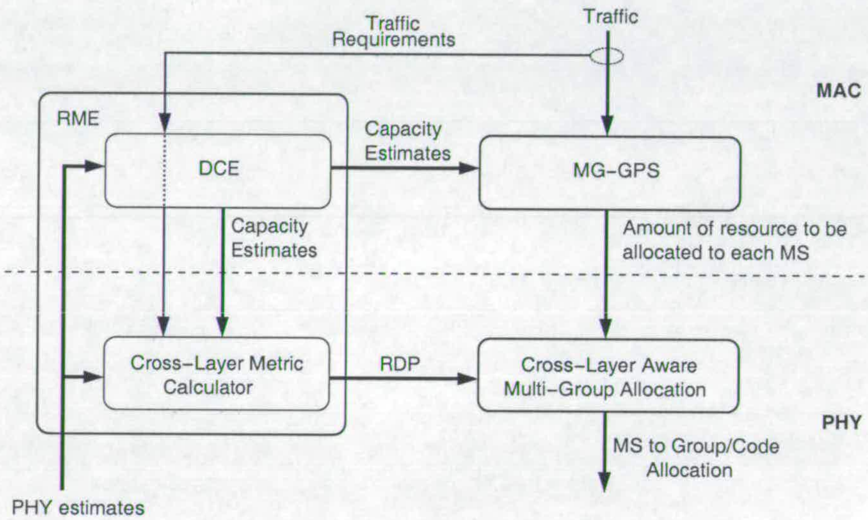


Figure 6.4: *Overview of the cross-layer resource management framework*

Resource Metric Estimator includes two units, the DCE and a cross-layer metric calculator. The DCE takes in information and metrics from the PHY layer and outputs estimates of the free capacity. In this particular embodiment, the capacity estimates are in terms of the number of usable code resources in each group. The capacity estimates are then passed on to the MG-GPS scheduler, which makes a decision as to what MS(s) should be scheduled for transmission at a particular time instance and how much of the free resource should be assigned to each such MS. The information at this point is purely in terms of number of resources to be allocated to each MS; a decision has not yet been made on which group and how many codes from a particular group an MS should allocate to fulfil its resource assignment. The decision on group/code allocation is taken by a cross-layer aware multi-group allocation algorithm. This algorithm uses the metrics calculated by the cross-layer metric estimator; these metrics take the form of a Relative Downlink Power (RDP) which is dependent on the PHY layer state of each group's subcarriers in conjunction with the estimated out-cell interference and maximum capacity per group from the DCE. The cross-layer aware multi-group allocation uses the RDP to optimise the allocation of MSs to groups and codes within each group.

Each of these elements is explained in further detail in the sub-sections below.

6.2.2 MG-GPS scheduling with DCE

Section 4.5 of Chapter 4 highlighted the fact that the choice of what per group capacity value to use impacted significantly on the performance of the MG-GPS scheduler. The best value of per code capacity was found to be 10 codes, however this value is not fixed and may vary for different propagation scenarios. The DCE of Chapter 5 provides a good framework to utilise the MG-GPS scheduler without any pre-existing assumptions on the free capacity. In this form the DCE is adapted to output per code estimates of the capacity in each group. The MG-GPS scheduler can then be easily modified to take into account these estimates. The MG-GPS scheduling equation (4.4) of chapter 4 is repeated below for ease of reference.

$$g_m = min\left(\left\lfloor\frac{\phi_m}{\sum_{i=1}^{M}\phi_i}\sum_{g=1}^{G}C_g\right\rfloor,\ \left\lfloor\frac{B_m}{R_{res}}\right\rfloor\right)$$

In this equation the assumption, C_g, of the per group capacity is replaced with the DCE estimate of the per group capacity $C_{DCE,g}$ which gives the following DCE aware MG-GPS scheduling equation:

$$g_m = min\left(\left\lfloor\frac{\phi_m}{\sum_{i=1}^{M}\phi_i}\sum_{g=1}^{G}C_{DCE,g}\right\rfloor,\ \left\lfloor\frac{B_m}{R_{res}}\right\rfloor\right) \tag{6.4}$$

6.2.3 Cross-layer aware multi-group allocation

In this section an algorithm that takes into account cross-layer information to allocate MSs to groups and codes[1] within those groups is presented.

The approach followed for implementing a cross-layer aware allocation algorithm is to define a metric that incorporates information from the different layers. A metric termed the Relative

[1]An assumption is made that all codes experience largely the same amounts of interference from each other. This is true for the case where the equalised channel coefficients are uncorrelated [119]. This can easily be assured if, for example, subcarriers are allocated to groups randomly, or with a fixed allocation whereby there are a large number of subcarriers between each successive subcarrier allocation to a group. Therefore, the allocation to codes is done in the sense that multiple codes from one group may be assigned to an MS. There is no allocation of MSs to a particular code in preference over another code in the same group.

Downlink Power (RDP) is proposed for this purpose. The RDP is the power required, relative to the current power allocation, for a given MS to meet its E_b/I_0 target.

Section 4.1.1 of Chapter 4 presented the Distributed Constrained Power Control algorithm. This algorithm is a special case of the general (and unconstrained) Distributed Power Control (DPC) [31] algorithm. The DPC updates the power according to the following rule:

$$p_{u,g}(n+1) = \frac{\gamma^t}{\gamma_{u,g}(n)} p_{u,g}(n) \tag{6.5}$$

where γ^t is the E_b/I_0 target, $\gamma_{u,g}(n)$ is the current E_b/I_0 and $p_{u,g}(n)$ is the current power for the u-th MS on the g-th group. This equation may be rewritten as

$$\frac{p_{u,g}(n+1)}{p_{u,g}(n)} = \frac{\gamma^t}{\gamma_{u,g}(n)} \tag{6.6}$$

Adding the amount of code resources assigned to an MS on a particular group, $c_{u,g}$, and substituting the current E_b/I_0, $\gamma_{u,g}(n)$, for the E_b/I_0 estimated by the cross-layer metric calculator, $\hat{\gamma}_{u,g}(n)$, an expression for the RDP may be obtained:

$$RDP_{u,g} = \frac{\gamma^t \, c_{u,g}}{\hat{\gamma}_{u,g}(n)} \tag{6.7}$$

From the above equation, the RDP may also be thought of as the resource product of the difference between the estimated and target E_b/I_0. The smaller the RDP, the better the quality an MS will experience, or equivalently, the smaller the power an MS will require to achieve a given target E_b/I_0.

The achievable E_b/I_0 for each group depends on a number of variables at the PHY and MAC layers. In calculating an estimate of the E_b/I_0, the cross-layer metric estimator must account for all these variables. The variables, and how the estimator accounts for them, are outlined in the following sub-sections.

6.2.3.1 Cross-layer metric: PHY layer effects

At the PHY layer the quality of the subcarriers assigned to a group will affect the frequency diversity and orthogonality achieved, in addition, use of a multicarrier detector will result in the enhancement of noise and out-cell interference. The in-cell received power level will be a function of the MS to BS gain.

Adopting the earlier notation of equation (4.7), the received E_b/I_0 at the output of a grouped multi-carrier receiver for the u-th MS attached to BS 0 and communicating on the g-th group may be expressed as:

$$\left(\frac{E_b}{I_0}\right)_{u,g} = \frac{\frac{W_{grp}}{R_{tm}} p_{0,u,g} L_{0,u}}{\alpha_{0,u,g} P_{0,g} L_{0,u} + I_{u,g}^{out} + N W_{grp}} \tag{6.8}$$

where the $I_{u,g}^{out}$ term accounts for the out-cell interference experienced at the u-th MS's location.

If PHY layer effects are accounted for and a single user receiver that performs chip-by-chip equalisation as outlined in chapter 2 is used, this equation may be re-written as equation (6.9). The latter equation involves two additional factors; a frequency diversity gain $\varphi_{0,u,g}$ and an interference plus noise enhancement factor $\kappa_{0,u,g}$ [119–121]. These two factors together with the orthogonality factor, $\alpha_{0,u,g}$, fully account for the effects of the PHY layer subcarrier quality on the received E_b/I_0.

$$\left(\frac{E_b}{I_0}\right)_{u,g} = \frac{\frac{W_{grp}}{R_{tm}} \varphi_{0,u,g} p_{0,u,g} L_{0,u}}{\alpha_{0,u,g} P_{0,g} L_{0,u} + \kappa_{0,u,g}(I_{u,g}^{out} + N W_{grp})} \tag{6.9}$$

The frequency diversity gain parametrises the diversity gain introduced by the channel, the orthogonality factor parametrises the degradation in orthogonality between orthogonal code sequences from the same BS caused by the channel after the MC-CDMA chip-by-chip equalisation is performed. Finally, the interference plus noise enhancement factor characterises the gain in out-cell interference plus noise caused by performing the chip-by-chip equalisation.

Expressions have been developed for each of these factors [119–121]. The diversity gain can be expressed as:

$$\varphi = \frac{1}{K}\Gamma_{\varrho_k}[0] + \frac{2}{K}\sum_{l=1}^{K-1}\left(1 - \frac{l}{K}\right)\Gamma_{\varrho_k}[l] \tag{6.10}$$

where K is the spreading gain and Γ_{ϱ_k} is the statistical correlation function of the the equalised channel coefficient ϱ_k in which k denotes the k-th subcarrier of a group:

$$\Gamma_{\varrho_k}[l] = \mathbb{E}\{\varrho_k[n]\varrho_k[n-l]\} \tag{6.11}$$

The equalised channel co-efficient is that channel coefficient after the channel has been equalised by applying a chip-by-chip equaliser (see Chapter 2 for further details).

The orthogonality factor can be written as:

$$\alpha = \Gamma_{\varrho_k}[0] - \mathbb{E}\{\varrho_k\}^2 \tag{6.12}$$

Finally, the interference plus noise enhancement factor can be expressed as:

$$\kappa = \mathbb{E}\{z_k[n]\}^2 \tag{6.13}$$

where z_k is the per chip channel equalisation coefficient.

The cross-layer metric calculates the effect of the PHY layer subcarrier channel coefficients by estimating the above three factors. However, these expressions utilise expectations and hence ensemble averages. For the estimation of each of the factors the cross-layer metric calculator replaces the ensemble average with sample averages. This begs an examination as to whether sample averages are adequate for this purpose; this question is answered by way of simulation.

A simulation corresponding to the set up in chapter 3 was run, with the exceptions that a spreading length of 16 instead of 8 was used and a single user MMSE detector was used in place of the ML-MUD receiver. The objective of the simulation was to verify that sample averages were sufficient to allow a good estimate of the degradation in the received E_b/I_0 caused by the channel. The calculated E_b/I_0 using sample averages to calculate the various factors was compared with the E_b/I_0 measured by the subspace-based SIR estimator algorithm developed by Ramakrishna *et al.* [122].

Figure 6.5 shows that the measured and calculated E_b/I_0, with the exception of a few outliers, show excellent agreement. Thus it is concluded that the use of sample averages by the cross-layer metric estimator is sufficient to characterise the PHY layer degradations caused by the channel in an MC-CDMA system.

6.2.3.2 Cross-layer metric: MAC layer information

The expression for the E_b/I_0 of equation (6.9) has as some of its variables in the denominator the total in-cell transmitted BS power, P_0 (which is related to the in-cell interference via the orthogonality factor), and the out-of cell interference I^{out}. These terms are related to MAC layer parameters, namely the number of users that are allowed to access the system in the cell

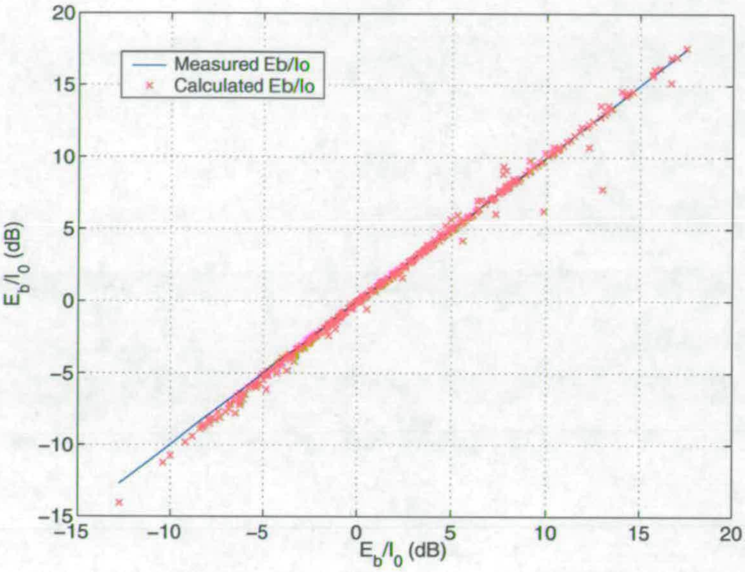


Figure 6.5: *Use of sample averages for calculating frequency factors and thereafter* E_b/I_0

of interest and in the surrounding, interfering, cells.

For the purposes of estimating the in-cell interference for the cross-layer metric calculation an assumption is made that all available codes are allocated. The number of codes available for allocation is drawn from the DCE capacity estimate. Although the assumption that all codes are utilised may not be met for a lightly loaded system, the assumption will generally be true for the most interesting cases of systems that have moderate to high loads. The cross-layer metric calculator also assumes that each of these codes is assigned unit power.

As was explained in chapter 5, one of the measurements required by the DCE is the other-cell downlink relative interference I_{out}/S_0. Combined with the knowledge of the total BS power and the MS-BS gain, use of this measurement allows the cross-layer metric estimator to obtain a good estimate of the out-cell interference received at a particular MS location:

$$\hat{I}_{u,g}^{out} = \hat{P}_{0,g} L_{0,u} \left(\frac{I_{out}}{S_{0,g}} \right) \tag{6.14}$$

where $\hat{P}_{0,g}$ is an Exponential Moving Average of the total BS power assigned to a particular group:

$$\hat{P}_{0,g} = \vartheta \hat{P}_{0,g} + (1 - \vartheta) P_{0,g} \tag{6.15}$$

in which ϑ is the forgetting factor.

In the event that an MS does not have a measurement of the other-cell relative downlink interference, cell wide averages for that particular group may be used.

