Power Control and Adaptive Resource Allocation in DS-CDMA Systems

FREDRIK BERGGREN
Power Control and Adaptive Resource Allocation in DS-CDMA Systems

FREDRIK BERGGREN

A dissertation submitted to the Royal Institute of Technology in partial fulfillment of the requirements for the degree of Doctor of Philosophy

June 2003

TRITA—S3—RST—0308
ISSN 1400—9137
ISRN KTH/ RST/ R—03/ 08—SE

RADIO COMMUNICATION SYSTEMS LABORATORY
DEPARTMENT OF SIGNALS, SENSORS AND SYSTEMS
Abstract

With an increased request for wireless data services, methods for managing the scarce radio resources become needful. Especially for applications characterized by heterogeneous quality of service (QoS) requirements, radio resource allocation becomes an extensive task. Due to the necessity of sharing the radio spectrum, mutual interference will limit system capacity. Transmitter power control is a well-known method for upholding required signal quality and reducing the energy consumption.

In this thesis, transmission schemes based on distributed power control are developed and analyzed for cellular DS-CDMA systems. The schemes are designed for providing various QoS, while assuring global stability and rapid convergence. First, we suggest an iterative power control algorithm which handles congested situations by autonomously removing radio connections. This algorithm is then extended with a greedy rate allocation procedure, with the purpose to maximize throughput in a multirate system, where users have a limited number of discrete transmission rates.

For supporting downlink nonreal time services with a required average data rate, joint power control and scheduling is investigated. It is found that time division has merits in terms of increased capacity and multiuser diversity effects. For this, we suggest and evaluate power control and channel adaptive schedulers of different adaptation rate.

Finally, a class of receivers, utilizing successive interference cancellation with soft feedback is investigated. We determine the optimal power control solution and characterize its user capacity region.
Acknowledgments

This work has been supported by many people to whom I wish to express my gratitude. I wish to thank my main adviser Prof. Jens Zander and Prof. Ben Slimane at KTH; Prof. Seong-Lyun Kim at ICU, Korea, and Prof. Riku Jäntti at HUT, Finland. Their positive attitude and encouragement provided inspiration to many ideas and undoubtedly gave dynamism to my research studies. Especially I want to thank Prof. Zander for the freedom I have been given in this research project. I would like to thank Prof. Roy Yates and Prof. Narayan Mandayam, both at the Wireless Information Network Laboratory (WINLAB), which I visited during the spring of 2002, for interesting discussions and feedback on selected parts of this thesis.

Special thanks go to my colleagues at the Radio Communication Systems group for many valuable discussions. Thereto did Magnus Lindström never hesitate to cure my computer problems. Lise-Lotte Wahlberg’s assistance with all practical matters is greatly appreciated.

A major part of this work has been financially supported by VINNOVA within the Explorative Systems-Integrated Technologies (EXSITE) project. My trips abroad have been supported by grants from Knut och Alice Wallenbergs Stiftelse, Pleijel fonden and the IEEE Communications Society.

Finally, the never-ending support from my parents made it all possible.
List of Abbreviations

AGC  Automatic Gain Control
ALP  Active Link Protection
ARQ  Automatic Repeat reQuest
ATM  Asynchronous Transfer Mode
BER  Bit Error Rate
CDMA Code Division Multiple Access
CDMA/HDR High Data Rate CDMA
CSMA Carrier Sense Multiple Access
DB   Distributed Balancing
DCPC Distributed Constrained Power Control
DPC  Distributed Power Control
DS-CDMA Direct Sequence CDMA
ET   Equal Time
FDMA Frequency Division Multiple Access
FEC  Forward Error Correction
FER  Frame Error Rate
FF   Fractional Fair
FM   Foschini-Miljanic
GDCPC Generalized DCPC
GRP  Greedy Rate Packing
GRR  Gradual Removals Restricted, Gradual Rate Removal
GSPC Generalized SPC
IS-95 Interim Standard-95
ISI  InterSymbol Interference
JOR  Jacobi Over-Relaxation
MC   Multi Carrier
MIMO Multiple Input Multiple Output
MPA  Minimum Power Assignment
MUX  Multiplexer
OFDM Orthogonal Frequency Division Multiplexing
PA   Power Amplifier
<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>PCMA</td>
<td>Power Controlled Multiple Access</td>
</tr>
<tr>
<td>PCS</td>
<td>Personal Communication System</td>
</tr>
<tr>
<td>PF</td>
<td>Proportional Fair</td>
</tr>
<tr>
<td>PRMA</td>
<td>Packet Reservation Multiple Access</td>
</tr>
<tr>
<td>OFDM</td>
<td>Orthogonal Frequency Division Multiplexing</td>
</tr>
<tr>
<td>QoS</td>
<td>Quality of Service</td>
</tr>
<tr>
<td>RAKE</td>
<td>Rake</td>
</tr>
<tr>
<td>RB</td>
<td>Relatively Best</td>
</tr>
<tr>
<td>RF</td>
<td>Radio Frequency</td>
</tr>
<tr>
<td>RR</td>
<td>Round Robin</td>
</tr>
<tr>
<td>RRM</td>
<td>Radio Resource Management</td>
</tr>
<tr>
<td>SAS</td>
<td>Soft And Safe</td>
</tr>
<tr>
<td>SCDMA</td>
<td>Scheduled CDMA</td>
</tr>
<tr>
<td>SDMA</td>
<td>Spatial Division Multiple Access</td>
</tr>
<tr>
<td>SIR</td>
<td>Signal-to-Interference Ratio</td>
</tr>
<tr>
<td>SMIRA</td>
<td>Stepwise Maximum Interference Removal Algorithm</td>
</tr>
<tr>
<td>SMS</td>
<td>Short Message Service</td>
</tr>
<tr>
<td>SNR</td>
<td>Signal-to-Noise Ratio</td>
</tr>
<tr>
<td>SPC</td>
<td>Selective Power Control</td>
</tr>
<tr>
<td>SRA</td>
<td>Stepwise Removal Algorithm</td>
</tr>
<tr>
<td>TDMA</td>
<td>Time Division Multiple Access</td>
</tr>
<tr>
<td>TPC</td>
<td>Transmitter Power Control</td>
</tr>
<tr>
<td>UMTS</td>
<td>Universal Mobile Telecommunications System</td>
</tr>
<tr>
<td>WCDMA</td>
<td>Wideband CDMA</td>
</tr>
<tr>
<td>WLAN</td>
<td>Wireless Local Area Network</td>
</tr>
</tbody>
</table>
# Contents

1 Introduction ................................................. 1
   1.1 Wireless Data ............................................ 2
   1.2 Radio Resource Management .............................. 5
      1.2.1 Transmitter Power Control ............................. 5
      1.2.2 Transmission Rate Control ............................ 6
      1.2.3 Admission Control ..................................... 7
      1.2.4 Congestion Control ..................................... 8
      1.2.5 Transmission Scheduling ............................... 8
   1.3 Wireless Communications ................................. 9
      1.3.1 Multiple Access ....................................... 9
      1.3.2 Cellular Radio Systems ............................... 10
   1.4 Previous Work .......................................... 12
      1.4.1 Fixed-Rate Power Control ............................. 12
      1.4.2 Variable-Rate Power Control .......................... 16
      1.4.3 Transmission Scheduling and Multiple Access ........ 18
   1.5 Scope of the Thesis ...................................... 20
      1.5.1 Problem Background ................................... 20
      1.5.2 Scope .................................................. 21
      1.5.3 Contributions .......................................... 22
   1.6 Thesis Outline .......................................... 25

2 Models and Performance Evaluation ....................... 27
   2.1 Radio Wave Propagation .................................... 27
   2.2 System Architecture ....................................... 29
      2.2.1 Receiver Model and Communication Quality ............ 29
      2.2.2 System Layout and Access Scheme ..................... 32
      2.2.3 Control Signaling ...................................... 33
      2.2.4 Power Control Functionality ........................... 33
   2.3 Performance Evaluation ................................. 34
      2.3.1 Performance Measures ................................. 35
3 Preliminaries
3.1 Noisy Power Control Problem ........................................ 37
3.2 Iterative Solution Methods ........................................ 41
   3.2.1 Matrix Norms ............................................... 41
   3.2.2 Iterative Power Control ................................... 41
   3.2.3 Rate of Convergence ...................................... 43
   3.2.4 Standard Interference Functions ......................... 46
3.3 Local Loop Stability ............................................... 46

4 Generalized DCPC ................................................ 49
4.1 Introduction ....................................................... 49
4.2 Distributed Power Control Algorithm ......................... 50
   4.2.1 Energy Conservation ..................................... 50
   4.2.2 Convergence in Feasible Systems ..................... 51
   4.2.3 Convergence in Infeasible Systems .................... 54
4.3 Numerical Results ............................................... 56
   4.3.1 Feasible System .......................................... 57
   4.3.2 Infeasible System ...................................... 59
4.4 Concluding Remarks ............................................. 60

5 Multirate Power Control .......................................... 63
5.1 Introduction ....................................................... 63
5.2 Transmission Rate Assignment ................................ 64
   5.2.1 Refined System Model ................................... 64
   5.2.2 Greedy Rate Packing .................................... 66
   5.2.3 Throughput Maximization ................................ 68
   5.2.4 Minimum Rate Requirements ......................... 69
   5.2.5 Downlink GRP ............................................ 70
5.3 Iterative Power Control ......................................... 71
5.4 Power Control with Gradual Rate Removal .................. 74
5.5 Numerical Results ............................................... 76
5.6 Concluding Remarks ............................................. 78

6 Joint Power Control and Intracell Scheduling ................. 81
6.1 Introduction ....................................................... 81
6.2 Minimum Energy Problem ..................................... 83
   6.2.1 Refined System Model ................................... 83
   6.2.2 Problem Definition .................................... 84
6.3 Intracell Scheduling ............................................. 84
   6.3.1 Energy Conservation ................................... 87
   6.3.2 Minimum Time Span ................................... 88
6.4 Distributed Power Control ..................................... 89
   6.4.1 Convergence ............................................. 90
   6.4.2 Rate of Convergence .................................... 92
7 Transmission Scheduling over Fading Channels 101
  7.1 Introduction ................................................. 101
    7.1.1 System Model ........................................ 102
  7.2 Transmission Scheduling ................................. 104
    7.2.1 One-by-One Scheduling ................................ 104
    7.2.2 Slow Scheduling ...................................... 105
    7.2.3 Fast Scheduling ...................................... 107
  7.3 Scheduling Gain ............................................ 109
    7.3.1 Unconstrained Bandwidth Case ....................... 110
    7.3.2 Constrained Bandwidth Case ......................... 112
  7.4 Multicellular Case ......................................... 116
  7.5 Measurement Delay ......................................... 118
  7.6 Multiuser Scheduling ...................................... 120
    7.6.1 Generalized Scheduler ................................ 121
    7.6.2 Multiuser Scheduling Gain ............................ 122
    7.6.3 System Multiuser Scheduling Gain .................... 125
  7.7 Concluding Remarks ....................................... 126

8 Linear Successive Interference Cancellation 129
  8.1 Introduction ................................................ 129
  8.2 Receiver Analysis ......................................... 130
    8.2.1 Single-rate System ................................... 131
    8.2.2 Multirate System .................................... 134
  8.3 Power Control and User Capacity .......................... 134
    8.3.1 Minimum Power Allocation ............................. 135
    8.3.2 Partial Successive Interference Cancellation ......... 136
    8.3.3 Limited Successive Interference Cancellation ......... 136
    8.3.4 Effective Bandwidth .................................. 137
  8.4 Decoding Order ............................................. 141
  8.5 Iterative Power Control .................................. 144
  8.6 Numerical Results ......................................... 145
  8.7 Concluding Remarks ....................................... 148

9 Conclusions 149
  9.1 Summary .................................................. 149
  9.2 Discussion ............................................... 152
  9.3 Further Studies .......................................... 153

A Multicellular Outage Probability 155

B Average Throughput with Time Delays 157
C Scheduling Gain Expressions  
  C.1 Equivalent Representations  
    C.1.1 Equation 7.12  
    C.1.2 Equation 7.26  
    C.1.3 Equation 7.43  
    C.1.4 Equation 7.45  
  C.2 Limit Values  

References
## List of Figures

1.1 Example of a relation between transmission rate and SIR .................. 7
1.2 Cellular system grid ........................................... 11

2.1 Block diagram of closed loop power control .......................... 34

4.1 Two-user example of fixed points .................................. 54
4.2 Outage probability of GDCPC .................................... 56
4.3 Average power of GDCPC ........................................ 57
4.4 Convergence rate of GDCPC ....................................... 58
4.5 Outage probability for infeasible system of GDCPC ................. 59
4.6 Sensitivity of the removal parameter ............................... 60

5.1 Two-user example with multiple rates ................................. 74
5.2 Average power of GSPC ........................................... 76
5.3 Average throughput of GSPC ........................................ 78

6.1 Two-user example of intracell scheduling ............................. 89
6.2 Relative energy efficiency .......................................... 96
6.3 Convergence rate of power control .................................. 97
6.4 Relative capacity increase ........................................... 98

7.1 Feasible time fractions of the fast scheduler ......................... 109
7.2 Scheduling gain for Ricean channels ................................ 111
7.3 Scheduling gain with RAKE receiver ................................ 112
7.4 Scheduling gain with rate constraint ................................ 114
7.5 Relative throughput with discrete rates .............................. 116
7.6 Scheduling gain for logarithmic rate ................................ 117
7.7 Scheduling gain with measurement delay ............................ 120
7.8 Multituser scheduling gain ......................................... 122
7.9 Multituser scheduling gain ......................................... 123
7.10 Multituser scheduling gain ........................................ 124
7.11 Average scheduling gain .......................................... 126

8.1 Block diagram of SIC detector ...................................... 133
8.2 Relative user capacity of SIC .............................. 138
8.3 Relative user capacity of SIC .............................. 139
8.4 Relative user capacity of SIC .............................. 140
8.5 Relative power reduction of SIC ............................ 145
8.6 Relative received power reduction of SIC ............... 146
8.7 Average power of SIC ................................. 147
List of Tables

5.1 Steady state performance of GRR-GPC . . . . . . . . . . . . . . . . . . 80
5.2 Steady state performance of GRR-GSPC2 . . . . . . . . . . . . . . . . . 80
5.3 Steady state performance of GRR-SPC . . . . . . . . . . . . . . . . . . 80
Chapter 1

Introduction

Recent years tremendous success and sky-rocketing growth of wireless personal communications have necessitated careful management of the common radio resources. As the demand for wireless speech- and wireless data services is likely to grow, developing transmission schemes that efficiently utilize the available radio resources is of major importance. In comparison to a wireline channel, the radio channel is time variant and transmissions should preferably be adapted to the channel quality as well as the service requirements. With user mobility, this imposes several challenges in providing tetherless communications. Mainly due to the scarcity in available frequency spectrum, some form of resource sharing must be considered. In practice, all sharing methods introduce some form of interference, impairing the ability to communicate. To ward off a complete breakdown of the system, interference control is a necessity. Transmitter power control is a key technique to combat and reduce unnecessary interference. By adjusting the transmit powers, a better balance between the desired signal and the interference can be achieved at the receiver. As a consequence, the capacity of the system may increase so that more users can be accommodated and more data be transmitted. Ensuing gains also result in reduced energy consumption of the transmitters. As wireless data services require different quality of service (QoS) as compared to pure speech services, a larger freedom in allocating the system resources arises. Nevertheless, the problem of supporting the respective QoS while maintaining low energy consumption and high capacity becomes complex and resource allocation in this new context needs to be readdressed.

For wireline communications, extended capacity can be obtained by adding more wires or fibers. Wireless bandwidth on the other hand, can usually not be easily expanded. To increase the capacity of a wireless system for a given bandwidth, more radio access ports could be added or methods for interference suppression could be deployed. In this thesis, we put focus on interference management for cellular systems and develop adaptive schemes mainly suited for
providing wireless data delivery. Among the possible approaches, power control is a versatile technique which can concurrently achieve many of the networking objectives. Therefore, distributed power control algorithms will be developed, serving as building blocks in the transmission schemes. To further exploit the resource allocation possibilities, the power control functions will be combined with adaptive methods for transmission rate allocation and transmission scheduling, in order to provide heterogeneous QoS in a suitable way.

The following of this chapter briefly recaptures radio resource management (RRM) techniques, the concepts of cellular radio systems and highlights some of the most relevant previous work. The scope and contributions of this thesis are listed at the end of the chapter.

1.1 Wireless Data

The evolution of the Internet has undoubtedly created a vibrant source for information retrieval and networking. Just a few clicks of the mouse away, is an enormous amount of data available, some of it useful. Motivated by the almost unprecedented success of the Internet, a great deal of research and recent product development have focused on making the data access wireless. Wireless local area networks (WLANs) provide access with a limited coverage to a relatively small deployment cost, whereas cellular systems tend to offer the opposite. As second generation cellular systems have limited capability of providing broadband\(^1\) wireless data, third generation cellular systems are currently being developed. The deployment of these systems has been spurred by the success of voice- and short messaging services (SMSs) in second generation systems. However, the once so rollicking mood of the communications industry faded as deployment plans slowed down when markets took a downturn. The decline was certainly fueled by the enormous costs of acquiring spectrum licenses or stipulating excessive coverage promises. Thereto, the business models for these systems have been outpaced by the technology advancements and so far lack driving “killer applications”. With more functionality, the price-alone competition of voice services has a compliment in value-added services as an alternate source of revenue, which could pedal these systems. To capture the attention of users satisfied with current speech services and for wireless data to take off, more attractive options or improved services may have to be marketed and supported in forthcoming wireless systems. Multimedia, i.e., simultaneous transmission of several types of information (e.g., voice, data and video) is such a direction that has given an impetus to communication networks in general. Typically a user could be interested in setting up a video or speech call, which requires that the time relation between the information entities of the data stream must be preserved. For non-

\(^1\)Several definitions of broadband exist, related either to the maximum data rate or the channel bandwidth. WCDMA [32] targets rates of 384 kbps in urban and suburban areas and peak rates of 2 Mbps for low-range and indoor environments.
real time data services such as file retrieval and email delivery, the user receives bandwidth when available, i.e., in a best effort fashion. Music and images, which can be characterized as some form of retrieval-type of service, have rather different requirements compared to the conversational-type, like video-telephony, which has interaction between several end users. Many of the services available through the Internet could indeed be interesting for wireless access but there are also possibilities of wireless specific services, including interactive gaming with location and positioning features. Another service in current development, is the use of video-telephony for telemedicine and remote diagnosing. The SMS in second generation cellular system, was initially believed to be attractive for business customers only. Remarkably, this service is now widely used, and serves as a nonvanishing source of income for the network operators. Therefore, at this point it is difficult to predict what services will be requested or successful in the future. Anyhow, different type of services will generate traffic with varying data rate and the network control must consider and take advantage of the mixed traffic properties for maximizing spectral efficiency. Spectral efficiency describes the ability of the system to deliver useful data, given a certain amount of dedicated radio spectrum. In cellular systems, spectral efficiency is often measured in the unit bits per second per Hertz per cell or Erlang per Hertz per cell. A spectrally efficient system, offers reduced need for radio spectrum and a more sparse network infrastructure. Together, this will reduce capital- and operating costs for the operator. Several factors play a role in the efficiency measure, e.g., the amount of overhead signaling and the features of the air interface, including coding and modulation formats. Moreover, which also is the main focus of this thesis, to enable spectral efficiency on a network level, suitable multiple access schemes and their resource allocation algorithms are essential.

An example is the universal mobile telecommunications system (UMTS) standard, where a layered architecture for QoS support is outlined. The QoS support structure relies on layered bearer services, which are involved to compose the end-to-end bearer service. To each bearer service, QoS attributes are defined. These attributes serve to map the end-to-end QoS requirements to appropriate requirements for each bearer service. The attributes typically describe requirements on bit rates, delays and priorities. To classify the services, the traffic is in UMTS divided into the classes; conversational, streaming, interactive and background [1]. Problems in multirate schemes are both related to the bearer mapping, i.e., how to choose rates and schedule the transmission for obtaining the QoS and more physical layer type of issues, such as; how to map the bit rates into the given bandwidth and how to inform the receiver about the characteristics of the signal. For a service where the delay requirement is loose, longer delays allows for longer interleaving, more retransmissions and therefore lower signal-to-interference ratio (SIR) that in turn could increase capacity. On the protocol level, it is necessary that the terminal capability can be matched to the content of the service as well as providing billing and security functionality.

Fueled by the growth of Internet use, applications for higher data rates and
advanced multimedia services can be expected. Thus the network should provide different transmission rates and/or QoS requirements, sharing the capacity in the most proper way. In contrast to speech services where the QoS metrics usually considered are call dropping- and blocking probability, other measures are more valid for multimedia services. Throughput, delay and service outage probability are typical such measures commonly used in that respect. Clearly, they may reflect the “speed” of the bits and therefore are naturally coupled to the perceived quality. However, in this thesis we will additionally consider the transmission power as part of the QoS metric. For example, portable devices rely upon a limited source of energy, the battery. Discouragingly, battery performance has so far improved relatively slowly, forcing the network control to consider judicious management of the energy resources, while taking into account the QoS requirements and the ever changing wireless environment. Energy consumption issues will become even more evident to the mobile terminal users if they will have the option to transmit larger amount of data more rapidly, experiencing faster battery drain. For such data service users, it could be so that the battery will literally be perceived as a black box “containing” a limited amount of bits rather than seeing it as an energy source, providing a certain operational time. Albeit the research in more efficient batteries and low-power electronics will result in products that can cope with such a situation, a relevant criteria for the network control still is to successfully deliver the data bits with as little energy as possible. Thus, ingenious resource allocation will help to “put” more bits into the battery and make the system energy-efficient. In the base station, the cost of energy is per se not that critical, as power is available from the mains outlet. Energy conservation in the base station, however, will lead to reduced interference and therefore higher capacity and spectral efficiency of the system.

Energy-efficient wireless communications is a topic that appears frequently in the research literature of today. The research focus is wide and typically spans from; how to design protocols that minimize energy consumption but at the same time are capable to drain the maximum energy out of the battery; to efficient signal processing implementation. Issues that are of concern in the lower layers include dynamic power management, modulation- and error control schemes and reduction of mode transitions of the receiver. Not only should energy conservation be considered at each single layer but also jointly for the whole system. In the higher layers, an objective is to schedule the data flow in order to minimize the time the radio needs to be powered. Also there is a tradeoff and close relation between throughput and energy consumption for the error recovery scheme in the transport protocols. That includes questions whether it would be beneficial with many low power retransmissions or fewer with higher power; automatic repeat request (ARQ) versus forward error correction (FEC) and to avoid retransmitting useless data. In this thesis, the aspect of energy efficiency will be limited to the transmitter power control and scheduling part.

---

2 A mobile terminal can be regarded to be in a mode of, e.g., transmit, receive, idle, sleep or off.
Interestingly it has been argued that a key issue and fundamental barrier for the success of wireless data is low data rate, high energy consumption and high cost per transmitted bit [124].

1.2 Radio Resource Management

One lesson to be drawn from the research efforts of cellular radio systems so far, is the importance of radio resource management. As was mentioned, when having different QoS requirements, controlling the radio resources may not be straightforward. However, suitable RRM tools can be identified that can constitute the basic foundation also in this case. Typical such methods previously suggested and successfully applied in single-rate systems are admission-, congestion- and power control. These methods aim to provide a benign interference level in the system. Therefore, they will be naturally connected to the case of wireless data systems. Furthermore, for wireless data, by exploiting the delay tolerance of certain services, the area of transmission scheduling becomes a possible interference management method to incorporate. It should be pointed out here, that there exist other interference suppression methods which would improve the energy efficiency and capacity, like adaptive antennas, multiple input multiple output (MIMO) systems as well as diversity schemes, but those are excluded in this work as we try to limit this thesis to the above RRM view.

1.2.1 Transmitter Power Control

One controllable radio resource, highly related to the network capacity, is the transmitter power. For example, delay sensitive users with stringent bit error rate (BER) requirements can be accommodated by adapting their transmit powers to the channel so as to increase their SIR, resulting in a lower bit error rate. However, this causes an increase in the interference experienced by the other users, in turn increasing their BER. Consider for example a DS-CDMA system, which is known to be interference limited [116]. The system resource a user occupies can be related to the generated interference spectral density level, which is generally proportional to the received power level and therefore the data rate. Consequently, a larger received, and thus transmitted power, implies that the mobile occupies a larger portion of the system resources. In this work will design efficient resource allocation policies by the use of transmitter power control as a common ingredient. Power control is an active area of research and much work has been performed already for the so called fixed-rate system. However, as variable transmission rates are introduced in the network, the dimension of rate allocation is added to the power control.

In a real system, the effective transmission rate is closely coupled to the SIR and power control has shown to be an efficient instrument for controlling the SIRs. Hence, it becomes intriguing and natural to investigate joint power- and rate control schemes. The problem formulation for the classical fixed-rate power
control is usually considered to be; to find the minimum power assignment that supports the required SIR for as many users as possible. Due to the different services in next generation personal communication systems (PCs) mentioned before, defining a general power control problem is not as straightforward and perhaps not that meaningful in a multirate system. Certain users may perhaps accept a varying data rate as long as the average rate is satisfactory. Others may primarily be interested in delivering their data with minimum energy, while sacrificing throughput and allowing large delays. Caused by the different purposes and performance measures, developing a generic power control algorithm seems to be cumbersome. The areas of fixed- and variable-rate power control are not disjoint. If the duration of the data message is sufficiently long and the power control frequency is high, quality measurements can be obtained, which paves the way for executing SIR-based power control. For short messages, or very bursty interference, measuring the channel is more troublesome and other types of power control [15, 76] should be considered, if any. If quality requirements dependent on, e.g., the SIR can be specified, SIR-based power control can be performed with those targets. When certain quantities like the SIR-targets, link gains, noise etc., are known or can somehow be estimated, the fixed-rate power control problem can be written as a linear programming problem. For a given instant, it can be shown that there exists a unique optimal solution, in terms of finding the componentwise smallest power vector, at which the SIR-targets are met. A vast number of distributed and centralized algorithms have been developed for solving this problem; consequently we may revisit them, or rather the principles upon which they rely, when the transmission rates have been assigned. Thus at this stage, we envisage that the problem of assigning transmitter powers for multirate systems is as much a transmission rate assignment problem, from which the power control problem is inherited.

1.2.2 Transmission Rate Control

In order to assign the data rates, one possibility is to relate them to SIR-targets. We may then for example assume access to infinitely many SIR-targets, resulting in a continuous relation between transmission rate and SIR. If instead the number is limited, it becomes a discrete relation. In Fig. 1.1, an example of a linear and a discrete relation between the effective data rate and SIR is depicted. A linear relation is motivated in DS-CDMA systems where the processing gain is adapted by the symbol duration and also the fact that the rate in wideband Gaussian channels is approximately linear (cf., footnote in Section 7.1.1). With adaptive modulation and coding, a linear growth can be too optimistic and a logarithmic mapping is often used. An immediate interpretation of the assumption of a relation between rate and SIR-target, is a guaranteed maximum BER. If the connection is established at a certain SIR, say the predetermined target, it can be assumed that there exists a one-to-one relation to a corresponding bit error probability. Thus, by choosing the proper SIR-targets, the data rate can
be offered with a specified bit error probability. In particular for a DS-CDMA system, a high SIR-target compensates for the low processing gain when using a high data rate. To design algorithms that assign the proper SIR-targets and rates to the users, is a key issue. Of practical interest in a wireless system are algorithms of fully distributed or semi-distributed form; relying on limited information, e.g., the link gains within the cell.

1.2.3 Admission Control

The purpose of admission control is to preserve the quality of ongoing connections while admitting new ones. A good admission scheme would grant access to as many users as possible without jeopardizing the maintenance of ongoing calls. Data rate assignment can be viewed upon as a more general form of connection admission, where each link has several SIR-targets to choose from. A very common type of problem definition for multirate systems is system throughput maximization. An approach for that problem is to admit as much data rate into the system as possible while meeting constraints of quality, data rate, maximum power etc. Several issues arise here though, since there must somehow be a decision of which user should be admitted to transmit and at what data rates. To use admission control for such a purpose, it is necessary to find constraints such that an admitted rate combination is feasible within the dynamic power
range. For analyzing CDMA systems, an often used simplification is to focus on a single cell only, while approximating interference coming from outside the cell to be constant. The single cell approach gives that closed form admission constraints can often be found rather straightforwardly, which will also be dealt with in this thesis.

1.2.4 Congestion Control

When a system is highly congested, there might occur situations where a subgroup of users cannot be supported with their required data rate. Congestion may occur when the aggregate resource requirements are close to the capacity limit, traffic exhibits bursty behavior, radio channels are poor or an unevenly distributed user population. Most power control algorithm would react to such a situation by increasing the powers. The result will be very high interference and lower system performance. The notion of “power warfare” has been used to characterize that phenomenon [44]. One cure for such an infeasible situation is reduce the load of the system. Typically this can be achieved by removing connections from the channel, decreasing data rates and QoS requirements of certain links. How to deal with such issues will be very important in a multirate system. As for admission control, congestion control could be used for throughput maximization by filling the system up with data rate and decrease until feasibility occurs. Yet from whom data rate is to be removed is an open control problem. One important issue in congestion control is also to detect the infeasible situations as fast as possible, so that actions can be quickly taken. Having only limited channel information of the users in the network makes the problem difficult though. The congestion control should also exhibit certain degree of system robustness, such that recovery from overload situations becomes smooth.

1.2.5 Transmission Scheduling

Scheduling of transmission attempts in time can be used to differentiate between service classes, sessions or to increase fairness among users. For CDMA, scheduling has an additional feature in interference management. In cellular system, interference can be decomposed into intracell interference, originating from simultaneous transmissions within the cell and intercell interference caused by terminals transmitting in other cells. If properly done, a selective choice of transmissions could therefore possibly reduce the interference amount without sacrificing throughput. Some services do not require an instantaneous data rate but rather an average rate, i.e., that an amount of data is delivered over a certain time interval, allowing for more flexible use of the channel. In addition to power control, utilizing the possibility of scheduling the transmissions within the cell, intracell interference can be efficiently mitigated. Scheduling is highly related to the rate control previously described and is a special case of it, since setting
the data rate and power to zero can be regarded equivalent to scheduling. Predictions suggest that future traffic will be highly asymmetric with the downlink carrying a major part of the load. For such a scenario, scheduling downlink traffic in each cell locally, might be a simple and attractive candidate scheme for supporting nonreal time data services.

1.3 Wireless Communications

By establishing a wireless connection, data from a transmitter can be delivered to a receiver by means of an information-bearing signal. The flexibility of setting up a nonwired link between the communicators has become the driving force of wireless information distribution and its widespread deployment. Early applications of which wireless technology became synonymous of, were broadcasting services such as radio and TV. As the radio channel is essentially a broadcast medium with high connectivity, radio and TV fit well as applications. These services mainly relied upon continuous time and amplitude varying signaling, also known as analogue communication systems. Much later, by the advent of more sophisticated transmission schemes processing the signals digitally, the PCSs as we know them today, were getting a reality. Driven by competition between service providers and supervised by government regulations, a great deal of interest has focused on enhancing the use of common radio resources.

1.3.1 Multiple Access

The wide coverage of the radio signal, which is a favorable property in broadcasting, is in a PCS instead problematic. When several users wish to access a common channel, some form of separation of their waveforms must exist to distinguish between them at the receiver. The basis for any air interface design is to determine how to share the common medium among users, i.e., the multiple access scheme. Depending on if the signaling bandwidth is chosen to be small or large compared to the channel coherence bandwidth, different narrowband or wideband techniques find their applications. The following are the most well-known.

- **Frequency Division Multiple Access (FDMA)**
  The separation is performed by dividing the radio spectrum into orthogonal channels. Each user is assigned a unique frequency channel upon request, which is not used by others when idle.

- **Time Division Multiple Access (TDMA)**
  The separation is performed by dividing the radio spectrum into time slots. In each slot, only one user may transmit at a time.

By reusing the frequencies and time slots in geographical locations separated sufficiently apart, more users can be accommodated. In spread spectrum mul-
multiple access, the information bearing signal is spread over a larger bandwidth than the signal itself. Not being spectrally efficient for one single user, spread spectrum systems become bandwidth efficient in the multiuser case since it is possible to share the same spreading bandwidth.

- **Code Division Multiple Access (CDMA)**
  The separation is performed by unique codes assigned to the users. The codes are used for either modulating the signal or alternating the carrier frequency.

Usually FDMA is used together with TDMA or CDMA to separate the spectrum into smaller bands, which are then divided in a time- or code division fashion. Utilizing the geographical separation of the receiver, spatial multiple access can be considered.

- **Space Division Multiple Access (SDMA)**
  The separation is performed by directing the emitted energy directly toward the receiver. Directional antennas support spatial separation.

Various hybrid schemes that consist of the aforementioned fundamental techniques also exist. Multicarrier (MC) techniques have also been outlined, e.g., MC-CDMA. For WLANs, orthogonal frequency division multiplexing (OFDM) has drawn some interest considering its suitability for high-speed data transmission. Several methods for providing multiple access with OFDM exist, designating the subcarriers by TDMA, CDMA or adaptively within each cell. In practice all these schemes require some form of “orthogonality” coordination, e.g., frequency/time/code allocation, if the communication should be successful. Another approach is to minimize the overhead information exchange by connection-free resource sharing. Schemes in this class, often referred to as packet data, offer more flexible use for shorter data deliveries but risk that the transmission fails due to simultaneous transmissions, which necessitates retransmission policies. Various schemes exist and they are usually divided into random-, scheduled-and hybrid access. Common methods include ALOHA, carrier sense multiple access (CSMA) and packet reservation multiple access (PRMA). For moderate load, these type of access schemes are well suited for packet data since a rather low delay can be offered. Increasing the load though, can make them collapse.