6.2.3.3 Summary of cross-layer multi-group allocation

The cross-layer aware multi-group allocation can now be summarised as follows:

1. The multi-group allocation receives from the MG-GPS scheduler information as to how much resource should be allocated to each MS.

2. For each group of the active MS, the cross-layer metric calculator takes per subcarrier channel estimates [2] and uses these to calculate the frequency diversity gain, orthogonality factor and interference plus noise gain. Information from the DCE is used in conjunction with these factors to calculate the in-cell and out-cell interference. All these estimates are then combined in equation (6.9) to provide an estimate of the received E_b/I_0, $\hat{\gamma}_{u,g}(n)$. That estimate is then used in equation (6.7) to calculate the cross-layer metric: the RDP.

3. The RDP is used in the optimisation (the general notation from chapter 3 is reused here)

$$\begin{aligned} minimise \qquad & \sum_{g=0}^{G-1}\sum_{u=0}^{U-1} RDP_{u,g}\left(=\frac{\gamma^t\, c_{u,g}}{\hat{\gamma}_{u,g}(n)}\right) \\ subject\ to: \quad \sum_{g=0}^{G-1} c_{u,g} &= \text{MG-GPS Allocation for the } u\text{-th MS} \\ & \quad \forall\, u \in [0,1,\ldots,U-1] \\ \sum_{u=0}^{U-1} c_{u,g} &\leq C_g \\ & \quad \forall\, g \in [0,1,\ldots,G-1] \\ c_{u,g} &\in [0,1,\ldots,C_g] \end{aligned} \tag{6.16}$$

As this can readily be seen to be an integer linear program, the mixed probabilistic-greedy algorithm of section 3.3.2 is applicable to this problem with some minor modifications, namely that the optimisation is changed from a maximisation to a minimisation and the algorithm changed to allow for non-binary (i.e. '0' or '1') integer allocations.

[2] As was explained in chapter 3, it is assumed that a TDD system is used and that subcarrier channel estimates are readily available by virtue of their being needed by the BS to decode data from each MS.

4. After the local search algorithm reaches a solution, the total RDP for each group is checked, if it is found to exceed a threshold RDP_{thresh} then the offending MS in that group has a resource unit deducted from its allocation. This continues until all groups have an RDP less than the threshold. The algorithm then goes back to step 3.

5. Once all groups have been allocated in such a manner that no group exceeds the RDP threshold, the cross-layer algorithm proceeds to inform the BS of the groups and codes that have been assigned to each MS. The BS will then transmit packets from each MS on the selected resources.

6.2.4 Performance Evaluation

A system level simulation with PHY layer interactions was performed to evaluate the performance of the cross-layer resource scheduler. The same model as presented in section 6.2.3.2 was used as a link between the system and PHY layers in the simulation. The simulation parameters were the same as those of Table 4.4 of section 4.3 with the exception that the mean time between packets was changed to 40 μs giving a peak bit rate of 100 Mbps. The PHY channel model used was the exponential model of section 3.4.1. The RDP_{thresh} was set to 20.

The cross-layer resource scheduler (referred to in the results as MG-GPS+Cross-Layer) is compared to the Fixed Resource (Fix-RR), Fair Resource (FaR) and basic MG-GPS schedulers introduced in chapter 4. In addition to these schedulers, it is also compared to the MG-GPS scheduler utilising the Local Search grouping and subcarrier allocation algorithm of chapter 3 (referred to in the results as MG-GPS+LocalSearch). The MG-GPS+LocalSearch is an example of an algorithm with loose coupling between the PHY and MAC levels.

Figure 6.6 shows the average packet call throughput of these scheduling algorithms. As in chapter 4, the various MG-GPS based schedulers outperform the other schedulers particularly at high loads. Focusing on the MG-GPS based schedulers, it can be seen that at low loads all these schedulers exhibit roughly the same level of performance, this is despite the MG-GPS+LocalSearch and the MG-GPS+Cross-Layer adding extra complexity by trying to optimise the allocation of MSs to groups. This can be explained by noting that at low loads there are very few users requesting channel access at any one scheduling instant, there is thus little opportunity for either of the algorithms to take advantage of multi-user diversity. In fact, for those cases where there is only one user requesting channel access, the allocation decision may, if the user has enough backlogged data, default to allocating most/all of the groups and codes to that user i.e. no optimisation is possible.

As the load increases the number of users requesting channel access at any one scheduling instant goes up proportionally, giving the optimisation based allocation algorithms leeway to take advantage of the multi-user diversity presented. The MG-GPS+LocalSearch algorithm is able to improve the throughput by several percentage points. On the other hand, the MG-GPS+Cross-Layer algorithm with its tighter coupling between PHY and MAC layers is able to offer a throughput improvement of between 10% to 25% at moderate to high loads.

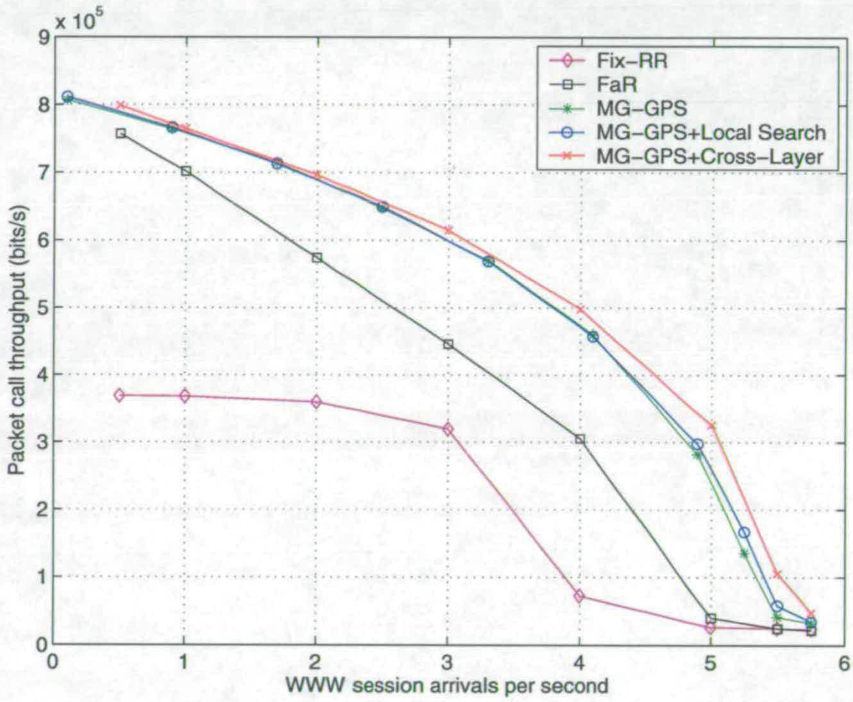


Figure 6.6: *Average packet call throughput performance for various scheduling algorithms*

Figure 6.7 shows the average packet delay for the scheduling algorithms. At low loads the non MG-GPS algorithms exhibit the same performance as the MG-GPS, this situation changes at loads of 3 and 4 session arrivals per second where the performance of the Fixed Resource and Fair Resource schedulers respectively, diverge rapidly from those of the MG-GPS based algorithms. From this figure it can be seen that the MG-GPS+Cross-Layer exhibits slightly better delay performance than either the MG-GPS+LocalSearch or plain MG-GPS schedulers.

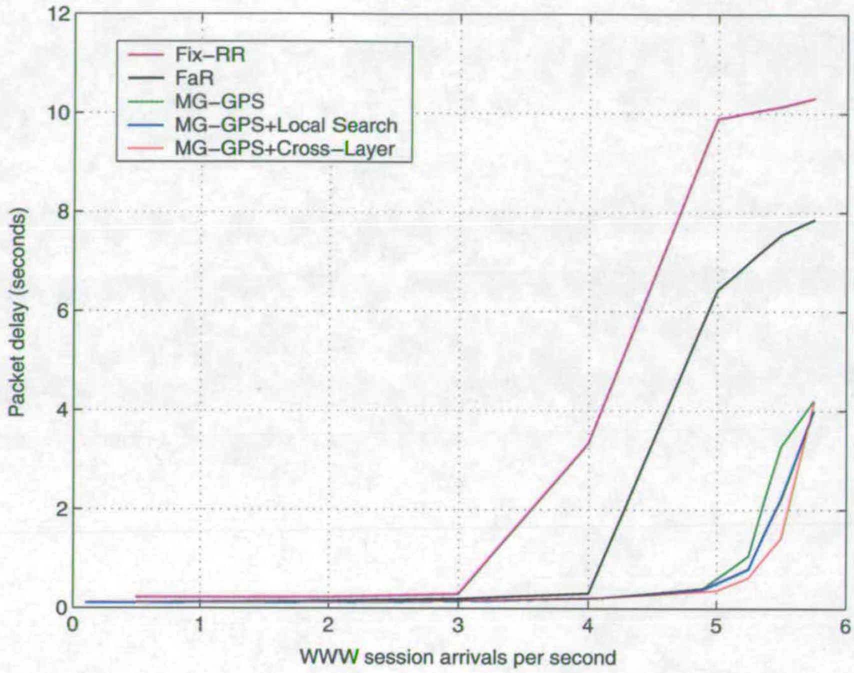


Figure 6.7: *Average packet delay performance for various scheduling algorithms*

6.3 Conclusions

In this chapter the concepts explored in earlier parts of the thesis were brought together and ways to integrate them were explored. For real-time traffic, a joint call admission control and multi-group allocation algorithm that based its decision on the capacity estimates delivered by the DCE was explored. This algorithm was found to perform better than using no admission control, and the same as a joint power threshold admission control and multi-group allocation algorithm. However the DCE algorithm required no setting of thresholds and is suitable for use in a variety of propagation environments due to its ability to make run-time capacity estimates.

The scheduling of data and resources for non-real time traffic was examined. A new algorithm that was based on the earlier proposed MG-GPS scheduler and that used various types of information from the DCE and PHY layers to schedule data and resources in a cross-layer aware manner was proposed. This algorithm was evaluated against other sets of scheduling algorithms, and in particular against another MG-GPS scheduler with looser cross-layer links. The cross-layer aware resource scheduler was found to improve the QoS in terms of packet call throughput and delay experienced by non-real time users at moderate to high traffic loads.

Chapter 7
Conclusions

In this chapter a detailed summary of the contents of the thesis is provided, some conclusions from the work are presented and finally limitations and areas for future research are highlighted.

7.1 Summary and Conclusions

7.1.1 Introduction

In the first Chapter, the key aim of the thesis was stated; to investigate the management of resources in an MC-CDMA system, leveraging estimates of the wireless resource to enhance the QoS experienced by users. This led to Chapter 2 in which the usage of multicarrier transmission in conjunction with CDMA multiple access techniques was explained in depth. Quality of service and methods to provide quality of service in the different layers of a wireless system was discussed; a common feature of all the methods is that they depend on estimates of the state of the wireless resource. The concept of the resource metric estimator (RME) and how it may aid in the provision of QoS was explained.

7.1.2 PHY layer subcarrier allocation

The major part of the thesis began with an examination of physical layer allocation for a grouped MC-CDMA system. In particular, the manner in which users should be allocated to groups and subcarriers in order to provide QoS in terms of the bit error rate was addressed. Several metrics upon which to base the allocation decision were presented, it was decided that the most suitable metric at the PHY layer was the received per subcarrier amplitudes. The introduction of a subcarrier and group sharing factor allowed the group and subcarrier allocation problem to be formulated as a multi-dimensional optimisation in which the allocation of subcarriers to groups and users to groups was simultaneously solved. This formulation was found to be impractical due to the large number of constraints it would entail. A sub-optimal approach was warranted that split the allocation into two steps; a grouping of subcarriers followed by an optimisation of the allocation of users to those groups. The resulting optimisation problem was recognisable as an integer linear program belonging to the class of $\mathcal{NP}$-complete problems.

Two algorithms were proposed to solve the integer linear program. Firstly a branch and bound based technique that used specialised heuristics in an attempt to reduce the size of the search space and the time taken to find a solution. The second algorithm was a local search based algorithm that used a mixed probabilistic-greedy algorithm to navigate the search space. These algorithms were compared to two others from the literature; a fixed allocation algorithm and a swapping algorithm. A simulation study of the different algorithms showed the integer linear program based algorithms were able to provide larger mean per subcarrier amplitudes and fewer low quality subcarriers than the other algorithms. The improvement in subcarrier amplitudes was due to the way the integer linear program based algorithms were able to successfully navigate the search space, unlike the case for the swapping algorithm which effectively resulted in a reduction in the quality of subcarriers available to users in a group. It was shown that the gains in subcarrier amplitude by the integer linear program based algorithms translated into better QoS in terms of BER for more users. Of note in the case of the mixed probabilistic-greedy local search algorithm is that this QoS improvement came with a complexity that was 20% of the complexity of the swapping algorithm and a minuscule 0.001% of the complexity of the branch and bound based algorithm.

7.1.3 MAC layer QoS provisioning: power control

The focus then moved to examine how QoS could be provided at the MAC layer by judicious management of resources in a multi-group MC-CDMA system. The first QoS provisioning technique examined was power control. Three power control variants based on distributed constrained power control (DCPC) were discussed; DCPC with maximum per MS power limits, DCPC with maximum total per group power limits and a hybrid DCPC with both maximum per MS and total per group power limits. The simulation results revealed that the best power control scheme was the hybrid DCPC due its flexibility in giving more power to MSs with a medium quality channel whilst at the same time ensuring that those MSs with a bad quality channel were cut off before they were able to instigate a power race. The power control algorithm with maximum per group total power constraints behaved poorly for precisely this reason, whereas the algorithm with per MS constraints lacked the flexibility to selectively assign more power to some deserving MSs.

7.1.4 MAC layer QoS provisioning: multi-group allocation

The mix of data services in next generation networks will require the provision of multiple data rates. A multi-rate allocation model termed multi-group allocation (MGA) was described and three multi-group allocation schemes outlined; random, least-resource and least-interference. The performance of these three algorithms was examined in terms of their per MS and per MS's group probability of outage. It was found that the least interference algorithm performed better out of the three due to its improved selection of groups that have the smallest amounts of interference. The random algorithm had a similar per MS, per group outage value to the least resource algorithm, however its per MS outage was vastly different to that of the least resource algorithm. This was due to the random algorithm's allocation procedure resulting in very different per group interference values. Thus it can be concluded that if an application e.g. a real-time service has strict bandwidth guarantees, then the best multi-group allocation model would be one where the groups are allocated on a super-group basis.

7.1.5 MAC layer QoS provisioning: scheduling

The final element of work on MAC layer QoS provisioning was an evaluation of scheduling for multi-group CDMA. A new multi-group generalised processor sharing scheduler was presented. The performance of this scheduler was compared to other schedulers, assumptions were made as to the maximum availability of codes in a group. The MG-GPS scheduler was shown to perform better than other schedulers due to its ability to reassign unused resource to needy users. The key finding of this section was that while in general the MG-GPS performed better than the other schedulers studied, the assumption as to how much code capacity was available in each group had a significant effect on the QoS in terms of packet call throughput and packet delay.

7.1.6 Dynamic Capacity Estimator

The interference in CDMA systems is time-varying and interference limited, proper estimates of the capacity can be a useful aid to radio resource management algorithms that seek to maximise their use of the available capacity. The issue of how to estimate the capacity in MC-CDMA systems was addressed. The first step in the investigation into capacity estimation was the estimation of the single-cell capacity for MC-CDMA using a constraint model to develop a feasible condition for the power solution for each user. Simulation results showed the achievable capacity for different traffic services and values of the power control error.

A drawback of this capacity model is that it is for a single-cell only and does not consider the statistical gain possible from the inactivity of various traffic sources. These issues were addressed in the subsequent sections where a capacity analysis was highlighted that took into account multiple cells and different traffic models. The capacity analysis made use of the other-cell downlink interference ratio as its primary metric; this is the ratio of the total interference received from surrounding cells versus the total in-cell received power. A new method to estimate these values called the null pilot assignment plan was proposed. This method required the insertion of null pilots into a transmit frame in specific places for each cell to allow all the MSs in a given cell to measure the other-cell interference. Another metric used in the capacity estimation scheme was the information on whether or not an MS was at the edge of a cell. A cell edge detection procedure was proposed that utilised the same metrics (the MS-BS gain) as used in making handoff decisions. To cope with the discrepancy introduced by taking edge results from a region wider than the very narrow cell edge assumed in the interference analysis a degree of confidence level (DCL) in the capacity estimate was proposed.

Simulations of the null pilot assignment plan and several averaging methods showed the plan was able to offer good estimates of the other-cell downlink relative interference and that the best averaging method was a median averaging which was able to cope well with outliers. This new framework for estimating the various metrics and using them in the equation derived from the traffic and interference analysis to estimate the capacity was termed the dynamic capacity estimator (DCE). The performance of the DCE was evaluated for different orthogonality factors and levels of physical layer imperfections introduced by the metric estimation procedure. The DCE was found to be able to estimate the capacity at time scales of greater than 1 second with the DCL improving the performance of the estimate, the exact improvement depended upon the orthogonality factor.

7.1.7 Cross layer resource management

The final contribution of the thesis was to integrate the disparate MAC and PHY techniques for QoS provisioning into one algorithm for the provisioning of QoS for real- and non-real time traffic across layers. For real-time traffic the DCE was integrated with the least resource multi-group allocation algorithm to provide a joint multi-group admission control algorithm that aimed to ensure the quality of ongoing calls was preserved. This algorithm was compared to not using any admission control and an integrated power threshold/least resource algorithm. The integrated DCE and power threshold algorithms were shown to have better performance than not using any admission control. In relation to each other, both the algorithms showed similar levels of performance, the DCE based algorithm having a slight edge in preserving the probability of ongoing calls. The DCE based algorithm also has the advantage in that it does not require the setting of an admission threshold.

For non-real time traffic the DCE was used to provide the MG-GPS scheduler with per group code capacity estimates. For the allocation of MSs to resources, a new metric termed the relative downlink power (RDP) was proposed. The RDP is the power required, relative to the current power, for an MS to meet a given E_b/I_0 target. The RDP incorporates various phenomena from the different layers; from the PHY layer it includes the MS-BS gain and information on the degradation caused by the subcarrier channel gains; from the MAC layer it includes the in-cell and out-cell interference, values which are dependent on the MAC decisions as to how many users are accessing the channel at a given time instance. Thus the RDP allows for a cross-layer aware allocation of resources.

The RDP was placed into an expression in which the allocation of MSs to group and code resources was optimised to minimise the total RDP. This optimisation turned out to be quite similar to that developed for the PHY layer grouping and subcarrier allocation algorithm of Chapter 3. As a result, with some few modifications, the mixed probabilistic-greedy algorithm of that chapter was easily applied to solve the RDP optimisation. The performance of the DCE enhanced MG-GPS scheduler with RDP based cross-layer aware multi-group allocation was evaluated and compared to schedulers previously introduced in Chapter 4, it was also compared to a MG-GPS scheduler utilising the grouping and subcarrier allocation algorithm of Chapter 3. The MG-GPS based schedulers were found to have significant performance improvements over the other schedulers tested. Out of the MG-GPS schedulers, it was seen that at low loads the optimised allocation algorithms were unable to take advantage of the multi-user diversity due to a dearth of multiple users requesting access to the system during any one time instance. As the load increased, there was an attendant increase in users simultaneously requesting channel access and thus better opportunities to exploit the multi-user diversity. It could be seen that cross-layer aware algorithm resulted in performance improvements over the other MG-GPS algorithms due to its tighter coupling between MAC and PHY layers giving it more opportunity to exploit the multi-user diversity.

7.1.8 Conclusion

Finally, in conclusion, there has been to date little work done on the management of resources in multicarrier CDMA systems, most work instead focusing on PHY layer activities. This thesis has made a contribution to the understanding of resource management in MC-CDMA. Several algorithms have been proposed to manage resources at the PHY and MAC layer to provide increased QoS. A capacity estimation scheme has been defined and techniques to measure metrics for this scheme investigated. The developed capacity estimation scheme has then been used in conjunction with multi-layer aware algorithms to make decisions across the MAC and PHY layers that result in better QoS performance.

7.2 Limitations and areas for future research

There are number of limitations to the work carried out in this thesis, these are primarily focused on the DCE. Some suggestions are given as to how future research work may address them and other interesting topics that may spawn from this research.

The DCE was found to need a filter length corresponding to a length of more than a second to produce a good capacity estimate. As has already been outlined before, the interference situation in CDMA systems can be quite fast-changing resulting in similar changes to the achievable capacity. Due to the length of this filter these changes may be averaged out, algorithms that use the resulting capacity estimate may thus not be able to take full advantage of short-lived favourable interference conditions when they occur. It should be noted that in the DCE scheme consideration was not given to collecting estimates of the other cell downlink relative interference for inactive groups (i.e. groups that an MS is not actively transmitting on). If a protocol was instituted whereby such groups periodically measured the ratio (say at an interval of every several hundred milliseconds) then the length of the averaging filter could be reduced as more ratio estimates would be available on which to estimate the capacity.

The traffic analysis does not include an expression for the delay a non-real time flow experiences, the traffic model for non-real time traffic could be modified to include a delay bound.

The degree of confidence was calculated off-line, future research may consider developing an on-line estimator of this value, perhaps by using a hidden Markov model to model the capacity states and associated degrees of confidence.

In actual systems, real- and non-real time traffic do not exist in isolation. The chapter on resource scheduling and control did not include an evaluation of the interaction between these traffic types. Future work could consider this issue and further how congestion control may be performed such that the amount of non-real time traffic allowed to access the network is controlled dynamically depending on the free capacity and the amount of real-time traffic in the network.

Appendix A
Publications

A.1 Published papers

M. Tabulo, D. Laurenson, S. McLaughlin and E. A. Al-Susa, *"A Linear Programming Algorithm for a Grouped MC-CDMA System,"* in Proc. of IEEE VTC Fall 2003, Orlando, USA, October 2003.

Y. Lee, M. Tabulo, D. Laurenson and S. McLaughlin, *"Radio Resource Allocation with Resource Metric Estimation for Multimedia CDMA Systems,"* in Proc. of the 8th International Conference on Cellular and Intelligent Communications(CIC), Seoul, Korea, October 2003.

A.2 Submitted manuscripts

Y. Lee, M. Tabulo, D. Laurenson, E. Al-Susa and S. McLaughlin, *"Radio Resource Allocation with Resource Metric Estimation for Wireless Multimedia CDMA Systems,"* submitted to Kluwer Journal on Wireless Networks.

A.3 Manuscripts in preparation for submission

M. Tabulo, D. Laurenson, S. McLaughlin and Y. Lee, *"Power Control and Dynamic Channel Allocation for Multi-Group MC-CDMA,"* in preparation for submission to IEE Electronics Letters.

M. Tabulo, D. Laurenson, S. McLaughlin and Y. Lee, *"A computationally efficient Mixed Probabilistic-Greedy algorithm for MC-CDMA subcarrier allocation,"* in preparation for submission to IEEE Communications Letters.

M. Tabulo, D. Laurenson, Y. Lee and S. McLaughlin, *"Resource Control and Scheduling for MC-CDMA,"* in preparation for submission to IEE Proceedings-Communications.

A Linear Programming Algorithm for a Grouped MC-CDMA System

Moti Tabulo, David Laurenson, Stephen McLaughlin and Emad Al-Susa

Institute for Digital Communications,
School of Engineering and Electronics
University of Edinburgh,
Edinburgh EH9 3JL, UK
Email: Moti.Tabulo@ee.ed.ac.uk

Abstract— **In this paper we present an adaptive linear programming based algorithm to exploit the multi-user diversity present when communicating to multiple users on a multicarrier CDMA (MC-CDMA) downlink. Utilising the received channel characteristics as the adaptation criteria, we adaptively allocate subcarriers to groups and thereafter adaptively assign users to those groups (the grouping is done to allow the use of a maximum likelihood multi-user detector – ML-MUD). The objective in performing this adaptive grouping and allocation is to improve the quality of service (QoS) – in terms of bit error rate (BER) – seen by users in the system. Employing a simulation study, we compare the results of the proposed linear programming algorithm with the results of other approaches from the literature. These results show that the proposed algorithm gives better QoS performance, for more users in the system, than previous approaches.**

I. Introduction

Recently, combined OFDM (Orthogonal Frequency Division Multiplexing) and DS-CDMA (Direct-Sequence Code Division Multiple Access) transmission systems have received significant interest as potential multiple access methods for the next generation of high data rate wireless systems. OFDM modulation provides robustness against multipath fading as well as good spectral efficiency, while the CDMA part allows the transmission of multiple users in parallel by allocating a distinct spreading code to each user.

There are two main types of combined OFDM/CDMA schemes. One converts the original data stream from serial-to-parallel, spreads the converted data using a spreading code and then modulates a different subcarrier with the resulting spread data stream [1](this is similar to a normal DS-CDMA system where the spreading operation is performed in the time domain). In the second type, the original data stream is spread using a spreading code and then each chip of the resulting spread data stream is used to modulate a different subcarrier [2], [3](this is equivalent to the spreading operation being performed in the frequency domain). These two schemes are commonly known as MC-DS-CDMA and MC-CDMA respectively. In this paper we consider the latter.

Orthogonal spreading codes can be used in an MC-CDMA system, but in a frequency selective channel the attenuations in subcarriers destroy the orthogonality of the codes [2] resulting in Multiple Access Interference (MAI). A Multi-User Detector (MUD) such as the Maximum-Likelihood MUD (ML-MUD) can be used to mitigate the effects of MAI. To reduce the complexity of the ML-MUD (which grows exponentially with the number of users), a grouped MC-CDMA scheme was proposed in [4] whereby the users in the system were subdivided into a set of smaller groups and each group transmitted on a subset of all the subcarriers. They employed a fixed allocation of users to groups and subcarriers, i.e. channel information was ignored and the users were simply divided amongst the groups/subcarriers.

In this paper we utilise the ML-MUD and grouping architecture, but we develop a linear programming (LP) algorithm to adaptively – according to the channel characteristics – perform the grouping and subcarrier allocation. By doing this, we are able to take advantage of the channel diversity that exists between users, resulting in significant performance improvements. The paper is organised as follows; in section II we describe grouped MC-CDMA in more detail and present our system model. Section III explains how the grouping and subcarrier allocation problem is fitted to a linear programming approach. The Branch and Bound technique used to solve the linear program is outlined in section IV. Section V presents simulation results showing the performance of the proposed linear programming approach. Conclusions are drawn in the final section.

II. System Model

Our system model is as shown in figure 1. We concentrate on the downlink. The transmitter comprises a user grouping algorithm that divides the U users in the system into G groups, spreaders, a subcarrier allocation algorithm that allocates K subcarriers to each group and finally an inverse discrete Fourier transform (IDFT) for OFDM multiplexing. The receiver for the u^{th} user belonging to the g^{th} group comprises a discrete Fourier transform (DFT) for OFDM demultiplexing, a subcarrier selection block to select the K subcarriers assigned to this user's group and an ML-MUD.

The grouping and allocation algorithm allocates subcarriers to groups resulting in a subcarrier grouping vector

$$SUBC_g = [subc_{0,g},\ subc_{1,g}, \ldots,\ subc_{K-1,g}] \quad (1)$$

for $g = 0, \ldots, G-1$ that shows which K subcarriers are assigned to the g^{th} group. The algorithm also assigns a grouping vector

$$GRP_g = [grp_{0,g},\ grp_{1,g}, \ldots,\ grp_{\lceil \frac{U}{G} \rceil -1,g}] \quad (2)$$

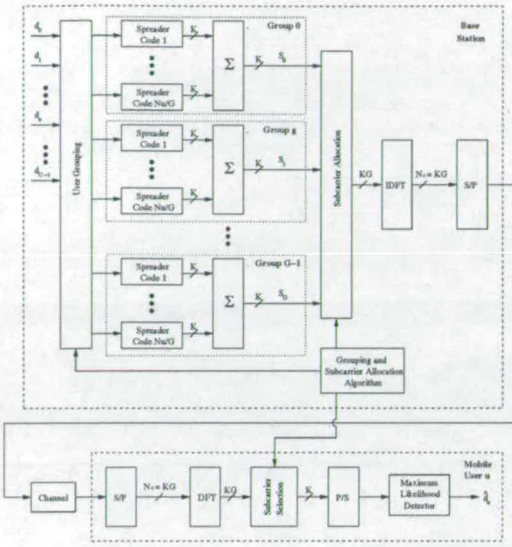


Fig. 1: Downlink adaptive Grouped MC-CDMA receiver

for $g = 0, \ldots, G-1$ that indicates which user u is in the g^{th} group. $\lceil x \rceil$ indicates that x is rounded up to the nearest integer.