### 1.3.2 Cellular Radio Systems

Inherent from the fact that the permitted spectrum to use is limited, the problem of supporting a larger population wireless communications arises. The concept of cellular radio, which dates back to the 1960s [34], allows channels to be reused over the geographical area having sufficient spatial separation. These systems are mostly interference limited implying that the capacity is limited even in the absence of maximum power constraints. Dividing the area into cells, provides
system planning with higher spectral efficiency. To plan such a system, a common way of breaking the area up, is to consider cells shaped as hexagons which will tessellate the service area. In each cell a portion of the available channels is allocated and neighboring cells are assigned different channels. In that way the spectrum can be reused as many times as necessary, if the interference can be kept below acceptable levels. For CDMA systems, frequency planning is unnecessary and a reuse of unity can be used. Channel assignment can be performed dynamically, adapting to the geographical load variance, or in a fixed fashion. In fixed channel assignment, the channels are assigned, independently of traffic load, to cells rather than users. In contrast, a more spectral efficient policy is dynamic channel assignment which assigns users the most appropriate channel. The price for this is more overhead signaling and channel measurements. Interference can also be suppressed by sectoring the antenna such that interference is not emitted in the direction at which the intended receiver is not located. Extending this principle, the concept of adaptive antennas combined with spatial- and temporal signal processing can further suppress interference. Let us henceforth denote by mobile or user, a member of the collective group that can communicate using some form of radio terminal. The radio access for a mobile is served by base stations, which usually are located in the middle or the corner of a cell, see Fig. 1.2. While the user is roaming around over the
area, base station selection, or a handover, is performed so that a mobile usually is connected to one single serving base station at a time. F/TDMA schemes usually perform hard handover to a neighboring cell, if its signal strength plus some threshold exceeds the current cell’s signal strength. Hard handover can be seamless or nonseamless. Seamless hard handover means that the handover is not perceptible to the user. In practice, a handover that requires a change of carrier frequency is usually performed as hard handover. In CDMA, due to the unity reuse, two or more base stations can receive the same signal so that a soft handover can be performed. In the CDMA downlink, macro diversity can be obtained by combining signals from different base stations, since they can be regarded as additional multipath components at the receiver. The gain of soft handover in the downlink is thus highly dependent on the ability of the receiver to resolve the multipath components. A special case of soft handover is the so-called softer handover, where the radio links that are added and removed belong to the same cell site of co-located base stations, from which several sector-cells are served. During softer handover, a mobile is in the overlapping cell coverage area of two adjacent sectors of a base station. Thus softer handover requires that the antennas cannot provide perfect sectorization. The communications between mobile and base station can take place concurrently via two air interface channels, one for each sector separately. In the downlink, there are very few differences between softer and soft handover. However, in the uplink direction, soft handover differs significantly from softer handover since the received data then typically undergoes selection combining rather than maximum ratio combining as for softer handover.

1.4 Previous Work

The work in this thesis touches upon different areas of power control and transmission schemes, therefore this section is divided into parts. As the work in these areas has been extensive, we limit this section to the most representative work in the respective direction.

1.4.1 Fixed-Rate Power Control

The major concern in this thesis is quality based power control. Early work on quality based power control for combating co-channel interference was performed by Bock and Elstein [26] already in the year 1964. They found that the power assignment problem can be formulated as a linear programming problem. Aein [2] investigated power assignment to mitigate co-channel interference in both noiseless and noisy satellite systems. It was found that the problem of SIR-balancing in noiseless systems, i.e., to obtain the same quality on all links, was reduced to an eigenvalue problem for nonnegative matrices. Existence and uniqueness of a feasible power solution associated with a link gain matrix were found as a consequence of Perron-Frobenius theorem. Nettleton and Alavi [3,86], extended the
concept of SIR-balancing to spread spectrum systems without background noise. The above concepts were further enhanced when applied by Zander to narrow-band systems [128]. Therein, the optimum power assignment for minimizing the outage probability in terms of finding the maximum achievable SIR that all links can simultaneously reach, is derived. If reciprocity in link gains can be assumed, it follows straightforwardly that downlink and uplink maximum achievable SIR are the same [3,131]. Grandhi et al., show that for a noiseless system, there exists only one common balanced SIR and only one positive power eigenvector, which also yields the maximum achievable SIR [37]. Recently, Wu [118] investigated balancing for heterogeneous SIR-targets. Differing from the noisy case, these targets cannot be chosen individually, rather they are dependent on the ratio with the minimum required SIRs. In [130], Zander extends the noiseless case to include nonzero background noise and show that the same SIR can be achieved arbitrarily close, if there are no constraints on transmitter power. In [39], Grandhi et al., introduce maximum transmitter power constraints which imply that in the noisy case, there will always be at least one user utilizing the maximum power.

Focus has also been on developing practical algorithms for solving the above problems, without the excessive effort of collecting necessary information to a centralized controller. With respect to that, more appealing and simple distributed schemes have drawn some attention. Based on Aein’s work, Meyerhoff [81] suggested an iterative procedure for finding the power vector. Moreover, it was shown that equalizing the SIRs is equivalent to maximizing the minimum SIR. In Zander’s companion paper [129], the distributed balancing algorithm (DB), which only requires that local measurements are available, is introduced and shown to converge to the power solution that gives the maximum achievable SIR. Simulations indicated faster convergence by the distributed power control (DPC) algorithm, suggested by Grandhi et al., in [38]. Common for both these algorithms is the exclusion of background noise which makes scaling of the power vector necessary. Lee and Lin [74] proposed an algorithm not using any centralized normalization. The convergence rate of the distributed schemes for noiseless systems depends on the magnitude of the ratio between the second largest and the largest eigenvalue. In [71,72,77], the technique of linear coordinate transformation was used to further improve the rate of convergence of SIR-balancing algorithms for noiseless systems.

The power control problem was extended to include preset SIR-targets and nonzero background noise by Foschini and Miljanic [35]. Their distributed algorithm (FM) was shown to converge to the preset target, conditioned on that it is less than the maximum achievable SIR. Furthermore, their algorithm can include user dependent targets, background noise and proportionality factors. In fact, the DPC algorithm can be considered as a special case of Foschini and Miljanic’s algorithm in the noisy case. Although derived from another context, the algorithm of Foschini and Miljanic fits nicely into the numerical linear algebra area, since it is equivalent to the simultaneous Jacobi over-relaxation method.
(JOR), cf., [125].

An asynchronous version of Foschini and Miljanic’s algorithm, where users update their powers in an uncoordinated fashion with possibly outdated measurements, was proven by Mitra [83] to converge to the fixed point that supports the preset SIR-targets under stationary link gains.

An extension of the DPC algorithm to include background noise and maximum power constraints called the distributed constrained power control (DCPC) algorithm was presented in [39]. It was shown to converge under both synchronous and asynchronous updates. The DCPC can also be interpreted as a constrained variant of the FM algorithm with relaxation parameter unity. The algorithms so far were of first order, i.e., they only used power values from previous iteration. In [55], Jäntti and Kim proposed second-order iterative power control having accelerated asymptotic rate of convergence as compared to the first-order DPC. A second-order algorithm has also been suggested by putting a difference equation obtained from a basic PI-controller on discrete form [112].

By applying iterative methods from numerical linear algebra, a general power control algorithm was suggested by Jäntti and Kim [33]. Their block-distributed algorithm is general in the sense that it allows different degree of distributiveness and availability and reliability of used link gains. The idea is that by including more information about the link gains in the power control, convergence rate is improved. From their framework, several algorithms can be deduced, e.g., the DCPC. Sung and Wong [104] suggested a semi-distributed algorithm for SIR-balancing which incorporated communication between neighboring base stations for providing the current minimum experienced SIR. Numerical results indicate that it can outperform both DB and DPC algorithms.

Kim considered downlink power allocation for CDMA in [62]. The algorithm takes into account a total sum power constraint and allocates the cell powers such that all users in the cell experience the same SIR during convergence to a preset target. However, this may cause a whole cell to have insufficient quality for some instances. The power control problem was written on a form with dimensionality equal to the number of base stations rather than the number of users. This technique was further generalized by Mendo and Hernando [80], which gave a more accurate interference description including orthogonality factors and unequal quality targets. Also Kotakis and Papavassiliou [68], give a system description with the same dimensionality. Their convergence analysis can be considered as a special case of [53].

The above work consider a continuous power domain, which is a relaxation of the quantized levels in a real system. Andersin et al., investigated the DCPC under discrete power levels [8]. It was found that convergence must be stated in a weaker sense, within an envelope, and oscillations may occur. Herdtner and Chong [47] analyze a single bit up/down power control algorithm and characterize its convergence region. They show that the convergence region grows exponentially with the measurement delay. In [103] Sung and Wong suggested, a two bit up/down power control algorithm which also exhibits active link pro-
tection so that supported links remain supported during convergence. Power control with adaptive step sizes has been investigated by Lee and Steele [73] and incorporated with estimation of the tap weights in a RAKE receiver. Song et al., [102] showed that a step size ranging from 0.5-1.3 dB, is sufficient for maintaining acceptably low variance of the power control error. Their analysis encompasses nonlinear feedback effects and channel fading.

Another direction in power control, in particular for CDMA, are schemes that offer constant received power at the base station [116]. For TDMA systems it has been shown though, that these type of schemes do not significantly have impact on co-channel interference [106].

Almgren et al., [4] and Yates et al., [123] propose algorithms that decrease the SIR-targets linearly when the system becomes congested in order to reduce the probability of an infeasible power control problem. The other direction for dealing with infeasible situations is to remove users from the current channel. The work on centralized removal algorithms was initiated by Zander’s work on SIR-balancing [128], where a stepwise removal algorithm (SRA) was proposed. Later, the stepwise maximum interference removal algorithm (SMIRA) was shown to outperform the SRA [72]. Andersin et al., suggested several removal algorithms in [6]. Among the algorithms therein, the gradual removal restricted (GRR) algorithm allows for removing connections during power updates. The combined removal/power control algorithm, GRR-DCPC, in each instance removes a connection with a certain probability, if the demanded power level exceeds the maximum power. Thus removals can be performed in a fully distributed way and it was shown that close to optimal performance can be obtained. The removal problem has also been studied from an optimization point of view by Kim and Zander using the Lagrangian relaxation technique [63]. In the paper they propose algorithms that give the same outage performance but with faster convergence than previous algorithms.

When new users are to access the network, admission control is required for avoiding system congestion. To prevent that already existing connections experience a drop in quality, the power of an admitted user has to be chosen so that the SIR does not fall below its target. Bambos et al., [11] suggest the concept of active link protection (ALP), in which active links are given a SIR protection margin while admitted users are only allowed to power up in certain fixed steps performed by the FM algorithm. The resulting algorithm, called DPC-ALP, guarantees the quality of active calls. The concept was further expanded to include maximum power constraints, which implies that some form of distress signaling has to be executed when users approach the maximum power in order to preserve the active link protection. The DPC algorithm was also used in [119] in a distributed admission scheme. The soft and safe (SAS) admission algorithm by Andersin et al., [7] avoids the loss of capacity due to operating with a fixed protection margin. Instead, they use DCPC where already admitted users can adapt their margin while newly admitted power up gradually.

The minimum power assignment (MPA) problem, which includes base station
selection for finding the lowest possible uplink power vector, has independently been solved by Yates and Huang [121] and Hanly [42]. Huang and Yates later verified a geometric rate of convergence for their algorithm [51]. For the downlink, Rashid-Farrukhi et al., [93] showed that there may not necessarily exist a Pareto optimal solution. However, they show that if the same base station assignment obtained from solving the uplink problem is used in the downlink, the sum of powers is minimized if the background noise level is the same at all receivers. In [122], Yates gives a general framework with sufficient conditions for proving convergence of a big class of power control algorithms. An algorithm exhibiting such conditions is referred to as a standard power control algorithm. Unfortunately, the rate of convergence is not easily determined from the framework.

Most of the above work consider time invariant models, which could be interpreted as the power control actions are performed much faster than the propagation situation changes. This has validated the use of snapshot evaluation having fixed link gains. However, Andersin and Rosberg found that snapshot evaluation under-estimates the outage probability and significant margins added to the SIR-target have to be used [5]. Adaptive SIR margins are investigated by Rosberg, utilizing the duration outage measure to the relate to the SIR-target [96]. A distributed power control algorithm was suggested that used the average SIR-level crossing rate for determining the SIR target. Kandukuri and Boyd [58] considered a Rayleigh fading channel and minimized the outage probability. This problem reduced to nonlinear convex optimization and a heuristic approach reduced to an eigenvalue problem. The work of Mitra and Morrison [84] takes into account the variance and mean of the interference due to randomness in transmissions but can also take into account variations in link gains. A power control algorithm based on direct measurements of the bit error rate has been suggested by Kumar et al., [69]. Perfect estimates of SIR, received power or bit error rates may be difficult to obtain though. Ulukus and Yates consider stochastic measurements in power control and study convergence in stochastic sense by means of mean square error [110].

The impact of time delays was studied by Gunnarsson et al., in [41]. They suggest time delay compensation by adjusting the powers taking into account previous power control commands that have been sent, but not yet been experienced by the receiver.

1.4.2 Variable-Rate Power Control

Maximum achievable channel capacity has traditionally been a well-examined topic in the literature. Various capacity measures are used for nonreal time services. The ergodic capacity [36] describes the maximum achievable rate over all fading states. The ergodic capacity could be less relevant for real time services over slowly fading channels. Delay-limited capacity has been used for such cases. A general information theoretic approach was considered by Hanly and
Tse [43,108]. Therein, they find the capacity regions for the uplink single-cell multiple access fading channel considering both delay tolerant and intolerant cases. Earlier, Knopp and Humblet [67] determined the optimal power control regime from an information theoretic aspect for the single-cell multiple access fading channel. The results show that to achieve capacity, only one user should transmit at any given time over the entire bandwidth. Tse arrived at a similar conclusion for a downlink case in [107]. Common for these approaches is that they consider constraints on average transmit power.

As a means of supporting multiple data rates, adaptive modulation has drawn some interest. In [91], Qiu and Chawla investigate joint optimization of modulation and powers to maximize the log-sum of the users' SIR. The resulted power control algorithms were not fully distributed but a large gain was also shown using adaptive modulation solely. Leung and Wang [76] investigate combined modulation and power control to achieve a specified range of packet error rate for real time applications, including Kalman filtering for accurate interference prediction.

Goldsmith and Varaiya [36] apply water filling in the time domain for power and rate to achieve capacity over a fading channel. The results illustrate the common principal solution characteristic of throughput maximization; to allocate resources to good channels. A similar power control scheme, truncated power control, was considered over fading channels by Kim and Lee [65]. There variable processing gain was used in DS-CDMA to adapt transmission rate with the objective of maximizing average transmission rate. The principle is that to avoid loss of capacity when compensating for deep fades, either transmission rate or power or both can be decreased. A truncated rate adaptation scheme, which suspends transmission when the link gain is below some threshold, was suggested for traffic tolerating longer delays. Analysis of truncated channel inversion power control was given by Ding and Lehner [33], which included queuing effects. It was illustrated that to the price of queuing delay, energy conservation can be achieved. In [66], Kim and Honig determined the processing gain that minimize average delay for a Poisson arrival packet process. It was found that multiple operating points can be found, of which one results in infinite delay.

Early work on rate adaption in DS-CDMA by Yun and Messerschmitt [126] considered minimization of the total downlink transmitted power given constraints on individual user data rates and the resulted problem was identified as a linear programming problem. In a sequel paper [127], they extended the analysis to include statistics of code correlations and packet arrivals. Sampath et al., considered a similar problem in [98] and obtained the instantaneous capacity region for a single cell. Further, they formulated the problem of maximizing the total transmission rate in the system, which reduced to a nonlinear programming problem with linear constraints. Subsequently, in [94], Rezaalifar and Holtzman show the convergence of a power control algorithm applied in a multicellular system updating the powers according to a linear approximation of the problem stated in [98]. The same problem was also treated by Ramakrishna and
Holtzman [92] with two classes of users being tolerant and intolerant to delays respectively. Sung and Wong considered to maximize the sum of the effective transmission rates subject to a constraint on total received power [105]. In [101], Soleimanpour et al., give a more general problem definition of throughput maximization in DS-CDMA, taking into account base station assignment, handover- and call-dropping cost. The problem was shown to belong to the set of NP-complete problems. Oh and Wasserman used a slightly different approach in [87], as compared to quality based power control, in the sense that no target SIRs were set. The proposed algorithm, denoted greedy power control, assigns high data rates starting with mobiles having high link gain. The power control exhibits a similar “bang-bang” type of behavior, where either maximum or zero power is used. A similar problem formulation was also investigated by Ulukus and Greenstein [111]. Kim [61] investigate regulation of transmitter rates jointly with power control for maintaining signal quality over a certain requirement. Most of the above work consider a continuous relation between throughput and transmission rate. However, in real systems the transmission modes, e.g., modulation level, code rate, processing gain, are limited to a small number of discrete values. Assuming a continuous relation greatly simplifies the problem and as a consequence, Kim et al., considered a discrete number of transmission rates available [64]. For the purpose of maximizing throughput, two distributed power/rate control algorithms were suggested, both of them heuristic due to the NP-complete problem structure. The first one was based on the Lagrangian relaxation technique while in the second, selective power control (SPC), every mobile chooses the maximum possible rate that can be achieved at every instant.

Utility based network control has also been used for wireless data. Xiao et al., [120] included the cost of energy in the objective function and relaxed the SIR-target from being a step-function to a sigmoid-function. A general utility function approach was taken by Lee et al., in [75], where also the total power constraint of the downlink was included.

1.4.3 Transmission Scheduling and Multiple Access

A large amount of work has been carried out for scheduling user’s transmission to obtain higher throughput, guarantee maximum delays or minimize the energy consumption. In particular when it comes to using scheduling as part of the interference management, different hybrid access schemes have been suggested. A fundamental issue for these nonreal time services is the multiple access scheme, whether simultaneous transmissions are beneficial or not [87,92,132]. Although different models are considered, the results of those works, to some extent exhibit a time division solution for maximizing the system throughput. In [50], Honig and Madhow discuss the concept of utilizing a hybrid of TDMA and CDMA, as a way of taking advantage of the high intracell capacity of TDMA and the intercell interference suppression ability of CDMA. The notion Scheduled CDMA (SCDMA) was used by Arad and Leon-Garcia for supporting QoS in a wireless
ATM network [10], however their focus was more on admission criteria rather than time scheduling.

In [13], Bedekar et al., found that throughput in the DS-CDMA downlink of a single cell with linear rate, is maximized when each base station transmits to at most one user at a time and uses maximum power. Schemes which coordinate transmissions between cells were also suggested for a cellular highway system. Inter-cell coordination for hexagonal systems was further considered by Maileander et al., in [79]. Recently, a similar concept to [13] based on current CDMA physical layer structure [14], where simultaneous transmissions within the cell are avoided, has been proposed for supporting high data rates (HDRs) in the downlink. In the CDMA/HDR system, due to the vastly different requirements, high data rate users are scheduled over time slots, where the slot lengths depend on channel conditions and transmission is executed with a constant maximum power. Correspondingly, the results in [67,87,92] all exhibit hybrid TDMA/CDMA behavior. UMTS has its counterpart in the high speed downlink packet access (HSDPA) mode [88]. Several schedulers have been suggested for these systems [57]. The proportional fair scheduler is one well known direction [48,49,114]. This scheduler allocate channel access for a user when its channel quality is good relative to its average quality. This creates multiuser diversity effects and channel variations are desirable in that respect.

The issue of when to induce time division on a multiple access channel was investigated by Rulnick and Bambos [97]. Their underlying idea is to induce TDMA without central control or synchronization when the channel conditions are bad. The result is that the total energy consumed by the system can be reduced. The incentive for distributed time division was interference and backlog, i.e., the number of packets in the queue waiting for transmission. Later, Bambos [12] coined the notion of Power Controlled Multiple Access (PCMA) for this type of autonomous power management/access control.

Aspects of power control and scheduling were investigated in [54] by Jäntti and Kim, where the problem of minimizing the time-span for emptying all users’ data buffers was addressed. This may be viewed upon as minimizing the maximum delay for the buffered data. A discrete time model was assumed and the results showed that the optimal solution may require that time division must be induced. However, their work did not directly suggest any such practical algorithm. In [92], it has also been found that higher throughput in the DS-CDMA uplink can be obtained by scheduling delay tolerant users and this gain does not necessarily require more average transmission power.
1.5 Scope of the Thesis

1.5.1 Problem Background

To provide heterogeneous services with high spectral efficiency, cellular systems will benefit from adaptive and integrated network control. Practical resource allocation for homogeneous services is fairly well understood in terms of, e.g., power control and a vast number of solution methods exist. When it comes to systems offering more functionality than voice, the actual meaning of the RRM mechanisms is not as clear and needs to be revisited. The scenario of wireless data described previously, indicates that new transmission schemes, including power control, offering high spectral efficiency should be considered. With this thesis we shall draw some insight on how to design such schemes. For this, we will mainly devote the study to DS-CDMA systems, which have multirate functionality and will be deployed in a near future. Of particular interest are distributed control architectures. The purpose of the work is to characterize the principal solution behavior and help understanding suitable algorithm design. In particular we shall limit the study to the following main issues.

Global Stability

Users will request certain amount of QoS from the network and it is incumbent on the network control mechanisms to grant the requests. It is of utmost importance to admit services, not making the system load reach beyond feasible system capacity. Since the resource allocations are interconnected through the interference on the channel, a failure in this respect, can make the channel quality oscillate and ultimately cause a system breakdown. Global stability will be a necessary requirement in the resource allocation.

Distributed Control

To be practical, a resource allocation scheme should be distributed and agile for fast tracking of channel variations. For wireless data, the more bursty interference will necessitate such a property even more than for speech services. The bursty interference, in that sense, introduces more dynamics into the system. If the schemes react slowly to changes in the interference, stability and performance will be affected and the system will stall and in the worst case, collapse. Therefore, in the analysis when we are able to determine existence of an optimal solution, for example in a power control problem, schemes should be designed that find and converge to it. Convergence and global stability are tightly connected and rate of convergence will be a measure used in the algorithm design and for benchmarking.
Energy Efficiency

By QoS provisioning we typically mean that information should be delivered at a certain rate, within a certain period of time and/or with a maximum error probability. As a complement to this, it is beneficial if the transmit power were low, limiting unnecessary interference. The benefit of low transmitter power is obvious for portable devices in terms of lower energy consumption. In addition, reducing the powers decreases the biological effect of the electromagnetic waves, which should not be ignored. Bringing the powers down in the downlink case for wireless data delivery leaves more of the total power budget for other services, e.g., voice and reduces the cost of electrical power. Hence, the transmitter power can be regarded as a component in the QoS management which preferably should be kept low. In this thesis, we will focus on proper assignment of the system resources in order to provide high energy efficiency; that is a high amount of data should be delivered by a certain energy input to the system.

1.5.2 Scope

A common problem of network control is some form of throughput maximization. This problem is considered also in this thesis, subject to varying fairness constraints and system models. First we wish to address the problem with discrete rates, a highly practical scenario. That is, a user can only transmit with a finite set of rates. Unfortunately this problem is NP-complete which leads us to design heuristic schemes. They should possess properties of achieving high throughput, low energy consumption and being distributed. Since this is a rather difficult problem, it is reasonable to attack the problem by considering first only a single-rate system and then generalize. For the throughput maximization, two constraints will be considered: with and without a minimum transmission rate requirement, respectively. The latter one covers a best effort type of service whereas the former addresses delay sensitivity. Throughout the thesis, there is also always a constraint of the received bit-energy-to-interference-spectral-density ratio, $E_b/\sigma^2$, being equal to some target. That is, data is guaranteed a maximum error rate. For these different cases, we seek to highlight tradeoffs between throughput and energy consumption. A common approach of power control research, partly also utilized in this thesis, is to analyze the system assuming that the environment changes slower than the updates of the resource algorithms. The benefit from this quasi-static model is that it is analytically tractable which renders in that, e.g., convergence analysis can be performed and feasibility conditions determined.

Next, we wish to approach the problem where users have a minimum average data rate requirement. That is a constraint somewhere in between the two former ones. Here, we focus exclusively on the DS-CDMA downlink and for analytical tractability, allow the rates to be continuous. The immediate application can be found in HDR [14] and HSDPA [88]. In contrast to the previous problem, an average rate constraint makes the actual power control problem different from
before. Therefore, the problem that needs to be solved should first be identified. The issues associated with this problem also include whether time division is favorable on the channel or if the usual pure CDMA offers better capacity.

Given that time division is used on the DS-CDMA downlink, transmission scheduling becomes an inherited problem. That is, not only does the data rate need to be determined but also should a proper user be granted to transmit. All the problems above considered static channels, i.e., a snapshot analysis. However, in this case it is very important to include fading into the analysis. This is due to the fact that if users have asynchronously varying channels, we may try to allocate transmissions when channels are near their peak quality. This option is similar to that of selection diversity and is in this context often referred to as multiuser diversity. Our main idea is to set a fairness constraint which will asymptotically provide all users with the same channel access time, i.e., a resource fair scheduler. The natural comparison can then be made with the common channel independent round robin scheduler.

In the last problem, we leave the rate allocation procedures and solely consider the power control of successive interference cancellation (SIC). The soft decision SIC receiver is adopted, which is rather simple and a possible candidate for real implementation. The basic problem is here how to control the powers for achieving the requirement on $E_b/I_0$ after cancellation. It is likely that if this can be done optimally, it will translate into significant system capacity gains. An issue here is to determine the decoding order of the users. Just ordering by the strength of the received powers may not make sense, if we at the same time use some form of power control.

The general methodology pursued, is to formulate resource allocation problems for which we suggest, analyze and evaluate algorithms. This thesis will provide a sound theoretical approach to these problems and thereto should contribute with schemes of reasonable complexity and distributiveness. The practical implication of the thesis is to address principles of resource allocation applicable to a cellular system rather than detailing implementation aspects.

1.5.3 Contributions

Chapters 4-8 in this thesis are based on submitted and published papers [16–18, 20–25]. In addition to this, the author has published a journal paper on power control for ALOHA [15], which will not be included in the thesis. Example 3.2.4 is based on contributions of [19]. In what follows, a brief overview of the contributions in the thesis is given.

Chapter 4

For a fixed rate system with low traffic load, there exists a unique power solution that supports all users with the minimum individual power as well as sum power. However, while converging to the fixed point, it may occur that more than the maximum feasible power is requested. Using the maximum power may not lead to sufficient improvement on signal quality. Moreover, severe interference will hit other users. A new class of algorithms that under such conditions rather decrease the powers is proposed. Properties of convergence are provided. For the infeasible case, where not all connected mobiles can be supported, temporary removals by means of shutting the power off until channel conditions improve, are suggested and shown to decrease the probability of having a nonsupported connection.

Chapter 5


The above power control concept is extended to a multirate system with a set of discrete rates. By using a semi-distributed scheme, sharing information within the cell but fully distributed between cells, we suggest an admission criteria called greedy rate packing, which can be used for the transmission rate allocation. We show that high rates should be assigned to mobiles with high link gains for maximizing capacity. The rate allocation is joined with a suggested power control algorithm, and convergence properties are provided. To ameliorate convergence and improve energy efficiency, the gradual removal is extended to multiple rates and is shown to bring the powers further down.

Chapter 6


To remedy intracell interference for nonreal time services, we suggest transmission in a one-by-one fashion in each cell. That is TDMA within the cells but CDMA between cells. It is proved for the downlink that if a data rate can be achieved instantaneously, it can also be achieved in average by intracell scheduling. An important issue is the interpretation and meaning of a minimum power solution; what is the minimum power solution for this form of discontinuous transmission? We address this issue and show that it can be written in
matrix form similar to the standard power control problem. An iterative convergent algorithm is suggested for finding the solution. By comparing certain matrix properties we find a gain in the capacity and rate of convergence of the T/CDMA scheme as compared to pure CDMA.

Chapter 7


Taking channel fading into account, multiuser diversity provides additional gains from one-by-one scheduling. We extend the previous properties on time division to the case of fading. A fast channel adaptive scheduler is suggested and its asymptotic performance evaluated. The scheduler is asymptotically resource fair and allocates equal channel access times to the users while using the channel on good instants. The scheduler is generalized to allocate several simultaneous users, which is shown to have merits when the average channel quality is good.

Chapter 8


With a slightly more complex receiver, capacity gains can be achieved by canceling detected signals. We analyze a class of successive interference cancellation receivers which uses soft feedback. The results are general and include effects of imperfect-, limited- and partial cancellation. The resulting user capacity region is determined on closed-form and power control algorithms suggested.

In the above papers, the first author has been the major contributor of the problem definitions, main concepts, analysis, numerical results and writing of the final paper. In Chapter 4, the convergence analysis was shared with R. Jäntti. In Chapter 5, S.-L. Kim contributed to the optimality issues. In Chapter 6, R. Jäntti contributed to the eigenvalue analysis. In Chapter 7, R. Jäntti
contributed to the proof of the time division property. In Chapter 8, S. B. Slimane contributed to the receiver analysis. Thereto, the clarity and accuracy of the chapters and manuscripts have been improved by feedback and contributions from the coauthors.

1.6 Thesis Outline

This thesis is written as a monograph. Chapter 2 describes and reviews the system and simulation models used. In Chapter 3, some tools from numerical linear algebra are applied to the model and introduced for subsequent use. Several examples illustrate their applicability to the power control problem. Chapters 4-8 are based on the material in [16–25]. The thesis is concluded and further work is outlined in Chapter 9.
Chapter 2
Models and Performance Evaluation

This chapter describes the assumptions and models used in the subsequent analysis and numerical evaluation of a wireless communication system. The performance evaluation methodology and corresponding measures are described and reviewed.

2.1 Radio Wave Propagation

In a mobile radio environment, radio wave propagation cannot be assumed to be characterized as free-space and line-of-sight. As the radio channel is generally regarded as being a hostile medium in comparison to a wireline channel, it is often also difficult to predict. The radio channel causes the received signal to suffer a path loss. Generally, the path loss is affected by antenna heights, carrier frequency, local reflectors, absorbers and obstacles. In principle Maxwell’s equations suffice to describe electromagnetic propagation but in practice, other more tractable methods, based on a statistical representation, are used. These models generally rely upon wave phenomena such as reflection, diffraction and scattering. Reflection from an object typically occurs when the wavelength of an impinging wave is much smaller than the object itself, creating multipath components. These components can have different phase, polarization and received power. Therefore, they may add either constructively or destructively, resulting in a signal fade. Diffraction causes the wave to bend around obstacles and can be explained by Huygen’s principle of seeing the wave front consisting of point sources. When a wave travels in a medium with a large number of elements having small dimensions compared to the wavelength, the energy is scattered. The signal fading in a wireless environment is often considered to be decomposed into three components with different time scales of variation: large-scale path
loss, medium-scale slow fading and small-scale fast fading. Decreased received power with distance, reflection and diffraction constitute the path loss. These are denoted large-scale since changes appear when moving over hundreds of meters. As the receiver can be shadowed by, e.g., buildings, trees and foliage, the local mean received power changes when moving just a few tens of meters, i.e., on a medium-scale. Small-scale fast fading, or multipath fading, captures the effect of multipath reflections by local scatterers and changes by the order of wavelengths.

Empirical studies by, e.g., Hata [43] have shown that the large-scale path loss in a cellular system can be modeled as,

\[ L_{ij} = K_{0} r_{ij}^{\alpha} \]  \hspace{1cm} (2.1)

where \( r_{ij} \) is the distance between transmitter \( j \) and receiver \( i \), \( K_{0} \) is a constant depending on antennas and wavelength and \( \alpha > 0 \) the attenuation exponent.

Motivated by measurements, a common method to take the shadow fading into account, is modeling with a lognormal distribution

\[ f_{s}(x) = \frac{10}{\sqrt{2\pi} \sigma_{s} x \log_{10} 10} e^{-\frac{10 \log_{10} x - \mu_{s}}{2 \sigma_{s}^{2}}}, \quad x \geq 0 \] \hspace{1cm} (2.2)

where \( \mu_{s} \) (dB) and \( \sigma_{s} \) (dB) are the mean and standard deviation of \( 10 \log_{10} S \) respectively. The moments of (2.2) are given by, (see, e.g., [100])

\[ \mathbb{E}[S^{k}] = e^{\frac{b_{s} k}{\sigma_{s}^{2}}} \mu_{s} + \frac{b_{s} k}{\sigma_{s}^{2}} \left( \frac{b_{s} k}{\sigma_{s}^{2}} \right)^{2} \sigma_{s}^{2} \] \hspace{1cm} (2.3)

where \( \mathbb{E}[X] \) denotes the expectation operator of the random variable \( X \). For generating the shadowing effects between transmitter \( j \) and receiver \( i \), we will use

\[ S_{ij} = 10^{\mathbb{E}[S_{ij}]} \] \hspace{1cm} (2.4)

in which \( X_{ij} \) is Gaussian with zero mean and unit variance.

Delayed replicas of the signal can cause intersymbol interference (ISI) and frequency selective fading. A time varying frequency selective channel can be modeled as a finite tapped-delay line. In general, the time varying lowpass equivalent channel impulse response can be defined as \( h(t; L) = \sum_{l=1}^{L} h_{l}(t) \), where \( L \) is the maximum number of resolvable multipath components, \( h_{l}(t) \) is a complex random process and \( \delta(\cdot) \) is the Dirac delta function. For each resolvable path \( l \), \( h_{l}(t) \) is comprised by a summation of many physical paths, making a Gaussian process assumption plausible. Under a slow fading assumption, \( L \) remains fixed over a longer period of time and \( \{ h_{l}(t) \}_{l=1}^{L} \) and \( \{ \tau(t) \}_{l=1}^{L} \) are constant over a symbol duration. If the multipath components are caused by different scatterers, the resolvable paths can be assumed mutually independent. If the small-scale fading is frequency nonselective, no ISI occurs and the channel is considered to be of flat fading type. Consider the variations of the signal power, or signal-to-noise ratio (SNR), \( \xi_{l} \propto |h_{l}|^{2} \) to be from small-scale
fading. The mean $\hat{\xi}_i$ can be regarded as being comprised by attenuation and shadowing effects, occurring on a much longer time scale. The channel fading can be assumed to follow the noncentral $\chi^2$ distribution with two degrees of freedom, defined by its density function

$$f_{\xi_i}(x) = \frac{\kappa + 1}{\hat{\xi}_i} e^{-\kappa - \frac{(\kappa + 1)x}{\hat{\xi}_i}} I_0 \left( \sqrt{\frac{4\kappa(\kappa + 1)x}{\hat{\xi}_i}} \right), \quad x \geq 0 \quad (2.5)$$

where $\kappa$ is the Rice factor\footnote{$\kappa$ is here given in linear units.} and $I_0$ the zeroth-order modified Bessel function. The variance of $\xi_i$ can be shown to be $\sigma_i^2 = (2\kappa + 1)\hat{\xi}_i^2 / (\kappa + 1)^2$ [100]. The Rice factor is the ratio of the power in the specular and scattered components. When $\kappa = 0$, the channel exhibits Rayleigh fading and when $\kappa \to \infty$ the channel is not fading at all. Moreover, it has been observed that the amplitude distribution of a despread DS-CDMA signal, which is exposed to multipath propagation, becomes more specular as the spreading bandwidth increases [82]. This is irrespective of the actual fading statistics of the received wave form. Thus its power statistics have been found to be closely described by a Ricean fading distribution, viz. (2.5).

Channel fading implies that diversity methods can be applied for taking advantage of instants with good quality. Combining signals from different such branches may therefore provide a better average quality had the channel been time invariant. Traditionally, this is achieved over space (multiple transmit/receive antennas), over time (channel interleaving) or over frequency (RAKE receivers, frequency hopping). Combinations of these techniques are also possible.

2.2 System Architecture

2.2.1 Receiver Model and Communication Quality

One of the most accepted and commonly used signal quality measures in the literature is the SIR, which will also be used in this thesis. The SIR is strongly dependent on the received power and thus related to power control. Therefore, we adopt it due to its analytical suitability for global analysis and to keep our work comparable with a major part of previous research on power control and resource allocation. Although some of the concepts presented in this thesis apply to other multiple access methods, the schemes are evaluated for a DS-CDMA system. In this section we give the background to the chosen quality measure.

Consider a complex equivalent baseband model of a DS-CDMA system where each data symbol is assumed to have unity amplitude, i.e., $|b_i(t)|^2 = 1$ and the spreading waveform is denoted by $a_i(t)$. The spreading bandwidth is $W \approx 1/T_c$ where $T_c$ is the chip duration. In the absence of multipath propagation, the
received lowpass equivalent signal can be expressed as

$$r(t) = \sum_{i=1}^{N} \sqrt{g_i p_i} a_i(t - \tau_i)b_i(t - \tau_i) + n(t) \quad (2.6)$$

where $n(t)$ is a complex Gaussian process describing noise and other interference sources, $p_i$ is the transmit power, $g_i$ is a constant path gain and $N$ is the number of users transmitting. We assume that coherent detection can be deployed. At the receiver, every user despreads the signal. Assuming that the delays $\{\tau_i\}_{i=1}^{N}$ are known, the output sample of user 1 becomes,

$$y_1 = \frac{1}{\sqrt{T}} \int_{\tau_1}^{T+\tau_1} r(t)a_1(t - \tau_1) dt$$

$$= \frac{1}{\sqrt{T}} \int_{0}^{T} \sqrt{g_i p_i} a_1(t)^2 b_1(t) dt + \frac{1}{\sqrt{T}} \int_{0}^{T} n(t + \tau_1) a_1(t) dt +$$

$$\sum_{i=2}^{N} \frac{1}{\sqrt{T}} \int_{0}^{T} \sqrt{g_i p_i} a_i(t - \tau_i + \tau_1) a_1(t) b_i(t - \tau_i + \tau_1) dt$$

$$\overset{\text{def}}{=} \sum_{i=1}^{N} \sqrt{g_i p_i} \hat{\rho}_{i1} + n_1$$

where $T = 1/R$ is the data symbol time and

$$\hat{\rho}_{ik} = \frac{1}{\sqrt{T}} \int_{0}^{T} a_i(t - \tau_i + \tau_k) a_k(t) b_k(t - \tau_i + \tau_k) dt \quad (2.7)$$

If $N$ is large, a Gaussian assumption\(^2\) is commonly made on the decision statistic as well as defining the individual interference terms being mutually independent. Then a communication quality measure that can be related to the BER is given by its second order statistics,\(^3\)

$$\left( \frac{E_b}{I_0} \right)_1 = \frac{\mathbb{E}[y_1 | b_1(t)] \mathbb{E}^*[y_1 | b_1(t)]}{\text{var}[y_1 | b_1(t)]} = \frac{g_1 p_1 / R}{\sum_{j=2}^{N} \hat{\rho}_{j1} g_j p_j + N_0}$$

with $\hat{\rho}_{ik} = \text{var}[\hat{\rho}_{ik}]$ and

$$\text{var}[n_i] = \frac{1}{T} \int_{0}^{T} \int_{0}^{T} \mathbb{E}[n(t + \tau_i)n^*(u + \tau_i)] dtdu$$

$$= \frac{1}{T} \int_{0}^{T} \int_{0}^{T} N_0 \delta(t - u) dtdu = N_0,$$

\(^2\)This can be shown in the limiting case using the central limit theorem. For finite $N$, a “fuzzy” central limit theorem says that data which are influenced by many small and unrelated random effects are approximately normally distributed.

\(^3\)This can be shown in the limiting case using the central limit theorem. For finite $N$, a “fuzzy” central limit theorem says that data which are influenced by many small and unrelated random effects are approximately normally distributed.
In the above we assumed that \( \int_0^T a_i(t)^2 \, dt = T \) for all \( i \). If the spreading waveforms are long, over several symbols, the correlation properties of (2.7) are the same as for

\[
\rho_{ik} = \frac{1}{\sqrt{T}} \int_0^T a_i(t - \tau_i + \tau_k) a_k(t) \, dt. \tag{2.8}
\]

For a spreading waveform defined as

\[
a_i(t) = \sum_{n=-\infty}^{+\infty} a_{i,n} \Gamma_{T_e}(t - nT_e) \tag{2.9}
\]

where \( a_{i,n} = \pm 1 \) with equal probability and \( \Gamma_{T_e}(t) \) is a rectangular pulse of length \( T \) and unity amplitude, the cross-correlations (2.8) can be determined using the results in [90]. Using our notation, it can be shown that when \( i \neq k \), \( \mathbb{E}[\rho_{ik}^2] = 1/W \) for synchronous pulses and \( \mathbb{E}[\rho_{ik}^2] = 1/3W \) for asynchronous pulses. In WCDMA, the spreading is carried out by first modulating the signal with orthogonal channelization codes and then cell-specific scrambling codes [32]. It is common to refer to the bandwidth expansion as the spreading factor, whereas the processing gain also includes the effect of FEC encoding.

Transmitted over a wideband radio channel, the signals will undergo time dispersion. If the spreading bandwidth is larger than the coherence bandwidth, the channel is considered frequency selective. A way of achieving diversity improvement on a frequency selective channel is by the use of a RAKE receiver. The RAKE receiver tries to draw benefits from the diversity of multipath components while suppressing the extra interference they may give rise to. Such a receiver structure consists of a bank of correlators, each correlating to a respective delay \( \tau_i(t) \) of the received signal. These so called finger outputs are then weighted and coherently combined to form a decision statistic. In this process, both \( h_i(t) \) and \( \tau_i(t) \) are needed and they are in practice typically estimated using pilot symbols or a pilot channel. The greater the signal bandwidth, the higher the order of diversity that could be obtainable by the RAKE. However, there is a tradeoff in the receiver design, since the energy-per-received resolvable path decreases as the spreading bandwidth increases [82]. This makes the channel estimation more difficult and suboptimal performance may occur. Recently, a generalized RAKE receiver was proposed that could employ more fingers than signal paths in order to further suppress the interference [28].

When dealing with stability analysis of the resource allocation schemes, we will disregard fast channel variations. This could translate into an assumption of using a RAKE receiver, where the estimated tap weights are perfect so that the intersymbol interference becomes negligible. This means that the signal bandwidth provides sufficient frequency diversity such that fast fading can be effectively averaged out by combining different independent multipath components. Furthermore, if we assume interleaving and channel coding so that possible fast fading can be considered to be averaged over the codewords, the local-mean
received power at receiver $i$ from transmitter $j$ using power $p_j$ for a given instant follows

$$p_{rx,i} = \frac{S_{ij}}{L_{ij}} p_j.$$  \hfill (2.10)

Denote by $g_{i,j} = \frac{S_{ij}}{L_{ij}}$ the power link gain at a certain instant from transmitter $j$ to receiver $i$. Assume without loss of generality that transmitter $i$ communicates with receiver $i$. We assume that each transmitter is connected to only one receiver at a time. With this notation, receivers $i$ and $j$ may denote the same physical one which may occur, e.g., in a CDMA system where several mobiles connect to the same base station. The signal-to-interference ratio at the receiver for transmitter $i$ can generally be defined as,

$$SIR_i = \frac{g_{ii}p_i}{\sum_{j \neq i}^N \theta_{ij} g_{ij} p_j + \nu_i}.$$  \hfill (2.11)

where $\nu_i > 0$ is uncontrollable interference in terms of background noise. For a DS-CDMA system where multirate functionality is supported by different data symbol duration $T_i = 1/R_i$, we shall assume that the quality measure of interest is given by

$$\left( \frac{E_b}{I_0} \right)_i = \frac{W}{R_i} \cdot SIR_i.$$  \hfill (2.12)

In the above definition, $0 \leq \theta_{ij} \leq 1$ denotes the normalized cross-correlation between signal $i$ and $j$ at receiver $i$; that is the effective fraction of the received signal power from transmitter $j$ that contributes to the interference experienced by user $i$. For orthogonal systems such as F/TDMA, $\theta_{ij}$ is often assumed to be unity. In DS-CDMA systems, the quantity is dependent on the spreading code characteristics. In the uplink, signals are generally asynchronous and do not exhibit orthogonality, unless the mobiles can be accurately synchronized and operated in a channel with no multipath components present. In the downlink, interference originates from the same point and the codes can be synchronized. Ideally the codes will be orthogonal but due to multipath components, interference among same-cell users is introduced at the receiver, therefore some positive correlation may occur. For the DS-CDMA model, unless not specified, we assume that orthogonality is lost in the uplink and $\theta_{ij} = 1$, whereas in the downlink, it can take values $0 \leq \theta_{ij} \leq 1$ if $i$ and $j$ are connected to the same base station, otherwise $\theta_{ij} = 1$. When the simplification $\theta_{ij} = \theta$ is assumed, we refer to $\theta$ as the orthogonality factor of the system.

### 2.2.2 System Layout and Access Scheme

The algorithms suggested in this thesis are evaluated in a macro-cell system, commonly deployed in rural areas. It should be pointed out though, that the applicability of the suggested transmission schemes is not strictly limited to this particular system setup, only the numerical evaluation. For the numerical
evaluation in Chapters 4-6 and 8, the system architecture shown in Fig. 1.2 is used: a hexagonal plan of two tiers where each cell has a radius of 1 km. In addition to the nineteen cells, the cell plan is wrapped around itself to avoid edge effects during simulations. The centrally located base stations are assumed to use omnidirectional antennas and each user is connected to the base station that offers the lowest signal attenuation. The underlying access method assumed is DS-CDMA, where variable processing gain is used for adapting the transmission rates.

### 2.2.3 Control Signaling

A cellular PCS can be divided into access and core network parts. In the access network, all air-interface related functions are maintained, whereas switching and call control is part of the core network. To distribute data and signaling information for setting up a connection, the base stations are connected via a wired network. Base stations interconnected by the wired network are assumed to be able to exchange parameters and performed measurements. In this thesis, the algorithms work under different restrictions of available measurements. In the fully distributed schemes, each mobile acts independently and its measurements are used solely for its adaption. In the semi-distributed schemes, it is assumed that each base station has knowledge about parameters, such as data rates used within the cell and measurements of interference and channel gains but no knowledge about those of the neighboring ones. In the downlink, pilot and measurement channel can cause extra intercell interference. For simplicity such effects are ignored in this work and could be part of the additive background noise level \( \nu \).

### 2.2.4 Power Control Functionality

In real cellular PCSs, power control often rely on three mechanisms.

**Open Loop**

The open loop power control adjusts the initial access power of a mobile station and compensates for abrupt changes in propagation conditions. The power updates are based on estimates of the propagation loss, which are obtained by measuring the received signal strength at the receiver.

**Closed Loop**

Since the fast fading is not fully correlated on up- and downlink, closed loop power control is used. In the uplink power control, the base station measures the quality and sends power control commands to the mobile. The closed loop power control is deteriorated by feedback delays, imperfect power estimates and errors in the feedback channel.
Figure 2.1: The figure shows a closed loop example where the TPC commands are fed via a multiplexer (MUX) and extracted at the mobile after the AGC.

**Outer Loop**

The closed loop compares an estimated SIR with a target SIR. This target is different for different propagation conditions, e.g., delay profile, number of resolvable paths, Doppler frequency etc. The outer loop sets the target in order to obtain a certain bit error rate, or a related quantity such as frame error rate (FER). In Fig. 2.1, the closed loop power control of a CDMA system is schematically illustrated. The automatic gain circuitry (AGC), controls the amplifier gain such that the input to the receiver analog/digital converters is kept constant. From the outer loop, the binary transmitter power control (TPC) command is generated and added to the transmitted data which is then extracted at the mobile and fed to the power amplifier (PA). This system can be viewed upon as a local loop interacting, through the interference on the channel, with the other loops.

In this work, we will limit our study and not fully model all parts of this system. We will exclude tracking-, stability- and performance issues of the local loops and rather focus on the global aspects, where the interference management is crucial. Therefore, any time delays will be neglected, continuous values for the TPC commands will be assumed and the preferred SIR-targets will be given.

**2.3 Performance Evaluation**

To evaluate the proposed schemes we are in certain situations obliged to system simulations. The simulation models in this thesis should closely correspond to the system models utilized for analysis. In Chapters 4-6 and 8, we use a time independent analysis, often referred to as snapshot analysis. In the snapshot
analysis, the link gains are frozen in time and neither measurement nor sampling errors occur. This could be interpreted as an observation of the system at a random instant of time and allowing the radio resource management functions perform infinitely fast updates. Clearly, the results of this kind of analysis provide upper-bounds on real performance in a time varying environment. However, according to Rosberg and Zander [95], the snapshot analysis could still be relevant when some favorable conditions occur. For instance, fast fading is averaged out, mobiles move slowly and power updates are relatively fast compared to the changes in link gains. Under such conditions, the controlled powers are drifted toward values which are in the vicinity of a temporary fixed point. Thus quasi-convergence and quasi-stability do occur. To assess the stability of a scheme, this model allows us to study convergence. In that sense, our algorithms will be deterministic, as we get the exact knowledge or perfect estimates of deterministic measures, e.g., SIR. An attempt of analyzing nondeterministic algorithms was carried out in [110]. If a scheme performs poorly or converges slowly even under stable conditions, the performance on more severe channels is likely to be bad. It should be noted though, that snapshot simulations cannot fully replace full-scale system simulations since time correlation properties of the traffic, such as call arrivals and call duration, and performance of handover schemes are not considered.

A way of taking the dynamical aspects, like fading, into account is to consider SIR-margins [5, 96]. If a proper SIR margin is utilized, then at least convergence in terms of low probability of having a nonsupported connection does occur. This technique is utilized in practical systems to adapt the deterministic power control algorithms to channel statistics. For instance, the outer loop power control function suggested for WCDMA systems [59] increases or decreases the SIR-target according to channel statistics derived from the decoder. Keeping in mind that the results of snapshot analysis can be related to the real performance of a system, we adopt it when analyzing SIR-based algorithms, because of its simplicity and in order to keep our results comparable with a major part of previous work on power control.

In Chapter 7 where convergence is not studied, we relax the the stationarity assumption to capture small-scale fading effects. Therein, we will consider an ergodic channel such that the time average can be evaluated by the ensemble average.

2.3.1 Performance Measures

Clearly there are many aspects of measuring performance of a cellular system, especially with diverse QoS objectives. Moreover, network operators and subscribers may have different perspectives on which is the primary performance measure. Users would typically be interested of the availability of the service along with its quality (data rates, delays, BERs etc.). Network operators seek to maximize revenue, consider network costs and performance of the system as
a whole (fraction of satisfied users etc.). In this thesis we shall not consider
economical aspects but focus on end-user performance measures.

If for some SIR-target $\gamma^d > 0$, the obtained $\text{SIR}_i \geq \gamma^d$, connection $i$ is
said to be supported, otherwise non-supported. A user having a non-supported
connection is referred to as being in outage. The most commonly used measures
in this thesis are:

- **Average transmitter power per user**
  Estimated as the sum of the total power used divided by the total number
  of users in the system.

- **Rate of convergence**
  Estimated as the normalized Euclidean distance to the fixed point.

- **Outage probability**
  Estimated as the number of users having a non-supported connection di-
  vided by the total number of users in the system.

- **Average throughput per user**
  Estimated as the total system throughput divided by the total number of
  users in the system.

The average power connects to the energy efficiency as it is proportional to the
energy consumption a user would experience in average. Rate of convergence
is measured as a geometric distance and is further motivated in Chapter 3.
The outage probability captures to what extent the users get served by the
system at a given instant. For fixed-rate systems, the main RRM objective is
to minimize this quantity. When there are multirate functionality, this may be
interpreted as a service outage, capturing the event where the allocated rate is
not delivered. For variable-rate systems, a measure of the system capacity is the
average throughput that a user could expect. Additional measures are defined
in the respective chapters.
Chapter 3

Preliminaries

In this chapter we exploit the model from Chapter 2. Using the assumption of stationary link gains, we state the power control problem and recapture some useful results to be used in subsequent global convergence analysis of iterative power control. Several examples show applications of the classical numerical linear algebra techniques. A note on the notation: throughout this thesis vectors and matrices are typeset in bold font. Moreover we shall between vectors and matrices use the following order relations. Let \( A = [a_{ij}] \) and \( B = [b_{ij}] \) be matrices. If \( a_{ij} \leq b_{ij} \) for all \( i, j \) then \( A \leq B \) while if \( a_{ij} < b_{ij} \) for all \( i, j \) then \( A < B \). A matrix \( A \) is referred to as nonnegative (positive) if \( A \geq 0 \ (A > 0) \). The same notation applies to a vector \( a = (a_i) \).

3.1 Noisy Power Control Problem

Let us consider the power control problem for \( N \) users when the transmission rates are assigned. Denote by \( H = [h_{ij}] \) an \( N \times N \) matrix such that

\[
h_{ij} = \begin{cases} 
\gamma_i^t \theta_{ij} \frac{\eta_i}{g_{ij}}, & i \neq j \\
0, & i = j.
\end{cases}
\]

Here, \( \gamma_i^t \) is a predetermined SIR-target of user \( i \) reflecting a required signal quality at the receiver. Furthermore, define the \( N \times 1 \) vector \( \eta = (\eta_i) \) where \( \eta_i = \gamma_i^t \frac{\eta_i}{g_{ii}} \). The power control problem is then to find a power vector such that \( SIR_i \geq \gamma_i^t \) for all \( i \), which by use of (2.11) can be equivalently expressed in matrix form as

\[
Ap \geq \eta \quad (3.1)
\]

37
for \( A = I - H \) where \( I \) denotes the identity matrix and \( p = (p_i) \) the power vector. Let us define the general problem of minimizing the sum of powers

\[
(P_{\geq}) \quad \min \limits_{\mathcal{P}} 1^T p
\]

\[
Ap \geq \eta \\
p \in \mathcal{P}
\]

where \( 1^T = (1, 1, \ldots, 1) \) and \( \mathcal{P} \) denotes the set of feasible power vectors.

**Definition 3.1** If there exists a vector \( p \in \mathcal{P} \) that solves \((P_{\geq})\), we say that the system is feasible at the instant. Otherwise, the system is infeasible.

If \( \mathcal{P} = \{p : p \geq 0\} \) we shall refer to \((P_{\geq})\) as an unconstrained power control problem. We make use of the following property of \( H \).

**Lemma 3.1** \( H \) is an irreducible nonnegative matrix.

**Proof:** We base the proof on the connectivity of graphs. By Theorem 1.6 in [113] we have to show that the corresponding directed graph of \( H \) is strongly connected. The graph is generated by marking a connection between node \( i \) (row \( i \)) and node \( j \) (column \( j \)) if \( h_{ij} > 0 \). We only need to consider the worst case in terms of connectivity, that is the downlink when \( \theta_{ij} = 0 \) if \( i \) and \( j \) are connected to the same base station. Let user \( i \) be connected to base station \( b_i \). Let \( N = \{i : 1 \leq i \leq N\} \), then for any node \( i \) there exists a connection to all nodes in the set \( C_i = \{j : b_j \neq b_i\} \). For any \( j' \in C_i \) we must have \( i \in C_{j'} \), therefore \( (N \setminus C_i) \subset C_{j'} \) and there is a connection \( i \to j' \to i' \) where \( i, i' \in (N \setminus C_i) \) and the graph is strongly connected. Since \( H \) is nonnegative, the lemma follows.

Alternative proof for the case when \( \theta_{ij} > 0 \) was given by Lemma 1 in [37]. Irreducibility is related to the solvability of linear equation systems and provides certain properties of the eigenvalues of \( H \), needful later on. Denote by the spectrum of the matrix \( H \), the set \( \mathcal{S}_H = \{\lambda_i\}_{i=1}^N \) containing the, in general complex, eigenvalues of \( H \). Let \( \rho(H) = \max_{i \in \mathcal{S}_H} |\lambda_i| \) denote the maximum modulus eigenvalue which is referred to as the spectral radius of \( H \). From the theory of nonnegative matrices we have:

**Proposition 3.1 (Perron-Frobenius Theorem)**

Let \( H \geq 0 \) be a square irreducible matrix. Then:

a. \( \rho(H) \) is an eigenvalue of \( H \).

b. A positive eigenvector \( e > 0 \) corresponds to \( \rho(H) \).

c. \( \rho(H) \) increases when any entry of \( H \) increases.

d. \( \rho(H) \) is a simple eigenvalue of \( H \).

**Proof:** See, e.g., Theorem 2.1 in [113].
The spectral radius plays a major role in the rate of convergence for various iterative methods. As the system feasibility is germane to the magnitude of the spectral radius, it has also been proposed as a measure of congestion [40,44]. The following two results make the connection clear.

**Proposition 3.2** If $\mathbf{H} \geq 0$ is a square matrix, then the following are equivalent:

a. $\alpha > \rho(\mathbf{H})$ and $\mathbf{H}$ is irreducible.

b. $\alpha \mathbf{I} - \mathbf{H}$ is nonsingular and $(\alpha \mathbf{I} - \mathbf{H})^{-1} > 0$.

**Proof:** See, e.g., Theorem 3.9 in [113].

**Proposition 3.3** The matrix $(\mathbf{I} - \mathbf{H})$ is nonsingular and the series $\mathbf{I} + \mathbf{H} + \mathbf{H}^2 + \cdots$ converges if and only if $\rho(\mathbf{H}) < 1$. Moreover if $\rho(\mathbf{H}) < 1$ then

$$(\mathbf{I} - \mathbf{H})^{-1} = \mathbf{I} + \mathbf{H} + \mathbf{H}^2 + \cdots = \sum_{i=0}^\infty \mathbf{H}^i.$$ 

**Proof:** See, e.g., Theorem 2-4.4 in [125].

Let us define the following problem.

$$\begin{align*}
\min_{\mathbf{p}} & \quad \mathbf{1}^T \mathbf{p} \\
\text{subject to} & \quad \mathbf{A}\mathbf{p} = \mathbf{\eta} \\
& \quad \mathbf{p} \in \mathcal{P}
\end{align*}$$

Then we can state the following property on $(\mathcal{P}_\omega)$.

**Corollary 3.1** If there exists a feasible solution to the problem $(\mathcal{P}_\omega)$, the optimum solution of $(\mathcal{P}_\omega)$ is equivalently obtained by solving $(\mathcal{P}_\omega)$ and is given by $\mathbf{p}^* = \mathbf{A}^{-1}\mathbf{\eta}$.

**Proof:** We show that solving $(\mathcal{P}_\omega)$ is equivalent to solving $(\mathcal{P}_\omega)$. We first show that if a solution to $(\mathcal{P}_\omega)$ exists, then it is the optimum solution to $(\mathcal{P}_\omega)$. Clearly, if there exists a feasible solution to $(\mathcal{P}_\omega)$, $\mathbf{A}$ must be nonsingular and there exists one to $(\mathcal{P}_\omega)$. Let $\mathbf{\Gamma}^* = \text{diag}[\gamma_i^{**}]$ and $\mathbf{F} = [\theta_{ij} \phi_{ij}]$ so that $\mathbf{A} = \mathbf{I} - \mathbf{\Gamma}^* \mathbf{F}$ and $\mathbf{\eta} = \mathbf{\Gamma}^* \mathbf{u}$ for the $N \times 1$ vector $\mathbf{u} = (u_i)$ where $u_i = \frac{a_i}{y_i}$. Let $(\mathbf{I} - \mathbf{\Gamma}^* \mathbf{F})\mathbf{p}^* = \mathbf{\Gamma}^* \mathbf{u}$. For any other feasible solution $(\mathbf{I} - \mathbf{\Gamma}^* \mathbf{F})\mathbf{p} \geq \mathbf{\Gamma}^* \mathbf{u}$, there exists a matrix $\mathbf{\Gamma}$ with at least one element $\gamma_i^{**} > \gamma_i^{**}, 1 \leq i \leq N$ such that $(\mathbf{I} - \mathbf{\Gamma} \mathbf{F})\mathbf{p} = \mathbf{\Gamma} \mathbf{u}$. We have that $\mathbf{I} - \mathbf{\Gamma} \mathbf{F} \leq \mathbf{I} - \mathbf{\Gamma}^* \mathbf{F}$ and therefore $(\mathbf{I} - \mathbf{\Gamma} \mathbf{F})^{-1} \geq (\mathbf{I} - \mathbf{\Gamma}^* \mathbf{F})^{-1}$, which by Proposition 3.3 exists for a feasible system.

Thus $\mathbf{1}^T \mathbf{p} = \mathbf{1}^T (\mathbf{I} - \mathbf{\Gamma} \mathbf{F})^{-1} \mathbf{\Gamma}^* \mathbf{u} > \mathbf{1}^T (\mathbf{I} - \mathbf{\Gamma}^* \mathbf{F})^{-1} \mathbf{\Gamma}^* \mathbf{u} = \mathbf{1}^T \mathbf{p}^*$ and we have shown that any other feasible solution than $\mathbf{p}^*$ will yield a larger objective value.
From Proposition 3.1 c., it follows that if \((P_m)\) has no feasible solution, neither does \((P_\geq)\).  \(\square\)

A similar property was alternatively proven in [130]. Corollary 3.1 and the above propositions say that we may equivalently solve

\[(I - H)p = \eta\]  \hspace{1cm} (3.2)

for finding the optimal power vector. By Proposition 3.3 the resulting power solution is unique. From Proposition 3.3 having \(\rho(H) < 1\) is a necessary condition for feasibility and also sufficient for the unconstrained problem. The resulting solution is the Pareto optimal solution to the system of inequalities in \((P_\geq)\) in the following sense: any other feasible power vector will have every element not less and at least one element greater than the corresponding in \(p^* = (I - H)^{-1}\eta\), which means that the optimum solution also minimizes every user’s power. With power constraints, the feasibility cannot be determined by the spectral radius solely. In general, determining feasibility of a large system is difficult. Chapter 4-5 will address these issues.

**Example 3.1.1** Consider a single cell with \(N\) users in a CDMA system, each requiring a SIR-target \(\gamma^i\). With the \(H\) matrix consisting of

\[h_{ij} = \begin{cases} \gamma^i g_{ij}, & i \neq j \\ 0, & i = j \end{cases}\]

it is easy to see that there exists a positive eigenvector \(e = (1/g_i)\) that gives an eigenvalue \(\lambda = \gamma^i/(N - 1)\). Since \(H\) is nonnegative and irreducible, \(e\) is the Perron eigenvector and \(\rho(H) = \lambda\). The optimal solution to this power control problem (3.2) with background noise \(\nu\), can be shown to equal \(p_i = \frac{\nu}{\gamma^i 1 - \gamma^i/(N - 1)}\). Hence, the user capacity is in the noisy unconstrained case given by \(N < 1 - 1/\gamma^i\). The limit \(1 - 1/\gamma^i\) is sometimes known as the pole capacity of the cell.

**Example 3.1.2** The previous example showed that the spectral radius in the single-cell case is independent of the link gains. Let us consider two cells, where user 1 and 2 reside in one cell; and user 3 and 4 in the other. By symmetry reasons, \(g_{33} = g_{22}, g_{14} = g_{34}, g_{31} = g_{41}, g_{21} = g_{12}, g_{32} = g_{43}, g_{44} = g_{44}\) and \(g_{32} = g_{42}\). If we consider \(\gamma^i = \gamma^j\) and \(\theta_{ij} = 1\), we can obtain the spectral radius

\[\rho(H) = \gamma^i \left(1 + \frac{g_{21}}{g_{11}} + \frac{g_{22}}{g_{22}} \sqrt{\frac{g_{33}}{g_{33}} + \frac{g_{44}}{g_{44}}}\right)\]

which now depends on the link gains. Naturally large \(g_{ij}\) and small \(g_{ij}\), will result in higher achievable \(\gamma^i\).
3.2 Iterative Solution Methods

3.2.1 Matrix Norms

Let us consider the norms \( \|p\|_\infty = \max_i |p_i| \) and \( \|H\|_\infty = \max_{i,j} \sum_{j=1}^N |h_{ij}| \)
for a vector and matrix, respectively. It can be easily proven that they are consistent, \( \|Hp\|_\infty \leq \|H\|_\infty \|p\|_\infty \). Further, define the weighted vector norm of \( p \) as \( \|p\|_W^W = \|WH^{-1}Wp\|_\infty \) for a nonsingular matrix \( W \). Then it follows from the consistency that,

\[
\|Hp\|_W^W = \|WH^{-1}Wp\|_\infty \leq \|WH^{-1}\|_\infty \|Wp\|_\infty \overset{\text{def}}{=} \|H\|_W^W \|p\|_W^W.
\]

Thus the consistent weighted matrix norm can be defined as \( \|H\|_W^W = \|WH^{-1}\|_\infty \). We shall also make use of the following result.

**Lemma 3.2** For any matrix \( B \) and any \( \epsilon > 0 \) there exists a nonsingular matrix \( W \) such that \( \|B\|_W^W \leq \rho(B) + \epsilon \).

**Proof:** By Theorem 2-1.15 in [125] there exists a nonsingular matrix \( P \) such that \( P^{-1}BP = J \) where \( J \) is the Jordan canonical form of \( B \). If we choose \( D = \text{diag}(1, \epsilon, \epsilon^2, \ldots, \epsilon^{N-1}) \) we can write \( D^{-1}JD = J^{(\epsilon)} \) where \( J^{(\epsilon)} \) is the same as \( J \) except that the off-diagonal elements are multiplied by \( \epsilon \). Therefore we obtain \( J^{(\epsilon)} = Q^{-1}BQ \) where \( Q = PD \) is nonsingular. Clearly, \( \|J^{(\epsilon)}\|_\infty \leq \rho(B) + \epsilon \) and \( \|B\|_W^{-1} = \|Q^{-1}BQ\|_\infty = \|J^{(\epsilon)}\|_\infty \), which means that there exists a nonsingular weight matrix \( W = Q^{-1} \). \( \square \)

Thus when \( \rho(H) < 1 \), we can take \( \epsilon = (1 - \rho(H))/2 \) and always find a weight matrix such that \( \|H\|_W^W < 1 \). In particular, if we let \( W = \text{diag}(e^{-1}) \) where \( e \) is the Perron vector of \( H \), we obtain

\[
\|H\|_W^W = \max_i \left\{ \frac{1}{e_i} \sum_{j=1}^N h_{ij} e_j \right\} \leq \rho(H).
\]

Theorem 2-3.4 in [125] states that, for any weighted matrix norm \( \rho(H) \leq \|H\|_W^W \). Hence, it follows that the weight matrix \( W = \text{diag}(e^{-1}) \) will give the smallest possible norm, which motivates its use in subsequent analysis.

3.2.2 Iterative Power Control

If \( N \) is large, inverting matrices like \( I - H \) in (3.2) may be cumbersome and computationally expensive. Therefore solving large size linear equation systems
iteratively, has become a key technique that appears in many practical problems, see [113, 125]. In addition, for power control, even if the inverse could be computed by efficient methods, having knowledge about all elements in $H$ may not always be possible. The iterative approach is further applicable from a practical point of view since ultimately, the environment will not be fully stationary and updates are required. Let us specifically consider a first-order basic stationary iterative method for solving $Ap = \eta$. Let $p(n)$ denote a power vector at iteration $n$. Then, update the powers according to

$$\begin{align*}
p(n+1) = M^{-1}Np(n) + M^{-1}\eta, \quad n = 0, 1, 2, \ldots
\end{align*}$$

where $M$ and $N$ are appropriately chosen matrices. How to choose them are dependent on two things, namely consistency and convergence. Let $M$ be a nonsingular matrix and $A = M - N$, which is known as a splitting of $A$. If we set $p(0) = A^{-1}\eta$ we get $p(n) = M^{-1}N\eta^n + M^{-1}\eta = M^{-1}(M - A)A^{-1}\eta + M^{-1}\eta = A^{-1}\eta$, for all $n > 0$, thus (3.3) is said to be consistent with (3.2). Usually $M^{-1}N$ is called the iteration matrix. For proving convergence we need to find a proper norm. Generally we have,

**Lemma 3.3** The sequence $p(0), p(1), \ldots$ converges to $p^*$ if and only if

$$\lim_{n \to \infty} \|p(n) - p^*\| = 0 \text{ for any vector norm } \| \cdot \|.$$ 

**Proof:** See, e.g., Theorem 2-4.1 in [125].

Unless not specified, we shall say that a sequence is convergent if for all $p(0) \geq 0$, the sequence converges to a limit independent of $p(0)$. A sequence exhibiting the property that $\|p(n+1) - p^*\| \leq \alpha \|p(n) - p^*\|$ with $\alpha < 1$, is referred to as a pseudo-contraction with rate $\alpha$. By Lemma 3.3, it can be shown that a pseudo-contraction is convergent. Then we have that the above method converges under the following condition.

**Proposition 3.4** The sequence of vectors $p(n)$ in (3.3) converges to the solution of $Ap = \eta$ for any $p(0)$ if and only if $\rho(M^{-1}N) < 1$.

**Proof:** Define the error vector $e(n) = p(n) - p^*$. Since $Mp^* = Np^* + \eta$ we get $Me(n+1) = M(p(n+1) - p^*) = Np(n) + \eta - Np^* - \eta = N(p(n) - p^*)$ or equivalently $Me(n+1) = Ne(n)$. By recursion, $e(m) = (M^{-1}N)^m e(0)$. If $\rho(M^{-1}N) < 1$, we can choose $\epsilon > 0$ such that $\rho(M^{-1}N) + \epsilon < 1$. From Lemma 3.2, there exists a matrix norm such that $\|M^{-1}N\|^W \leq \rho(M^{-1}N) + \epsilon < 1$. Hence $\|M^{-1}N\|^W \leq \rho(M^{-1}N)^m e(0)$ as $m \to \infty$. Therefore, $\lim_{m \to \infty} \|e(m)\|^W \leq \lim_{m \to \infty} \|M^{-1}N\|^W \|e(0)\|^W = 0$ and by Lemma 3.3, the iteration converges. Moreover, if it converges we must have $\lim_{m \to \infty} (M^{-1}N)^m = 0$. So for the Perron solution we must have $\rho(M^{-1}N)^m e = (M^{-1}N)^m e = 0$ as $m \to \infty$, which means that $\rho(M^{-1}N) < 1$. If $\rho(M^{-1}N) \geq 1$, we can choose an initial starting vector $p(0)$ such that $e(0)$ becomes the eigenvector corresponding to $\rho(M^{-1}N)$. 

Then, \( \lim_{m \to \infty} ||e(m)|| = ||\rho(M^{-1}N)^m e(0)|| \neq 0 \), and (3.3) cannot converge. This concludes the proposition. \( \square \)

**Example 3.2.1** Depending on how to split the matrix \( A \), different power control algorithms of different distributiveness can be derived. If we let \( 0 < \beta \leq 1 \) and choose \( M = \frac{1}{\beta}I \) and \( N = (\frac{1}{\beta} - 1)I + H \), we obtain the componentwise update form

\[
p_k(n + 1) = (1 - \beta)p_k(n) + \beta \frac{\gamma_1}{SIR_r(n)} p_k(n)
\]

which is the Foschini-Miljanic algorithm [35]. Furthermore, \( \rho(M^{-1}N) = \rho((1 - \beta)I + \beta H) = 1 - \beta + \beta \rho(H) \). Thus, if \( \rho(H) < 1 \) it converges for all \( 0 < \beta \leq 1 \) and \( \min_{\beta} \rho(M^{-1}N) = \rho(H) \) for \( \beta = 1 \). When \( \beta = 1 \) and power constraints are added, the resulting mapping is known as the DCPC algorithm [39].