After grouping and allocation, the data bits d_u of each user u in the g^{th} group are spread by a spreading code vector C_u of length K. The chip streams of all users in the g^{th} group are then added synchronously in the base station to yield the vector $S_g,\ g = 0, 1, \ldots, G-1$, with

$$S_g = \sum_{u \in GRP_g} d_u C_u = [s_{g,0},\ s_{g,1},\ \ldots,\ s_{g,K-1}] \quad (3)$$

The S_g is OFDM modulated using an IDFT and passed through a frequency selective Rayleigh fading channel. We assume that the duration of the cyclic prefix is greater than the multi-path spread of the channel and that the channel varies slowly compared to the symbol duration. We also assume the presence of additive white gaussian noise (AWGN). The received signal is then

$$\begin{aligned} r(t) &= \sum_{g=0}^{G-1} \sum_{k=0}^{K-1} \alpha_{u,subc_{g,k}} s_{g,k} \exp(j2\pi f_{subc_{g,k}}) \\ &\quad + n(t), \qquad t \in [0, T] \end{aligned} \quad (4)$$

where $\alpha_{u,subc_{g,k}}$ accounts for the effects of frequency non-selective fading for the k^{th} subcarrier, of the desired user u that is in the g^{th} group. $f_{subc_{g,k}}$ indicates the frequency of the k^{th} subcarrier in the g^{th} group. T is the OFDM symbol duration.

At the receiver, the signal is OFDM demodulated using a DFT. We assume that the receiver has knowledge of which group it belongs to and which subcarriers are allocated to that group. With this knowledge, the receiver can extract the desired subcarriers from the set of all received subcarriers. Therefore, the received signal, R_g, can be written as a vector of K components with

$$R_g = H_g S_g^T + N \quad (5)$$

where H_g is a diagonal matrix that describes the flat Rayleigh fading on the subcarriers assigned to group g. S_g^T represents the transposed transmitted sequence and N is the noise vector.

As mentioned earlier, a ML-MUD detector is used in the receiver. We denote $V_j, j = 1, \ldots, 2^K$ as the set of all possible transmitted sequences. Then the estimate of the received signal $\widehat{S}_g = V_j$ if the Euclidean distance e_j^2 between the received and all possible transmitted sequences is minimised. i.e.

$$e_j^2 = min|R_g - H_g V_j^T|^2 \quad (6)$$

The estimate of the data bit of the desired user, $\widehat{d}_u$, can be obtained from $\widehat{S}_g$.

III. Grouping and Subcarrier Allocation

There are several criteria that may be used to execute the grouping and allocation of subcarriers, these include the Global Received Amplitude and Bit Error Rate (BER). The rationale behind the Global Received Amplitude is that choosing those subcarriers that are least faded (and hence have higher amplitude) will aid in the diversity reception of the chips on those subcarriers. Correspondingly a low received amplitude will hinder diversity reception of the transmitted chips. With BER as a criterion, the grouping and allocation could be performed to maximise the BER seen by users in a cell. As for the availability of the two estimates; the received amplitude per subcarrier should already be available in the base station as these values (in the form of channel estimates) are required by the ML-MUD for coherent detection. The BER on the other hand needs to be estimated and it is computationally intensive to do so. Hence we select the Global Received Amplitude as the criterion for grouping and allocation due to the ready availability of estimates.

The optimal approach would be to optimise the grouping and subcarrier allocation simultaneously (i.e. a multi-dimensional optimisation), but this is computationally infeasible due to the large number of constraints required.

We implement a two stage sub-optimal, but computationally feasible optimisation algorithm. Firstly, subcarriers are clustered into groups, the second stage is to assign users to these groups. Each of these stages is explained further below.

The subcarriers are clustered into groups either randomly, or, based on how different the subcarrier amplitude profiles across users are. Figure 2 shows an example of subcarrier amplitudes for an 8 user, 8 subcarrier system, it is used to illustrate the concept of the subcarrier amplitude across users. The conventional subcarrier amplitude profile (for one user, or, equivalently across subcarriers) is that taken along the y-axis, while the profile we consider, the subcarrier amplitude profile across users, is that amplitude profile taken along the x-axis.

As an alternative to grouping randomly, or based on how different the subcarriers are, we could have decided to cluster subcarriers into groups based on how similar their subcarrier amplitude profiles were. Such a scheme may be attractive from the point of view that it would be easier for the second stage to converge to a solution if the subcarriers in a group were similar. However, using this scheme would have meant that some groups would be composed entirely of subcarriers that have low amplitudes (i.e. highly faded) and thus whatever algorithm employed in the second stage to assign users, it would be unable to surmount the poor initial selection of subcarriers available to those groups.

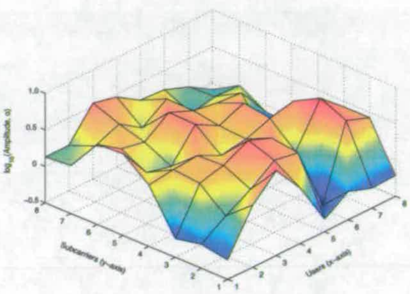


Fig. 2: Mesh showing subcarrier amplitudes across both users and subcarriers

In the second stage, using the subcarrier groups from the first stage, we assign users to these groups by solving the following maximisation problem:

$$maximise \quad \sum_{u=0}^{U-1} \sum_{k=0}^{GK-1} \alpha_{u,k}\rho_{u,k} \tag{7}$$

$$subject\ to: \quad \sum_{k=0}^{GK-1} \rho_{u,k} = K \qquad \forall\, u \in [0,1,\ldots,U-1] \tag{8}$$

$$\sum_{u=0}^{U-1} \rho_{u,k} \leq \left\lceil \frac{U}{G} \right\rceil \qquad \forall\, k \in [0,1,\ldots,GK-1] \tag{9}$$

$$\sum_{k \in SUBC_g} \rho_{u,k} = (1 - b_{k+uGK}) \left\lceil \frac{U}{G} \right\rceil \qquad \forall\, u \in [0,1,\ldots,U-1] \quad \forall\, g \in [0,1,\ldots,G-1] \tag{10}$$

$$\rho_{u,k} \in [0,1] \tag{11}$$

$$b_{k+uGK} \in [0,1] \tag{12}$$

$\rho_{u,k}$ is a sharing factor that can only take the integer values '0' or'1'. If it is equal to '1' it indicates that the the u^{th} user is assigned the k^{th} subcarrier. The first constraint (equation (8)) limits the number of subcarriers allocated to a user, while the second constraint (equation (9)) limits the total number of users that can be allocated to a group. The purpose of the third constraint (equation (10)) is to ensure that the subcarriers from stage 1 of the algorithm are always grouped together. Essentially this constraint says that if a user u is to be assigned a subcarrier that is in a group $SUBC_g$, then that assignment can only proceed if the user is assigned to all the subcarriers in $SUBC_g$. This constraint makes use of slack variables b_{k+uGK} that can only take the integer values '0' or '1'.

Equations (7-12) form a linear program with integer variables. The procedure used to solve it is described in the next section.

IV. Solving the Linear Program

Several methods exist for solving linear programs of the type outlined in section III such as Branch and Bound, Dual Ascent and Simulated Annealing [5]. These algorithms may be broadly divided into two categories; general purpose algorithms and special purpose algorithms. General purpose algorithms are fairly generic in nature and applicable to a wide set of problems. Special purpose algorithms may outperform their general purpose counterparts in terms of computational complexity with the caveat that to achieve the best performance they have to be designed to fit a particular problem.

We select the linear programming based Branch and Bound algorithm as the method to solve the linear program of section III due to the generic nature of the Branch and Bound algorithm. However we introduce some specialisations to make the algorithm more efficient.

Branch and Bound is a general purpose linear programming based tree search that systematically enumerates feasible solutions such that the optimal integer solution is found without having to enumerate all the possible solutions [5]. In the Branch and Bound approach, the integrality constraints are relaxed and variables are allowed to take on values in the range of 0 to 1. With the integrality constraints relaxed, the optimisation problem of section III becomes a straightforward linear program with linear constraints that can be solved using conventional techniques such as the simplex algorithm [6]. This may result in a solution with values that are non-integer and therefore we now seek to add tighter bounds to the problem to force those variables with non-integer values to take on integer values. To tighten the bound at a node, we select a variable with a non-integer value and force it to take on integer parts. As we are unable to tell which integer value the variable must have to give an optimal solution, we examine all the possible integer values that variable can take. In the case of a problem with '0' or '1' values, this will result in the root node branching into two descendant nodes, the left descendant with the constraint that the variable takes a '0' value and the right descendant with the constraint that the variable takes a '1' value. These two nodes are then solved, the descendant node with the lowest solution is pruned i.e. it is not considered anymore in the Branch and Bound process. This branching process can be carried out recursively; each of the descendant nodes resulting in two additional nodes. Eventually after enough bounds are placed on the variables, a solution to the relaxed maximisation problem whose variables are all integers with '0' or '1' values is obtained. The value of the best integer solution found so far is retained. If we have no other nodes to examine then this is the solution to the maximisation problem.

The specialisations we introduce to the Branch and Bound procedure aim to minimise the time required by the technique to converge to a good solution, and are as follows:

1) *Active node selection:* To select which node is selected for processing next, we always select the node with the highest bound; this is also known as a 'best-first' search.
2) *Branching variable selection:* To choose branching variables, we always select the variable whose value in the solution is furthest from being integral. The direction of processing is always the left branch.
3) *Additional branch pruning rule:* We define a grouped MC-CDMA feasible solution as one in which subcarriers exist in only one group and all the constraints about users/subcarriers per group are met. Then, as an addition to pruning branches based on their bounds, we also prune those branches that result in a grouped MC-CDMA feasible solution regardless of whether or not all the slack variables are integers. (For a grouped MC-CDMA feasible solution it is not necessary for the slack variables, b_{k+uGK}, to be integers so long as the main constraints are met and the main variables, $\rho_{u,k}$, take on integer values).

V. Simulation Results

Table I outlines the parameters used in the simulation.

Parameter	Value
Total Number of Users, U	64
Number of Groups, G	8
Number of Subcarriers per Group K	8
Total Number of Subcarriers	64
Channel	Six tap Frequency Selective
Spreading Code Type	PN

TABLE I: Simulation Parameters

We simulate four grouping/allocation policies. Firstly, a fixed allocation as described in section I.

The second grouping/allocation policy we simulate is the swapping algorithm of [7] in which they aim to maximise global received power; users are first clustered into groups based on their channel impulse responses. Then, each user's subcarriers are sorted by their amplitudes. The subcarriers that have high amplitudes for the majority of users in a group are allocated to that group. The algorithm swaps users amongst groups until no improvement in global received power is seen. It also swaps the weakest users amongst the groups to improve their performance.

The third policy simulated is the proposed linear programming algorithm with subcarriers assigned randomly (referred to as LP-Ran-Subc) to groups regardless of channel information.

The final policy is the proposed linear programming algorithm with subcarriers assigned to groups according to how different (referred to as LP-Diff-Subc) the subcarrier amplitude profiles across users are.

We firstly examine the effect of the four grouping/allocation policies on the user's subcarrier amplitudes. Figure 3 shows the cumulative distribution function (CDF) of the mean of each user's subcarrier amplitudes for each of the four grouping/allocation policies. We see that of all the policies the fixed allocation has the highest percentage of users with a low mean subcarrier amplitude. This is hardly surprising since the fixed allocation does not make any attempt to allocate users to good subcarriers. Comparing the two proposed algorithms (LP-Ran-Subc and LP-Diff-Subc) with the swapping algorithm, we see that initially the LP-Ran-Subc and the LP-Diff-Subc algorithms have a larger percentage of users with low mean subcarrier amplitudes than the swapping algorithm. The proposed LP-Ran-Subc algorithm then goes on to have more users with larger subcarrier amplitude means in the 20% to 70% range.

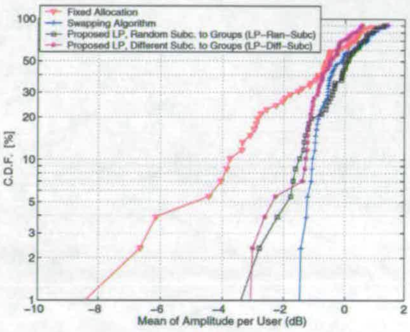


Fig. 3: CDF of mean of user's subcarrier amplitudes

We then focus on the quality of service (QoS) – in terms of BER – experienced by users in the system. The majority of users will experience good QoS if the system has a low average user BER coupled with a low variation between individual user's BER's. The variation of BERs in the system is measured by calculating the standard deviation of user's BERs. Figure 4 shows the average user BER for the four grouping/allocation policies. Figure 5 shows the logarithm of the standard deviation of all the user's BERs for the four policies.