### 3.2.3 Rate of Convergence

Once convergence is established, the rate of convergence is a useful measure to compare iterative algorithms for different choices of \( M \) and \( N \). As can be seen from the proof of Proposition 3.4, the norm of the iteration matrix \( M^{-1}N \) can be taken for such a measure as it determines how fast the error vector decreases. Example 3.2.1 shows that the relaxation parameter \( \beta \) in the Foschini-Miljanic algorithm should be taken to unity for minimizing the spectral radius of the iteration matrix. In a matrix norm, \( ||(M^{-1}N)^m|| \) is called the convergence factor for \( m \) iterations and \( ||(M^{-1}N)^m|| \# \) the average convergence factor per iteration for \( m \) iterations respectively. Unfortunately, these numbers may be difficult to obtain in a computationally cheap manner. Also, \( ||(M^{-1}N)^m|| \# \) may not have to converge monotonically to zero which motivates the focus on asymptotic measures for rate of convergence. As a motivation for an estimate of the average convergence factor, \( \lim_{m \to \infty} ||(M^{-1}N)^m|| \# = \rho(M^{-1}N) \), which can be proven (see, e.g., Corollary 3.7.3 [125]).

**Definition 3.2** Denote by \( r_m = -\log_{10} ||(M^{-1}N)^m|| \# \) the average rate of convergence and \( r_{\infty} = -\log_{10} \rho(M^{-1}N) \) the asymptotic average rate of convergence.

If for two different splittings \( A = M_1 - N_1 = M_2 - N_2 \) the convergence factors \( -\log_{10} ||(M_1^{-1}N_1)^m|| \# < -\log_{10} ||(M_2^{-1}N_2)^m|| \# \), then the iteration matrix \( M_2^{-1}N_2 \) is said to be by definition iteratively faster [113] for \( m \) iterations than \( M_1^{-1}N_1 \). We will refer to the splitting \( M_2 - N_2 \) as having faster asymptotic average rate of convergence if \( \rho(M_2^{-1}N_2) < \rho(M_1^{-1}N_1) \).

**Example 3.2.2** We have \( e(m) = (M^{-1}N)^m e(0) \) and by Lemma 3.2, for a feasible system we can always find a norm such that the normalized error decays as \( \frac{||e(m)||}{||e(0)||} \leq \rho(M^{-1}N) + \epsilon)^m \). In particular when \( M^{-1}N \) is nonnegative and irreducible, for some weighted maximum norm, the normalized error decays as \( \rho(M^{-1}N)^m \). To reduce the
error norm by a factor $10^{-p}$, $p > 0$, the number of iterations required is the least value $m$ such that $\rho(M^{-1}N)^m \leq 10^{-p}$ or $m \geq \frac{-\log_{10} \rho(M^{-1}N)}{-p} \approx \frac{2}{r_{\infty}}$ holds. Therefore, it also follows that $r_{\infty}$ finds its practical interpretation as it asymptotically is an estimate of the number of new correct decimals per iteration.

**Definition 3.3** A sequence of vectors $p(n)$ is said to converge geometrically at a rate $\alpha$ to $p^*$ if there exist nonnegative constants $A$ and $\alpha < 1$ such that $\|p(n) - p^*\| \leq A\alpha^n$.

**Example 3.2.3** Consider the iterative method (3.3) with the choice of $M = I$ and $N = H$. Expanding (3.3), we get $p_i(n+1) = \frac{1}{\sum_{j=1}^{N} h_{ij}(n)} p_i(n)$ which is the DPC algorithm executed in a noisy system with $c_i = \gamma_i^2$ or the Foschini-Miljanic algorithm with $\beta = 1$. Proposition 3.4 says that it converges iff $\rho(H) < 1$ which was also a necessary condition from Proposition 3.3 for finding a feasible solution. Thus it converges with the asymptotic average rate of $r_{\infty} = -\log_{10} \rho(H)$. Moreover, we have

$$\|p(n) - p^*\|_{\infty} = \ldots = \|H^*(p(0) - p^*)\|_{\infty} \leq \rho(H)^n \|p(0) - p^*\|_{\infty}$$

for $W = \text{diag}[e_i^{-1}]$ where $e$ is the Perron vector. Thus, for the weighted maximum norm it converges geometrically at rate $\rho(H)$.

**Example 3.2.4** Consider SIR-balancing in a noisless system and an $N \times N$ matrix $H^*$ with elements $h_{ij} = g_{ij}/g_i$ for $i \neq j$ and 0 otherwise. For solving the resulting eigenvalue problem, i.e., to find the Perron vector, consider the following algorithm

$$p(n+1) = (\delta(n)H + \omega(n)I)p(n)$$

(3.4)

$$\lambda(n+1) = \|\tilde{p}(n+1)\|_{\infty}$$

(3.5)

$$p(n+1) = \lambda(n+1)^{-1}\tilde{p}(n+1)$$

(3.6)

where $p(n)$ is the power vector at iteration $n$, $I$ an identity matrix and $\lambda(n+1)$ is a value distributed to each transmitter. If $\omega(n) = 1$ or 0 and $\delta(n) = 1$, we get the DB and DPC algorithms respectively. The procedure (3.4)-(3.6) can be expressed for each user $i$ as

$$p_i(n+1) = \lambda(n+1)^{-1}\left(\frac{\delta(n)}{\sum_{j=1}^{N} h_{ij}(n)} p_i(n) + \omega(n) p_i(n)\right).$$

(3.7)

Define the diagonal matrix $A = \text{diag}[\lambda_i]$ and let us assume that there exists a nonsingular matrix $E$ such that $HE = EA$. Let the column vectors in $E$ be denoted as $e_i$. From the assumption, they must be the right-hand eigenvectors of $H$ and we can order them so $e_1$ is the Perron vector. Since $E$ is nonsingular, they are linearly independent. It follows that the rows in $E^{-1}$ are the left-hand eigenvectors of $H$. Since the eigenvectors span the $N$ dimensional space, we can express any positive starting vector in the right-hand eigenvectors as $p(0) = c_1 e_1 + c_2 e_2 + \cdots + c_N e_N$ where it is assumed that $\|e_i\|_{\infty} = 1$ for all $i$. For any such vector we can obtain the coefficients as $c = E^{-1}p(0)$. By Perron-Frobenius theorem there must exist one positive row, namely the first, in $E^{-1}$ consisting of the left-hand Perron vector. Therefore $c_1 > 0$ for any positive $p(0)$. Let us order the in general complex eigenvalues as $\rho(H) = \lambda_1 > |\lambda_2| \geq \ldots \geq |\lambda_N|$ and
define $\zeta_n = \prod_{k=1}^{n} \lambda(k)$. Iterating, we get

$$
\mathbf{p}(n) = \zeta_n^{-1} \sum_{j=1}^{N} c_j \prod_{k=0}^{n-1} (\delta(k) \lambda_j + \omega(k)) e_j
= \zeta_n^{-1} c_1 \prod_{k=0}^{n-1} (\delta(k) \lambda_1 + \omega(k)) \left( e_1 + \sum_{j=2}^{N} c_j \prod_{k=0}^{n-1} (\delta(k) \lambda_j + \omega(k)) e_j \right).
$$

Define $\bar{\omega}(k) = \omega(k) / \delta(k)$, then there exist constants $A > 0$, $0 < \alpha < 1$ so that the norm of the error vector can be bounded as

$$
\| (\zeta_n^{-1} c_1 \prod_{k=0}^{n-1} (\delta(k) \lambda_1 + \omega(k)))^{-1} \mathbf{p}(n) - e_1 \| \leq \sum_{j=2}^{N} \left| \frac{c_j}{c_1} \prod_{k=0}^{n-1} |(\lambda_j + \bar{\omega}(k))| \right| \leq A n^\alpha.
$$

Clearly the right hand side goes to zero for any sequence $0 \leq \{ \bar{\omega}(k) \} < \infty$ which implies a geometric rate of convergence. Define

$$
\beta_n = \zeta_n^{-1} c_1 \prod_{k=0}^{n-1} (\delta(k) \lambda_1 + \omega(k)),
$$

so from the above, we have $\mathbf{p}(n) \to \beta_n e_1$, and it follows that

$$
\lim_{n \to \infty} \beta_n = \lim_{n \to \infty} \frac{c_1 \prod_{k=0}^{n-1} (\lambda_1 + \bar{\omega}(k))}{\| \sum_{j=2}^{N} c_j \prod_{k=0}^{n-1} (\lambda_j + \bar{\omega}(k)) e_j \|} = 1
$$

which concludes geometric convergence with respect to the maximum norm. The spectral radius can be estimated observing that

$$
\lambda(n) \to \| (\delta(n) \mathbf{H} + \omega(n) \mathbf{I}) e_1 \| = \delta(n) \rho(\mathbf{H}) + \omega(n)
$$
as $n \to \infty$. For a given matrix $\mathbf{H}$, the optimal $\omega$ and $\delta$ values in terms of offering the fastest rate of convergence, depend on the location of the eigenvalues, i.e., the system environment. Since time variant parameters $\omega$ and $\delta$ can be used, different values can be utilized during iterations to enhance the rate of convergence. If all the eigenvalues of $\mathbf{H}$ are real and ordered as $\lambda_1 > \lambda_2 \geq \ldots \geq \lambda_N$, it is easy to see that the asymptotically fastest convergence is obtained for $\omega / \delta = -(\lambda_2 + \lambda_N) / 2$. A limitation is that, for practical reasons, $\omega / \delta$ must also be chosen so that a positive power value is guaranteed.

It should be noted that the approach of designing algorithms that minimize the norm of the error vector is not always strictly optimal for power control applications. Even though the error may be arbitrarily small, in the pathological case, all users could still be in outage. However, rate of convergence clearly relates to how “contractive” an algorithm is and could serve as comparison between algorithms, reflecting their ability to perform in a nonstationary environment and rapidly bring the powers to the neighborhood of a quasi-stationary fixed point.
3.2.4 Standard Interference Functions

A framework for distributed power control, not being based on matrix splittings, has been proposed by Yates [122]. Let \( \mathcal{T} : \mathbb{R}_+^N \rightarrow \mathbb{R}_+^N \) be a mapping of a power vector and consider the following definition.

**Definition 3.4** A power control mapping \( \mathcal{T} \) is called a standard interference function if for all \( p > 0 \) the following properties are satisfied.

a. Positivity. \( \mathcal{T}(p) > 0 \)

b. Monotonicity. If \( p \geq p' \), then \( \mathcal{T}(p) \geq \mathcal{T}(p') \).

c. Scalability. For all \( \alpha > 1, \alpha \mathcal{T}(p) > \mathcal{T}(\alpha p) \).

If \( \mathcal{T} \) is a standard interference function, the iteration \( p(n+1) = \mathcal{T}(p(n)) \) is called a standard power control algorithm. Then it follows that convergence is guaranteed if there exists a feasible solution.

**Proposition 3.5** If there exists a feasible solution, then for any initial power vector \( p(0) \), the standard power control algorithm converges to a unique fixed point \( p^* \).

**Proof:** See, e.g., Theorem 2 in [122].

It can easily be verified that the mapping \( \mathcal{T}(p(n)) = M^{-1}Np(n) + M^{-1}\eta \) is also a standard power control algorithm. In a real system there are delays and not all power control actions can be synchronized. Therefore it is important to establish stability when the powers are updated at different rates and under bounded delays. Convergence in the asynchronous case is covered by the framework.

**Proposition 3.6** If there exists a feasible solution, then for any initial power vector \( p(0) \), the asynchronous standard power control algorithm converges to a unique fixed point \( p^* \).

**Proof:** See, e.g., Theorem 4 in [122].

It should be noted that the conditions stated in Definition 3.4 are only sufficient conditions. Convergent power control algorithms that are not standard may be found and will be exemplified in the next chapter.

3.3 Local Loop Stability

In addition to the issue of global feasibility discussed hitherto, the stability of the system is tightly intertwined with the dynamics of the power control loop. A system with closed loop power control may be regarded as a collection of local loops, interchanging information about the system state via the interference. It is well-known from control theory that, local loops may become unstable subject
to time delays, forcing the control actions to be based on outdated information. As a consequence, this may give rise to oscillations, severely hampering the tracking capability and lowering signal quality. In this thesis, we shall focus on the resource allocation part, and therefore the global stability issues. A comprehensive stability investigation of local loops can be found in [40].
Chapter 4

Generalized Distributed Constrained Power Control

The DISTRIBUTED constrained power control algorithm (DCPC) [39] has become one of the most widely accepted practical power control algorithms by the academic community. It frequently appears incorporated in radio resource management functions of related areas [6, 7, 11, 119, 121]. The DCPC algorithm has a property that the power could reach its maximum level when a user is experiencing degradation of channel quality. Unfortunately, using maximum transmitter power may not necessarily lead to sufficient improvement of channel quality and will thereby generate severe interference, hitting other users. This undesirable phenomenon happens more often when the system is congested.

4.1 Introduction

In this chapter, we revisit and generalize the DCPC algorithm in order not to necessarily use the maximum power when the channel quality is poor. The suggested generalized algorithm is general in the sense that DCPC is a special case of it. Moreover, for convergence analysis it does not fit directly into the frameworks of matrix splittings or standard interference functions. Based on the generalization, we suggest two power control algorithms and compare them with DCPC. We prove that, for the feasible system, the generalized algorithm converges to the fixed point that supports every active transmitter, as DCPC does. Under poor conditions, rather than combating the interference by maximizing the power, we suggest that the power may instead be temporarily set to zero, which we will call temporary removal. This is a softer form of the permanent removal [6], which disconnects the user from the channel. In our case the user stays on the same channel and transmission will be resumed if the interference situation becomes favorable again. This concept of high interference
inducing aborted transmission is particularly applicable for nonreal time data services where the delay tolerance is higher. Because of the decreased interference power, the capacity of the system is expected to increase and the energy consumption decrease.

4.2 Distributed Power Control Algorithm

Let us consider a power constrained case with \( \mathcal{P} = \mathcal{P}_1 \times \ldots \times \mathcal{P}_N \) where \( \mathcal{P}_i = \{ p_i : 0 \leq p_i \leq \bar{p}_i \} \). The DCPC algorithm suggested by Grandhi et al., [39] has the form;

\[
\text{(DCPC)} \quad p_i(n + 1) = \mathcal{T}_i(p(n)) \overset{\text{def}}{=} \min \{ T_i(p(n)), \bar{p}_i \}, \quad n = 0, 1, \ldots ,
\]

(4.1)

for the mapping \( T_i(p(n)) \overset{\text{def}}{=} \frac{1}{\sum_{j=1}^{N} p_j(n)} \sum_{j=1}^{N} h_{ij} p_j(n) + \eta_i \). Now consider the following generalized algorithm for constrained power control;

\[
\text{(GDCPC)} \quad p_i(n + 1) = \bar{T}_i(p(n)) \overset{\text{def}}{=} \begin{cases} 
T_i(p(n)), & \text{if } T_i(p(n)) \in \mathcal{P}_i \\
\bar{p}_i(n), & \text{if } T_i(p(n)) \notin \mathcal{P}_i 
\end{cases}, \quad n = 0, 1, \ldots ,
\]

(4.2)

where the power vector \( \vec{p}(n) \in \mathcal{P} \). If we choose \( \vec{p}(n) = \bar{p} \), GDCPC is reduced to the DCPC algorithm. When setting \( \bar{p}_i(n) = 0 \), it can be interpreted as a temporary connection removal, allowing the removed user to stay on the channel and automatically power up again if the interference has decreased. By setting the transmitter power to zero, the user will not waste energy mitigating bad channel conditions and other users will benefit from lower interference. When \( \vec{p}(n) \neq \bar{p} \), GDCPC violates the monotonicity property, and thus it is not a standard power control algorithm [122] that guarantees convergence. In some cases, it can also violate the scalability and positivity properties. However, we can prove the convergence of GDCPC to the optimal solution in the feasible case, which will be given later in this section. To begin with, we describe the energy conservation property of GDCPC. Note that from now on, we refer to the case \( \bar{p}_i(n) \neq \bar{p}_i \) for some \( i \) as GDCPC and the case \( \bar{p}_i(n) = \bar{p}_i \) for all \( i \) as DCPC.

4.2.1 Energy Conservation

Let \( \overline{T}_i^n(p) \) and \( \overline{T}_i^n(p) \) respectively, denote the power level of mobile \( i \) of DCPC and GDCPC at iteration \( n \), starting with a power vector \( p \). Then we can prove the following property on the generated powers of GDCPC.

**Proposition 4.1** For any \( p \geq 0 \), \( \overline{T}_i^n(p) \leq \overline{T}_i^n(p) \) for all \( i \) and \( n \).
Proof: From the definition of GDCPC, we have \( \bar{T}_i(p) \leq \bar{T}_i(p) \) for any nonnegative power vector \( p \). Also, from the definition of \( \bar{T}_i(p) \), we find that if \( 0 \leq p_1 \leq p_2 \) then \( \bar{T}_i(p_1) \leq \bar{T}_i(p_2) \). Therefore, if \( 0 \leq p_1 \leq p_2 \), then \( \bar{T}_i(p_1) \leq \bar{T}_i(p) \leq \bar{T}_i(p_2) \). Using these relations, we have:

\[
\begin{align*}
\bar{T}_i(p) & \leq \bar{T}_i(p) \\
\bar{T}_i(\bar{T}_i(p)) & \leq \bar{T}_i(\bar{T}_i(p)) \\
& \vdots \\
\bar{T}_i^n(p) & \leq \bar{T}_i(p)
\end{align*}
\]

which concludes the proof. \( \square \)

Furthermore, the total system power becomes strictly smaller if \( \bar{p}_i(n) \) is utilized, which is described by the next result.

Proposition 4.2 Let us assume that there exists an iteration \( n_0 \) and user \( i \) such that \( T_i^{n_0}(p) > \bar{p}_i \) and that \( \bar{p}_i(n) < \bar{p}_i \) for all \( i \) and \( n \). Then, \( \sum_{i=1}^{N} T_i^n(p) < \sum_{i=1}^{N} T_i^n(p) \) for all \( n \geq n_0 \).

Proof: Since \( \bar{T}_i^n(p) \leq \bar{T}_i^n(p) \) from Proposition 4.1, it is sufficient to show that there exists at least one user \( j \in J_n \) for any iteration \( n \geq n_0 \) where \( J_n = \{j : T_j^n(p) < \bar{T}_j^n(p)\} \). At iteration \( n = n_0 \), user \( i \in J_n \) because \( T_i^{n_0}(p) = \bar{p}_i(n_0) < \bar{p}_i = \bar{T}_i^{n_0}(p) \). At iteration \( n_0 + 1 \), from the definition of \( T_i(p) \), we can find that \( T_i^{n_0+1}(p) < T_i^{n_0+1}(p) \) for any \( j \neq i \). In the same manner, we can see that there exists a user \( j \in J_n \) for all \( n > n_0 + 1 \). \( \square \)

Note that Proposition 4.1 and 4.2 are general and hold for both feasible and infeasible systems. Proposition 4.1 says that when starting from the same initial power vector, the power value from GDCPC is not greater than that of DCPC in any iteration. Further, if there is an event that the required power is greater than the maximum allowed level, then from Proposition 4.2, we can expect a certain amount of energy conservation from GDCPC, compared to DCPC. For \( \bar{p}_i(n) \) of GDCPC, we can use any nonnegative value less than or equal to \( \bar{p}_i \). However, from the proof of Proposition 4.2, setting \( \bar{p}_i(n) = 0 \) will lead to the most energy-saving results. For simplicity, we will denote this version of GDCPC by GDCPC(I) throughout the chapter.

### 4.2.2 Convergence in Feasible Systems

As was mentioned before, GDCPC is not a standard interference function but convergence is still guaranteed for a feasible system by the following property.

Proposition 4.3 Starting with any power vector \( p \in P \), GDCPC converges to \( p^* \) of a feasible system.
Proof: From the definition of GDCPC, we have \( \tilde{T}_i(p) \leq T_i(p) \) for any non-negative power vector \( p \). Also, from the definition of \( T_i(p) \), we find that if \( 0 \leq p_1 \leq p_2 \) then \( T_i(p_1) \leq T_i(p_2) \). Therefore, if \( 0 \leq p_1 \leq p_2 \), then \( \tilde{T}_i(p_1) \leq \tilde{T}_i(p_2) \). Using these relations, we have:

\[
\tilde{T}_i(p) \leq T_i(p) \\
\tilde{T}_i(\tilde{T}_i(p)) \leq T_i(\tilde{T}_i(p)) \\
\vdots \\
\tilde{T}_i^n(p) \leq T_i^n(p)
\] (4.4)

From Proposition 3.1, it is known that \( 0 < \rho(H) < 1 \) for a feasible system. Let us consider the weighted maximum norm of \( H \) with the nonsingular weight matrix \( W = \text{diag}[e^{-1}] \) where \( e \) is the Perron eigenvector of \( H \). Then the convergent mapping \( \mathcal{T} = (\tilde{T}_i) \) fulfills \( \|T_i^n(p) - p^*\|_\infty \leq \rho(H)^n \|p - p^*\|_\infty \) (Proposition 3.1 [83] and Example 3.2.3). Therefore, for a feasible system\(^1\), there exists an integer \( n_1 \) such that

\[
\mathcal{T}^n(p) < \hat{p} \text{ for all } n > n_1.
\] (4.5)

By denoting \( \mathcal{T} = (\tilde{T}_i) \), it follows from Equation (4.4) that \( \tilde{T}_i^n(p) \leq \mathcal{T}_i^n(p) \), thus for \( n > n_1 \) we can write:

\[
\tilde{T}_i^{n+1}(p) = T_i(\tilde{T}_i^n(p)) \leq T_i^{n+1}(p) \\
\vdots \\
\tilde{T}_i^{n+m+1}(p) = T_i^m(\tilde{T}_i^n(p)) \leq T_i^{n+m+1}(p)
\] (4.6)

Therefore, \( \lim_{m \to \infty} \tilde{T}_i^{n+m+1}(p) = p^* \), due to the convergent mapping \( \mathcal{T} \). \(\square\)

Remark 4.1 So far we have focused on generalizing DCPC but Propositions 4.1-4.3 will still hold even if we choose other standard interference functions [51,122] for \( \mathcal{T} \) in (4.2).

Remark 4.2 The asynchronous DCPC, which is standard [122], can be generalized to asynchronous GDCPC. By noting that the power vectors generated by the asynchronous GDCPC are always less than or equal to the corresponding power vectors generated by the asynchronous DCPC and by changing the weighted maximum norm to the norm utilized by Mitra [83], its convergence can be proved as in the proof of Proposition 4.3.

\(^1\)To be strict, we do not allow any \( i \) such that \( p_i^* = p_i \). In reality this event would occur with zero probability. However, convergence for this pathological case can be shown using the pseudo-contractive property of Proposition 4.4.
From Proposition 4.2, we may have that \( \sum_{i=1}^{N} \tilde{T}^n_i(p) < \sum_{i=1}^{N} T^n_i(p) \) for all \( n \geq n_0 \). However, Proposition 4.3 says that

\[
\lim_{n \to \infty} \sum_{i=1}^{N} T^n_i(p) = \lim_{n \to \infty} \sum_{i=1}^{N} \tilde{T}^n_i(p) = \sum_{i=1}^{N} p^*_i,
\]

in the feasible case.

Besides the energy consumption, it is important how fast the power value will converge. It has been reported that DCPC converges to \( p^* \) at a geometric rate \([39, 51]\). So far the convergence rate of GDCPC is an open issue. However, let us choose in (4.2)

\[
\hat{p}_i(n) = \max \{ \bar{p}_i - (T_i(p(n)) - \bar{p}_i), 0 \}
\]

and denote this choice by GDCPC(II). In GDCPC(II), if the required power is larger than the maximum power \( \bar{p}_i \), a power lower than \( \bar{p}_i \) by the amount of the gap between the required power and \( \hat{p}_i \) is used. If the required power is twice larger than the maximum power, the transmitter power is set to zero. Then we have the following:

**Proposition 4.4** Starting with any power vector \( p \in \mathcal{P} \), GDCPC(II) converges to \( p^* \) of a feasible system with the same geometric rate as DCPC.

**Proof:** If \( T_i(p(n)) \leq \bar{p}_i \) for user \( i \), then \( |p_i(n + 1) - p^*_i| = |T_i(p(n)) - p^*_i| \).

However, if \( T_i(p(n)) > \bar{p}_i \) for user \( i \), then we have

\[
|p_i(n + 1) - p^*_i| = \max \{ 2\bar{p}_i - T_i(p(n)), 0 \} - p^*_i.
\]

When \( \bar{p}_i < T_i(p(n)) < 2\bar{p}_i \), we have

\[
|p_i(n + 1) - p^*_i| = |2(\bar{p}_i - T_i(p(n))) + T_i(p(n)) - p^*_i| \leq |T_i(p(n)) - p^*_i|
\]

and if \( T_i(p(n)) \geq 2\bar{p}_i \), it follows that

\[
|p_i(n + 1) - p^*_i| = |p_i(n) - p^*_i| \leq |T_i(p(n)) - p^*_i|.
\]

Therefore, we can say that, for any user \( i \),

\[
|p_i(n + 1) - p^*_i| \leq |T_i(p(n)) - p^*_i|.
\]

Consider the consistent matrix norm \( ||H||_W \). If we choose the Perron eigenvector \( e \) of \( H \) for the nonsingular weight matrix \( W = \text{diag}[e^{-1}] \), we have

\[
||p(n + 1) - p^*||_W \leq ||T(p(n)) - p^*||_W \leq ||H(p(n)) - p^*||_W \leq \rho(H)||p(n) - p^*||_W.
\]
Therefore,
\[ ||p(n) - p^*||_\infty^W \leq \rho(H)^n ||p(0) - p^*||_\infty^W.\]

If the system is feasible, it can be shown (see, e.g., Proposition 3.3) that \( \rho(H) < 1 \). Thus, we conclude that GDCPC(II) is a pseudo-contraction mapping with a geometric rate \( \rho(H) \), which is same as that of DCPC [39, 51]. \( \Box \)

With respect to Proposition 4.4, we conclude that within the class of GDCPC, there exists at least one \( \tilde{p}(n) \neq \tilde{p} \) which gives the same asymptotic convergence rate as DCPC and possibly increases the energy efficiency. Note that \( \tilde{p}(n) < \tilde{p} \) in GDCPC(II), and Proposition 4.2 is applicable to GDCPC(II).

### 4.2.3 Convergence in Infeasible Systems

Compared to DCPC, which also converges to a fixed point in an infeasible system, the convergence properties of GDCPC(I) and GDCPC(II) will differ. Let us consider an example in which two users are currently active. We can denote this case by the two-dimensional power vector space in Fig. 4.1. In the figure, the \( p_1 \) and \( p_2 \) axes denote the power values of user 1 and 2, respectively. The maximum power value that a user can transmit with, is limited to \( \bar{p} \). The SIR constraints are given by (3.2) and are displayed by the solid lines in the figure. The dashed lines represent (4.7), which we will refer to as virtual targets. So, if
adjusting the power toward the real target with GDCPC(II) would require more power than the maximum, the virtual target becomes active. In this example, the intersection between the solid lines, i.e., the fixed point that would support both users, is outside the feasible power region so the system is infeasible. Hence no more than one user can be supported. For fixed points \( p_c = (\eta_1, 0) \) and \( p_d = (0, \eta_2) \), user 1 and 2 are supported with their minimum power respectively. The fixed point of DCPC will in this case be \( p_f = (\tilde{p}, \tilde{p}) \) due to the power constraints. Clearly this point is the worst, considering no user is supported while the power usage is maximized. From the figure we find that GDCPC(I) will, depending on the starting point, oscillate between the points \( p_d = (0, 0) \) and \( p_b = (\eta_1, \eta_2) \) or converge to \( p_d \) or \( p_c \). In GDCPC(II), the fixed point may be the intersection between the virtual targets, \( p_v \). However, GDCPC(II) does not always necessarily converge to a fixed point in an infeasible system, which is exemplified next.

**Example 4.2.1** Let \( e \) be the eigenvector corresponding to \( \rho(H) \). Assume that there exists a \( p_0 \) such that \( \mathcal{T}(p_0) = p_0 \) and \( \tilde{p} < \mathcal{T}(p_0) < 2\tilde{p} \). For example the point \( p_c \) in Fig. 4.1 would be the only such point. If we choose \( p_0 + \alpha e \leq \tilde{p} \) where \( e \) is the Perron eigenvector satisfying \( He = \rho(H)e \) and \( \alpha > 0 \) is a constant, then we have

\[
|\mathcal{T}(p_0 + \alpha e) - \mathcal{T}(p_0)|_\infty = \|2\tilde{p} - H(p_0 + \alpha e) - \eta - 2\tilde{p} + HP_0 + \eta\|_\infty = \|H\alpha e\|_\infty = \|\rho(H)\alpha e\|_\infty.
\]

If \( \rho(H) > 1 \), by Proposition 3.1, \( e \) is still positive and we can write

\[
|\mathcal{T}(p_0 + \alpha e) - \mathcal{T}(p_0)|_\infty > \|p_0 + \alpha e - p_0\|_\infty, \text{ i.e., a noncontractive mapping that would yield a diverging sequence.}
\]

The characteristics of \( p_0 \) could be described as an unstable fixed point to which convergence is guaranteed only if the starting point is the point itself, \( p_0 \).

As illustrated in Fig. 4.1, GDCPC(I) and GDCPC(II) may converge to a fixed point but the dynamics are more unpredictable. Due to the possibility of oscillating powers, each user may generally expect a more varying SIR and its impact on the BER is not clear. Depending on power control interval, coding and interleaving strategies, the oscillation of SIR may or may not cause problems.

For an infeasible system, permanent connection removal has been utilized to increase system capacity [6, 72, 128]. In GDCPC(I) and GDCPC(II), the power may oscillate, and thus certain removal algorithms relying on convergence to some fixed points may not be utilized. For the purpose of permanent removal, we can extend the previously suggested, GRR-DCPC (Gradual Removal Restricted) [6], which is an “on-the-fly” gradual removal combined with DCPC. Instead of DCPC, we combine the gradual removal procedure with GDCPC(I) and GDCPC(II). That is, our modified gradual removal algorithm, which incorporates both temporary- and permanent removal, GRR-GDCPC, identifies user \( i \) as a candidate for permanent removal at iteration \( n_0 \) if \( T_i(p(n_0)) > \tilde{p}_i \) and sets \( p_i(n) = 0 \) for all \( n > n_0 \), with a given probability \( \delta > 0 \). Otherwise, \( p_i(n_0 + 1) = T_i(p(n_0)) \) and the power control proceeds with the next power iteration. In order to maximize system capacity and not over-aggressively remove connections, the removal probability \( \delta \), should be taken so that in each iteration, a single removal at a time is more probable than multiple removals.
Remark 4.3 It has been shown that GRR-DCPC converges to a stationary power vector [6]. Since GRR-GDCPC uses the same decision procedure as GRR-DCPC, it is clear that GRR-GDCPC will also converge.

4.3 Numerical Results

The main purpose of the numerical evaluation is to draw insight on how GD-CPC(I) and GDPC(II) perform in terms of ability to increase the number of supported users, energy conservation and convergence. To compare the performance of our proposed algorithms, we use DCPC as a reference algorithm throughout. Evaluation is performed in the DS-CDMA system described in Chapter 2 and illustrated in Fig. 1.2. The cell radius is set to 1 km and a wrap-around technique is used along with omni-directional antennas. For a given instance, 10 mobiles/cell are generated, the locations of which are uniformly distributed over the cells. At any given instance, the link gain is modeled by \( g_{ij} = S_{ij} \cdot r_{ij}^{-4} \), where \( S_{ij} \) is the shadow fading factor and \( r_{ij} \) is the distance between base \( i \) and mobile \( j \). The shadow fading factor is generated from a lognormal distribution with \( \mu_s = 0 \) dB, and \( \sigma_s = 8 \) dB, where the \( S_{ij} \)'s are mutually independent. The base station receiver noise is taken to be \( \nu_i = 10^{-12} \) W and the maximum mobile power is set to unity. The initial power for each mobile is randomly chosen from the interval \([0,1]\). Each user is assigned to the
Figure 4.3: Average transmitter power per user in a feasible system.

base station that provides the lowest signal attenuation. The data rate is set to \( R = 9.6 \) kbps and the spreading bandwidth \( W = 1.2288 \) MHz. The received \( E_b/I_0 \) from user \( i \) at the corresponding base is calculated by adding the processing gain to the received SIR (in dB). The target \( E_b/I_0 \) is set to 8 dB and 12 dB for each user when analyzing a feasible and an infeasible system, respectively. In the infeasible case, the higher target value was chosen to represent higher traffic load. It could be interpreted as users demanding higher data rates or lower BERs, resulting in an infeasible system.

We have considered one thousand independent feasible and infeasible snapshots of user locations and shadow fading factors respectively and the results are obtained as an average over these snapshots. To evaluate the impact on the system capacity, the outage probability is used as a performance measure. A connection is considered to be supported if the received \( E_b/I_0 \) is above 7.5 dB and 11.5 dB, respectively. That is, a 0.5 dB fade margin is utilized.

### 4.3.1 Feasible System

In Fig. 4.2, we see that the outage probability in the feasible system is lower over the whole range of iterations considered (except for the initial iterations) for both GDCPC(I) and GDCPC(II), compared to DCPC. Since we are starting from a random power vector, a large portion of the users are temporarily
removed the first iteration. This causes the crossover between DCPC and GDCPC. However, when they start powering up, the outage decreases. To assess the energy conservation, the average power, that is the total power used divided by the total number of mobiles in the system, is plotted. The energy conservation property is shown in Fig. 4.3, where GDCPC(I) obviously gives the best performance. The curves of GDCPC(I) and GDCPC(II) indicate that there are many users for which the required power at iteration 1 is greater than the maximum and the rest of the iterations follow Proposition 4.2. It can also be verified that the absolute gap among the three algorithms is decreasing with increasing iteration number as the power vectors converge to the fixed point for the feasible system. In Fig. 4.4, the convergence rate, measured as $||e(n)||_2/||e(0)||_2 = ||p(n) - p^*||_2/||p(0) - p^*||_2$, where $||.||_2$ denotes the Euclidean norm, empirically shows that GDCPC(I) is faster than GDCPC(II) and DCPC, both of which are proved to converge with a geometric rate. In conclusion, GDCPC(I) shows the best performance in terms of supporting more users, energy conservation and convergence rate while converging to the fixed point in a feasible system.
4.3.2 Infeasible System

Now, let us consider the outage probability for infeasible systems. In Fig. 4.5, the outage probability is plotted for each iteration when combining the algorithms with gradual removals using the removal probability $\delta = 0.01$. The suggested algorithms GRR-GDCPC(I) and GRR-GDCPC(II) reach a lower outage probability and they also achieve it much faster than the GRR-DCPC. To verify that the above behavior is not specifically related to that particular $\delta$, we perform simulations for various removal probabilities and collect data when the iteration has reached 300. From Fig. 4.6, it can be seen that a lower outage is obtained for a range of removal probabilities with $\delta = 0.01$ being the best one. We see that if no permanent removal is used, $\delta = 0$, the outage of GRR-GDCPC(I) and GRR-GDCPC(II) is much lower than that of GRR-DCPC, suggesting that the temporary removals alone, find the proper users to be removed. The figure also shows the tradeoff between a low $\delta$, offering lower outage but higher power and a high $\delta$ offering higher outage but lower power. Therefore we conclude that both GRR-GDCPC(I) and GRR-GDCPC(II) seem to be superior to GRR-DCPC that was known to be the best distributed removal algorithm so far [6].
4.4 Concluding Remarks

In this chapter, we have proposed a general scheme for distributed power control, from which algorithms that consume less power and support more users than the DCPC can be derived. The idea is that, when a user requires more power than the available, the power will be decreased to benefit other users experiencing favorable situations, increasing the overall capacity. It was shown that our algorithms converge to the fixed point of a feasible system, supporting every active user. For an infeasible system, convergence to a fixed point was exemplified not to necessarily occur. For that case, power oscillations may cause a rapidly varying SIR. However, this may not be a major obstacle, since the power control must be combined with a permanent removal algorithm. The difficulty with the proposed algorithms is that infeasibility may not be detected as for DCPC. This raises the question of how to combine a permanent removal algorithm with the proposed algorithms. We propose one possible approach by modifying a distributed gradual removal algorithm that was originally designed for use with DCPC.