In line with the results from figure 3, the fixed allocation exhibits the worst performance out of all of the four grouping/allocation policies due to it not being able to select which subcarriers to use. As for the algorithms that do try to optimise the allocation; initially the proposed LP-Diff-Subc and LP-Ran-Subc algorithms perform roughly the same as the swapping algorithm. At SNRs above 8 dB, the two proposed (LP-Ran-Subc and LP-Diff-Subc) algorithms outper-

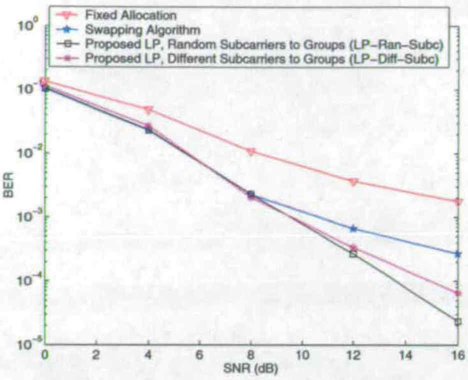


Fig. 4: Average user BER for various grouping/allocation policies

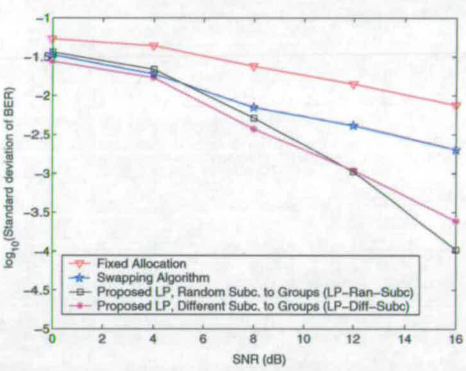


Fig. 5: Logarithm of the standard deviation of user's BER for various grouping/allocation policies

form both the swapping algorithm and the fixed allocation. For higher SNRs, as can be predicted from figure 3, the LP-Ran-Subc algorithm outperforms all the other grouping/allocation policies. However, figure 3 does not account for why the LP-Diff-Subc algorithm performs better than the swapping algorithm as the SNR increases. This discrepancy is most likely due to some underlying unfairness in the algorithms. To investigate further we examine the allocation, to each user, of subcarriers with low amplitudes. Figure 6 shows the CDF of the number of subcarriers per user that have a low amplitude (i.e. amplitude of less than –8 db). From this figure, with the swapping algorithm, 50% of the users will have no low amplitude subcarriers, that figure is 45% and 40% for the LP-Diff-Subc and LP-Ran-Subc algorithms respectively. Therefore, the proposed algorithms tend to be unfair to some users by allocating them more lower amplitude subcarriers than others, correspondingly some users get more high amplitude subcarriers than others. At low SNRs the low amplitude subcarriers tend to dominate, as the SNR increases the dominance of these low amplitude subcarriers diminishes.

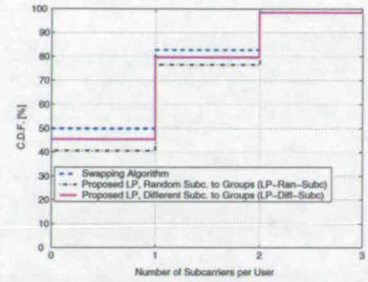


Fig. 6: CDF of the number of subcarriers per user that have low amplitude

VI. Conclusion

This work investigated a new linear programming based approach to the grouping and subcarrier allocation problem for a grouped MC-CDMA system. A sub-optimal approach was suggested whereby the subcarriers are first grouped and then users are assigned to groups using a linear programming algorithm. The subcarrier grouping method that performed best was found to be grouping subcarriers randomly. This subcarrier grouping in conjunction with the linear programming algorithm to allocate users to groups (LP-Ran-Subc) resulted in improved QoS performance, for more users, in the MC-CDMA downlink.

Acknowledgment

This work is supported by the Virtual Centre of Excellence in Mobile & Personal Communications (Mobile VCE, http://www.mobilevce.com) under the Wireless Access work area of the Core 2 Research Programme.

References

[1] V. M. DaSilva and E. S. Sousa, "Performance of Orthogonal CDMA Codes for Quasi-Synchronous Communication Systems," in *Proc. of IEEE ICUPC '93*, Ottawa, Canada, Oct. 1993, pp. 995-99.

[2] N. Yee and J.-P Linnartz, "BER of multi-carrier CDMA in an indoor Rician fading channel," in *Proc. of The Twenty-Seventh Asilomar Conference on Signals, Systems and Computers, 1993*, Pacific Grove, USA, Nov. 1993, pp. 426-30.

[3] K. Fazel and L. Papke, "On the performance of Convolutionally-Coded CDMA/OFDM," in *Proc. of IEEE PIMRC '93*, Yokohama, Japan, Sept. 1993, pp. 468-72.

[4] S. Kaiser, "OFDM-CDMA versus DS-CDMA: performance evaluation for fading channels," in *Proc. of IEEE ICC '95*, Seattle, USA, June 1995, pp. 1722-26.

[5] L.A. Wolsey, "Integer Programming", New York, NY: John Wiley and Sons, 1998.

[6] V. Chvatál, "Linear Programming", New York, NY: W. H. Freeman and Co., 1983.

[7] E. Al-Susa and D. Cruickshank, "An Adaptive Orthogonal Multi-carrier Multi-user CDMA Technique for a Broadband Mobile Communication System", in *Proc. of the 3rd International Workshop on Multi-Carrier Spread Spectrum & Related Topics*, Oberpfaffenhofen, Germany, Sept. 2001.

Radio Resource Allocation with Resource Metric Estimation for Multimedia CDMA Systems

Yeonwoo Lee, Moti Tabulo, Dave Laurenson, and Steve McLaughlin

Institute for Digital Communications (IDCOM), School of Engineering and Electronics,
The University of Edinburgh, Edinburgh EH9 3JL, U.K.
Tel: +44-131-650-5659, Fax: +44-131-650-6554, e-mail: {yl, mmt, dil, sml}@ee.ed.ac.uk

bstract: **In this paper, the radio resource allocation with source metric estimation (RME) technique is presented and scussed, which is applicable to wireless multimedia CDMA stems such as W-CDMA and single-code or multi-code multi-rrier ($(MC)^2$)-CDMA system. Based on call admission criterion, e admissible residual resource region in which QoS and ceptable link quality for NRTT data users can be guaranteed e demonstrated. It is shown that the resource allocation with e predictive RME can deliver exact status of residual ADR and lmissible capacity region, which in turn can decide appropriate lowable rates for each data users. Thus, the full utilisation of ft resource, i.e. the residual capacity, can be achieved in a ynamic and predictive manner.**[†]

I. Introduction

BEYOND third generation (3G) mobile wireless communication systems mobile information appliances ill be ubiquitous providing multiple multimedia services to ers. The visions of "Beyond 3G" and 4G throughout the orld are quite consistent as exhibited by "The book of sions" compiled within the Wireless Word Research Forum WWRF) [1]. Next generation wireless communication stems will require inter-disciplinary systems-level research order to provide more appropriate means to optimise and ploit scarce wireless resources, while maintaining true to e Internet model. In future evolutions toward 4G the timisation of capacity in the air-interface for the lowest cost ill gain momentum. This optimisation can only be made ssible with the adoption of efficient radio network chitectures and novel algorithms for radio resource and uality of Service (QoS) management, all of which are tegrated into a complete system solution.

The aim of any advanced radio resource management RRM) is to provide the requested QoS and high capacity for l active users in the network. One of the most pertinent aracteristics of multi-media services is the traffic ymmetry between uplink and downlink and from cell to cell ue to unbalanced data traffic generated by IP-based formation retrieval applications. This has seen the mobile dustry move towards a cellular IP solution for 3G. It is clear at in 4G this trend will continue. Also of interest is which r interface is likely to be adopted as the solution for 4G stems. From the RRM point of view, a multi-carrier based CDMA system is a good air interface candidate since it has the flexibility of a common pool of available resource units (RU) such as frequency unit and code slot unit. The definition of these RU can be enlarged to include spatial unit and time slot unit, i.e. a multi-dimension resource space, if the system is combined with multiple antenna technique, and time duplex division (TDD) techniques. In this paper, we present and consider a generic multi-carrier CDMA system with multiple codes and spreading gain cross frequency domain, which has a flexibility of variable rate transmission and of rate control feature. With this system, the resource allocation and management task needs an advanced approach to utilise the multi-dimensional resource plane, ultimately pursuing the optimal network capacity and QoS provision.

Since current RRM functionality in 2G and 3G systems acts from the network layer down to the data link and physical layer, the wireless physical resources are utilised directly without intermediate layers. Therefore, research work incorporating an RRM that takes account of the conditions of the physical layer resources suggests the need to focus on its potential impact in the intermediate layer, i.e. radio resource metric estimation (RME), which can be aware of the cross layer capabilities and states. Based on the knowledge of the desired loads and channel and radio resources, the RRM in cooperation with the RME can manage both up and down the protocol stack. Thus, it can decide and control the parameters and functions required to optimise the desired features such as QoS, throughput, power utilisation and overall system capacity. In this article, we describe and present some of these concepts and techniques. The rest of this article is organized as follows. The RME technique and its role in RRA are described in the following section. The call admission control (CAC) criteria for multimedia CDMA systems are discussed. The multimedia traffic behavior with RRA using RME are demonstrated by simulation results, followed by concluding remarks.

II. Radio Resource Metric Estimation

A. Radio Resource Metric Estimation in RRA Algorithm

Resource metric estimation (RME) is a crucial part of the radio resource allocation (RRA) algorithm that performs call admission control (CAC), resource scheduling and power/rate scheduling tasks, which provides the following control tasks:

[†] This work is supported in part by the WA3 area of the Core2 Research ogramme of the Virtual Centre of Excellence in Mobile & Personal mmunications, Mobile VCE, www.mobilevce.com.

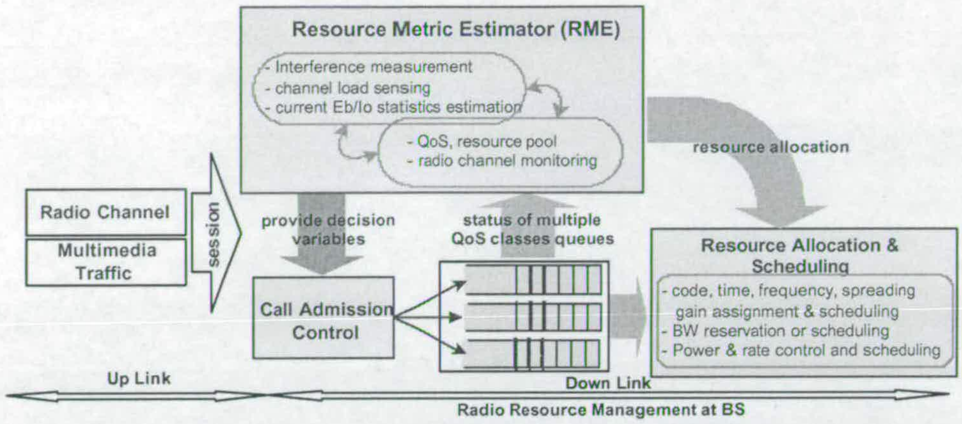


Figure 1: Resource metric estimator in RRA algorithm at the BS

- The radio channel characteristics and session quality requirements are used for optimal power and rate allocation.
- The current channel load in the BS, session traffic characteristics and quality requirements are used for controlling the resource scheduler.
- With the built-in capacity models, the RME assists CAC in accepting or rejecting new (or handoff) sessions.

The question of how to combine the interference measurements with the current load situation and QoS requirements of the existing traffic classes to control CAC, or channel allocation is a very interesting issue [2]. The most appropriate resource metrics permit efficient inter-working between the physical layer and higher layers in the protocol stack and thus, it is essential to optimize the overall system performance. In this paper, as a solution for efficient cross-layer inter-working, the radio resource metric region (RMR) concept has been introduced, which can provide the acceptable resource region where QoS and acceptable link quality can be guaranteed with an achievable resource margin to be utilised in terms of the resource metric mapping function (RMMF) [3].

B. RRA with Predictive Load Based RME

The predictive RME method proposed in [4, 5] can provide the acceptable resource region where QoS and acceptable link quality can be guaranteed with an achievable resource margin to be utilised in terms of the RMMF [3]. In higher layers, this mapping function can be used as a resource look-up table having actual radio mobile channel and interference characteristics and as a monitor of resource availability or efficiency. With this mapping function, the predictive RME can maximize the resource utilisation by delivering the information of resource availability of the current and the next time-stage to resource allocation or call admission control algorithm. This concept is depicted in Figure 2, where the measurement of SIR and traffic followed by a Kalman filter based prediction is performed [4, 5]. This provides measured or predicted resource parameters for the resource scheduler to search their current position on the surface of the RMMF and to deliver the actual status of the resource usage. All of these are integrated by establishing RMR in which the calculation and estimation of resource availability are performed based on the resource region decision criteria and help the call admission decision on a call, a packet, or a time slot basis.

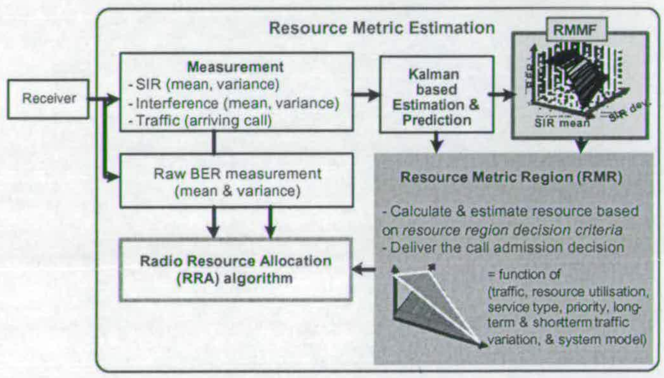


Figure 2: The predicted load based resource metric estimation technique [4,5].

The proposed predicted RME is dependent on the standard deviation of capacity and the degree of confidence (DCL[‡]) of user's status that can be driven from second-order statistics of E_b / I_0. Note that the higher the variance of the E_b / I_0 the lower maximum capacity, capacity margin, and resource availability. For an estimation of the maximum available number of users and available resource margin, both the received *Eb/Io* and the variance of E_b / I_0 have to be taken into account. By using this DCL parameter, the current and predicted capacity statuses of users are delivered to RRA algorithm to accept or reject a new coming call.

‡ DCL is defined by the reliability change between the current measured capacity margin and the predicted capacity margin, which are dependent on the status of users.