In the numerical evaluation of the congested case, the resulting outage probability is unrealistically high, around 40%. In a practical scenario, the admission control would prohibit such heavy congestion. The reason for evaluating the
algorithm under this load is motivated for the approach of multirate systems, taken in next chapter. For a multirate system, the outage probability could be interpreted as the fraction of users not being served with the maximum data rate. This level could be high and of the same order as here. The practical applicability of the concept of temporary removals, which GDCPC(I) and GD-CPC(II) benefit from, could for example be non-real time data traffic where the flexibility of handling the transmission attempts is larger. This will be utilized in the next chapter Finding necessary and sufficient conditions for convergence in infeasible systems is still an open issue. Also, there is a possibility of designing more sophisticated removal algorithms suitable for our framework.
Chapter 5
Multirate Power Control

QUALITY-OF-SERVICE in DS-CDMA systems can be controlled by a suitable selection of processing gain and transmitter powers. In this chapter, distributed control of rate and power for best effort data services is considered. In particular, we elaborate on the problem of how to control the transmission rates for maximizing system throughput while simultaneously minimizing the transmitter powers. We assume a practical scenario, where every user has a finite set of discrete transmission rates and propose a simple heuristic rate allocation scheme, greedy rate packing (GRP), applicable in both up- and downlink. The scheme can be interpreted as a practical form of water-filling, in the sense that high transmission rates are allocated to users having high link gains and low interference. We show that GRP will under certain conditions, give maximum throughput. The GRP is then extended to guarantee a minimum data rate while maximizing network excess capacity. A distributed power control algorithm, managing intercell interference, is suggested and analyzed for applying GRP to a multicellular system.

5.1 Introduction
The service classes defined for multimedia services in UMTS include data rate, BER and delay constraints [1]. Our main focus will be on the throughput performance; resource allocation with delay considerations can be found in, e.g., [33, 66]. Since the effective transmission rate and QoS objective are closely related to the SIR, which can be efficiently controlled by SIR-based power control, it becomes natural to investigate a joint scheme for rate and power control. One approach in previous work on resource allocation for multi QoS support, is to specify no requirements on BERs and therefore no SIR-targets [87, 91, 105]. Another direction of resource allocation is to specify a minimum SIR [92, 98, 126]. The problem of throughput maximization was also studied in [94] for the multicellular case and for different user classes in [92]. Some of the solutions in the
aforementioned work, set some transmitter powers to zero. This can be denoted as a hybrid of TDMA and CDMA, which also bears resemblance to our concept of GRP. The main idea of the GRP is that the rates (and thus power values) of users with unfavorable link conditions are decreased, even to zero, while the users with good channels transmit data with higher rates.

The practical applicability of the algorithms in [87, 91, 92, 94, 98, 105, 126] is not clear because, either they require global information, consider only a single cell or use continuous transmission rates. In practical systems, the feasible transmission modes are discrete and finite; the number of link adaptation schemes, code rates, processing gain, etc. Furthermore, the resource allocation should preferably be distributed and function in multicellular systems. We shall adopt the discrete model for feasible transmission rates used in [64]. Therein a suggested distributed SPC algorithm was shown to converge to the power assignment that supports every active user with its maximum possible rate, using the minimum total transmitter power, if such a power assignment exists. However for a congested system, if no such assignment exists, little is known about its properties since convergence is not guaranteed. Due to the discreteness of rates, there may exist multiple rate combinations giving the same system throughput but with different total transmitter power. Therefore, an important issue is how to assure the maximum total throughput to the system while using the minimum total transmitter power. For a single-rate system, the question is equivalently reduced to maximizing the number of supported users with the minimum total transmitter power, which was shown to be NP-complete [5]. Therefore the problem with multiple rates, which is the focus in this chapter, is in the general multicellular case also NP-complete.

5.2 Transmission Rate Assignment

5.2.1 Refined System Model

Consider a single cell in a DS-CDMA cellular radio system where $N$ mobiles are using the same frequency channel. We will mainly consider the uplink throughout this chapter. Extension to the downlink will be discussed in Section 5.2.5. Each mobile $i$ ($1 \leq i \leq N$) can transmit with a power from the set $P_i = \{p_i : 0 \leq p_i \leq \bar{p}\}$. We consider a short time interval such that the link gain between each mobile $i$ and the base station is stationary and given by $g_i$. Given a power vector $p = (p_i)$, the received SIR of mobile $i$, is for the uplink defined by

$$SIR_i(p) \overset{\text{def}}{=} \frac{g_i p_i}{\sum_{j\neq i} g_j p_j + I + \nu}, \quad (5.1)$$

where $I > 0$ is the intercell interference power and $\nu > 0$ is the background noise power at the base station. For the single cell analysis, $I$ is considered to

\footnote{We will generalize the definition of a cell to a block in Section 5.3.}
be fixed. Suppose that mobile $i$ can utilize $K > 1$ discrete transmission rates $R_i \in \{ r^{(1)} < r^{(2)} < \ldots < r^{(K)} \}$ by varying its processing gain. Then, we can describe the bit-energy-to-interference-power-spectral-density ratio of mobile $i$ by

$$\left( \frac{E_{b,i}}{I_0} \right)_i = \frac{W}{R_i} \cdot SIR_i(p),$$

(5.2)

where $W$ denotes the spreading bandwidth. The realized data rate of a mobile depends on many factors such as receiver structure, user mobility, etc. Nevertheless, SIR is a reasonable measure to match effective data rates. To properly receive information at transmission rate $r^{(k)}$, mobile $i$ is required to attain an $SIR_i(p)$, not less than $\gamma^{(k)}$. Let $\gamma_i$ be the SIR-target of mobile $i$ such that $\gamma_i \in \{ \gamma^{(1)} < \gamma^{(2)} < \ldots < \gamma^{(K)} \}$, where each target element corresponds to each rate, respectively. In the sequel, we will use rate assignment and target assignment interchangeably. For setting the relationship among different $\gamma^{(k)}$, we assume a common target $\Gamma$ being the same for all users and rates. This ensures a constant BER regardless of the data rate. Then using (5.2), we can confine SIR-targets to the following relationship:

$$\frac{r^{(1)}}{\gamma^{(1)}} \cdot \frac{r^{(2)}}{\gamma^{(2)}} \cdot \ldots \cdot \frac{r^{(K)}}{\gamma^{(K)}} = \frac{W}{\Gamma}$$

(5.3)

This essentially describes $K$ points from a linear relation between transmission rate and SIR. Nonlinear relations can be considered by not keeping $\Gamma$ uniform over all users. With a certain assignment of SIR-targets, we can find the power vector that supports every mobile with the minimum power by solving the linear equation system (3.1), where $H = [h_{ij}]$ is an $N \times N$ matrix with elements

$$h_{ij} = \begin{cases} \gamma_i g_{ij}, & i \neq j \\ \eta_i, & i = j \end{cases}$$

and $\eta = (\eta_i)$ is an $N \times 1$ vector with components $\eta_i = \frac{2^{(I+\nu)}}{g_i}$. For a feasible system, the solution to (3.1) is found to be:

$$p_i^* = \frac{I + \nu}{g_i} \frac{\gamma_i}{1 + \gamma_i} \left[ 1 - \frac{1}{\sum_{j=1}^{N} \frac{\gamma_j}{\gamma_i + \gamma_j}} \right]$$

(5.4)

By Proposition 3.3, the spectral radius $\rho(H) < 1$ when $\sum_{j=1}^{N} \frac{\gamma_j}{\gamma_i + \gamma_j} < 1$. Since $p_i^*$ should also be less than or equal to $\bar{p}$ for all $1 \leq i \leq N$, we can express the $N$ constraints similar to [98]:

$$\sum_{j=1}^{N} \frac{\gamma_j}{1 + \gamma_j} \leq 1 - \max_{1 \leq j \leq N} \left[ \frac{\gamma_j}{\frac{\gamma_i}{\gamma_i + \gamma_j}} \right]$$

(5.5)
Constraint (5.5) can be regarded as an instantaneous user capacity region for given link conditions. Considering $\gamma_i/(1 + \gamma_i)$ as an effective bandwidth [109], the rightmost term in (5.5) can be interpreted as a relative utilization of the maximum achievable SIR\(^2\). We will assume that the maximum SIR, $g_j\bar{p}_j/(I + \nu)$, is available at the base station and that the feedback channel is error free and without any delay.

As there can be a large number of feasible rate combinations that give the same throughput for the cell, finding one which yields low total transmitted power is desirable, especially for multiclass cellular systems. In particular, we wish to design a scheme that finds solutions where the total throughput takes the maximum feasible value. Utilizing (5.3), (5.4) and (5.5), for a given throughput $\bar{\gamma}$ the power minimization problem can be formulated in terms of SIR-target assignment,

$$\min_{\gamma_j} \frac{I + \nu}{1 - \sum_{j=1}^{N} \frac{\gamma_j}{1 + \gamma_j}} \sum_{j=1}^{N} \frac{\gamma_j}{g_j(1 + \gamma_j)}$$ \hspace{1cm} (5.6)

s.t. \hspace{1cm} $\sum_{j=1}^{N} \gamma_j = \bar{\gamma}$ \hspace{1cm} (5.7)

$$\sum_{j=1}^{N} \frac{\gamma_j}{1 + \gamma_j} \leq 1 - \max_{1 \leq j \leq N} \left[ \frac{\gamma_j}{g_j\bar{p}_j/(I + \nu)} \right]$$ \hspace{1cm} (5.8)

$\gamma_j \in \Upsilon$ \hspace{1cm} (5.9)

where the set $\Upsilon = \{0, \gamma^{(1)}, \gamma^{(2)}, \ldots, \gamma^{(K)}\}$ contains the possible targets. When $\gamma_j = 0$, no transmission is performed for mobile $j$ and it does not contribute in (5.8). For a given $\bar{\gamma}$, we will mean by feasible target assignment, an SIR-target assignment that satisfies constraints (5.7)-(5.9). Given a feasible target assignment, the corresponding power value of each user can be computed by (5.4).

### 5.2.2 Greedy Rate Packing

From the structure of the objective function (5.6), it follows that when feasible SIR-targets are assigned, their order does not change the value of the denominator. However, for the sum in the numerator, the total transmitted power will be affected by which target is assigned to which mobile, due to the link gains. Obviously, the total transmitted power of a given cell can be minimized by assigning high targets to mobiles with high link gains, which we state in the proposition below.

\(^2\)The quantity $\gamma_i/(1 + \gamma_i)$ is also the SIR-target if own-interference is included. Rewriting (5.1), we have

$$\frac{g_{j}p_{j}}{\sum_{j}g_{j}p_{j} + I + \nu} = \frac{\gamma_j}{1 + \gamma_i}.$$
Proposition 5.1 Let us assume that \( g_1 \geq g_2 \geq \ldots \geq g_N \) and that there exists a feasible SIR-target assignment that satisfies (5.7)-(5.9). Among the feasible reassignments of the SIR-targets, the total power is minimized by reassigning them such that \( \gamma_1 \geq \gamma_2 \geq \ldots \geq \gamma_N \).

Proof: Among the feasible target reassignments, let us assume there exists an assignment that minimizes the objective function (5.6) but does not satisfy
\[
\gamma_1 \geq \gamma_2 \geq \ldots \geq \gamma_N. \tag{5.10}
\]
Then, we can find at least one pair of \( i \) and \( j \) (\( i < j \)) such that \( \gamma_i < \gamma_j \). Now, since \( g_i \geq g_j \), we find from (5.5) that a new feasible target assignment can be constructed by switching assignment \( \gamma_i \) and \( \gamma_j \). Since \( \frac{\gamma_i}{1+\gamma_i} > \frac{\gamma_j}{1+\gamma_j} \), we can define \( \frac{\gamma_i}{1+\gamma_i} + \Delta \), where the number \( \Delta \) is positive. For any reassignment, the denominator of (5.6) remains the same, therefore it is sufficient to compare
\[
\frac{\gamma_i}{g_i(1+\gamma_i)} + \frac{\gamma_j}{g_j(1+\gamma_j)} - \left( \frac{\gamma_i}{g_i(1+\gamma_i)} + \frac{\gamma_j}{g_j(1+\gamma_j)} \right) = \frac{\gamma_i}{g_i(1+\gamma_i)} + \frac{\Delta}{g_j} + \frac{\gamma_i}{g_j(1+\gamma_i)} - \frac{\gamma_i}{g_i(1+\gamma_i)} + \Delta \left( \frac{1}{g_j} - \frac{1}{g_i} \right) \geq 0.
\]
From the objective function (5.6), we have a contradiction that the new assignment has a lower objective function value than the optimal assignment. Thus for any \( i,j \) with \( g_i \geq g_j \), an optimal assignment must fulfill \( \gamma_i \geq \gamma_j \). \( \square \)

To keep the analogy to water-filling principles, we can equivalently order the mobiles by their link gains normalized with \( I + \nu \). Henceforth, we assume that the users are ordered such as \( \frac{g_1}{I+\nu} \geq \frac{g_2}{I+\nu} \geq \ldots \geq \frac{g_N}{I+\nu} \). In our case, they could also be ordered by their maximum SIRs, \( \frac{g_i}{I+\nu} \). The proposition says that we can reduce the total power of the mobiles while maintaining the same throughput, by reassigning SIR-targets in a decreasing manner. It should be pointed out though, that such a solution does not necessarily minimize the powers in a multicellular system. Since the proposition holds for any feasible value of \( \tilde{\gamma} \), an interesting task is to find a feasible target assignment that will achieve the maximal throughput while satisfying condition (5.10). For this purpose, we will, based on (5.5), suggest the following heuristic algorithm:

Greedy Rate Packing (GRP)
\[
\gamma_i = 0, \quad \forall i
\]
For \( i = 1 \) to \( N \) do
\[
\gamma_i^* = \max \left\{ \gamma_i \in \mathcal{Y} : \sum_{j=1}^{i} \frac{\gamma_j}{1+\gamma_j} \leq 1 - \max_{1 \leq j \leq i} \left[ \frac{\gamma_j}{1+\gamma_j} \right] \right\}
\]
\[
\gamma_i = \gamma_i^*
\]
The GRP procedure allocates the maximum feasible rate to each mobile, starting with the mobile with the highest link gain, or equivalently the maximum SIR. Further, it never reallocates rates of mobiles previously assigned, taking no more than \( N \) steps. It is easy to see that \( \gamma_i^* \geq \gamma_j^* \geq \ldots \geq \gamma_N^* \). However, the GRP is heuristic and does not necessarily find the solution to (5.6)-(5.9) but the result in the next section shows that high throughput solutions may be found.

### 5.2.3 Throughput Maximization

With the feasible target assignment generated from GRP denoted by \( \gamma_i^* \), we define \( \gamma^* = \sum_{i=1}^{N} \gamma_i^* \). A question at hand is how close the value \( \gamma^* \) is to the maximal throughput. This is generally difficult to answer, but for the special case where the upper bound on transmitter power is large enough compared to the interference plus noise, the uplink pole capacity is solely governed by the sum of the effective bandwidths

\[
\sum_{j=1}^{N} \frac{\gamma_j}{1 + \gamma_j} < 1. \tag{5.11}
\]

The following observation, which indicates that splitting a rate over several users consumes more effective bandwidth, can be made.

**Lemma 5.1** If \( \gamma \leq \sum_{j} \gamma_j \), \( 0 \leq \gamma_j < \gamma \) and \( \theta > 0 \), then \( \frac{\theta \gamma}{1 + \theta \gamma} < \sum_{j} \frac{\theta \gamma_j}{1 + \theta \gamma_j} \).

**Proof:** Since \( \frac{\theta \gamma}{1 + \theta \gamma} \) is a monotonously increasing concave function,

\[
\frac{\theta \gamma}{1 + \theta \gamma} \leq \frac{\sum_{j} \theta \gamma_j}{1 + \sum_{j} \theta \gamma_j} < \sum_{j} \frac{\theta \gamma_j}{1 + \theta \gamma_j}.
\]

Let us assume that the rates, and therefore the corresponding SIR-targets, are geometrically related as \( r^{(i+1)} = \mu r^{(i)} \) with \( 1 \leq i \leq K - 1 \), \( \mu > 1 \) and \( r^{(1)} > 0 \). Hence, it follows that \( r^{(i)} = \mu^{i-1} r^{(1)}, 1 \leq i \leq K \). Further, constrain \( \mu \) to be an integer, which is often the case in practical systems. The following result says that high rates achieve the most efficient resource allocation.

**Lemma 5.2** For a rate allocation that maximizes throughput subject to (5.11), each of the rates \( r^{(1)}, r^{(2)}, \ldots, r^{(K-1)} \) can be used at most \( \mu - 1 \) times.

**Proof:** Assume that \( r^{(i)} \), \( 1 \leq i < K \) is assigned to \( \mu \) users. Obviously \( \mu r^{(i)} = r^{(i+1)} \), so by using Lemma 5.1 with \( \theta = 1 \), the corresponding SIR-targets fulfill \( \frac{\gamma^{(i+1)}}{1 + \gamma^{(i+1)}} < \frac{\mu \gamma^{(i)}}{1 + \mu \gamma^{(i)}} \). Therefore, the same total throughput \( \mu r^{(i)} \) can be assigned with less utilization of the slack in the capacity constraint (5.11) by allocating
\( r^{(i+1)} \) to one single user. Hence, splitting the rate \( r^{(i+1)} \) over \( \mu \) or more users cannot maximize throughput. \( \square \)

Then, subject to (5.11) and the above assumptions, we can find that maximum throughput is obtained from the GRP.

**Proposition 5.2** If \( \bar{p} \to \infty \), \( r^{(i+1)} = \mu r^{(i)} \) with \( 1 \leq i \leq K - 1, \mu > 1 \) and \( r^{(1)} > 0 \), GRP gives the maximum throughput.

**Proof:** Assume a stepwise rate allocation procedure starting with its highest rates. Further, assume that at some point of this procedure, user \( i \) is assigned a rate \( R_i = r^{(k)}, k < K \) which is lower than the one of the GRP allocation, say \( R_i^* = r^{(p)}, k < p \leq K \). By Lemma 5.2, the rates \( r^{(1)}, r^{(2)}, \ldots, r^{(k)} \) in a throughput maximizing solution can be only used at most \( \mu - 1 \) times each. Therefore the remaining throughput of this rate allocation procedure is bounded by

\[
\sum_{j=i}^{N} R_j \leq (\mu - 1) \sum_{j=1}^{k} r^{(j)} = (\mu - 1) \sum_{j=1}^{k} \mu^{j-1} r^{(1)} = (\mu^{k-1}) r^{(1)} < r^{(k+1)} \leq R_i^* \leq \sum_{j=i}^{N} R_j^*,
\]

which means that, if \( R_i = r^{(k)} \) the total throughput of the remaining users will be lower than if \( R_i = r^{(p)} \). Therefore, to achieve a total throughput not less than the GRP solution, user \( i \) must be assigned a rate \( R_i > r^{(k)} \). Recursively applying this result, it follows that \( k \) must equal \( p \), the GRP assignment. Thus in any step, GRP assigns the optimal rate. \( \square \)

An interpretation of Proposition 5.2 is that, for an isolated or remote cell, where the intercell interference is small enough compared to the mobile’s maximum power, we can guarantee a maximum throughput solution from GRP.

### 5.2.4 Minimum Rate Requirements

So far, the focus has been exclusively on best effort services, with maximum BER requirements but without delay constraints. It is possible to utilize the concept of GRP for delay sensitive users with minimum rate requirements. For this purpose, a modified GRP procedure can be defined, aiming to support as many users as possible with their minimum rate, while maximizing the network excess capacity in a best effort manner. Henceforth, we consider \( \bar{r}^{(1)} \) to be the minimum rate requirement for all users.
GRP2
\[ \gamma_i = 0, \quad \forall i \]

**Step 1.**
For \( i = 1 \) to \( N \) do
\[ \gamma_i^* = \max \left\{ \gamma_i \in \{0, \gamma^{(1)}\} : \sum_{j=1}^{i} \frac{\gamma_j}{1+\gamma_j} \leq 1 - \max_{1 \leq j \leq i} \left[ \frac{\gamma_j}{1+\gamma_j} \right] \right\} \]
\[ \gamma_i = \gamma_i^* \]

**Step 2.**
For \( i = 1 \) to \( N \) do
\[ \gamma_i^* = \max \left\{ \gamma_i \in \mathcal{Y} : \sum_{j=1}^{N} \frac{\gamma_j}{1+\gamma_j} \leq 1 - \max_{1 \leq j \leq N} \left[ \frac{\gamma_j}{1+\gamma_j} \right] \right\} \]
\[ \gamma_i = \gamma_i^* \]

The GRP2 has the property that it supports as many connections as possible with the minimum rate.

**Proposition 5.3** The GRP2 procedure maximizes the number of users supported with the minimum rate.

**Proof:** It suffices to show that the proposition holds after Step 1. Let \( \mathcal{M} \) denote a partition of the users such that users in \( \mathcal{M} \) obtain rate \( \gamma^{(1)} \) while users not in \( \mathcal{M} \) obtain rate zero. Let \( m = |\mathcal{M}| \) denote the number of supported users. In this case, the inequalities (5.5) can be expressed in the form:
\[ m \frac{\gamma^{(1)}}{1+\gamma^{(1)}} + \frac{\gamma^{(1)}}{g+\nu} \leq 1 \quad (5.12) \]
Since \( g_1 \geq g_2 \geq \ldots \), we must have that for \( i = 1, 2, \ldots, m \), \( g_i \geq \min_{j \in \mathcal{M}} g_j \). It follows that (5.12) must hold \( i = 1, 2, \ldots, m \) and that Step 1 for GRP2 will assign nonzero rate to at least the first \( m \) users. Hence Step 1 of GRP2 supports at least as many users as any other assignment. \( \square \)

### 5.2.5 Downlink GRP

For the downlink case, the equation system (3.2) takes this form
\[ \gamma_i = \frac{g_i p_i}{\sum_{j \neq i} \theta g_i p_j + I_i + \nu} \quad (5.13) \]
where \( I_i \) is the intercell interference at mobile \( i \) and \( \theta \in [0, 1] \) is the orthogonality factor. Straightforwardly, it can be shown that the minimum power solution equals,
\[ p_i^* = \frac{\gamma_i^2}{1 + \theta \gamma_i^2} \left( \sum_{j=1}^{N} \frac{g_j I_i + \nu}{1 + \theta g_j} \right). \quad (5.14) \]
If in (5.13) $\theta$ depends on $i$, the power solution is given by (5.14) with $\theta = \theta_i$. In contrast to the uplink, usually the sum of the users’ powers in the cell is limited, rather than the maximum power of each user. If the sum of all transmitter powers in the cell is considered to be limited to $\bar{P}$, based on (5.14) the following feasibility constraint can be derived

$$\sum_{j=1}^{N} \frac{\theta_i g_j}{1 + \theta_i g_j} \left( \frac{I_j + \nu}{g_j P} + 1 \right) \leq 1$$

(5.15)

which could be used for a downlink greedy rate packing procedure. If (5.15) holds, it follows from (5.14) that $p_i > 0$ and $\sum_{i=1}^{N} p_i \leq \bar{P}$. Studying (5.15), it suggests that also in this case, to minimize the sum of powers, high rates should be assigned if $g_j/(I_j + \nu)$ is large. Here it is required that $g_j \bar{P}/(I_j + \nu)$, the maximum achievable SIR for user $j$, can be perfectly estimated at the base station. We will not elaborate further on the downlink in this chapter. However, we can generalize Lemma 5.2 to cover also the condition (5.15), if we consider users to be ordered such that $g_1/(I_1 + \nu) \geq g_2/(I_2 + \nu) \geq \ldots \geq g_N/(I_N + \nu)$. Downlink throughput optimality can then be shown using the same technique as for Proposition 5.2. In the downlink case, (5.15) is a true single-cell feasibility condition and optimality does not require an assumption that $\bar{P} \to \infty$. Hence, GRP will achieve maximum throughput in the downlink, given the same geometric rate relations as in Section 5.2.3.

### 5.3 Iterative Power Control

In this section, the GRP will be applied to a multicellular system. As the intercell interference will vary after each round of GRP assignment, we approach the problem by an iterative power/rate control algorithm. Consider the following algorithm for the combined rate and power assignment in a multicellular system.

**Generalized Selective Power Control (GSPC)** Each cell updates its SIR-target assignment for the users in the cell using GRP and correspondingly updates the power of each user using (5.14) in every iteration.

So far, we have assumed that $N$ is the number of users in a given cell and that they all require the same maximum BER. Now, let us assume that the users in a cell can be partitioned into a number of subsets, referred to as blocks, and consider $N$ to be the number of users in a given block. Within a block we require that $\Gamma$ is constant. Consequently, users within a block are assumed to require an identical service. In this setting, $I_i$ denotes the interference coming from outside of the given block. Then, our discussion so far is still valid even if we replace the terminology “cell” by “block”. For example, in the GSPC description above, we may use GSPC for the block-wise rate and power update. Therefore, we will
hereafter use the term block, instead of cell. For the special case where there is just a user in each block, GSPC is reduced to the following fully distributed SPC algorithm for combined rate and power control, which was suggested in [64],

\[
p_i(n+1) = \max_{\gamma_i \in \mathcal{T}} \left\{ \frac{\gamma_i}{SIR_i(p(n))} p_i(n) \cdot 1\{ \sum_{j=1}^{N} \frac{\gamma_j}{\gamma_j + 1} \} \right\},
\]

(5.16)

where

\[
1_A = \begin{cases} 1, & \text{if } A \text{ occurs} \\ 0, & \text{otherwise.} \end{cases}
\]

is the indicator function of the event \( A \). The SPC selects independently for each mobile, the maximum feasible SIR-target that it can achieve against the present interference. The GSPC on the other hand, can be interpreted as a locally centralized algorithm which requires that the power and rate assignments are executed at the base station and delivered to the mobile. Then the base station to which mobile \( i \) is connected, could estimate \( g_i/(I + \nu) \) needed for the GRP.

Let us define individual sets of transmission rates such that for user \( i \), the set \( \mathcal{T}_i \subseteq \mathcal{T} \) contains its SIR-targets and let \( \gamma_i = \sup \mathcal{T}_i \) be its maximum SIR-target. When the system is feasible for the targets \( \gamma_i \), it was given in [64] that the SPC algorithm converges to a state where every active mobile is supported with its maximum rate and the sum of the powers is minimized. For GSPC, we can prove similar properties which will be numerically verified in Section 5.5.

**Proposition 5.4** For a feasible system, GSPC converges to the power vector where every mobile is supported with its maximum rate and the sum of powers is minimized.

**Proof:** Let us first consider an unconstrained power control algorithm given by

\[
p_i(n+1) = \mathcal{T}_i(p(n)) = \frac{I(p(n)) + \nu \gamma_i}{\gamma_i} \quad \frac{1}{1 + \gamma_i} - \frac{1}{\sum_{j=1}^{N} \gamma_j},
\]

(5.17)

where \( p(n) \) and \( I(p(n)) \) denote the power vector and intercell (interblock) interference at iteration \( n \). It has been proven (see, e.g., Chapter 5 [56]) that the above algorithm converges to the power vector that achieves \( \gamma_i \) for every user using minimum total transmitter power, if such a power vector exists\(^3\). Now let us consider the generalized constrained power control algorithm suggested in Chapter 4,

\[
\tilde{p}_i(n+1) = \tilde{\mathcal{T}}_i(p(n)) \left\{ \begin{array}{ll}
\mathcal{T}_i(p(n)), & \text{if } \mathcal{T}_i(p(n)) \in \mathcal{P}_i \\
\tilde{p}_i(n), & \text{if } \mathcal{T}_i(p(n)) \not\in \mathcal{P}_i
\end{array} \right.
\]

(5.18)

\(^3\)The proof of convergence in [56] was carried out with the technique of matrix splittings. Convergence could alternatively be performed by verifying consistency and show that (5.17) is a standard power control algorithm.
where \(0 \leq \hat{p}_i(n) \leq \bar{p}_i\). By letting \(\gamma_i = \tilde{\gamma}_i\), we can regard GSPC as a special case of the above algorithm (5.18), where \(\hat{p}_i(n)\) is utilized if lower targets are selected. Since Proposition 4.3 states that (5.18) converges, so will GSPC. \(\Box\)

**Proposition 5.5** For a feasible system, SPC gives the slowest asymptotic rate of convergence among the class of GSPC algorithms.

**Proof:** Consider a power control algorithm given by

\[
p_i(n + 1) = \bar{T}_i(p(n)) = \frac{\gamma_i}{SIR_i(p(n))} p_i(n). \tag{5.19}\]

It is well known (see, e.g., Chapter 3) that the above iteration converges to the power vector, say \(p^*\) that achieves \(\gamma_i\) for every user \(i\) with minimum total transmission power, if such a power vector exists. Define the error vector as

\[
e_1(n) = \mathcal{T}(p(n)) - p^*, \quad n = 0, 1, \ldots \tag{5.20}\]

In the same way, we can define an error vector for the algorithm (5.17) by

\[
e_2(n) = \mathcal{T}(p(n)) - p^*, \quad n = 0, 1, \ldots \tag{5.21}\]

For a given number \(\delta > 0\), let \(n_1(\delta)\) and \(n_2(\delta)\) denote the minimum number of iterations that give \(\|e_1(n)\| \leq \delta \cdot \|e_1(0)\|\) and \(\|e_2(n)\| \leq \delta \cdot \|e_2(0)\|\), respectively. It has been proven (see, e.g., Proposition 5.2 [56]) that for a suitable norm, \(n_1(\delta) > n_2(\delta)\) as \(\delta\) approaches zero, which means the mapping \(\mathcal{T}\) is asymptotically faster than \(\mathcal{T}\). It is easy to see that SPC is a special case of (5.18), equivalently expressed as

\[
\hat{p}_i(n + 1) = \hat{T}_i(p(n)) = \begin{cases} 
\bar{T}_i(p(n)), & \text{if } \bar{T}_i(p(n)) \in \mathcal{P}_i \\
\hat{p}_i(n), & \text{if } \bar{T}_i(p(n)) \notin \mathcal{P}_i
\end{cases} \tag{5.22} \]

where \(\hat{p}_i(n) \in \mathcal{P}_i\). Asymptotic rate of convergence is then reduced to if \(\bar{T}\) (GSPC) is asymptotically faster than \(\hat{T}\) (SPC). Since both \(\bar{T}\) and \(\hat{T}\) will converge, each will follow the dynamics of \(\mathcal{T}\) and \(\mathcal{T}\), respectively, as \(\delta\) approaches zero. Therefore, we can say that \(\mathcal{T}\) is asymptotically faster than \(\hat{T}\). Applying Proposition 5.2 in [56], it follows that considering only one user per block (SPC) is the worst choice with respect to asymptotic average rate of convergence. \(\Box\)

The latter proposition is based on results from [56], saying that the more users the blocks contain, the faster the asymptotic rate of convergence becomes. Larger block size can be interpreted as a higher degree of centralization.
Figure 5.1: A two-user example when the maximum rate cannot be supported for both users simultaneously within the power constraint (dashed). The two solutions \( p_a \) and \( p_b \) give the same system throughput but the sum of powers is minimized for \( p_a \).

### 5.4 Power Control with Gradual Rate Removal

If the maximum targets are set too high, the system becomes infeasible, and convergence of GSPOC is not guaranteed. Even if high throughput may be obtained, infeasibility could cause unstable behavior where the power levels reach unnecessarily high levels and therefore is a problem. Consider the snapshot of a two-user case in Fig. 5.1, where each user can utilize any of two rates. In this figure, we assume that the users are connected to different base stations so that there is only one user in each block. The straight lines represent the linear equation system (3.1) with different rate assignments, i.e., \( h_{ij}^{(t)} = \gamma_i^{(t)} \frac{2\pi}{n} \)

and \( \eta_i^{(t)} = \gamma_i^{(t)} \frac{2\pi}{n} \). As the maximum power constraint prohibits both users from using the maximum rate simultaneously, it is clear that the system throughput is maximized when only one user uses the maximum rate while the other is using the lower rate. Obviously there are two such solutions \( p_a \) and \( p_b \) but as the total used power is less for \( p_a \), we wish to allocate those targets and make GSPOC converge to that point. An underlying idea here is to constrain some

\footnote{In [56] it is shown that the spectral radius of the iteration matrix resulting in \( \mathbf{P} \) is smaller than the one from the splitting resulting in \( \mathbf{P} \).}
users to lower its maximum rate so that the induced system becomes feasible. This can be done distributively by using the transmitter removal concept previously applied in fixed rate systems in Chapter 4, which will be extended to rate removal in this chapter. As we stated in the introduction, maximizing the system throughput in a multicellular system with maximum transmitter powers is NP-complete. Therefore, convergence to the optimal power and rate assignment cannot be guaranteed within a short computational time, which calls for a heuristic algorithm. Based on Proposition 5.1, which states that users with favorable link gains should utilize high data rates, we can construct the following rate removal algorithm.

Let $\Upsilon_i(n)$ be the set of possible SIR-targets at iteration $n$ for user $i$ and $\bar{\gamma}_i(n) = \sup \Upsilon_i(n)$ be its maximum target. Define in each block at each iteration, a set of mobiles $\mathcal{V}(n) = \{i : \gamma_i < \bar{\gamma}_i(n), \bar{\gamma}_i(n) > 0\}$. This set contains the mobiles not being assigned their maximum achievable rate during the GRP. Then consider:

**Gradual Rate Removal-GSPC (GRR-GSPC)** After each iteration of GSPC and for each block; if $\mathcal{V}(n) \neq \emptyset$, then with a given probability $\delta$, remove one rate from user $i_0$ such that $\Upsilon_{i_0}(n+1) = \Upsilon_{i_0}(n) \backslash \bar{\gamma}_{i_0}(n)$ where $i_0 = \arg\min_{i \in \mathcal{V}(n)} \gamma_i / (1 + \nu)$.