III. Power Controlled Admission Control for Variable QoS CDMA Systems

A. Call Admission Control for W-CDMA

The maximum achievable capacity can be achieved by following call admission control (CAC) criterion in [6, 7], assuming the existence of the optimal power level for each user:

$$\sum_{i=1}^{K}\frac{e_i}{e_i+G_i}+\frac{h_0 W}{\min\limits_{1\le i\le K}\left[P_{\max,i}h_i\frac{e_i+G_i}{e_i}\right]}<1 \qquad (1)$$

where W is the system bandwidth, G_j is the processing gain, the channel coefficient h_j, and h_0 is the noise spectral density. K is the total number of mobile user admitted in the system, which is the total sum of each users supporting corresponding data rate classes $K=\sum_{i=1}^{N}K_i$ where N is the total data rate services classes. In our study, we consider three kinds of service classes, voice, video, and data, of which number of users are denoted as K_v, K_{vid}, K_d, respectively. e_i is the required E_b/I_0 for the i-th user, which is defined by

$$e_i=\frac{G_i h_i P_i}{\sum_{j\neq i}^{K}h_j P_j+h_0 W} \qquad (2)$$

where P_i is the transmitted power for the i-th user.

B. Call Admission Control for Multi-Carrier CDMA

- *Single-code multi carrier CDMA system*

As shown in Figure 3, the single-code MC-CDMA system is the special case of the multi-code MC ($(MC)^2$)-CDMA system and this use only one code branch, i.e. one code serial-to-parallel (S/P) branch. Thus, the data coming in is of a rate NR bps using N rate S/P channels rather than MNR bps using M code S/P branches as like in a $(MC)^2$-CDMA system. A single-code MC-CDMA system block consists of N rate S/P channels of rate R, each of these is spread by a spreading code of length G and passed to a NG-point IDFT. In a SC MC-CDMA system, the E_b/I_0 for the i-th the desired user and the n-th rate S/P channel is given as follows:

$$e_{i,n}=\frac{G_{i,n}h_{i,n}P_{i,n}}{\sum_{j\neq i}^{K}h_{j,n}P_{j,n}+h_0 W}. \qquad (3)$$

Assuming that E_b/I_0 in a MC-CDMA system is the averaged value cross all sub-carriers, this equation can be reduced to e_i similar to that for conventional CDMA system as in Eq. (2). So, this system has the same CAC criterion to W-CDMA system.

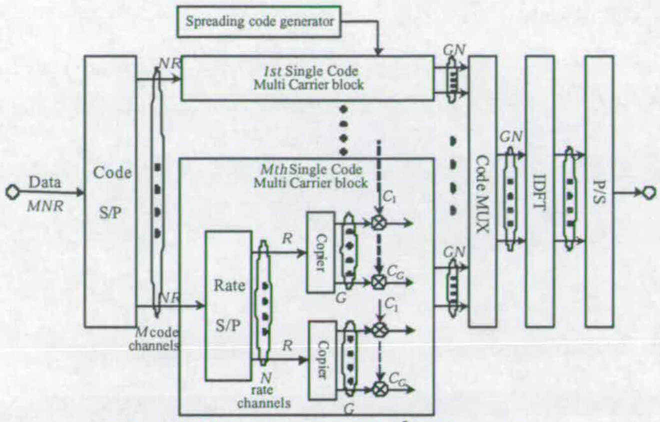


Figure 3: Block structure of $(MC)^2$-CDMA system

- *Multi-code multi-carrier ($(MC)^2$) CDMA system*

The processing gain G for each code remains the same, and the supported data rate for each user is decide by the number of multi-codes and rate channels, i.e. a rate of MNR bps. Assuming that for a multicode system, the received power for each of the M codes (of the same multicode user) is the same and the multiuser interference experienced by each of the M codes is also the same [8], the $h_{i,m,n}$ and $P_{i,m,n}$ for each code channel is the same and thus, reduced $h_{i,n}$ and $P_{i,n}$, respectively. The E_b/I_0 for the n-th rate S/P channel of the i-th user in a $(MC)^2$-CDMA system can be given by

$$e_{i,n}=\frac{G_i h_{i,n} P_{i,n}}{\sum_{j\neq i}^{K}M_j h_{j,n}P_{j,n}+(M_i-1)P_{i,n}+h_0 W} \qquad (4)$$

where M_j is the number of codes assigned to the j-th user. The denominator includes the multi-codes interference by other users and the self-interference caused by the imperfect orthogonality. Thus, the overall E_b/I_0 for the i-th user can be obtained by

$$e_i=\frac{1}{N_i}\sum_{n=1}^{N_i}e_{i,n} \qquad (5)$$

where N_i are the number of rate S/P channels assigned to the i-th user. Now, the CAC criterion for $(MC)^2$-CDMA system with a maximum of K users is as follows

$$\sum_{i=1}^{K}\frac{M_i e_i}{e_i+G_i}+\frac{h_0 W}{\min\limits_{1\le i\le K}\left[P_{\max,i}h_i\frac{e_i+G_i}{M_i e_i}\right]}<1\cdot \qquad (6)$$

The maximum number of supportable user K is determined by the predefined outage probability P_{out}, which is set to as 0.01 in our study.

IV. Multimedia Traffic Simulation Model and Results

A single micro cell configuration in which a set of power controlled mobile terminals are transmitting packets to a BS, in a CDMA system, is considered. With uplink channel, RTT (voice call and video-phone calls) and NRTT mixed traffic are considered. We generated the video-phone traffic model as in [8] for H.263 video sequence of which distribution is assumed to be Gamma distribution. The call arriving of each service is modeled as Poisson random variables with corresponding exponential distributed holding time. For data traffic, packet switched Web browsing traffic is considered. The system level multimedia traffic parameters are shown in Table I. In order to fully utilise available resource (or residual capacity), the monitoring RTT load and controlling NRTT packet data traffic is crucial in multimedia traffic call admission control. Since RTT calls are always given the highest priority and they are allowed to transmit without delay, in our study, NRTT calls are allowed to transmit according to the available residual capacity (here the residual aggregated data rate (ADR[§])) obtained by subtracting RTT contribution from the total system capacity. Thus, the resource availability for data service is determined by estimating and predicting RTT load contribution for the next time state by Kalman filtering and the predictive load based RME [4, 5]. This predictive RME method can deliver the fidelity information of resource availability of the current and the next time-stage to resource allocation algorithm, taking the variance of *Eb/Io* into account.

TABLE I: MULTIMEDIA TRAFFIC SIMULATION PARAMETERS

Common Parameters	Value
System bandwidth, W [MHz]	3.84
Noise spectral density, h_0 [dBm/Hz]	-174
Max. MS power limit (voice, data, video) [dBm]	5, 17.1, 14.1
Standard deviation of shadowing [dB]	10
Mean of required Eb/Io for voice, data, video; e_v^{req}, e_d^{req}, e_{vid}^{req} [dB]	5, 7, 5
Data rate for voice, packet data, video; R_v, R_d, R_{vid} [Kbps]	8,128, 64
Outage probability constraint, P_{out}	0.01
Parameters for (MC)^2-CDMA	**Value**
System bandwidth, W [MHz]	4.096
No. of rate S/P channels, N	2
No. of code S/P channels for voice, data, video (M)	1, 16, 8
Spreading gain, G	512
Total number of sub-carriers (IDFT point)	1024
Base data rate of rate S/P channel, R [Kbps]	4

Simulated and admission controlled multimedia traffic for both W-CDMA and (MC)^2-CDMA system is shown in Fig. 4 where the systematic error, the standard deviation of received *Eb/Io* target ($e_v^s = e_d^s = e_{vid}^s$) are all assumed to be 2 dB. This shows the achievable ADRs for each systems controlled by CAC criterion as in Eq. (1) and (6), respectively, with parameters given in Table I. In Fig. 5, DCL and admissible residual ADR with/without DCL for each system are shown in terms of frame index. It is demonstrated that DCL of (MC)^2-CDMA is relatively higher than that of W-CDMA, which means that for (MC)^2-CDMA system, the fidelity of information of the capacity status is more reliable than W-CDMA since long-term fading is assumed in our study. Thus, compared to the optimistically estimated residual ADR (Opt. resADR in Fig. 5), the residual ADR with DCL weighting given by the predictive RME gives almost the same ADR unlike W-CDMA.

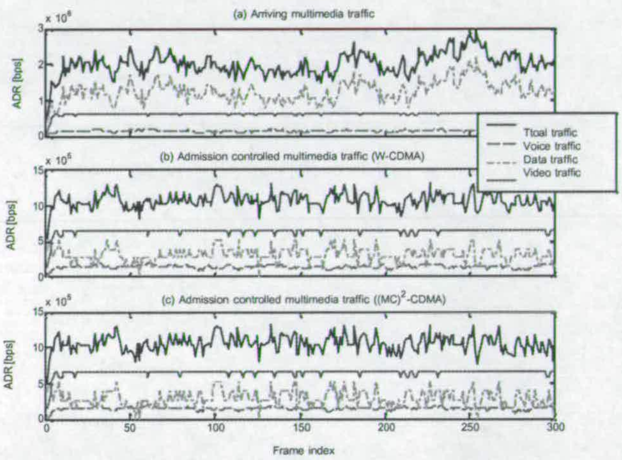


Figure 4: Simulated and admission controlled multimedia traffic for W-CDMA and (MC)^2-CDMA system; [I_v, I_d, I_{vid}]=[3, 2, 1] call/frame with 2 dB standard deviation of required *Eb/Io* target.

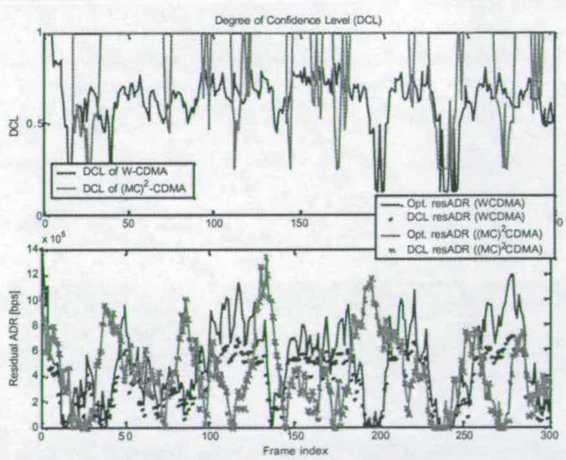


Figure 5: DCL and admissible residual ADR for W-CDMA and (MC)^2-CDMA system.

With the same parameters, Fig. 6 and 7 show the admissible residual capacity plane with the predictive RME for W-CDMA and (MC)^2-CDMA, respectively, for a multimedia traffic scenario. Here, the residual ADR for NRTT is estimated and predicted by Kalman filtering after

[§] ADR is defined by the total sum of the number of users multiplied by corresponding data rate, i.e. $ADR = (K_v R_v + K_{vid} R_{vid} + K_v R_v)$.

measuring the current ADR for RTT, and this is converted into available data users versus voice or video users. As shown in Fig. 6 and 7, the admissible residual capacity region is quite dependent on the selection of system, i.e. the choice of resource dimension, which in turn delivers the exact capacity status to CAC algorithm and resource allocation algorithm to control the packet service data rate in a dynamic fashion. The adaptive rate control mechanism incorporating with the admissible residual capacity will be our future work. Compared to Fig.6, it is noticeable that Fig. 7 can achieve a larger admissible residual capacity area for NRTT users, which of course depending on the fidelity of between the estimated and predicted residual capacity and the selection of system with different resource dimension. For example a $(MC)^2$-CDMA system has a three-dimensional resource plane such as code, frequency and spreading gain, even though this could be constrained by the channel fading and the receiver structure. With these results, the resource allocation with the predictive RME can deliver exact status of residual ADR and admissible capacity region, which in turn can decide appropriate allowable rates for each data users. Thus, the full utilisation of left resource, i.e. the residual ADR, can be achieved in a dynamic and predictive manner. Furthermore, if combined with adaptive modulation, or rate control, the proposed resource allocation method becomes important to provide the exact resource availability.

V. CONCLUSIONS

In this paper, the resource allocation with a predicted load based RME method has been investigated for multimedia CDMA systems such as W-CDMA and $(MC)^2$-CDMA. With aid of the predictive load based RME technique, the acceptable residual resource region can be achieved with QoS and acceptable link quality for NRTT data users guaranteed with an achievable resource margin in terms of capacity margin, the DCL of the system, and the second-order statistics of E_b / I_0. With the DCL weighted admissible residual capacity plane, $(MC)^2$-CDMA system with more flexibility of resource dimension can deliver more reliable and exact admissible residual ADR in our study, compared to W-CDMA system. This result could be reversed if a more severe frequency-selective fading is considered. However, with an adaptive control of spreading gain corresponding to its channel characteristics such as coherence bandwidth, the severe frequency-selective fading would not be a major drawback.

REFERENCES

[1] B. G. Evans and K. Baughan, "Visions of 4G," *IEE Electronics & Communication Engineering Journal, Special Issue on Mobile VCE*, Vol. 12, No. 6, Dec. 2000.

[2] L. Jorguseski *et al.*, "Radio Resource Allocation in Third-Generation Mobile Communication Systems," *IEEE Communications Magazine*, pp. 117–123, Feb. 2001.

[3] Y. Lee, S. McLaughlin and D. Jeon, "Resource Metric Mapping Function for a TDD-CDMA System Supporting WWW Traffic", *in Proc.* IEEE VTC2003 Spring, Jeju, April 2003.

[4] Y. Lee, S. McLaughlin and S. Yun, "Radio Resource Metric Estimation for Wireless CDMA Communication Systems Maximising Radio Resource Utilisation," *in Proc.* IEEE VTC2003 Spring, Jeju, April 2003.

[5] Y. Lee and S. McLaughlin, "Method of Operating A Telecommunications System," UK Patent Application No.: GB 0219978.4, August 28 2002.

[6] A. S. Sampath, P. S. Kumar, and J. M. Holtzman, "Power control and resource management for a multimedia CDMA wireless system," in *Proc. IEEE PIMRC' 95*, 1995.

[7] O. Gurbuz and H. Owen, "Power Control Based QoS Provisioning for Multimedia in W-CDMA," *Wireless Networks* 8, pp. 37-47, 2002.

[8] 3GPP, Technical Specification TS 23.107, V.4.0.0, QoS Concept and Architecture, Oct. 2000.

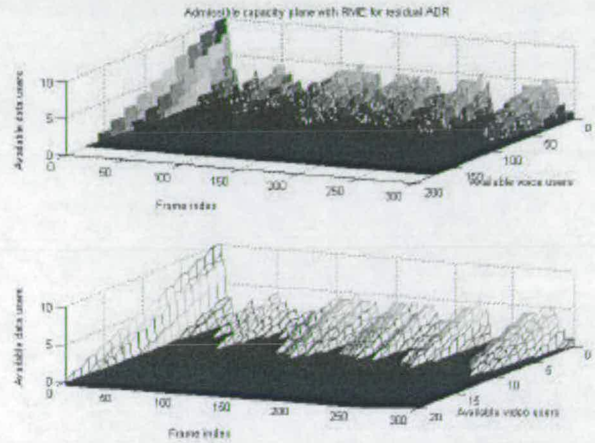


Figure 6: Admissible residual capacity plane with RME for residual ADR in a W-CDMA aggregated multimedia traffic.

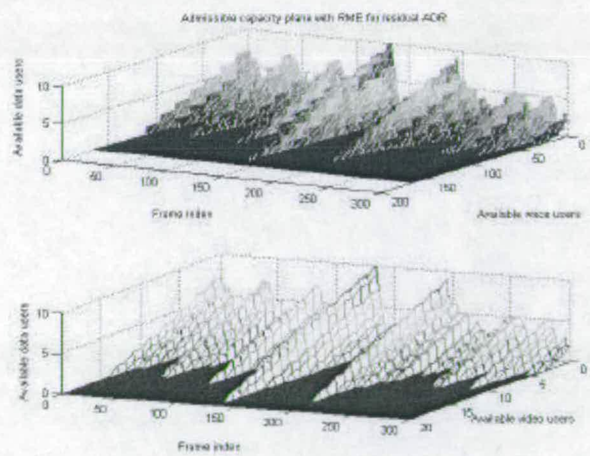


Figure 7: Admissible residual capacity plane with RME for residual ADR in a $(MC)^2$-CDMA aggregated multimedia traffic.

References

[1] W. H. Press, S. A. Teukolsky, W. T. Vetterling, and B. P. Flannery, *Numerical Recipes in C*. Cambridge, MA, USA: Cambridge University Press, 1992.

[2] FFTW, "The Fastest Fourier Transform in the West (FFTW)." http://www.fftw.org.

[3] LP-Solve, "The Linear Programming Solver." http://www.geocities.com/lpsolve.

[4] H. Sari, Y. Levy, and G. Karam, "An analysis of orthogonal frequency-division multiple access," in *Proc. of the Global Telecommunications Conference, 1997*, (Phoenix, USA,), pp. 1635–1639, Nov. 1997.

[5] N. Yee and J. Linnartz, "BER of multi-carrier CDMA in an indoor rician fading channel," in *Proc. of The Twenty-Seventh Asilomar Conference on Signals, Systems and Computers, 1993*, (Pacific Grove, USA), pp. 426–430, Nov. 1993.

[6] K. Fazel and L. Papke, "On the performance of convolutionaly-coded CDMA/OFDM," in *Proc. of IEEE PIMRC '93*, (Yokohama, Japan), pp. 468–72, Sept. 1993.

[7] H. Atarashi, S. Abeta, and M. Sawahashi, "Variable spreading factor-orthogonal frequency and code division multiplexing (VSF-OFCDM) for broadband packet wireless access," *IEICE Trans. Commun.*, vol. E86-B, pp. 291–299, Jan. 2003.

[8] L. Jorguseski, J. Farserotu, and R. Prasad, "Radio resource allocation in third-generation mobile communication systems," *IEEE Communications Magazine*, vol. 2, pp. 117–123, Feb. 2001.

[9] R. W. Chang, "Synthesis of band-limited orthogonal signals for multichannel data transmission," *Bell Syst. Tech. J.*, vol. 45, pp. 1775–1796, Aug. 1968.

[10] R. W. Chang and R. A. Gibby, "A theoretical study of performance of an orthogonal multiplexing data transmission scheme," *IEEE Trans. Commun. Tech.*, vol. COM-16, pp. 529–540, Aug. 1968.

[11] S. B. Weinstein and P. M. Ebert, "Data transmission by frequency division multiplexing using the discrete Fourier transform," *IEEE Trans. Commun. Tech.*, vol. COM-19, pp. 628–634, Oct. 1971.

[12] A. Peled and A. Ruiz, "Frequency domain data transmission using reduced computational complexity modems," in *Proc. of ICASSP*, pp. 964–967, 1980.

[13] R. van Nee and R. Prasad, *OFDM for Wireless Multimedia Communications*. Norwood, MA, USA: Artech House, 2000.

[14] A. J. Paulraj, D. A. Gore, R. U. Nabar, and H. Bolcksei, "An overview of MIMO communications - a key to gigabit wireless," *Proceedings of the IEEE*, vol. 92, pp. 198 – 218, Feb. 2004.

[15] L. Hanzo, W. Webb, and T. Keller, *Single- and Multi-Carrier Quadrature Amplitude Modulation*. New York, NY, USA: John Wiley and Sons, 2000.

[16] R. Prasad, W. Mohr, and W. Konhäuser, *Third Generation Mobile Communication Systems*. London: Artech House Publishers, 2000.

[17] T. S. Rappaport, *Wireless Communications: Principles and Practice*. New Jersey, USA: Prentice-Hall, 1996.

[18] J. G. Proakis, *Digital Communications*. Singapore: McGraw Hill, 2001.

[19] T. Chebaro and P. Godlewski, "About the CDMA capacity derivation," in *Proc. of the International Symposium on Signals, Systems and Electronics*, (Paris, France), pp. 36–39, Sept. 1992.

[20] G. E. Corazza, G. De Maio, and F. Vatalaro, "CDMA cellular systems performance with fading, shadowing and imperfect power control," *IEEE Trans. Veh. Tech.*, vol. 47, pp. 450–459, May 1998.

[21] D. G. Jeong and W. S. Jeon, "CDMA/TDD system for wireless multimedia services with traffic unbalance between uplink and downlink," *IEEE J. Sel. Areas Commun.*, vol. 17, pp. 939–946, May 1999.

[22] D. Kim and D. G. Jeong, "Capacity unbalance between uplink and downlink in spectrally overlaid narrow-band and wide-band CDMA systems," *IEEE Trans. Veh. Tech.*, vol. 49, pp. 1086–1093, July 2000.

[23] S. Hara and R. Prasad, "Overview of multicarrier CDMA," *IEEE Commun. Mag*, vol. 35, pp. 126–133, Dec. 1997.

[24] V. M. DaSilva and E. S. Sousa, "Performance of orthogonal CDMA codes for quasi-synchronous communication systems," in *Proc. of IEEE ICUPC '93*, (Ottawa, Canada), pp. 995–999, Oct. 1993.

[25] S. Abeta, H. Atarashi, M. Sawahashi, and F. Adachi, "Performance of coherent multi-carrier/DS-CDMA and MC-CDMA for broadband packet wireless access," *IEIECE Trans. Commun.*, vol. E84-B, pp. 406–414, Mar. 2001.

[26] S. Kaiser, "OFDM-CDMA versus DS-CDMA: performance evaluation for fading channels," in *Proc. of IEEE ICC '95*, (Seattle, USA), pp. 1722–26, June 1995.

[27] S. Kaiser, "On the performance of different detection techniques for OFDM-CDMA in fading channels," in *Proc. of IEEE Globecom '95*, vol. 3, pp. 2059–2063, Nov. 1995.

[28] A. Dutta-Roy, "The cost of quality in internet-style networks," *IEEE Spectrum*, pp. 57–62, Sept. 2000.

[29] P. J. Ameigeiras-Gutiérrez, *Packet scheduling and quality of service in HSDPA*. PhD thesis, Aalborg University, Denmark, Oct. 2003.

[30] K. Baum, T. Kostas, P. Sartori, and B. Classon, "Performance characteristics of cellular systems with different link adaptation strategies," *IEEE Trans. Veh. Tech.*, vol. 52, pp. 1497– 1507, Nov. 2003.

[31] J. Zander and S. L. Kim, *Radio Resource Management for Wireless Networks*. Artech House Publishers, 2001.

[32] Y. Cao and V. O. K. Li, "Scheduling algorithms in broad-band wireless networks," *Proceedings of the IEEE*, vol. 89, pp. 76–87, Jan. 2001.

[33] P. S. Chow, J. M. Cio, and J. A. C. Bingham, "A practical discrete multitone transceiver loading algorithm for data transmission over spectrally shaped channels," *IEEE Trans. Commun.*, vol. 43, pp. 773–775, Feb. 1995.

[34] T. Alen, A. Madhukumar, and F. Chin, "Capacity enhancement of a multi-user OFDM system using dynamic frequency allocation," *IEEE Transactions on Broadcasting*, vol. 49, pp. 344–353, Dec. 2003.

[35] C. Y. Wong, R. S. Cheng, K. B. Letaief, and R. D. Murch, "Multiuser OFDM with adaptive subcarrier, bit and power allocation," *IEEE Journal Sel. Areas Commun.*, vol. 17, pp. 1747–1757, Oct. 1999.

[36] D. Kivanc, G. Li, and H. Liu, "Computationally efficient bandwidth allocation and power control for OFDMA," *IEEE Trans. Wireless Commun.*, vol. 2, pp. 1150–1158, Nov. 2003.

[37] S. Moshavi, "Multi-user detection for DS-CDMA communications," *IEEE Commun. Mag.*, pp. 124–136, Oct. 1996.

[38] H. Holma and A. Toskala, *WCDMA for UMTS, 2^{nd} edition*. New York, NY, USA: John Wiley and Sons, 2002.

[39] T. Ojanperä and R. Prasad, *WCDMA: Towards IP Mobility and Mobile Internet*. London, UK: Artech House, 2001.

[40] V. Espinosa, L. Jorguseski, R. Litjens, E. R. Fledderus, and R. Prasad, "Downlink radio resource estimation and control in WCDMA cellular system with voice and data users," in *Proc. of Wireless Personal Multimedia Communications (WPMC)*, vol. 2, (Aalborg, Denmark), pp. 997–1002, Sept. 2001.

[41] P. Patel and J. Holtzmann, "Performance comparison of a DS/CDMA system using a successive interference cancellation (IC) scheme and a parallel IC scheme under fading," in *Proc. of ICC 1994*, pp. 510–514, May 1994.

[42] S. Kaiser, "OFDM code-division multiplexing in fading channels," *IEEE Trans. Commun.*, vol. 50, pp. 1266–1273, Aug. 2002.

[43] J. G. Andrews and T. H. Y. Meng, "Performance of multicarrier CDMA with successive interference cancellation with estimation error in a multipath fading channel," in *Proc. of 7th IEEE Symposium on Spread Spectrum Techniques and Applications (ISSSTA)*, pp. 150–154, Sept. 2002.

[44] Z. Li and M. Latva-Aho, "Multicarrier CDMA system with parallel interference cancellation in fading channel," in *Proc. of Third International Workshop on Multi-Carrier Spread Spectrum (MC-SS)*, (Oberpfaffenhoffen, Germany), pp. 273–280, Sept. 2001.

[45] E. Al-Susa and D. Cruickshank, "An adaptive orthogonal multi-carrier multi-user CDMA technique for a broadband mobile communication system," in *Proc. of the 3rd International Workshop on Multi-Carrier Spread Spectrum & Related Topics*, (Oberpfaffenhofen, Germany), Sept. 2001.

[46] P. Robertson and S. Kaiser, "Analysis of the loss of orthogonality through Doppler spread in OFDM systems," in *Proc. of IEEE Globecom '99*, pp. 701–706, Dec. 1999.

[47] C. H. Papadimitriou and K. Steiglitz, *Combinatorial Optimization: Algorithms and Complexity*. New Jersey, USA: Prentice-Hall, 1982.

[48] C. Blum and A. Roli, "Metaheuristics in combinatorial optimization: Overview and conceptual comparison," *ACM Computing Surveys*, vol. 35, pp. 268–308, September 2003.

[49] L. A. Wolsey, *Integer Programming*. New York, NY: John Wiley and Sons, 1998.

[50] V. Chvatál, *Linear Programming*. New York, NY: W. H. Freeman and Co., 1983.

[51] W. Wang and C. K. Rushforth, "An adaptive local-search algorithm for the channel-assignment problem (CAP)," *IEEE Trans. Veh. Tech.*, vol. 45, pp. 459–66, Aug. 1996.

[52] A. Capone and M. Trubian, "Channel assignment problem in cellular systems: a new model and a tabu search algorithm," *IEEE Trans. Veh. Tech.*, vol. 48, pp. 1252–1260, July 1999.

[53] M.-J. Kim and K.-H. Ryu, "A comparison of two search algorithms for solving the channel assignment problem," in *Proc. of 5th IEEE International Conference on Universal Personal Communications*, vol. 2, (Cambridge, USA), pp. 681–685, Sept. 1996.

[54] M. Duque-Anton, D. Kunz, and B. Ruber, "Channel assignment for cellular radio using simulated annealing," *IEEE Trans. Veh. Tech.*, vol. 42, pp. 14–21, Feb. 1993.

[55] M. Kobayashi, S. Haruyama, R. Kohno, and M. Nakagawa, "Optimal access point placement in simultaneous broadcast system using OFDM for indoor wireless LAN," in *Proc. of PIMRC 2000*, vol. 1, (London, UK), pp. 200–204, Sept. 2000.

[56] A. Koutsaftiki, C. Matrakidis, P. Lane, and J. J. O'Reilly, "Improved dynamic channel allocation for satellite systems," in *VTC 2000-Spring*, vol. 3, (Tokyo, Japan), pp. 1965–1969, May 2000.

[57] A. Al-Rustamani and B. Vojcic, "A new approach to greedy multiuser detection," *IEEE Trans. Commun.*, vol. 50, pp. 1326–1336, Aug. 2002.

[58] M. Andrews, B. Awerbuch, A. Fernandez, J. Kleinberg, T. Leighton, and Z. Liu, "Universal stability results for greedy contention-resolution protocols," in *Proc. of 37th Annual Symposium on Foundations of Computer Science, 1996*, pp. 380–389, Oct. 1996.

[59] J. Mayer, "Greedy quantization of control points for 2D and 3D data using blending surfaces representation," in *Conference Record of the Thirty-Third Asilomar Conference on Signals, Systems, and Computers, 1999*, vol. 2, 1999.