In GRR-GSPC, if in each block, there exists any user that has been assigned a rate less than its maximum rate, then the maximum rate of the user with the lowest link gain in the block will be decreased (one-step down) with a given probability $\delta$. This will shape the rate distribution such that the system will eventually become feasible. Based on Theorem 6 in [5], it can be noted that for $\delta > 0$, GRR-GSPC will converge to a stationary feasible power vector where all remaining users are supported with their resulting maximum rate. The corresponding concept of GRR-SPC yields:

$$p_i(n + 1) = \max_{\gamma_i \in \Upsilon_i(n)} \left\{ \frac{\gamma_i}{SIR_i(p(n))} p_i(n) \cdot 1 \left\{ \frac{\gamma_i}{SIR_i(p(n))} p_i(n) \in P_i \right\} \right\}$$

If $\frac{\gamma_i(n)}{SIR_i(p(n))} p_i(n) \not\in P_i$ and $\bar{\gamma}_i(n) > 0$, then $\Upsilon_i(n+1) = \Upsilon_i(n) \backslash \bar{\gamma}_i(n)$ with a given probability $\delta$.

For GRR-SPC, where each user is its own block, there exists only one possible $i = i_0$ and hence it is a special case of GRR-GSPC and convergence applies to it as well. The reason for adopting a probabilistic removal procedure, $\delta < 1$, is that if too many users simultaneously remove their maximum rate, capacity may get wasted, since removing only a subgroup of these rates might give a feasible system. By defining GSPC2 as GSPC but with GRP2 instead of GRP, we have the same type of block distributed algorithm and the rate removal procedure is applicable also to GSPC2.

\*In Chapter 4 and [5], GRR denoted Gradual Removals Restricted
5.5 Numerical Results

The algorithms are evaluated by simulations in a system consisting of 19 hexagonal cells with centrally located base stations using omni-directional antennas. The cell radius is set to 1 km and a wrap-around technique is used. Evaluation is performed in the DS-CDMA system described in Chapter 2 and illustrated in Fig. 1.2. The cell radius is set to 1 km and a wrap-around technique is used along with omni-directional antennas. For a given instance, 10 mobiles/cell are generated, the locations of which are uniformly distributed over the cells. At any given instance, the link gain is modeled by $g_{ij} = S_i r_{ij}^4$, where $S_i$ is the shadow fading factor and $r_{ij}$ is the distance between base $i$ and mobile $j$. The shadow fading factor is generated from a lognormal distribution with $\mu_S = 0$ dB, and $\sigma_S = 8$ dB, where the $S_i$'s are mutually independent. The base receiver noise is taken to be $n_i = 10^{-15}$ W and the maximum mobile power is set to unity. The initial power for each mobile is randomly chosen from the interval $[0,1]$. Each user is assigned to the base station that provides the lowest signal attenuation. We consider the uplink of a system that has a spreading bandwidth $W = 1.2288$ MHz and assume that the radio link can support four transmission rates, $r^{(k)} = 9.6 \cdot \frac{1}{2^{k-1}}$ kbps ($k = 1, 2, 3, 4$) and the zero rate if necessary. As in
(5.3), the required minimum SIR is assumed to be $\gamma^{(t)} = \gamma \cdot \frac{1}{10}$ for supporting each 9.6 \cdot \frac{1}{10} kbps. Different values of $\gamma$ can be considered for representing different congestion scenarios. Two values $\gamma = -14$ dB and $\gamma = -5$ dB are set to represent low and high system loads respectively. Adding the processing gain, this is equivalent to 7 and 16 dB as the corresponding $\Gamma$ values. In general the effective transmission rate for a given channel quality is dependent on the receiver structure. In our simulation, we use a pessimistic effective rate. That is, either the attains SIR is above the required threshold and the throughput is the selected rate or otherwise, no data is received. In each iteration, the average throughput per user, average transmitter power per user and outage probability are used as performance measures. The average throughput per user is computed by dividing the total throughput by the total number of users in the system, i.e., in this case 100. The average power is obtained similarly. The outage probability is here defined as the fraction of users not receiving any data, which occurs when the obtained SIR is lower than the selected target or the zero rate is selected. In the rate and power assignment, a slight fade margin of 0.5 dB was added, implying that an outage event is declared if the received $E_b/I_0$ is below 6.5 and 15.5 dB, respectively. We assume that all users in a cell constitute a block in GSPC. To evaluate performance, we have taken 100 independent instances of mobile locations and shadow fading. For each instance, we have performed the algorithms until the iteration number reached 500, a number that yielded stable solutions. In Fig. 5.2, where all users can be supported with the maximum data rate, Proposition 5.5 is illustrated by that the average of the power vector reaches its steady state faster for GSP than SPC. When $\Gamma = 16$ dB, the system becomes infeasible and the throughput curves in Fig. 5.3 indicate an oscillating behavior. It can be found by studying data from one single snapshot at a time, that SPC has a much more oscillative behavior for infeasible situations which also gives larger throughput variations than of GSPC. Clearly the throughput penalty of GSPC2 for trying to guarantee the minimum rate is seen as well. As a benefit from the faster convergence, the GSPC algorithms also reach their maximum throughput levels much faster than the SPC.

In the first columns of Tables 5.1-5.3, the resulting values after 500 iterations are contained for the case without removal, $\delta = 0$. We see that GSPC gives about 9 % higher throughput than SPC and at the same time about half the energy consumption. The price of the greedy behavior of GSPC is the resulting higher outage probability. In the other columns, the result of applying the GRR procedure can be seen for the respective $\delta$ values. When applying the rate removal, the probability $\delta$ should be chosen moderately so that the set of rates removed in each iteration becomes relatively small. Too aggressive a removal strategy might eventually result in an underutilization of the spectrum. For GRR-GSPC, rates may be removed from mobiles within the block that are not transmitting, so the removal probability may be chosen much larger than for GRR-SPC. There is obviously a tradeoff between choosing a high removal probability, making the system feasible rather quickly, which would give fast
Figure 5.3: Average throughput per user of GSPC, SPC and GSPC2 under heavy load, $E_b/I_0 = 16$ dB, without gradual rate removal.

convergence but low throughput and power, and a lower one which would give the opposite. From Tables 5.1-5.3 we draw the following conclusions: The GRP procedure seems to be superior in offering the highest throughput. The penalty from guaranteeing a minimum rate, like GSPC2, is that it requires much more power and lowers the total throughput but reduces outage. The gain from GRR is mainly that the powers will be decreased. Thus, even though high throughput can be maintained for infeasible situations, infeasibility is undesirable from an energy consumption point of view. The SPC is more sensitive to the choice of $\delta$ in terms of throughput losses.

5.6 Concluding Remarks

We considered a DS-CDMA system where every user had a discrete set of predetermined SIR-targets corresponding to different transmission rates. Since there are many combinations of transmission rates giving the same system throughput, we formulated the problem of minimizing the total transmission power while trying to maximize the throughput. For a single cell, we suggested that high data rates should be allocated to users with high link gains and low interference. That is, the area at which the service can be provided, will be located close to the radio access point. We may think of this as a spatial dif-
ferentiation. However, if users are mobile, all users could expect this service level in average. The characteristics of the single cell solution were used as a building block for a simple greedy rate and power assignment algorithm, called GSPC, for a multicellular system. The aim was to maximize the throughput but minimum rate requirements could be included as well. Numerical results showed that GSPC can attain higher throughput while the transmitter powers are considerably decreased, compared to a reference algorithm SPC [64]. Furthermore, GSPC has a merit of fast convergence and can significantly decrease the energy consumption, when it is combined with rate removal. The cost of this concept is a slight increase of overhead information, as it is not fully distributed but the gain seems to be large in terms of energy conservation as well as increased system capacity. We exemplified that, when trying to assure each user a minimum rate, both system throughput and energy conservation will be sacrificed for the fairness among the users. Further studies are needed on how to apply gradual removal for time varying channels. Somehow, some rates must be admitted again if the channel conditions have changed significantly.
Table 5.1: Data for GRR-GSPC at iteration 500 for different removal probabilities $\delta$. The first row is average throughput per user [kbps], the second row is the average power per user [W] and the last row is the outage probability.

<table>
<thead>
<tr>
<th>$\delta$</th>
<th>0</th>
<th>0.1</th>
<th>0.2</th>
<th>0.3</th>
<th>0.4</th>
<th>0.5</th>
<th>1</th>
</tr>
</thead>
<tbody>
<tr>
<td>Throughput</td>
<td>3.56</td>
<td>3.57</td>
<td>3.56</td>
<td>3.57</td>
<td>3.56</td>
<td>3.56</td>
<td>3.56</td>
</tr>
<tr>
<td>Power</td>
<td>0.18</td>
<td>0.07</td>
<td>0.02</td>
<td>0.02</td>
<td>0.02</td>
<td>0.02</td>
<td>0.02</td>
</tr>
<tr>
<td>Outage</td>
<td>0.56</td>
<td>0.55</td>
<td>0.55</td>
<td>0.55</td>
<td>0.55</td>
<td>0.55</td>
<td>0.55</td>
</tr>
</tbody>
</table>

Table 5.2: Data for GRR-GSPC2 at iteration 500 for different removal probabilities $\delta$.

<table>
<thead>
<tr>
<th>$\delta$</th>
<th>0</th>
<th>0.1</th>
<th>0.2</th>
<th>0.3</th>
<th>0.4</th>
<th>0.5</th>
<th>1</th>
</tr>
</thead>
<tbody>
<tr>
<td>Throughput</td>
<td>2.65</td>
<td>2.61</td>
<td>2.58</td>
<td>2.57</td>
<td>2.56</td>
<td>2.56</td>
<td>2.53</td>
</tr>
<tr>
<td>Power</td>
<td>0.27</td>
<td>0.14</td>
<td>0.03</td>
<td>0.03</td>
<td>0.02</td>
<td>0.02</td>
<td>0.02</td>
</tr>
<tr>
<td>Outage</td>
<td>0.02</td>
<td>0.02</td>
<td>0.02</td>
<td>0.02</td>
<td>0.02</td>
<td>0.02</td>
<td>0.02</td>
</tr>
</tbody>
</table>

Table 5.3: Data for GRR-SPC at iteration 500 for different removal probabilities $\delta$.

<table>
<thead>
<tr>
<th>$\delta$</th>
<th>0</th>
<th>0.02</th>
<th>0.04</th>
<th>0.06</th>
<th>0.08</th>
<th>0.1</th>
<th>1</th>
</tr>
</thead>
<tbody>
<tr>
<td>Throughput</td>
<td>3.26</td>
<td>3.34</td>
<td>3.33</td>
<td>3.32</td>
<td>3.32</td>
<td>3.21</td>
<td>2.95</td>
</tr>
<tr>
<td>Power</td>
<td>0.36</td>
<td>0.21</td>
<td>0.11</td>
<td>0.08</td>
<td>0.05</td>
<td>0.04</td>
<td>0.0005</td>
</tr>
<tr>
<td>Outage</td>
<td>0.39</td>
<td>0.36</td>
<td>0.36</td>
<td>0.36</td>
<td>0.36</td>
<td>0.36</td>
<td>0.35</td>
</tr>
</tbody>
</table>
Chapter 6

Joint Power Control and Intracell Scheduling

The performance of DS-CDMA systems highly depends on the success in managing interference arising from both intercell and intracell transmissions. Interference management in terms of power control for real-time data services has been widely studied and has shown to be a crucial component for the functionality of such applications. In this chapter we consider the problem of supporting downlink non-real time data services, where in addition to the power control, there is also the possibility of controlling the interference by means of transmission scheduling. One such decentralized schedule is to use time division so that users transmit in a one-by-one fashion within each cell. We show that this has merits in terms of energy conservation and increased system capacity. We combine this form of intracell scheduling with a suggested distributed power control algorithm for the intercell interference management. We address its rate of convergence and show that the algorithm converges to a power allocation that supports the non-real time data users using the minimum required power while meeting requirements on an average data rate.

6.1 Introduction

In the DS-CDMA cellular system, one adopted multiple access method for future PCSs [32], the total received interference can be divided into intracell interference, originating from simultaneous transmissions within the cell and intercell interference caused by terminals transmitting in other cells. To alleviate the impact of intercell and intracell interference, transmitter power has turned out to be an important controllable resource for balancing the desired signal- and interference power at the receiver, rendering in increased capacity and lower energy consumption. Much work has been performed in the area of designing
quality based power control algorithms for finding the minimum power allocation when users require fixed data rates at a given target SIR, see Chapter 1 and 4. Typically this direction sees applications for voice services (see [95, 122] and references therein). The optimal power solution for this problem has been shown to be Pareto optimal [83], i.e., minimizing each user’s transmitter power also minimizes the total sum power. Integrated over time, this solution also minimizes the energy consumption. However, the requirement on a constant data rate and thus SIR, prohibits any interruption of the data stream and therefore the degree of freedom to co-ordinate the transmissions in time is limited. Herein, we consider the problem to support a given set of nonreal time service users with a certain required average data rate with minimum energy consumption obtained via a minimum power solution. With average data rate, we mean that a least amount of data has to be delivered over a specified time interval, a more relaxed constraint than a fixed data rate, which increases the options of adaptively planning the transmissions. One scheme that does not require any co-ordination between base stations, is to plan the transmissions independently in a cell-wise manner. Thus, in addition to power control, by utilizing the possibility of scheduling the transmissions within the cell, the intracell interference can be efficiently avoided. Since DS-CDMA systems are considered to be interference limited, less interference allows for reduced transmission power, in turn increasing the capacity and lowering the energy consumption. Hence it suggests that, if the same amount of data can be delivered, the impact will be higher energy efficiency since no power is wasted on combating the intracell interference.

In the chapter, we consider efficient use of the radio resources by exploiting the delay tolerance of nonreal time service users. Particularly we elaborate on the one-by-one scheduling (TDMA) for the downlink transmission in a DS-CDMA system. A crucial component is quality based power control but rather than focusing on minimum power solutions as in Chapter 4, 5 and [83,95,122], we will direct this work toward minimum energy solutions, a more natural measure for an average data rate seen over a time interval. The main focus herein is to offer a given average data rate with minimum energy consumption while allowing intracell scheduling, which is a form of time multiplexing. We show that intracell scheduling in terms of time division within the cells merits of higher energy efficiency and system capacity than continuous transmission at a fixed data rate. For finding the optimal powers, a distributed power control algorithm is suggested and proven to drive the transmit power to the minimum power level that guarantees the average data rate to the nonreal time users. By minimizing their power usage, interference is minimized and also more of the cell’s available transmitter power is left for supporting real time services or more nonreal time users. With strong resemblance to the continuous transmission problem, the power allocation can be found from a linear equation system, as in (3.2) but expressed in the cell-site powers. It is found that this power solution minimizes the energy consumption. Moreover, it is shown that the power control can converge at a faster rate in the average data rate case than for the fixed data
rate case. Based on the spectral radius of the power control iteration matrix, we can identify a capacity gain of the proposed power control/scheduling concept.

6.2 Minimum Energy Problem

6.2.1 Refined System Model

For treatment of the downlink, we refine the general system model of Chapter 2 as follows. Consider the downlink of a DS-CDMA cellular system with $M$ base stations where $N$ nonreal time service users can access a common frequency channel. Let user $i$ be connected to base station $b_i$ and the set $B_k = \{i: b_i = k, 1 \leq i \leq N\}$ contain the users connected to base station $k$. Consider a time instant where the link gain between base station $j$ and user $i$ is stationary and given by $g_{ij}$. Denote by $\mathbf{p} = (p_i)$ the $N \times 1$ downlink powers dedicated to the users. If $i \in B_k$, the SIR for user $i$ can alternatively be written as

$$SIR_i(\mathbf{p}) \overset{\text{def}}{=} \frac{g_{ik} p_i}{\sum_{j \in B_k} \theta_{ij} g_{ik} p_j + I_i(\mathbf{p}) + \nu_i}$$  \hspace{1cm} (6.1)

where $I_i(\mathbf{p})$ is the intercell interference for user $i$ and $\nu_i$ is the thermal noise. The quantity $\theta_{ij} \in (0, 1]$ is the normalized cross-correlation between $p_i$ and $p_j$ at the receiver of user $i$. In this chapter, we exclude $\theta_{ij} = 0$, since intracell scheduling would in that ideal case be obsolete. Define $\mathbf{P} = (P_k)$ to be the $M \times 1$ vector of base station powers where $P_k = \sum_{i \in B_k} p_i$. We assume that transmission rates are adapted by varying the processing gain and that the effective data rate of user $i$ is given by the continuous and linear relation

$$R_i = \frac{W}{\Gamma_i} SIR_i(\mathbf{p}).$$  \hspace{1cm} (6.2)

In (6.2), $W$ is the common spreading bandwidth and $\Gamma_i$ a required bit-energy-to-noise-spectral-density ratio. Therefore, $R_i$ is achieved with a tolerable maximum bit error probability which is assumed to be a function of $\Gamma_i$, the desired $E_b/K_0$. Due to hardware constraints, the output power of the power amplifier is limited. In the downlink, the power constraint is typically related to the sum of the cell’s power rather than individual powers of the users. Therefore, if the cells have the maximum power constraints given by a vector $\mathbf{P}$, any feasible power assignment must then fulfill $0 \leq \mathbf{P} \leq \mathbf{\bar{P}}$. The minimum power assignment for achieving the data rates $R_i$, is obtained by putting (6.1) and (6.2) on matrix form. Hence, solving the linear equation system, which has the same form as (3.2),

$$\mathbf{(I - H)} \mathbf{p} = \mathbf{\eta}$$  \hspace{1cm} (6.3)
where $\mathbf{H} = [h_{ij}]$ denotes an $N \times N$ matrix with elements

$$h_{ij} = \begin{cases} \frac{R_i}{W} \theta_{ij}, & i \neq j, b_i = b_j \\ \frac{R_i}{W} \frac{g_{ij}}{g_{ii}}, & i \neq j, b_i \neq b_j \\ 0, & i = j \end{cases}$$

and $\boldsymbol{\eta} = (\eta_k)$ is an $N \times 1$ vector with $\eta_k = \frac{R_i}{W} \frac{P_k}{g_{ii}}$, giving the componentwise smallest power vector. In this notation, the indices $j$ may correspond to the same physical base station.

### 6.2.2 Problem Definition

Assume now that there exists a specified scheduling interval, $0 \leq t \leq T_s$, over which a nonreal time service user $i$ requires a minimum average data rate $R_i$. That is, at least $T_s R_i$ bits have to be received over the scheduling interval at the quality specified by $\Gamma_i$. Consider the following problem of finding the transmission powers during $T_s$ that minimize the total energy consumed.

$$\min_{\mathbf{p}(t)} \sum_{i=1}^{N} \int_0^{T_s} p_i(t) \, dt \quad \text{(6.4)}$$

subject to

$$\frac{1}{T_s} \int_0^{T_s} r_i(t) \, dt = R_i, \quad \forall i \quad \text{(6.5)}$$

$$\sum_{i \in B_k} p_i(t) = P_k, \quad \forall k, 0 \leq t \leq T_s \quad \text{(6.6)}$$

$$p_i(t) \geq 0, \quad \forall i, 0 \leq t \leq T_s \quad \text{(6.7)}$$

$$P_k \leq \bar{P}_k, \quad \forall k \quad \text{(6.8)}$$

The instantaneous data rate is assumed to follow

$$r_i(t) = \frac{W}{\Gamma_i} \cdot SIR_i(p(t)), \quad \text{(6.9)}$$

the same linear relation as in (6.2). Thus we seek the optimum stationary base station power allocation which delivers the data within the interval $T_s$ with minimum energy. Replacing constraint (6.6) with $p_i(t) = p_i$, the minimum energy problem reduces to the minimum power assignment problem for continuous transmission considered in [55, 83, 122], that is (6.3). Constraint (6.6) also restricts the problem formulation from the use of intercell coordination by requiring all base stations to transmit with constant powers over the whole scheduling interval.

### 6.3 Intracell Scheduling

As the service requirement is specified only in terms of an average rate over a certain interval, described by constraint (6.5), it opens possibilities of more flexible use of the spectrum, e.g., to schedule transmissions. One possible strategy
to provide the average data rate is continuous transmission with the rate \( R_i \), i.e., reducing the problem to (6.3). Another way would be to transmit with a higher rate but for a shorter time. This may not come for free though. When a user utilizes a higher data rate, a high transmission power might be needed to compensate for the decreased processing gain, which causes increased interference for other users. Alternatively a user could transmit at a higher rate using the same power, if the experienced interference can be decreased by transmission scheduling. Our approach is to decrease the interference by proper scheduling. To elaborate on this further, let us use the following definitions similar to those in [132]:

**Definition 6.1** A rate vector \( \mathbf{R} = (R_1, R_2, \ldots, R_N) \) is instantaneously achievable if there exists a feasible power vector \( \mathbf{p} \) such that \( R_i \leq \frac{P_i}{P_k} SIR_i(\mathbf{p}) \) for all \( 1 \leq i \leq N \).

**Definition 6.2** A rate vector \( \mathbf{R} = (R_1, R_2, \ldots, R_N) \) is achievable in average if it can be expressed as \( \mathbf{R} = \sum_i \Phi_i \mathbf{R}^{(i)} \) where \( \sum_i \Phi_i \leq 1 \) and \( \Phi_i \in (0, 1) \).

To interpret Definition 6.2, suppose \( T_s R_i \) bits are transmitted during an interval of length \( T_s \) using a sequence of rate vectors \( \{\mathbf{R}^{(i)}\} \). Since \( T_s \mathbf{R} = T_s \sum_i \Phi_i \mathbf{R}^{(i)} \), \( \Phi_i \) represents the fraction of the scheduling interval \( T_s \) the rate vector \( \mathbf{R}^{(i)} \) is used. Now, consider rate vectors \( \mathbf{R}^{(i)} \), such that not more than one user per cell has nonzero rate simultaneously. With these vectors, we can define intracell one-by-one scheduling as follows:

**Definition 6.3** If transmission is performed using a sequence of rate vectors \( \mathbf{R}^{(i)} \), where for every \( t \) and \( k \), there exists at most one \( R_i^{(t)} > 0 \) for all \( i \in B_k \), then it is referred to as (intracell) one-by-one scheduling.

Further let

\[
\phi_i \overset{\text{def}}{=} \sum_{t: R_i^{(t)} > 0} \Phi_i
\]

denote the fraction of the scheduling interval user \( i \) transmits in the one-by-one schedule. Define for all \( i \in B_k \)

\[
\Delta_i(\mathbf{p}) \overset{\text{def}}{=} \frac{P_i}{P_k}, \tag{6.10}
\]

\[
f_i(\mathbf{p}) \overset{\text{def}}{=} \frac{\sum_{j \in B_k \setminus \{i\}} \theta_{ij} g_{ij} p_j}{I_i(\mathbf{p}) + v_i}, \tag{6.11}
\]

where \( f_i \) denotes the other-cell interference ratio for user \( i \). We can then prove the following proposition.

**Proposition 6.4** If \( \mathbf{R} = (R_1, R_2, \ldots, R_N) \) is instantaneously achievable, it is achievable in average by one-by-one scheduling.
Proof: We prove the proposition by showing the achievability of one-by-one scheduling when the total power that is used for all the users in the cell is instead assigned to one single user at a time in the cell. Consider an arbitrary cell $k$. When assigning $P_k$ to one single user in cell $k$, the generated intercell interference remains the same as if the cell power were divided on several users. Thus other cells are not influenced by the intracell scheduling in cell $k$. Therefore it is sufficient to prove the proposition for one single cell. For the proposition to be true, there must exist $\phi_i$s such that for every user $i \in B_k$

$$R_i = \frac{W}{\Gamma_i} \sum_{j \neq i} \frac{g_{ik} p_i}{\theta_{ij} b_{ik} p_j + I_i(p) + \nu_i} = \phi_i \frac{W}{\Gamma_i} \frac{g_{ik} P_k}{I_i(p) + \nu_i}$$  \hspace{1cm} (6.12)

where $\phi_i > 0$. Using (6.10) and (6.11), we obtain

$$\phi_i = \frac{\Delta_i(p)}{f_i(p) + 1}$$

Then the proposition is true for cell $k$ if the average rate can be achieved for all users in the cell within the interval $T_s$ by one-by-one scheduling, i.e., $\sum_{i \in B_k} \phi_i \leq 1$. Since $\theta_{ij} > 0$ it implies that $1/(f_i(p) + 1) < 1$ and we have

$$\sum_{i \in B_k} \phi_i = \sum_{i \in B_k} \frac{\Delta_i(p)}{f_i(p) + 1} < \sum_{i \in B_k} \Delta_i(p) = 1$$  \hspace{1cm} (6.13)

Hence $\sum_{i \in B_k} \phi_i < 1$ and all users in cell $k$ can be scheduled within $T_s$. \hfill \square

Proposition 6.1 says that capacity is not lost for scheduling nonreal time services. In fact, since $\sum_{i \in B_k} \phi_i < 1$, we can choose either to admit more nonreal time data traffic or to lower the transmit power so that more of the cell’s power is left for real time traffic. Given a rate vector $R$, to determine if it is instantaneously achievable is rather difficult with limited information of the link gains. However, in the case where $\theta_{ij} = \theta_i$ applies, the following proposition gives a necessary condition.

**Proposition 6.2** If $\theta_{ij} = \theta_i$, a necessary condition for a rate vector $R$ to be instantaneously achievable is $\sum_{i \in B_k} \frac{R_i \Gamma_i \phi_i}{W + R_i \Gamma_i \theta_i} < 1$ for all cells $k$.

**Proof:** Let $\theta_{ij} = \theta_i$, then the power can be solved for from Equations (6.1) and (6.2),

$$p_i = \frac{\Gamma_i R_i}{W + \Gamma_i R_i \theta_i} \cdot \left( \theta_i \sum_{j \in B_k} p_j + \frac{I_i(p) + \nu_i}{g_{ik}} \right)$$

and therefore

$$\sum_{i \in B_k} p_i = \sum_{i \in B_k} \frac{R_i \Gamma_i}{W + R_i \Gamma_i \theta_i} \cdot \left( \theta_i \sum_{i \in B_k} p_i \right) \frac{I_i(p) + \nu_i}{g_{ik}} \frac{R_i \Gamma_i}{W + R_i \Gamma_i \theta_i}.$$
Clearly the total power of any cell must be positive and limited, from which the condition follows.

A more conservative condition applicable to a single cell, can be found by including the maximum power constraint $\sum_{i \in B_k} p_i \leq \bar{P}_k$. Following the proof above, this can equivalently be expressed as

$$\sum_{i \in B_k} \frac{R_i G_i \theta_i}{W + \bar{R}_i G_i \theta_i} \left( \frac{I_i(p) + \nu_i}{\theta g_{ik} \bar{P}_k} + 1 \right) \leq 1. \quad (6.14)$$

Therefore, if $\theta_i$ and the maximum achievable SIR, $g_{ik} \bar{P}_k/(I_i(p) + \nu_i)$, can be measured for all users in the cell, the relation (6.14) can serve as an admission criteria in cell $k$ for a given intercell interference level $I_i(p)$.

### 6.3.1 Energy Conservation

A major benefit from using one-by-one scheduling is that it requires less energy for delivering the same amount of data as for continuous transmission, which we state below.

**Corollary 6.1** For any single cell, one-by-one scheduling requires the least energy of all schedules for providing an average data rate.

**Proof:** Assume that the minimum energy solution requires that at least two users transmit simultaneously between the times $t_1$ and $t_2$, $0 \leq t_1 < t_2 \leq T$, and that the sum power of the users is $P_k$. Then the proof of Proposition 6.1 gives that if $P_k$ is assigned to one single user in the cell at a time, the same data amount transmitted in $[t_1, t_2]$ can be delivered in $\delta < t_2 - t_1$ units by one-by-one scheduling. Since the energy consumption in $[t_1, t_2]$ of one-by-one scheduling is $\delta P_k < (t_2 - t_1)P_k$, simultaneous transmissions cannot minimize the consumed energy. \hfill $\square$

Corollary 6.1 says that if no coordination between cells is considered, which by constraint (6.6) is our assumption, one-by-one scheduling is necessary for minimizing the energy consumption. In other words, the solution of (6.4)-(6.8) will use time division within the cells. The energy consumed in a cell by one-by-one scheduling is $\sum_{i \in B_k} \phi_i P_k T_s$, when $P_k$ is assigned to a single user. For a single cell $k$, let us define the relative energy efficiency of one-by-one scheduling as compared to continuous transmission

$$E \overset{\text{def}}{=} \frac{P_k T_s}{\sum_{i \in B_k} \phi_i P_k T_s} = \frac{1}{\sum_{i \in B_k} \phi_i} > 1$$

where the last inequality follows from Proposition 6.1. For user $i$ the relative energy efficiency is

$$E_i \overset{\text{def}}{=} \frac{\phi_i T_s}{P_k T_s} = \frac{f_i(p)}{\phi_i P_k T_s} + 1 > 1.$$
This means that the whole system will use relatively less energy, as will every user in it. Assuming that \( \theta_{ij} = \theta_i \), from the definition of \( f_i(\mathbf{p}) \), the relative energy efficiency is approximately given by \( E_i \approx 1 + \theta_i g_{ij} P_k / (I_i(\mathbf{p}) + \nu_i) \). This is intuitively understood as if there is much intracell interference (\( \theta_i g_{ij} P_k \) high) and less intercell interference (\( I_i(\mathbf{p}) + \nu_i \) low), there is large possible gain from one-by-one scheduling in terms of improved energy efficiency. Closed form expressions of \( E \) and \( E_i \) are given in [16].

Since \( \sum_{i \in B_k} \phi_i < 1 \), it is possible to lower the transmission power and compensate the lower data rate by prolonging the duration of the transmission so that users are scheduled over the whole interval \( T_s \). For that purpose, we may normalize the \( \phi_i \)'s such that \( \phi'_i = \frac{\phi_i}{\sum_{i \in B_k} \phi_i} \) which gives \( \sum_{i \in B_k} \phi'_i = 1 \).

From (6.12), when replacing \( \phi_i \) with \( \phi'_i \), we must consequently adjust the power to \( P'_k = \sum_{i \in B_k} \phi'_i P_k \) for maintaining the same average data rate. Therefore \( \sum_{i \in B_k} \phi'_i P'_k T_s = \sum_{i \in B_k} \phi_i P_k T_s \) and the energy consumption remains the same, thus \( \tilde{E} \) and \( E_i \) are unchanged. This is illustrated in Fig. 6.1.

### 6.3.2 Minimum Time Span

For some services, the quality guarantee is in the form of a maximum message delay. The problem of minimizing the maximum message delay given a certain amount of data in each user's buffer was studied in [54] and referred to as the minimum time span problem. An analysis was carried out but practical algorithms for achieving the optimal schedule were not suggested. Consider any single cell in the system and the case where each user \( i \) has \( R_i T_s \) bits to transmit. Then one-by-one scheduling minimizes the total time it takes for all users in the cell to transmit their data.

**Corollary 6.2** For any single cell, the minimum time span for transmitting the data of all users, is achieved by one-by-one scheduling using \( P_k = \tilde{P}_k \) and is given by \( T_s \cdot \sum_{i \in B_k} \phi_i \).

**Proof:** Assume that the minimum time span solution requires that at least two users transmit simultaneously between the times \( t_1 \) and \( t_2 \), \( 0 \leq t_1 < t_2 \leq T_s \), and that the sum power of the users is \( P_k \). Then the proof of Proposition 6.1 gives that if \( P_k \) is assigned to one single user at a time in the cell, the same data amount transmitted in \([t_1, t_2]\) can be delivered in \( \delta < t_2 - t_1 \) units by one-by-one scheduling. This contradicts to the assumption, thus simultaneous transmissions cannot minimize the time span. From (6.2) it follows that the rate is maximized and thus time span minimized when using maximum power, \( \tilde{P}_k \). \( \square \)

If no coordination between cells is considered, which is our assumption, one-by-one scheduling is necessary for minimizing the time span.
Figure 6.1: An example of two users transmitting in a one-by-one fashion using the power $P_k$ (solid). By utilizing the whole scheduling interval $T_s$, the power can be decreased to $P_k^*$ (dashed), where $\phi_1 = \ell_k / t_s$, $\phi_2 = 1 - \phi_1$ and $P_k^* T_s = P_k t_2$.

### 6.4 Distributed Power Control

Since $\sum_{i \in B_k} \phi_i < 1$ in all cells when using the same total transmission power as for continuous transmission, it is possible to use one-by-one scheduling but with less transmission power. The lowest such power assignment will yield $\sum_{i \in B_k} \phi_i = 1$ for all cells $k$, which is found by utilizing the whole scheduling interval for transmission. For that purpose we use (6.12) expressed in $P$ to solve for a state such that,

$$\sum_{i \in B_k} \phi_i = \frac{1}{W} P_k \sum_{i \in B_k} \frac{R_i \Gamma_i}{g_k} (I_i(P) + \nu_i) = 1$$

for all $M$ cells. With $I_i(P) = \sum_{i=1}^{M} g_k P_i$, this is a linear equation system in $P$ and evaluating for $k = 1$ we get

$$P_1 - \left( \frac{1}{W} \sum_{i \in B_1} \frac{g_2 R_i \Gamma_i}{g_{i_1}} \right) P_2 - \ldots - \left( \frac{1}{W} \sum_{i \in B_1} \frac{g_i M R_i \Gamma_i}{g_{i_1}} \right) P_M = \frac{1}{W} \sum_{i \in B_1} \frac{R_i \Gamma_i \nu_i}{g_{i_1}}.$$ 

Proceeding in the same manner for the other rows, we can write the problem in matrix form as

$$(I - \tilde{H})P = \tilde{\eta} \quad (6.15)$$
where the matrix elements of the $M \times M$ iteration matrix $\tilde{\mathbf{H}}$ are

$$
\tilde{h}_{ij} = \begin{cases} 
\sum_{k \in \mathcal{B}} r_k \frac{\Gamma_i}{\gamma_{ik}} g_{ik} & i \neq j \\
0, & i = j
\end{cases}
$$

and $\tilde{\eta}$ is a $M \times 1$ vector with $\tilde{\eta}_i = \sum_{k \in \mathcal{B}} \frac{r_k \Gamma_i}{\gamma_{ik}} g_{ik}$.