[60] K. Kim, H. Kim, Y. Han, and S.-L. Kim, "Iterative and greedy resource allocation in an uplink OFDMA system," in *Proc. of IEEE PIMRC 2004*, vol. 4, pp. 2377– 2381, Sept. 2004.

[61] N. Maeda, H. Atarashi, S. Abeta, and M. Sawahashi, "Pilot channel assisted MMSE combining in forward link for broadband OFCDM packet wireless access," *IEICE Trans. Fundamentals*, vol. E85-A, pp. 1635–1646, July 2002.

[62] A. K. Parekh and R. G. Gallagher, "A generalised processor sharing approach to flow control in integrated services networks: the single node case," *IEEE/ACM Trans. Networking*, vol. 1, pp. 344–357, June 1993.

[63] A. Demers, S. Keshav, and S. Shenker, "Analysis and simulation of a fair queueing algorithm," in *Proc. of ACM SIGCOMM*, (Texas, USA), pp. 1–12, Sept. 1989.

[64] A. Stamoulis, N. D. Sidiropoulos, and G. B. Giannakis, "Time-varying fair queueing scheduling for multi-code CDMA based on dynamic programming," *IEEE Trans. Wireless Comm.*, vol. 3, pp. 512–523, Mar. 2004.

[65] P. Ramanathan and P. Agrawal, "Adapting packet fair queueing algorithms to wireless networks," in *Proc. of IEEE/ACM MOBICOM'98*, (Dallas, USA), pp. 1–9, Oct. 1998.

[66] S. Lu, V. Bharghavan, and R. Srikant, "Fair scheduling in wireless packet networks," *IEEE/ACM Trans. Networking*, vol. 7, pp. 473–489, Aug. 1999.

[67] M. A. Arad and A. Leon-Garcia, "A generalized processor sharing approach to time scheduling in hybrid CDMA/TDMA," in *Proc. of IEEE INFOCOM'98*, (San Francisco, USA), pp. 1164–1171, Mar. 1998.

[68] Ö. Gürbüz and H. Owen, "Dynamic resource scheduling schemes for W-CDMA systems," *IEEE Commun. Mag.*, vol. 38, pp. 80–84, Oct. 2000.

[69] L. Xu, X. Shen, and J. W. Mark, "Dynamic bandwidth allocation with fair scheduling for WCDMA systems," *IEEE Wireless Commun.*, vol. 9, pp. 26–32, April 2002.

[70] L. Xu, X. Shen, and J. W. Mark, "Dynamic fair scheduling with QoS constraints in multimedia wideband CDMA cellular networks," *IEEE Trans. Wireless Commun.*, vol. 3, pp. 60–73, Jan. 2004.

[71] H. Harada and R. Prasad, *Simulation and Software Radio for Mobile Communications*. Artech House Publishers, 2002.

[72] ETSI, "Universal Mobile Telecommunications System (UMTS); selection procedures for the choice of radio transmission technologies of the UMTS," Tech. Rep. TR 101 112 v3.2.0, ETSI, April 1998.

[73] M. Gudmundson, "Correlation model for shadow fading in mobile radio systems," *IEE Electronics Letters*, vol. 27, pp. 2145–2146, Nov. 1991.

[74] ITU, "Reduced complexity 8 kbit/s CS-ACELP speech codec." ITU Rec. G.729 Annex A, Nov. 1996.

[75] 3GPP, "3[rd] Generation Partnership Project: Technical Specification Group Radio Access Networks: Radio Frequency (RF) system scenarios (Release 6)." 3GPP TR 25.942 v6.3.0, 2004.

[76] Y. Kishiyama, N. Maeda, K. Higuchi, H. Atarashi, and M. Sawahashi, "Field experiments on throughput performance above 100 Mbps in forward link for VSF-OFCDM broadband wireless access," *IEICE Trans. Commun.*, vol. E88-B, pp. 604–614, Feb. 2005.

[77] H. Atarashi and M. Nakagawa, "An efficient ARQ scheme for multi-carrier modulation systems based on packet combining," *IEICE Trans. Commun.*, vol. E82-B, pp. 731–739, May 1999.

[78] N. Miki, H. Atarashi, S. Abeta, and M. Sawahashi, "Comparison of hybrid ARQ packet combining algorithm in high speed downlink packet access in a multipath fading channel," *IEICE Trans. Fundamentals*, vol. E85-A, pp. 1557–1568, July 2002.

[79] D. Ayyagari and A. Ephremides, "Optimal admission control in cellular DS-CDMA systems with multimedia traffic," *IEEE Trans. Wireless Commun.*, vol. 2, pp. 195–202, Jan. 2003.

[80] Ö. Gürbüz and H. Owen, "Power control based QoS provisioning for multimedia in W-CDMA," *Wireless Networks*, vol. 8, pp. 37–47, 2002.

[81] Y. Lee, S. McLaughlin, and S. Yun, "Radio resource metric estimation for wireless CDMA communication systems maximising radio resource utilisation," in *Proc. of IEEE VTC 2003-Spring*, vol. 3, (Jeju, Korea), pp. 1940–1944, April 2003.

[82] A. J. Viterbi and R. Padovani, "Implications of mobile cellular CDMA," *IEEE Communications Magazine*, vol. 30, pp. 38–41, Dec. 1992.

[83] A. J. Viterbi, *CDMA: Principles of Spread Spectrum Communication*. Reading, Massachusetts: Addison-Wesley Publishing Company, 1995.

[84] W. Choi and J. Y. Kim, "Forward-link capacity of a DS/CDMA system with mixed multirate sources," *IEEE Trans. Veh. Tech.*, vol. 50, pp. 737–749, May 2001.

[85] S. C. Schwartz and Y. S. Yeh, "On the distribution function and moments of power sums with lognormal moments," *Bell Syst. Tech. Journal*, vol. 61, pp. 1441–1462, Sept. 1982.

[86] A. Sampath, N. B. Mandayam, and J. M. Holtzmann, "Erlang capacity of a power controlled integrated voice and data CDMA systems," in *Proc. of IEEE VTC'97*, (Phoenix, AZ), pp. 1557–1561, May 1997.

[87] S. S. B. Moore, "Blocking probability analysis for packet circuit emulation services," *IEEE Comms Letters*, vol. 8, pp. 268–270, May 2004.

[88] T. G. Robertazzi, *Computer Networks and Systems*. Springer-Verlag, 3[rd] ed., 2000.

[89] M. Molina, P. Castelli, and G. Foddis, "Web traffic modelling exploiting TCP connections' temporal clustering through HTML-REDUCE," *IEEE Network*, vol. 14, pp. 46–55, May – June 2000.

[90] H. Andoh, M. Sawahashi, and F. Adachi, "Channel estimation filter using time-multiplexed pilot channel for coherent Rake combining in DS-CDMA mobile radio," *IEICE Trans. Commun.*, vol. E81-B, pp. 1517–1526, July 1998.

[91] Y. Lee, S. McLaughlin, and E. A. Al-Susa, "Pre-selection tentative decision device-based SIR estimator for a UTRA-TDD system in time varying channels," in *Proc. of ICC 2002*, vol. 3, pp. 1829–1833, April 2002.

[92] Z. Liu and M. E. Zarki, "SIR-based admission control for DS-CDMA cellular systems," *IEEE Journal Sel. Areas in Commun.*, vol. 12, pp. 638–644, May 1994.

[93] D. Kim, "Efficient interactive call admission control in power controlled mobile systems," *IEEE Trans. Veh. Tech.*, vol. 49, pp. 1017–1028, May 2000.

[94] W. S. Jeon and D. G. Jeong, "Call admission control for CDMA mobile communications systems supporting multimedia services," *IEEE Trans. Wireless Commun.*, vol. 1, pp. 649–659, Oct. 2002.

[95] F. Gunnarsson, E. Geijer-Lundin, G. Bark, and N. Wiberg, "Uplink admission control in WCDMA based on relative load estimates," in *Proc. of ICC 2002*, vol. 5, (New York, USA), pp. 3091–3095, April 2002.

[96] N. Dimitriou, R. Tafazolli, and G. Sfikas, "Quality of service for multimedia CDMA," *IEEE Communications Magazine*, vol. 38, pp. 88–94, July 2000.

[97] C. Y. Huang and R. D. Yates, "Call admission in power controlled CDMA systems," in *Proc. of IEEE VTC 1996*, (Atlanta, GA, USA), May 1996.

[98] J. Knutsson, P. Butovitsch, M. Persson, and R. D. Yates, "Downlink admission control strategies for CDMA systems in a Manhattan environment," in *Proc. of IEEE VTC '98*, (Ottawa, Canada), pp. 1453–1457, May 1998.

[99] K. Kim and Y. Han, "A call admission control scheme for multi-rate traffic based on total received power," *IEICE Trans. Commun.*, vol. E84-B, pp. 457–463, March 2001.

[100] J. Perez-Romero, O. Sallent, R. Agusti, and G. Pares, "A downlink admission control algorithm for UTRA-FDD," in *Proc. of 4th International Workshop on Mobile and Wireless Communications Networks, 2002*, (Stockholm, Sweden), pp. 18–22, Sept. 2002.

[101] S. Aissa, J. Kuri, and P. Mermelstein, "Call admission on the downlink of a CDMA system based on total transmitted power," in *Proc. of IEEE VTC 2003-Fall*, (Orlando, USA), pp. 933–937, Oct. 2003.

[102] M. A. Arad and A. Leon-Garcia, "Hybrid multiple access for wireless ATM networks," in *Proc. of PIMRC' 96*, (Taipei, Taiwan), Oct. 1996.

[103] A. Sampath, P. S. Kumar, and J. M. Holtzmann, "Power control and resource management for a multimedia CDMA wireless system," in *Proc. of IEEE PIMRC'95*, pp. 21–25, Sept. 1995.

[104] Y. Lee, M. Tabulo, D. I. Laurenson, and S. McLaughlin, "Radio resource allocation with resource metric estimation for multimedia cdma systems," in *Proc. of the 8th International Conference on Cellular and Intelligent Communications(CIC)*, (Seoul, Korea), Oct. 2003.

[105] D. E. Everitt, "Traffic engineering of the radio interface for cellular mobile networks," *Proc. IEEE*, vol. 82, pp. 1371–1382, Sept. 1994.

[106] Y. Ishikawa and N. Umeda, "Capacity design and performance of call admission control in cellular CDMA systems," *IEEE Journal on Sel. Areas. in Commun.*, vol. 15, pp. 1627–1635, Oct. 1997.

[107] J. S. Evans and D. Everitt, "Effective bandwidth-based admission control for multiservice CDMA cellular networks," *IEEE Trans. Veh. Tech.*, vol. 48, pp. 36–46, Jan. 1999.

[108] Q. Liu and G. Shengli Zhou; Giannakis, "Cross-layer combining of adaptive modulation and coding with truncated arq over wireless links," *IEEE Trans. Wireless Commun.*, vol. 3, pp. 1746–1755, Sept. 2004.

[109] V. Friderikos, L. Wang, and A. H. Aghvami, "TCP-aware power and rate adaptation in DS/CDMA networks," *IEE Proceedings-Communication*, vol. 151, pp. 581–588, Dec. 2004.

[110] M. Realp and A. I. Prez-Neira, "Cross-layer MAC scheduling for multiple antenna systems," in *Proc. of IEEE Globecom 2004*, (Dallas, TX, USA), pp. 3352–3356, Dec. 2004.

[111] J. López-Vicario and C. Antón-Haro, "Transmit antenna selection for rate adaptation in HSDPA systems," in *Proc. Wireless World Conference (WWC-04)*, (San Francisco, CA, USA), May 2004.

[112] O. S. Shin and K. Lee, "Antenna-assisted round robin scheduling for MIMO cellular systems," *IEEE Commun. Letters*, vol. 7, pp. 109–111, Mar. 2003.

[113] H. Rohling and R. Grunheid, "Adaptive coding and modulation in an OFDM-TDMA communication system," *Proc. VTC 1998*, vol. 2, pp. 773–776.

[114] J. Cai, X. Shen, and J. W. Mark, "Downlink resource management for packet transmission in OFDM wireless communication systems," in *Proc. of IEEE Globecom '03*, vol. 6, (San Francisco, CA, USA), pp. 2999–3003, Dec. 2003.

[115] Y. J. Zhang and K. B. Letaief, "Adaptive resource allocation and scheduling for multiuser packet-based OFDM networks," in *Proc. of IEEE ICC'04*, vol. 5, (Paris, France), pp. 2949–2953, June 2004.

[116] L. Li and Z. Niu, "A multi-dimensional radio resource scheduling scheme for MIMO-OFDM systems with channel dependent parallel weighted fair queueing (CDPWFQ)," in *Proc. of PIMRC 2004*, vol. 4, (Barcelona, Spain), pp. 2367–2371, Sept. 2004.

[117] G. Song and Y. G. Li, "Cross-layer optimization for OFDM wireless networks–Part I: Theoretical framework," *IEEE Trans. Wireless Commun.*, vol. 4, pp. 614–624, Mar. 2005.

[118] G. Song and Y. G. Li, "Cross-layer optimization for OFDM wireless networks–Part II: Algorithm development," *IEEE Trans. Wireless Commun.*, vol. 4, pp. 625–634, Mar. 2005.

[119] A.-M. Mourad, A. Guéguen, and R. Pyndiah, "Impact of the spreading sequences on the performance of forward link MC-CDMA systems," in *Proc. of IEEE Eigth International Symposium on Spread Spectrum Techniques and Applications*, (Sydney, Australia), pp. 683–687, Sept. 2004.

[120] D. Mottier and D. Castelain, "A spreading sequence allocation procedure for MC-CDMA transmission systems," in *Proc. of IEEE VTC 2000-Fall*, vol. 3, (Boston, MA, USA), pp. 1270–1275, Sept. 2000.

[121] A.-M. Mourad, A. Guéguen, and R. Pyndiah, "MAI analysis for forward link mono-dimensionally spread OFDM systems," in *Proc. of IEEE VTC 2004-Spring*, vol. 3, (Milan, Italy), pp. 1528–1533, May 2004.

[122] D. Ramakrishna, N. B. Mandayam, and R. D. Yates, "Subspace-based SIR estimation for CDMA cellular systems," *IEEE Trans. Veh. Tech.*, vol. 49, pp. 1732–1742, Sept. 2000.