Now for the purpose of solving (6.15) in a distributed fashion, consider the following iterative power control procedure,

$$
\mathbf{P}(n + 1) = \tilde{\mathbf{H}} \mathbf{P}(n) + \tilde{\eta}
$$

(6.16)
or equivalently for each cell $k$,

$$
P_k(n + 1) = \sum_{i \in \mathcal{B}_k} r_k \frac{\Gamma_i}{\gamma_{ik}} \left( I_i(\mathbf{P}(n)) + \nu_i \right).
$$

With $\gamma_i = r_i \Gamma_i / W$ and $SIR_i(\mathbf{P}(n)) = g_{ik} P_k(n) / (I_i(\mathbf{P}(n)) + \nu_i)$ being the SIR a user would experience during its transmission, an equivalent representation is

$$
P_k(n + 1) = \sum_{i \in \mathcal{B}_k} \frac{\gamma_i}{SIR_i(\mathbf{P}(n))} P_k(n).
$$

Taking into account the power constraint we can execute intracell scheduling power control by defining the mapping

$$
\mathcal{T}(\mathbf{P}) \overset{\text{def}}{=} \min \{ \tilde{\mathbf{H}} \mathbf{P}(n) + \tilde{\eta}, \mathbf{P} \}.
$$

(6.17)

### 6.4.1 Convergence

The following result guarantees that the optimal power allocation will be found.

**Proposition 6.3** For any initial $\mathbf{P}(0) \geq \mathbf{0}$, the mapping $\mathcal{T}$ converges with geometric rate to a power vector that supports $\sum_i \phi_i = 1$ for all cells, if such a power allocation is feasible.

**Proof:** Consider the weighted maximum norm $\|\mathbf{P}\|_W = \|\mathbf{W}\mathbf{P}\|_\infty$ and the consistent weighted matrix norm $\|\tilde{\mathbf{H}}\|_W = \|\mathbf{W}^{-1}\tilde{\mathbf{H}}\|_\infty$ where $\mathbf{W} = \text{diag} \{ e_i^{-1} \}$.

It can be shown that $\tilde{\mathbf{H}}$ is irreducible and nonnegative and therefore has a positive real eigenvalue equal to its spectral radius $\rho(\tilde{\mathbf{H}})$. By Proposition 3.3, a necessary condition for a feasible solution of (6.15) to exist is $\rho(\tilde{\mathbf{H}}) < 1$. Let $\mathbf{e} = (e_i)$ be the Perron eigenvector, $\tilde{\mathbf{e}} = \rho(\tilde{\mathbf{H}}) \mathbf{e}$, then

$$
\|\mathcal{T}(\mathbf{P}(n)) - \mathbf{P}^*\|_W^\infty \leq \|\tilde{\mathbf{H}} \mathbf{P}(n) + \tilde{\eta} - \mathbf{P}^*\|_W^\infty \\
= \|\tilde{\mathbf{H}} \mathbf{P}(n) + \tilde{\eta} - (\tilde{\eta} + \tilde{\mathbf{H}} \mathbf{P}^*)\|_W^\infty \\
\leq \|\tilde{\mathbf{H}}\|_W^\infty \cdot \|\mathbf{P}(n) - \mathbf{P}^*\|_W^\infty.
$$
Therefore,
\[ ||P(n) - P^*||_W^\infty \leq \rho(\mathbf{H})^n ||P(0) - P^*||_W^\infty, \]
so if \( \rho(\mathbf{H}) < 1 \) and the fixed point is within the feasible power range, \( \mathcal{T} \) is a pseudo-contractive mapping with respect to the weighted maximum norm and converges to a unique fixed point given by \( P^* = \mathcal{T}(P^*) \) which is the solution to \( (I - \mathbf{H})P^* = \tilde{\eta} \). \( \square \)

**Remark 6.1** It is also possible to prove convergence of \( \mathcal{T} \) by verifying that it is a standard interference function [122], which also guarantees its asynchronous convergence.

It should also be pointed out that (6.17) will converge to a fixed point even if \( R \) is not achievable in average. In that case, at least one cell will use \( P_h = \tilde{P}_h \). This phenomenon could be utilized as an incentive for removing users or decreasing some users’ data rates. The power control algorithm (6.17) also exhibits the property that if all cells are supported, the powers decrease monotonically and all cells stay supported. This is stated below, where \( \phi_i(n) \) denotes the value at iteration \( n \).

**Corollary 6.3** If there exists an iteration \( n \) where \( \sum_{i \in B_k} \phi_i(n) \leq 1 \) for all cells \( k \), then
\[ P_k(n + l) \leq P_k(n + l - 1) \]
and
\[ \sum_{i \in B_k} \phi_i(n + l) \leq 1 \]
for all \( k \) and \( l \geq 1 \).

**Proof:** Using (6.12), we can equivalently express the power control algorithm as
\[ P_k(n + 1) = \sum_{i \in B_k} \phi_i(n)P_k(n). \]
Let \( l = 1 \) and assume that \( \sum_{i \in B_k} \phi_i(n) \leq 1 \) which gives gives \( P_k(n + 1) \leq P_k(n) \) for all \( k \). We have,
\[ \sum_{i \in B_k} \phi_i(n + 1) = \frac{1}{WP_k(n + 1)} \sum_{i \in B_k} \frac{R_i\Gamma_i}{g_{ik}}(I_i(P(n + 1)) + \nu_i) \]
\[ = \frac{1}{WP_k(n + 1)} \sum_{i \in B_k} \frac{R_i\Gamma_i}{g_{ik}}(I_i(P(n)) + \nu_i) \]
\[ \leq 1. \]
By mathematical induction, the relation holds for \( l \geq 1 \). \( \square \)
6.4.2 Rate of Convergence

The proof of Proposition 6.3 illustrates how the rate of convergence is strongly related to the spectral radius of the iteration matrix. Now, let us compare with the continuous transmission problem in (6.3). A related way to distributively find that optimal power allocation, neglecting the power constraints, is to use a mapping \( T(p(n)) = H p(n) + \eta \) suggested in [35], also see Example 3.2.1. In a similar manner, it can be shown that this so-called DPC algorithm, will asymptotically converge at the rate \( \rho(H) \). The issue of faster convergence is then reduced to relating the different spectral radii of \( H \) and \( \hat{H} \). Define a new \( N \times N \) matrix \( \hat{H} \) with elements

\[
\hat{h}_{ij} = \begin{cases} 
0, & i \neq j, b_i = b_j \\
\frac{\Gamma_k}{W} \frac{g_{kj}}{g_{kj}}, & i \neq j, b_i \neq b_j \\
0, & i = j.
\end{cases}
\]

Theorem 2.1-16 in [125] implies that \( \rho(\hat{H}) \leq \rho(H) \) with equality only if there exists exactly one user per cell. The \( \hat{H} \) matrix can be interpreted as the iteration matrix of a problem where no intracell interference exists, for example by fully orthogonal spreading sequences at the receiver but it is also useful for describing intracell time division. We can relate its spectral properties to \( H \) by a following proposition. First we observe the following on its matrix spectrum.

**Lemma 6.1** \( \hat{H} \) has less than or equal to \( M \) distinct eigenvalues.

**Proof:** Consider the transposed matrix \( \hat{H}^T = [\hat{h}_{ij}] \). As all rows \( i \in B_k \) in \( \hat{H}^T \) are equal for a given \( k \) it implies that it can have up to \( M \) distinct nonzero eigenvalues. Using the well-known property that \( \hat{H} \) shares the same matrix spectrum as \( \hat{H}^T \), the lemma follows. \( \square \)

**Proposition 6.4** The spectral radius fulfills \( \rho(\hat{H}) = \rho(H) \leq \rho(H) \) where the inequality becomes equality only if there is exactly one user per cell. In particular, \( \rho(\hat{H}) < 1 \) iff \( R \) is instantaneously achievable.

**Proof:** Let \( e = (e_i) \) be an eigenvector such that \( \hat{H} e = \lambda e \). Then for any \( k \in B_i \)

\[
\sum_{j=1}^{N} \hat{h}_{kj} e_j = \frac{\Gamma_k}{W} \sum_{j=1}^{M} \frac{g_{kj}}{g_{kj}} \sum_{i \in B_j} e_i = \lambda e_k \\
\sum_{k \in B_i} \frac{\Gamma_k}{W} \sum_{j=1}^{M} \frac{g_{kj}}{g_{kj}} \sum_{i \in B_j} e_i = \lambda \sum_{k \in B_i} e_k. \tag{6.18}
\]
Define
\[ F_i \overset{\text{def}}{=} \sum_{k \in B_i} e_k \] (6.19)
so that (6.18) becomes
\[ \sum_{j \neq i} \sum_{k \in B_i} \frac{R_k \Gamma_k}{W} \frac{g_{kj}}{g_{ki}} F_j = \lambda F_i. \]

Therefore we have shown that \( F = (F_i) \) is an eigenvector of \( \hat{H} \) with eigenvalue \( \lambda \). By Lemma 6.1, \( \hat{H} \) has \( \leq M \) distinct eigenvalues which therefore also are eigenvalues of \( H \). It is clear that if zero is an eigenvalue of \( \hat{H} \), it is also an eigenvalue of \( H \). We next show that \( \hat{H} \) does not have any other nonzero eigenvalues. If \( \hat{H} \) has exactly \( M \) distinct eigenvalues then it is trivial but if \( \hat{H} \) has \( M' < M \) distinct eigenvalues it must be shown that \( H \) does not have any additional eigenvalues. Assume that \( \lambda' \neq 0 \) is an eigenvalue of \( H \) but not of \( \hat{H} \). Let \( F = (F_i) \) be the eigenvector of \( H \) corresponding to \( \lambda' \). If we choose
\[ e_k = \frac{1}{\lambda'} \frac{R_k \Gamma_k}{W} \sum_{m \neq i} \frac{g_{km}}{g_{ki}} F_m, \quad \forall k \in B_i \]
we obtain for row \( k \)
\[ \sum_{j \neq k} h_{kj} F_j = \frac{R_k \Gamma_k}{W} \sum_{i \neq k} \frac{g_{kj}}{g_{ki}} \sum_{i \in B_j} \frac{1}{\lambda'} \frac{R_i \Gamma_i}{W} \sum_{m \neq j} \frac{g_{jm}}{g_{ij}} F_m \]
\[ = \frac{R_k \Gamma_k}{\lambda'} \frac{1}{W} \sum_{i \neq k} \frac{g_{kj}}{g_{ki}} \sum_{m \neq j} \sum_{i \in B_j} \frac{R_i \Gamma_i}{W} \frac{g_{jm}}{g_{ij}} F_m \]
\[ = \frac{R_k \Gamma_k}{\lambda'} \frac{1}{W} \sum_{i \neq k} \frac{g_{kj}}{g_{ki}} F_j = \lambda' e_k, \]
which contradicts the assumption that \( \lambda' \) is not an eigenvalue of \( \hat{H} \). Therefore \( \hat{H} \) and \( H \) have the same nonzero eigenvalues. Since \( \rho(\hat{H}) \) is the maximum modulus eigenvalue of \( \hat{H} \), it follows that it is also the spectral radius of \( H \) and the corresponding Perron eigenvector can be found from (6.19). A necessary condition for \( R \) to be instantaneously achievable is that (6.3) has a feasible solution. By Proposition 3.3, (6.3) has a unique and positive solution if \( \rho(H) < \)
1. Therefore, we can conclude that $\rho(\mathbf{H}) < 1$ if $\mathbf{R}$ is instantaneously achievable. □

Hence, if $\mathbf{R}$ is instantaneously achievable, the mapping (6.17) will always converge. If the mobiles connect to the base station with lowest attenuation, a condition for determining if $\mathbf{R}$ is achievable in average sense, is given by the following result.

**Corollary 6.4** When $M \geq 2$, a sufficient condition for (6.15) to have a positive solution $\mathbf{P} > 0$ is that

$$\sum_{i \in B_k} \frac{R_i \Gamma_i}{W} < \frac{1}{M - 1}$$

for all cells $k$.

**Proof:** From Corollary 1, p. 17 in [113], $\rho(\mathbf{H})$ is less than or equal to the maximum modulus row sum of $\mathbf{H}$. Since $g_{ij} \leq g_{ik}$ when $i \in B_k$ and from Proposition 3.3, the spectral radius must be strictly less than unity.

$$\rho(\mathbf{H}) \leq \max_{1 \leq k \leq M} \left\{ \frac{1}{W} \sum_{j \neq i}^{M} \sum_{i \in B_k}^{g_{ij}} g_{ik} R_i \Gamma_i \right\}$$

$$\leq \max_{1 \leq k \leq M} \left\{ \frac{1}{W} \sum_{j \neq i}^{M} \sum_{i \in B_k}^{R_i} \Gamma_i \right\}$$

$$= \max_{1 \leq k \leq M} \left\{ \frac{M - 1}{W} \sum_{i \in B_k}^{R_i} \Gamma_i \right\}.$$

So, if $\frac{M - 1}{W} \sum_{i \in B_k}^{R_i} \Gamma_i < 1$ for all cells $k$, the corollary follows. □

It should be pointed out that Corollary 6.4 could be very conservative, especially if $g_{ij} \ll g_{ik}$. This is likely to occur when $M$ increases, implying that Corollary 6.4 has its practical applicability limited to low $M$.

Proposition 6.4 shows that the power control for the scheduled system (6.15) may be performed faster than for (6.3), which practically is important in an ever changing radio environment. Not only is the spectral radius important from a rate of convergence point of view. It has also been identified as a measure strongly related to congestion and capacity [44]. For example, an $\alpha \geq 1$ may always be found so that $\rho(\alpha \mathbf{H}) = \rho(\mathbf{H})$. For upholding the same congestion, spectral radius in this case, all users in the one-by-one scheduled system may therefore increase their data rates $\alpha = \rho(\mathbf{H})/\rho(\mathbf{H})$ times. Thus the capacity is not less by this transmission scheme, which was also shown by Proposition 6.1.
A quantitative measure of $\alpha$ can be found by the following approximation.

$$\alpha \approx \frac{\text{SR}_t(\mathbf{p})_{\theta=0}}{\text{SR}_t(\mathbf{p})_{\theta=0}} = 1 + f_i(\mathbf{p})$$

(6.20)

That is, the capacity gain is linear in the orthogonality factor.

Since the power vector obtained from (6.15) is the smallest possible and the necessary condition from Corollary 6.1 of one-by-one scheduling is fulfilled, we can conclude that this cell-site power allocation is optimum and solves problem (6.4). There are however different orders in which the users in each cell could be scheduled but that does not violate any of the above results. Next chapter will deal with this further. The practical version of the suggested scheduling and power control procedure is to in each cell independently let one user $i$ transmit until $R_i T_i$ bits have been delivered, then user $i$ is shut off until all other users in the cell have delivered their data in the same manner.

6.5 Numerical Results

The algorithms are evaluated by simulations in a system consisting of 9 hexagonal cells with centrally located base stations using omni-directional antennas. The cell radius is set to 1 km and a wrap-around technique is used. Evaluation is performed in the DS-CDMA system described in Chapter 2 and illustrated in Fig. 1.2. The cell radius is set to 1 km and a wrap-around technique is used along with omni-directional antennas. For a given instance, 10 mobiles/cell are generated, the locations of which are uniformly distributed over the cells. At any given instance, the link gain is modeled by $g_{ij} = S_{ij} \cdot r_{ij}^{-4}$, where $S_{ij}$ is the shadow fading factor and $r_{ij}$ is the distance between base $i$ and mobile $j$. The shadow fading factor is generated from a lognormal distribution with $\mu_s = 0$ dB, and $\sigma_s = 8$ dB, where the $S_{ij}$'s are mutually independent. The base receiver noise is taken to be $\nu_i = 10^{-15}$ W. The required bit-energy-to-noise-spectral-density ratio is set to $\Gamma_i = 8$ dB for all users with a spreading bandwidth $W = 1.2288$ MHz. In Fig. 6.2, the result of one thousand independently generated $\mathbf{H}$ matrices is depicted for different orthogonality factors $\theta$ and the same data rate $R_i = R$ is used by all users. The figure shows the overall system relative energy efficiency of the suggested power control and scheduling concept for the cases where the rates are instantaneously achievable. That is, we plot $\mathbf{P}^* \mathbf{T}_m / \mathbf{P}^* \mathbf{T}_n$, where $\mathbf{p}^*$ is the power vector for supporting $\mathbf{R}$ instantaneously which can be solved from (6.3) and $\mathbf{P}^* = (\mathbf{I} - \mathbf{H})^{-1} \eta$. The plot shows that for high data rates and orthogonality factors, the same data amount can be delivered with approximately four times less energy if we are to use the proposed transmission scheme. For high data rates, the gain strongly varies with the orthogonality factor. Fig. 6.3 shows the convergence rate of the power
control algorithm, measured as the normalized Euclidean distance to the fixed point, \[ \frac{\|\epsilon(n)\|_2}{\|\epsilon(0)\|_2} = \frac{\|P(n) - P^*\|_2}{\|P(0) - P^*\|_2} \] in each iteration for different data rates. The initial power vector is chosen from a uniform distribution, \( P(0) \in [0, 1] \) and \( \hat{P} \) is set to 5 W. It is for example seen that for \( R = 10 \) kbps at 20 iterations, the normalized distance is about \( 10^{-5} \), which in turn suggests that the algorithm converges rapidly and is able to quickly react to changes in traffic. As the data rate is decreased by half to 5 kbps, the convergence is faster, which is expected since the spectral radius of the iteration matrix decreases by the same portion.

The capacity increase is illustrated by Fig. 6.4 with a solid line where \( \rho(H)/\rho(H) \) is plotted as function of the orthogonality factor. The relation is verified to be linear with a maximum gain of approximately 2.7 times when the orthogonality factor is unity. In [116], values of \( f^{-1} \) have been obtained for the uplink using constant received power control. For small orthogonality factors, the optimal power solution contained in Proposition 6.2, quantitatively resembles a constant received power. Therefore, an approximation can be to consider the \( f^{-1} \) values also for the downlink. It was shown in [116] that when the propagation attenuation is of fourth power and lognormal fading has standard deviation 8 dB, the factor is \( f^{-1} = 0.55 \). In Fig. 6.4, we plot the approximation (6.20) using \( f^{-1} = 0.55 \). For low \( \theta \), the curve shows a reasonable approximation.
Figure 6.3: The convergence rate measured as normalized Euclidean distance for $R = 10$ and $5$ kbps.

to the simulated results.

The relation between the relative energy efficiency and capacity gain can be found by simultaneously studying Fig. 6.2 and 6.4. For example, if $\theta = 1$ and the data rate is larger than $9$ kbps, the relative energy efficiency is larger than $\approx 2.7$, the relative capacity gain at $\theta = 1$, whereas for lower data rates there is a larger gain in the relative capacity than for the relative energy efficiency. For small values of the orthogonality factor, e.g., $\theta = 0.2$, the relative capacity gain is always larger than the relative energy efficiency.

6.6 Concluding Remarks

In this chapter, we considered the problem of delivering data with minimum energy to users with a required average data rate for a downlink DS-CDMA system. It is observed that having a QoS constraint based on an average- rather than an instantaneous data rate opens possibilities of flexible transmission scheduling. We proposed a form of combined intracell scheduling and power control, where only one user in the cell is allowed to transmit at a time. By such a scheduling, it was shown that the same data can be delivered with less energy. For the purpose of supporting the average data rate with as small energy consumption as possible, we combined the scheduling with a suggested convergent dis-
Figure 6.4: The relative capacity increase as function of the orthogonality factor \( \theta \). The dashed curve refers to the approximation (6.20).

Distributed power control algorithm adjusting cell-site powers. Another advantage of decreasing the power level to the minimum, is that the intercell interference becomes stable and less bursty. The power control problem was reduced to a familiar linear equation system and by comparing the different spectral radii of the iteration matrices, it was found that the system capacity was increased from the proposed scheduling/power control scheme as compared to the normal power control problem. This is due to a smaller spectral radius of the scheduled system which also means that the power control can converge faster. Numerical results indicate that with the proposed concept the transmit powers can be significantly decreased, which opens possibilities for supporting more real time users or admitting more nonreal time traffic. The issue of determining the order in which intracell transmissions should be performed was not addressed here. This topic is given a further study in next chapter. In addition, it is expected that there may be gains which are not captured in this chapter. For example, in each cell \( k \), if the users are scheduled in the order of the factors \( g_{ik}/I_i(\mathbf{P}) \), it is likely that if \( T_i \) is sufficiently long, the time varying channel may improve the link gains of the worst mobiles. Future work could include other models than (6.2), e.g., the nonlinear relations described in [97] and upper bounds on the data rates. Other mappings than the one suggested can be used on the \( \mathbf{H} \) matrix as well. Interestingly an iteration matrix of the similar condensed form has appeared in
[62,80] but in a different context.
Chapter 7

Transmission Scheduling over Fading Channels

SCHEDULING is a viable option in the provisioning of delay insensitive data services. Due to the delay tolerance, there exists a freedom in adapting the transmission attempts to both the channel quality and the user QoS requirements. In this chapter, we consider scheduling over a DS-CDMA downlink fading channel. We begin by building on the result from previous chapter, saying that by avoiding intracell interference, letting only one single user in each cell access the channel at a time, the average throughput will increase. This result was previously obtained for a nonfading channel. With fading, the channels can be considered to vary asynchronously among the users and additional multiuser diversity gains can be obtained by exploiting channel adaptive scheduling. To harness varying channel conditions, schedulers of different adaptation rate are suggested and compared. In particular we analyze a fast scheduler that schedules a user to transmit when its channel is good relative to its mean. This principle can asymptotically provide the same fairness, here defined as access to the channel, as a round robin (RR) scheduler but with significantly improved throughput. The chapter contains evaluation of this scheduler for varying channel models.

7.1 Introduction

The varying channel conditions in a wireless system imply that, two users allocated the same amount of resources (bandwidth or power), could experience different performance. To schedule users for maximum channel utilization, resources should be allocated to the users with the highest channel quality. This form of water filling principle however, in terms of scheduling, could cause severe fairness problem and potentially starve certain users. A well designed scheduler would allow for long-term fairness and if the latency requirements are tight,
also short-term fairness. In conjunction with a fairness constraint, a scheduler could prioritize users with temporarily better channels. In the literature, several schedulers and fairness criteria have been given. In [9], schedulers which are throughput optimal were found. Such a scheduler, is able to keep all packet queues stable, if it is at all feasible to do with any scheduler. Another direction is Kelly’s proportional fairness [60]. There the objective is to maximize the product of the individual user throughputs, or their log sum. It was shown that the proportionally fair solution $\mathbf{R}$ is unique and characterized by, for any other feasible allocation $\mathbf{R'}$, $\sum_i (R'_i - R_i)/R_i \leq 0$. A proportional fair (PF) scheduler to address this criterion was given in [48, 49, 114]. With $r_i(t)$ being the maximum supportable data rate of user $i$, the algorithm schedules according to

$$i^* = \arg \max_i \frac{r_i(t)}{T_i(t)}$$

where $T_i(t)$ is the average throughput, updated through a filter

$$T_i(t + 1) = \left(1 - \frac{1}{t_c}\right) T_i(t) + \frac{1}{t_c} r_i(t) 1_{i=i^*}.$$

It was shown [114] that, if $t_c = \infty$, this algorithm is proportionally fair and will maximize $\sum_{j=1}^N \log T_j$, where $T_j$ is the limit of the long-term average throughput. For long averaging windows users get scheduled the same amount of time [48].

Here, we will consider two classes of schedulers: slow (nonadaptive) and fast (adaptive). The slow scheduler base its decisions on slow channel variations, like attenuation and shadowing, or does not adapt to the channel at all. Clearly, this form is robust and requires small overhead, to the price of performance. Similarly to [27, 48, 49], an asymptotic analysis is considered. That is long-run performance is evaluated. A simple type of resource fair channel nonadaptive scheduling, known to allocate the access times fairly, is the RR scheduler. However, although the time allocation of RR is fair, the performance is mostly not very encouraging in fast fading environments. Therefore, a fast channel adaptive scheduler is suggested, taking into account the average and instantaneous channel conditions, where the user with the “relatively best” channel state is scheduled. As for the PF scheduler, the underlying principle is that a user which has a good quality relative to its expected quality, transmits. This concept very much resembles the principle of selection diversity. For a background on multiuser diversity, [114] provides additional references. In the analysis, absolute delay requirements and queue stability [9] are not considered. The throughput performance with such additional constraints will likely be less than the results reported herein.

### 7.1.1 System Model

Consider a single cell in a DS-CDMA system with a spreading bandwidth $W$. The required bit-energy-to-interference-spectral-density ratio $E_b/I_0$ is set to $\Gamma_i$. 


for any user \(1 \leq i \leq N\). The data arrives to buffers of infinite length, which are to be full all the time, i.e., aspects of queuing are not considered. No power control is exercised and the total downlink transmission power is fixed at \(P\). With \(I_i + \nu\) denoting the fixed intercell plus background interference level and \(g_i(t)\) the time varying link gain, we define the instantaneous channel state as the SIR without intracell interferers, given by \(\xi_i(t) = g_i(t)/P/(I_i + \nu)\). Different transmission rates are provided by adapting the processing gain. Here, we assume this is achieved by varying the symbol time and keeping a fixed chip rate and modulation format. Without restriction, we assume that the instantaneous transmission rate of user \(i\) will follow a linear relation of its SIR\(^1\)

\[
    r_i(t) = \frac{W}{\Gamma_i} \frac{\xi_i(t) \Delta_i}{\theta_i(1 - \Delta_i) \xi_i(t) + 1}
\]

(7.1)

The factor \(\Delta_i \in [0, 1]\) denotes the fraction of the total transmission power allocated to user \(i\) and \(\theta_i \in (0, 1]\) represents the orthogonality factor, which is assumed to be time-invariant. If one-by-one transmission is employed, all the transmission power of the base station can be allocated to one particular user at a time. Consequently, \(\Delta_i = 1\) and its instantaneous transmission rate during transmission becomes

\[
    r_i(t) = \frac{W}{\Gamma_i} \xi_i(t).
\]

(7.2)

Time is assumed to be a continuous variable without any restriction on scheduling instants or frame size granularity. Also, the rates are continuous and we do not impose any constraints on available bandwidth. Such restrictions are further investigated in Section 7.3.2. We assume that \(\xi_i(t)\) can be accurately estimated and that it is a wide sense stationary and ergodic stochastic process for all users. These processes are assumed to be mutually independent of each other, for all users. The channel state of each user has at any instant a mean \(\mathbb{E}[\xi_i(t)|t] = \bar{\xi}_i\) for all \(t\). From the these assumptions, we find the asymptotic data rate

\[
    R_i = \lim_{T \to \infty} \frac{1}{T} \int_0^T r_i(t) dt = \frac{W}{\Gamma_i} \bar{\xi}_i
\]

(7.3)

where the limit is stochastic and denotes convergence almost surely\(^2\). In this chapter, we will consider the variations of \(\xi_i(t)\) to be from small-scale fading. The mean \(\bar{\xi}_i\) can be regarded as being comprised by attenuation and shadowing effects, occurring on a much longer time scale. The channel fading is mainly assumed to follow the noncentral \(\chi^2\) distribution with two degrees of freedom, defined by its density function (2.5).

\(^1\)The linear relation can be further motivated for large bandwidth channels, since the capacity of a Gaussian channel with interference spectral density \(I_0\) Watt/Hz approaches

\[
    \lim_{W \to \infty} W \log_2 \left( 1 + \frac{P_{ex}}{I_0 W} \right) = \frac{P_{ex}}{I_0} \log_2 e \text{ bps.}
\]

\(^2\)In general, we shall require \(\Pr\left[ \lim_{T \to \infty} \frac{1}{T} \int_0^T r_i(t) dt = \int_0^\infty r_i(\xi) f_\xi(\xi) d\xi \right] = 1\).
7.2 Transmission Scheduling

7.2.1 One-by-One Scheduling

It has been found, for both uplink [67] and downlink [107], that the information-theoretic capacity of a time-varying channel is maximized by letting the user with the best channel transmit. In previous chapter, we found that for a nonfading channel with linear rate, simultaneous transmission is outperformed in terms of throughput by letting users transmit one-by-one. This can be interpreted as a static gain from the multiple access technique itself, independent of any multiuser diversity benefits. An interpretation thereof, is that we can always define a RR scheduler that will provide higher average throughput than simultaneous transmissions. Hence a static gain would relax the requirements on designing a high throughput scheduler. However, that result was based on a continuous and linear rate relation like (7.1), reasonable for systems with variable processing gain. For other nonlinear rate relations, implied by adaptive modulation and coding, a single user may not necessarily be able to utilize the whole channel capacity solely and it may be better to schedule users simultaneously, see, e.g., [52]. The GRP procedure in Chapter 5 is another consequence of this. However, even if there would not be any static gain from time division, such a system would still exhibit the gain from multiuser diversity. Hence there could still be a net gain compared to the usual pure CDMA. Considering (7.1), we can easily verify that there is a gain from time division also if we let the link gains be time varying and stochastic. This justifies our choice to focus on scheduling transmissions one-by-one. Let \( \phi_i \) be the fraction of time user \( i \) transmits, where \( 0 \leq \phi_i \leq 1 \) and \( \sum_{i=1}^{N} \phi_i = 1 \). For a single cell, we can state the following property.

**Proposition 7.1** The total throughput \( \sum_{i=1}^{N} R_i \) of a one-by-one transmission scheme is greater than that of simultaneous transmission.

**Proof.** Consider the asymptotic fraction of time \( \phi_i \) used in the one-by-one scheme, that is required to obtain the same average data rate as using simultaneous transmission:

\[
R_i = \phi_i \frac{W}{\Gamma_i} \xi_i = \frac{W}{\Gamma_i} \mathbb{E} \left[ \frac{\xi_i \Delta_i}{\theta_i (1 - \Delta_i) \xi_i + 1} \right]
\]

Since the channel state \( \xi(t) \) is a stationary ergodic process, it follows that the channel at certain time instant \( t \) is determined by a density function \( f_{\xi_i}(\xi_i) \) which is independent of \( t \). Furthermore, it is reasonable to assume that \( \xi_i(t) \geq 0 \) and the probability that \( \xi_i = 0 \) is strictly less than one. That is, \( f_{\xi_i}(\xi_i) = 0 \) for all \( \xi_i < 0 \) and \( f_{\xi_i}(\xi_i) \neq \delta(\xi_i) \) where \( \delta(\cdot) \) denotes the Dirac delta function. It
follows for any $i$ with $0 < \Delta_i < 1$ that

$$
E \left[ \frac{\xi_i \Delta_i}{\theta_i(1 - \Delta_i)\xi_i + 1} \right] = \int_0^\infty \frac{\xi_i \Delta_i}{\theta_i(1 - \Delta_i)\xi_i + 1} f_{\xi_i}(\xi_i)\,d\xi_i < \int_0^\infty \xi_i \Delta_i f_{\xi_i}(\xi_i)\,d\xi_i = \Delta_i \tilde{\xi}_i.
$$

whence $R_i = \phi_i \frac{W}{\Gamma_i} \tilde{\xi}_i < \frac{W}{\Gamma_i} \Delta_i \tilde{\xi}_i$. Solving for the sum of $\phi_i$ yields $\sum_{i=1}^N \phi_i < \sum_{i=1}^N \Delta_i = 1$. The excess time fraction $1 - \sum_{i=1}^N \phi_i > 0$ could be utilized to transmit more data. Consequently, the throughput of one-by-one scheduling scheme is greater than that of simultaneous transmission. □

The proposition could easily be extended to a multiecellular case by letting also $I_i$ be a random variable and observing that the average generated intercell interference from a cell, is the same as for simultaneous transmission. Hence it suffices to prove the proposition for one single cell separately, see Proposition 6.1. Proposition 7.1 implies that the increased throughput obtained by avoiding the intracell interference, sufficiently compensates for the silent periods, resulting in a higher average data rate than a simultaneous transmission. It should also be noted that since the proof is based on time averaging, ordering of the users is irrelevant, which means that multiuser diversity aspects are not taken into account. Hence, we may expect additional gains from channel adaptive schedulers. Now, consider two different classes of schedulers.

### 7.2.2 Slow Scheduling

A simple and robust type of scheduling policy is to ignore the instantaneous channel variations and determine a transmission schedule in advance. Thus, the input parameters to the scheduler are measured on long term basis and do not adapt to the small-scale variations. As the time average of $\xi_i(t)$ equals its ensemble mean $\tilde{\xi}_i$, the expected throughput of a user will asymptotically be

$$
R_i = \phi_i \frac{W}{\Gamma_i} \tilde{\xi}_i.
$$

(7.4)

A slow scheduler is defined by the different allocations of $\phi_i$, which could be assigned arbitrarily, subject to $\sum_{i=1}^N \phi_i = 1$ and some underlying fairness philosophy. Among them is the well known resource fair approach of round robin,

$$
\phi_i = \frac{1}{N}
$$

(7.5)

where the channel access is evenly allocated over all users. As comparison, a fully performance fair allocation would give equal throughput rather than time fractions to all users. This can be achieved by an equal throughput (ET)
scheduler

\[ \phi_i = \frac{\Gamma_i / \xi_i}{\sum_{j=1}^{N} \frac{\Gamma_j}{\xi_j}} \]  

(7.6)

which gives each user a throughput of \( R_i = W / \sum_{j=1}^{N} \frac{\Gamma_j}{\xi_j} \). The ET scheduler can be interpreted as a weighted RR scheduler, and in comparison allocates more channel time to users with unfavorable conditions in order to equalize the throughputs, to the price of a total throughput decrease. A more greedy strategy for allocating the \( \phi_i \), opposite to ET, is to allocate more time to users with favorable conditions and low quality requirements. Such a fractionally fair (FF) assignment could yield

\[ \phi_i = \frac{\xi_i / \Gamma_i}{\sum_{j=1}^{N} \frac{\xi_j}{\Gamma_j}} \]  

(7.7)

The next proposition follows straightforwardly.

**Proposition 7.2** The total throughput \( \sum_{i=1}^{N} R_i \) of RR is greater than that of ET but smaller than that of FF.

**Proof.** First we show that \( ET \leq RR \). By observing that \( f(x) = x + x^{-1} \geq 2 \) for nonnegative \( x \), we can write

\[ \sum_{i=1}^{N} x_i \sum_{i=1}^{N} \frac{1}{x_i} \geq N + 2 \sum_{i=1}^{N} (N - i) = N^2 \]

which also is the well-known arithmetic-harmonic mean inequality. This means that

\[ N \leq \sum_{i=1}^{N} \frac{\Gamma_i}{\xi_i} \frac{1}{W} \sum_{i=1}^{N} \frac{1}{N} \frac{W}{\Gamma_i} \xi_i \]

or equivalently

\[ NR_0 = \frac{NW}{\sum_{i=1}^{N} \frac{\Gamma_i}{\xi_i}} \leq \sum_{i=1}^{N} \frac{1}{N} \frac{W}{\Gamma_i} \xi_i. \]

To show that \( RR \leq FF \), the Chebyshev sum inequality\(^3\) gives

\[ \sum_{i=1}^{N} \frac{1}{N} \frac{W}{\Gamma_i} \xi_i \sum_{i=1}^{N} \frac{\xi_i}{\Gamma_i} \leq \sum_{i=1}^{N} \frac{\xi_i / \Gamma_i}{W} \xi_i \]

\(^3\)The Chebyshev sum inequality states that, if \( a_1 \leq a_2 \leq \ldots \leq a_N \) and \( b_1 \leq b_2 \leq \ldots \leq b_N \) are arbitrary real numbers, then

\[ \sum_{i=1}^{N} a_i \sum_{i=1}^{N} b_i \leq N \cdot \sum_{i=1}^{N} a_i b_i. \]

Equality holds iff, either \( a_1 = a_2 = \ldots = a_N \) or \( b_1 = b_2 = \ldots = b_N \).
which can be rewritten as

\[
\sum_{i=1}^{N} \frac{1}{N} \frac{W}{T_i} \xi_i \leq \sum_{i=1}^{N} \frac{\xi_i / T_i}{\sum_{j=1}^{N} \xi_j / T_j} \frac{W}{T_i} \xi_i.
\]

Equality is achieved iff, for all \( i, \xi_i / T_i = C \), where \( C > 0 \) is an arbitrary constant. □

In the following sections we will focus on a fast scheduler that will be resource fair, i.e., comparable to RR. Equivalently, since users get the same transmission power, the fairness criteria can be interpreted as every user is asymptotically allocated the same amount of energy. It can be noted that among the possible slow schedulers, the aggregate of the proportional changes in throughput of RR is not positive. For any feasible allocation of \( \phi_i' \neq 1/N \), the proportional fairness condition of [60] holds, \( \sum_{i=1}^{N} \frac{\xi_i'}{R_i} = \sum_{i=1}^{N} \frac{\phi_i - \phi_i'}{R_i} = 0 \).

### 7.2.3 Fast Scheduling

Now, consider a situation where for any time instant the vector of samples \( \xi = (\xi_i) \) is perfectly available for the scheduling decision. The suggested relatively best (RB) scheduler decides from a normalized channel state such that the preferred user is

\[
i^*(\xi) = \arg \max_{1 \leq i \leq N} \frac{\xi_i - \xi_i'}{c_i}
\]

Ties are broken with equal probability. The \( c_i \) are positive control parameters, resulting in different time fractions \( \phi_i \). Thus, the scheduling decision is made from a normalized channel state value, referred to as the “relatively best”. The consequence is that each user transmits only on good instants, but with a certain asymptotic channel access, controlled by the \( c_i \). Assume that the channels are mutually independent. For a Rayleigh fading channel, viz. a Rice factor \( \kappa = 0 \), the conditional probability that user \( i \) has a relatively better channel than user \( j \) is given by:

\[
\Pr \left[ \xi_j \leq \frac{c_i}{c_j} (\xi_i - \xi_i') + \xi_i' \mid \xi_i \right] = \begin{cases} 0, & \xi_j < 0 \\ 1 - e^{-\frac{c_i (\xi_i - \xi_i') + \xi_j}{\xi_j}}, & \frac{c_i (\xi_i - \xi_i') + \xi_j}{\xi_j} \geq 0. \end{cases}
\]

If we let \( \xi_{ii} = \max_{j \neq i} \{0, \xi_j - \frac{c_i}{c_j} \xi_j \} \), the resulting asymptotic time fraction assignment can be obtained as

\[
\phi_i = \int_{\xi_{ii}}^{\infty} \prod_{j=1}^{N} (1 - e^{-\frac{\phi_j (\xi_i - \xi_j) + \xi_j}{\xi_j}}) \frac{1}{\xi_i} e^{-\frac{\phi_j}{\xi_i}} d\xi_i. \tag{7.8}
\]
The feasible time fractions depend on the control parameters $c_i$ and the fading characteristics and cannot take any values.

**Example 7.2.1** Consider an example with $N = 2$, $c_1 = a\xi_1$ and $c_2 = \xi_2$. Evaluating the integral (7.8) we obtain

$$
\phi_1 = \begin{cases} 
(1 - \frac{a^{1-1}}{1+a^{(1-1)}}) & , 0 < a < 1 \\
1 & , a \geq 1.
\end{cases}
$$

(7.9)

Thus, the feasible time fraction in this case, becomes limited to $e^{-1} \leq \phi_1 \leq 1 - e^{-1}$. The time span can be extended with a slight modification of the scheduling rule. If we let $i^*(\xi) = \max_{1 \leq i \leq N} \xi_i/c_i$ and consider the same parameters,

$$
\phi_1 = \frac{1}{a + 1},
$$

Hence, the feasible time fractions in this case can take any value $0 \leq \phi_1 \leq 1$.

For fair comparison between schedulers, their asymptotic time fractions should be the same. Importantly, it can easily be shown that when $c_i = \xi_i$, the integral (7.8) reduces to $\phi_i = 1/N$ for any user $i$. From the variance of $\xi_i$ it is evident that the means and also variances of $(\xi_i - \xi)/\xi_i$ are the same for all users, also when $\kappa > 0$. Therefore, the RB algorithm can provide the same asymptotic fairness as RR for channels described by (2.5).

**Example 7.2.2** For further generalization, consider an example of two classes of users which could have different priorities. Let there be $N_A$ users with $c_i = a\xi_i$ and $N - N_A$ users with $c_j = \xi_j$. For $0 < a < 1$, the resulting time fractions become for any of the $N_A$ users

$$
\phi_i = \sum_{k=0}^{N_A-1} \sum_{l=0}^{N - N_A} \binom{N_A - 1}{k} \binom{N - N_A}{l} (-1)^{k+l} \frac{a^{-(1-1/a)^l} e^{-(1-1/a)^k}}{k + 1 + l/a}
$$

and for $a \geq 1$

$$
\phi_i = \sum_{k=0}^{N_A-1} \sum_{l=0}^{N - N_A} \binom{N_A - 1}{k} \binom{N - N_A}{l} (-1)^{k+l} \frac{a^{-(1-1/a)^l}}{k + 1 + l/a}
$$

where a binomial expansion is performed in (7.8). In Fig. 7.1, this is plotted for $N = 10$ and $N_A = 1, 2, \ldots, N$. The arrows denote increasing values of $N_A$. As expected, $\phi_i = 0.1$ at $a = 1$. It should be pointed out that the characteristic of the time fraction curve is very much dependent on the parameters $N$ and $N_A$, cf., the two dimensional case (7.9) above.

In [78], a suggested scheduler was shown to maximize the expected throughput subject to allocated time fractions. These time fractions could be assigned arbitrarily. In the case of $\phi_i = 1/N$ the RB scheduler is equivalent
Figure 7.1: Feasible time fractions for users in class A for different number of users $N_A = 1, 2, \ldots, 10$. The total number of users are $N = 10$ and the channel is Rayleigh fading.

to $i^*(\xi) = \arg\max_{i} (\log \xi_i - \log \xi_{i^*})$. According to Proposition 1 in [78], the RB would then be optimal in the sense it maximizes

$$\sum_{i=1}^{N} R_i = \sum_{i=1}^{N} \mathbb{E}[\log \xi_i 1_{i^*(\xi)=i}]$$

where the event indicator was defined in Chapter 5. Thus RB maximizes the expected throughput for $r_i(\xi_i) = \log \xi_i$ and can be regarded as a proportional fair scheduling rule. In comparison with the scheduler in [78], the RB scheduler, as defined by (7.2.3), has a smaller feasible region of time fractions. On the other hand the scheduler in [78] is more complex since it uses an iterative approximation algorithm for achieving the desired time allocation.

### 7.3 Scheduling Gain

Let $r_i(\xi_i)$ be the instantaneous throughput of user $i$. The asymptotic throughput is defined as

$$R_i = \mathbb{E}[r_i(\xi_i) 1_{i^*(\xi)=i}].$$

(7.10)
With an assumption of independent fading among users, we can express
\[ \Pr[s^*(\xi) = i] = \Pr\left[ \frac{\xi_i - \bar{\xi}_i}{c_i} > \max_{j \neq i} \frac{\xi_j - \bar{\xi}_j}{c_j} \right] \]
\[ = \prod_{j \neq i} F_{\xi_j} \left( \frac{c_j}{c_i} (\xi_i - \bar{\xi}_i) + \bar{\xi}_j \right) \]
where \( F_j(\cdot) \) is the cdf corresponding to \( f_j(\cdot) \). The scheduling gain can be computed from
\[ G_i = \frac{\int_{-\infty}^{\infty} r_i(\xi_i) \prod_{j \neq i} F_{\xi_j} \left( \frac{c_j}{c_i} (\xi_i - \bar{\xi}_i) + \bar{\xi}_j \right) f_{\xi_i}(\xi_i) d\xi_i}{\phi_i \int_{-\infty}^{\infty} r_i(\xi) f_{\xi_i}(\xi) d\xi} \]

(7.11)

Throughout the chapter, we will evaluate (7.11) by numerical integration and derive supplementary closed-form expressions, if possible. Let us henceforth consider \( \phi = 1/N \), that is \( c_i = \bar{\xi}_i \) and the throughput gain of RB compared to RR.

7.3.1 Unconstrained Bandwidth Case

Ricean Fading

The scheduling gain for the Rayleigh fading channel can, after a binomial expansion and some algebra, be written as
\[ G = \sum_{k=1}^{N} \binom{N}{k} (-1)^{k+1} \frac{1}{k} = \sum_{k=1}^{N} \frac{1}{k} \]

(7.12)

where the last equality is contained in Appendix C. From [31], \( G \approx \ln N + \gamma + 1/2N \) where \( \gamma \approx 0.577 \) is the Euler constant. That is, the more users to choose from, the higher benefit from multiuser diversity. This is in accordance with the simulation studies of [48, 49, 114]. It can be noted that (7.12) describes the total throughput gain in the cell as well the individual user gain. In comparison, this scheduling gain is exactly as that of selection diversity on a Rayleigh fading channel [100] with \( \bar{\xi}_i/N \) as the SNR in each diversity branch. Although \( G \) is increasing in \( N \), the average throughput per user decreases, and asymptotically \( R_i = O(\ln N/N) \). The closed form expression (7.12) accounts for the Rayleigh fading case. If we consider general \( \kappa \) factors, it can be shown that [100]
\[ F_{\xi_i}(x) = 1 - Q_1 \left( \sqrt{2\kappa}, \sqrt{\frac{2(\kappa + 1)}{\bar{\xi}_i}x} \right) \]

(7.13)

where \( Q_1(\cdot, \cdot) \) is the first-order Marcum Q-function\(^4\). The scheduling gain can be found by numerical integration. In Fig. 7.2, we plot the scheduling gain as

\(^4\)A common form equals \( Q_1(\alpha, \beta) \equiv \int_{\alpha}^{\infty} xe^{-\frac{x^2}{2}}I_0(\alpha x) dx \).
Figure 7.2: The scheduling gain for different Ricean factors $\kappa$ (linear) for a single cell. The dashed line shows the gain in the multicellular case with Rayleigh fading ($\kappa = 0$).

function of $N$ for different values of $\kappa$. The result of (7.12) corresponds to the curve of $\kappa = 0$. The results show that, the larger Rice factor, the less gain from the proposed scheduler but the gain is still significant for the depicted values.

**RAKE Receiver**

If a RAKE receiver is employed, combining signals from $L$ independent identically distributed Rayleigh fingers, the output channel state follows a $\chi^2$ distribution [100],

$$f_{\xi_i}(x) = \frac{x^{L-1}}{(L-1)^{\frac{L}{2}}} e^{-\frac{x}{L}}, \quad x \geq 0$$  \hspace{1cm} (7.14)

having a mean $L\xi_i$ and variance $L\xi_i^2$. Alternatively, (7.14) can be interpreted as capturing the maximum ratio combined channel state resulting from transmit diversity of $L$ independent Rayleigh fading paths. The cdf is given by

$$F_{\xi_i}(x) = 1 - e^{-\frac{x}{L}} \sum_{i=1}^{L} \frac{1}{(i-1)!} \left(\frac{x}{\xi_i}\right)^{i-1}. \hspace{1cm} (7.15)$$

The scheduling gain obtained by numerical integration is contained in Fig. 7.3. In this case, the scheduling gain decreases significantly as $L$ increases. This is
Figure 7.3: The scheduling gain on a Rayleigh fading channel for different number of RAKE fingers \( L \).

expected, since multipath combining can be regarded as time averaging of the channel quality and will therefore reduce the scheduling gain [114].

### 7.3.2 Constrained Bandwidth Case

The assumed rate relation (7.1) can be rather optimistic as it assumes a linear behavior between rate and SIR, i.e., effectively an infinite bandwidth. Such a relation may not be accurate if adaptive modulation- and coding techniques are employed.

**Maximum Rate Constraint**

Let us consider (7.1) but imposed with a maximum achievable instantaneous transmission rate. This may penalize one-by-one transmission, as the reduction in intracell interference may not directly translate to a higher transmission rate. If it is assumed that the transmission rate is constrained to \( r_t \leq \tilde{R} \), the Rayleigh
fading case (7.4) becomes

\[ R_i = \phi_i \frac{W}{\Gamma_i} \int_0^\infty \min \left\{ \frac{1}{\xi_i} \right\} d\xi_i \]

\[ = \phi_i \frac{W}{\Gamma_i} \xi_i \left( 1 - e^{-\frac{\nu_i \xi_i}{w_i}} \right). \] (7.16)

That is, the rates are scaled by the factor \(1 - e^{-\frac{\nu_i \xi_i}{w_i}}\), which simply is the probability that \(\xi_i\) is less than \(\Gamma_i \tilde{R} / W\). By redefining (7.6) and (7.7) with new mean values \(\xi_i (1 - e^{-\frac{\nu_i \xi_i}{w_i}})\), the relative performance of the slow schedulers would still hold, i.e., Proposition 7.2 would still hold. A general claim such as Proposition 7.1, seems difficult to verify in the constrained case. That is, without multiuser diversity based schedulers like RB, higher throughput than simultaneous transmissions may not necessarily be guaranteed. To find the time fractions, we have

\[ \phi_i \frac{W}{\Gamma_i} \xi_i \left( 1 - e^{-\frac{\nu_i \xi_i}{w_i}} \right) = E \left[ \min \left\{ \frac{W}{\Gamma_i} \frac{\xi_i \Delta_i}{\theta_i (1 - \Delta_i) \xi_i + 1}, \tilde{R} \right\} \right]. \] (7.17)

An upper bound on the time fractions can be obtained as

\[ \sum_{i=1}^N \phi_i \leq \sum_{i=1}^N \frac{\Delta_i}{(1 - e^{-\frac{\nu_i \xi_i}{w_i}})(\theta_i (1 - \Delta_i) \xi_i + 1)} \] (7.18)

\[ \leq \sum_{i=1}^N \Delta_i \cdot \max \left\{ \frac{W}{\Gamma_i \theta_i (1 - \Delta_i)}, 1 \right\} \] (7.19)

where (7.18) is due to Jensen’s inequality and (7.19) follows from \((1 - e^{-x})(k/x + 1) \geq \min\{k, 1\}\). If \(W / (\Gamma_i \theta_i) \leq 1 - \Delta_i\), there will be a static gain. The interpretation of this is that one-by-one transmission with a slow scheduler, not relying on multiuser diversity, could be favorable if \(\Gamma_i\) and \(\theta_i\) are large. That is, if the system works under low BERs and high intracell interference. For the RB scheduler, the asymptotic transmission rate can be obtained from

\[ R_i = \frac{W}{\Gamma_i} \int_0^\infty \min \left\{ \frac{\Gamma_i}{W} \tilde{R}, \xi_i \right\} \left( 1 - e^{-\frac{\xi_i}{\xi_i}} \right)^{N-1} \frac{1}{\xi_i} e^{-\frac{\nu_i \xi_i}{w_i}} d\xi_i \]

\[ = \frac{W \xi_i}{\Gamma_i N} \sum_{k=1}^N \binom{N}{k} (-1)^{k+1} \frac{1}{k} \left( 1 - e^{-\frac{\nu_i \xi_i}{w_i}} \right) \]

which gives

\[ G_i = \sum_{k=1}^N \binom{N}{k} (-1)^{k+1} \frac{1}{k} \left( 1 - e^{-\frac{\nu_i \xi_i}{w_i}} \right). \] (7.20)
Figure 7.4: The scheduling gain on a Rayleigh fading channel with a constrained maximum rate for different values of \( S = \Gamma_i \tilde{R}/W \tilde{\xi}_i \).

Compared to the unconstrained case, the scheduling gain here depends on the absolute value of \( \tilde{\xi}_i \). It can be seen that as \( \Gamma_i \tilde{R}/W \tilde{\xi}_i \to \infty \), the scheduling gain approaches the unconstrained case (7.12). For \( \Gamma_i \tilde{R}/W \tilde{\xi}_i \to 0 \), it follows that \( G \to \sum_{k=1}^{N} \binom{N}{k} (-1)^{k+1} = 1 \). Therefore, we conclude that an upper constraint on the instantaneous transmission rate still renders higher throughput for RB than RR. In Fig. 7.4, (7.20) is plotted as function of \( N \) for different ratios of \( \Gamma_i \tilde{R}/W \tilde{\xi}_i \). To obtain a measure of the average system gain, \( \tilde{\xi}_i \) can be regarded as a random variable comprised of shadowing and attenuation and could thus be averaged over.

Discrete Transmission Rates

In practice, the transmission rates are not only constrained by an upper level, but also confined to a set of discrete values. The consequence thereof is that improved channel states do not necessarily translate into higher rates, as the results in the foregoing section showed. Assume a model where there are \( K \)
discrete transmission rates, assigned to different intervals.

\[
    r_i(\xi_i) = \begin{cases} 
    0, & \xi_i < x_1 \\
    \frac{W}{\Gamma_i} x_k, & x_k \leq \xi_i < x_{k+1}, 1 \leq k \leq K - 1 \\
    \frac{W}{\Gamma_i} x_K, & \xi_i \geq x_K 
    \end{cases}
\]  

(7.21)

This can be seen as a quantization of the linear model (7.1). We use the model of Chapter 5, where the rates are geometrically related, such that for \( \mu > 1 \), \( x_k = \mu x_{k-1} \). Thus it follows that \( x_k = \mu^{k-1} x_1 \). If \( \mu \) is small, a fine granularity can be obtained, which will let performance approach the continuous case. However, this requires a large number of available rates \( K \). In real systems, with variable processing gain, the rates are often related with \( \mu = 2 \). Define \( x_{K+1} = \infty \) so that the asymptotic throughput can be found by evaluating

\[
    R_i = \frac{K}{N} \sum_{k=1}^{N} \int_{x_k}^{x_{k+1}} r_i(\xi_i)(1 - e^{-\frac{\xi_i}{\xi_i}}) e^{-\frac{\xi_i}{\xi_i}} \frac{1}{\xi_i} d\xi_i. 
\]

(7.22)

Since \( \int_{x_k}^{x_{k+1}} (1 - e^{-\frac{\xi_i}{\xi_i}}) e^{-\frac{\xi_i}{\xi_i}} d\xi_i = \sum_{k=1}^{N} \left( \begin{array}{c} N \\ k \end{array} \right) \frac{(-1)^{k+1}}{N} (e^{-\frac{\xi_i}{\xi_i}} - e^{-\frac{x_{k+1}}{\xi_i}}) \), it can be found that (7.22) equals

\[
    R_i = \frac{W}{\Gamma_i N} \sum_{k=1}^{N} \left( \begin{array}{c} N \\ k \end{array} \right) (-1)^{k+1} x_1 \frac{\xi_i}{\xi_i} \left( e^{-\frac{\xi_i}{\xi_i}} + \sum_{m=2}^{K} \mu^{m-2} (\mu - 1) e^{-\frac{x_{k+1}}{\xi_i}} \right). 
\]

(7.23)

The scheduling gain can be obtained by normalizing (7.23) with its first term \((k = 1)\) divided by \( N \). Clearly, \( R_i \) becomes a function of \( x_1 / \xi_i \) and for a fixed \( N \) it can be found that the scheduling gain does not necessarily attain its maximum when the absolute level of (7.23) does. To compare throughput performance with the RB scheduler in the continuous linear rate case, we plot (7.23) normalized with \( (W / \Gamma_i) / (N \cdot G) \)· \( G \) where \( G \) is defined by (7.12). In Fig. 7.5, the case of \( N = 10 \) and \( \mu = 2 \) is considered for different values of \( K \) and plotted as function of \( x_1 / \xi_i \). It can be seen that the maximum throughput in the discrete case is here only about 72% of the continuous case. This value is reached by quite a small number of rates though, and for the larger number of rates, the peak is rather insensitive to the choice of \( x_1 / \xi_i \).

**Logarithmic Rate**

If the intercell interference can be considered Gaussian and multiuser detection is not employed, the information theoretic-rate is given by

\[
    r_i(\xi_i) = W \log_2 (1 + \xi_i). 
\]

(7.24)

Consider a Rayleigh fading channel, then we can define

\[
    H(k, \xi) \overset{\text{def}}{=} \int_{0}^{\infty} \log_2 (1 + \xi) \frac{1}{\xi} e^{-\xi k} d\xi = \frac{e^{\frac{k}{\xi}}}{\xi k \log_2 2} \Gamma \left( \frac{k}{\xi} \right) 
\]

(7.25)
Figure 7.5: The relative throughput compared to a continuous linear rate on a Rayleigh fading channel with $N = 10$ users. The number of discrete rates is set to $K$. The x-axis denotes the relative threshold $x_i/\xi_i$.

where the last equality follows from

$$
\int_0^\infty \log_2(b + a\xi)e^{-\xi} \, d\xi = \frac{1}{c\log_2 2} \left( e^{\log_2 b} \Gamma \left( \frac{b}{a} \right) + \log_2 b \right)
$$

and we define the incomplete gamma function

$$
\Gamma(x) \overset{\text{def}}{=} \int_x^\infty t^{-1} e^{-t} \, dt.
$$

After some manipulations, see Appendix C, the scheduling gain can be shown to be

$$
G_i = \frac{1}{H(1,\xi_i)} \sum_{k=1}^N \binom{N}{k} (-1)^{k+1} kH(k,\xi_i).
$$

(7.26)

As in the previous case with an upper bounded rate, the scheduling gain depends on the absolute value of $\xi_i$. In Fig. 7.6, it can be seen that the scheduling gain is significantly lower than in the unconstrained case. A small $\xi_i$, results in that the log function is close to linear, in the region of interest.

### 7.4 Multicellular Case

In (7.1), it was assumed that $I_i + \nu$ was fixed and a single cell analysis was performed. Assuming that $I_i$ is also varying, the variance of $\xi_i$ increases, which
Figure 7.6: The scheduling gain on a Rayleigh fading channel for a logarithmic rate relation with different mean SIR.

suggests higher scheduling gain. Consider the following extensions to cover a varying intercell interference. Assume that the intercell interference is due to the first tier of base stations and that for any user $i$ in the desired cell, the intercell interferers $k$ have the same mean received power, $\bar{C}_k = C$. If, for simplicity, we let the background noise be $\nu = 0$, the result from Appendix A gives the distribution and density functions for $x \geq 0$

\[
F_i(x) = \Pr \left[ \frac{C_i}{\sum_{j \neq i} C_j} \leq x \right] = 1 - \left( \frac{\bar{C}_i}{Cc_i + C_i} \right)^6 \quad (7.27)
\]

\[
f_i(x) = \frac{6C\bar{C}_i^6}{(Cc_i + C_i)^7} \quad (7.28)
\]

where $\bar{C}_i$ is the mean received power of user $i$. We find that $\mathbb{E} \left[ \frac{C_i}{\sum_{j \neq i} C_j} \right] = \frac{C_i}{5C} \overset{\text{def}}{=} \xi_i$ and $\mathbb{E} \left[ \left( \frac{C_i}{\sum_{j \neq i} C_j} \right)^2 \right] = \frac{C_i^2}{10C^2}$. Thus the variance of $\xi_i$ becomes $\frac{3}{2} \xi_i^2$, which is
higher than for the single cell Rayleigh case. Then, for any user $1 \leq i \leq N$,

$$
F_\xi(x) = 1 - \left(\frac{5\xi_i}{x + 5\xi_i}\right)^6, \quad x \geq 0 \quad (7.29)
$$

$$
f_\xi(x) = \frac{6(5\xi_i)^6}{(x + 5\xi_i)^7}, \quad x \geq 0. \quad (7.30)
$$

Again consider the decision variable $\frac{\xi_j}{\xi_i}$, which can be shown to give $\phi_i = 1/N$. Since

$$
\prod_{j=1}^{N} \Pr[\xi_j \leq \frac{\xi_j}{\xi_i} \xi_i | \xi_i] = \left(1 - \left(\frac{1}{\frac{\xi_i}{5\xi_i} + 1}\right)^6\right)^{N-1}
$$

the asymptotic data rate can be obtained as

$$
R_i = \frac{W}{F_i} \int_0^\infty \xi_i \left(1 - \left(\frac{1}{\frac{\xi_i}{5\xi_i} + 1}\right)^6\right)^{N-1} \frac{6(5\xi_i)^6}{(\xi_i + 5\xi_i)^7} d\xi_i
$$

$$
= \frac{W}{F_i} \frac{5}{6N} \sum_{k=1}^{N} (-1)^{k+1} \frac{N}{k} \frac{1}{k - 1/6}. \quad (7.31)
$$

The scheduling gain can be expressed as

$$
G = \frac{5}{6} \sum_{k=1}^{N} (-1)^{k+1} \frac{N}{k} \frac{1}{k - 1/6} \quad (7.32)
$$

This is plotted in Fig. 7.2 as a dashed line, where it can be seen that the gain is larger than for the single cell case.

### 7.5 Measurement Delay

The gain of the fast scheduler relies, to a large extent, on that it can base its actions on reliable channel measurements. The previous analysis assumed perfect channel estimates. In a more practical scenario, the estimates must be obtained in the presence of noise and time delays. Depending on the channel variation, outdated measurements may cause erroneous scheduling actions and will therefore degrade performance. Let us consider the RB scheduler subject to a time delay $\delta$ in the measurements as follows:

$$
i^*(\xi(t)) = \arg \max_{1 \leq i \leq N} \frac{\xi_i(t - \delta) - \tilde{\xi}_i}{c_i}
$$

The implication is that the scheduling decision lags $\delta$ time units to the channel. Clearly the performance will depend on the correlation of the channel fading. If
the correlation is high, the delay tolerance is higher since the correct user is likely to be scheduled. If the correlation is zero and $\delta > 0$, the scheduler effectively makes a random pick and performance will reduce to the RR scheduler. If $\delta = 0$, we have the case previously analyzed.

Consider the following stochastic process for modeling the fast channel variation. For every user $i$, define two independent Gaussian processes $\{Z_i^{(k)}(t)\}$, $k = 1, 2$, with stationary means and variances, 0 and $\overline{\xi_i}/2$, respectively. Each process evolves independently according to the following correlated process,

$$Z_i^{(k)}(t + \delta) = \rho Z_i^{(k)}(t) + \sqrt{1 - \rho^2} N_i^{(k)}(t)$$

(7.33)

where $\{N_i^{(k)}(t)\}$ are independent normal random variables with zero mean and variance $\overline{\xi_i}/2$. The correlation can for a mobile radio channel be related to the radio environment by Clarke’s model, $\rho = J_0 \left( \frac{2\pi \nu f_0}{c} \right)$, where $f$ is the carrier frequency, $\nu$ the velocity, $c$ the speed of light and $J_0(\cdot)$ is the zeroth-order modified Bessel function of the first kind. The channel fading process is obtained by

$$\xi_i(t) = (Z_i^{(1)}(t))^2 + (Z_i^{(2)}(t))^2$$

(7.34)

which is a correlated exponential process with mean $\overline{\xi_i}$. Assume for simplicity that all users have the same correlation $\rho$. The throughput for the model (7.1), is obtained by solving the following integral (see Appendix B),

$$R_i = \frac{W}{\Gamma_i} \int_0^\infty \xi_i(t + \delta) \prod_{j=1}^N (1 - e^{-\frac{\xi_j(t)}{\overline{\xi}_j}}) f_{\xi_i(t)}(\xi_i(t)) d\xi_i(t)$$

$$= \frac{W \overline{\xi_i}}{\Gamma_i N} \left( \rho^2 \sum_{k=1}^N \frac{1}{k + 1 - \rho^2} \right).$$

(7.35)

Thus the scheduling gain, $G = \rho^2 \sum_{k=1}^N 1/k + 1 - \rho^2$ is a linear function of the power correlation $\rho^2$. To numerically verify this, we can use the joint probability density function (see, e.g., [100]) of $\xi_i(t)$ and $\xi_i(t + \delta)$,

$$f_{\xi_i(t), \xi_i(t+\delta)}(x, y) = \frac{1}{(1 - \rho^2)\overline{\xi_i}^2} e^{-\frac{x+y}{(1-\rho^2)\overline{\xi_i}}} \cdot I_0 \left( \frac{2\sqrt{\rho^2 xy}}{(1 - \rho^2)\overline{\xi_i}} \right), \quad 0 \leq \rho < 1$$

where the mean is constant during $[t, t + \delta]$ due to the stationarity assumption. The pdf can also be used to represent estimation errors, where $\rho$ is the correlation between actual and measured value. The asymptotic data rate can be evaluated by computing the integral,

$$R_i = \frac{W}{\Gamma_i} \int_0^\infty \int_0^\infty \xi_i(t + \delta) \left( 1 - e^{-\frac{\xi_i(t)}{\overline{\xi}_i}} \right)^{N-1} \times$$

$$f_{\xi_i(t), \xi_i(t+\delta)}(\xi_i(t), \xi_i(t + \delta)) d\xi_i(t) d\xi_i(t + \delta).$$

(7.36)
In Fig. 7.7 we plot the scheduling gain based on (7.35) for $N = 10$. For comparison, we evaluate the integral (7.36) numerically and plot the corresponding scheduling gain. It can be seen that the closed form expression (7.35) is matched by the numerical integration.

### 7.6 Multiuser Scheduling

Results in this chapter, and other investigations [49], show that only one user should be allocated to the channel at any instant, if the rate is growing linearly with channel quality. With a nonlinear relation between effective transmission rate and channel quality, it is not certain that scheduling one single user may achieve highest throughput. The HSDPA mode will primarily use time multiplexing of the channelization codes but code multiplexing is also supported [88]. In Section 7.3.2 it was shown that under a logarithmic relation, the RB scheduler still outperformed RR. However, this result is not tantamount to that RB also is better than conventional CDMA transmission with all users on the channel. Therefore, we generalize the RB scheduler to allocate more than one user on the channel. For pertinent comparison, we keep the fairness requirement that every user asymptotically shall be given the same amount of energy. For simplicity, we assume that $P$ is shared equally among the users. Hence, we do not adapt.
the transmit power to the channel conditions, only to the number of users on the channel. Consider the case where there are \( k \) users allocated to the channel giving each user a power of \( P/k \). In the absence of multiuser detection, the instantaneous transmission rate of user \( i \) can be assumed to follow a logarithmic relation

\[
    r_i(\xi_i, k) = W \log_2 \left( 1 + \frac{\xi_i}{\theta(k-1)\xi_i + k} \right). \tag{7.37}
\]

### 7.6.1 Generalized Scheduler

Now, let us generalize the RB scheduler by in any instant ordering the users such that their decision variables follow

\[
    \frac{\xi_1 - \xi_1}{\xi_1} \geq \ldots \geq \frac{\xi_N - \xi_N}{\xi_N}
\]

and schedule the \( L + 1 \) first users. Ties are broken with equal probability. The values \( L = 0, \ldots, N-1 \) span one-by-one transmission to regular CDMA with \( N \) users on the channel.

For a Rayleigh fading channel, the conditional probability that user \( i \) is scheduled will then equal

\[
    \Pr[i^*(\xi) = i] = \frac{L}{j} \left( 1 - e^{-\xi_i} \right)^{N-1-j} e^{-\xi_i j}. \tag{7.38}
\]

Averaging (7.38) over (2.5), it can be found that any user will access the channel a fraction \( (L+1)/N \) of the time. Hence, with a fixed power \( P/(L+1) \) per user, all users will asymptotically be allocated the same amount of energy. For a given number of users \( k \) the asymptotic throughput is again defined as

\[
    R_i(k) \overset{\text{def}}{=} \mathbb{E}[r_i(\xi_i, k)1_{i^*(\xi) = i}]. \tag{7.39}
\]

After some further manipulations, the asymptotic throughput of (7.38)-(7.39) can be obtained on closed-form

\[
    R_i(L + 1) = \int_0^\infty r_i(\xi, L + 1) \Pr[i^*(\xi) = i] f_\xi(\xi) \, d\xi

\[
    = \frac{L+1}{N+1-j} \sum_{k=1}^{L+1} \binom{N}{j} \binom{N+1-j}{k} \left( \frac{1}{N \log_2 2} \right) \times

    \frac{j}{(N+1-j)(N+1-k)} \times

    \left( V\left( \frac{(L+1)(N+1-j)}{\xi_i(\theta L + 1)} \right) - V\left( \frac{(L+1)(N+1-k)}{\xi_i \theta L} \right) \right)
\]

where \( V(x) = e^x \Gamma(x) \) and \( \Gamma(x) \) is the incomplete gamma function.
Figure 7.8: The scheduling gain as function of channel quality $\tilde{\xi}_i$ for different orthogonality factors $\theta = 0, 0.1, \ldots, 1$. The number of scheduled users equals 1 ($L = 0$) and the total is $N = 10$.

### 7.6.2 Multiuser Scheduling Gain

To compare with regular CDMA, $L = N - 1$, we define the multiuser scheduling gain as the normalized throughput

$$\tilde{G}_i \overset{\text{def}}{=} \frac{R_i(L + 1)}{R_i(N)}, \quad (7.41)$$

where we can express

$$R_i(N) = \int_0^\infty r_i(\xi_i, N) f_i(\xi_i) d\xi_i$$

$$= \frac{1}{\log_2 2} \left( V \left( \frac{N}{\xi_i(\theta(N - 1) + 1)} \right) - V \left( \frac{N}{\xi_i(\theta(N - 1))} \right) \right). \quad (7.42)$$

In Fig. 7.8 we plot the scheduling gain for $L = 0$, i.e., one user is scheduled at a time and $N = 10$. We observe that the larger $\theta$, the more gain from scheduling. It is also noteworthy that for perfectly orthogonal codes, $\theta = 0$, there exists a region ($\xi_i \gtrsim 2.2$) where regular CDMA transmission offers higher average throughput than the scheduled system. Consequently, in that case it is more
Figure 7.9: The scheduling gain as function of channel quality $\tilde{\xi}_i$ for different orthogonality factors $\theta = 0, 0.1, \ldots, 1$. The number of scheduled users equals 2 ($L = 1$) and the total is $N = 10$.

favorable to let the user transmit the whole time with $1/N$th of the power than $1/N$:th of the time with full power. This is mainly due to that (7.37) flattens out for large $\xi_i$. To fully assess performance, we study the limiting cases of high and low average channel quality. Using relations from Appendix C, the scheduling gain (7.41) reduces to the form

$$\lim_{\xi_i \to \infty} \hat{G} = \frac{L + 1}{N} \frac{\log_e \left( 1 + \frac{1}{\theta L} \right)}{\log_e \left( 1 + \frac{1}{\theta(N-1)} \right)}$$

(7.43)

Further, it can be shown that for any $L$,

$$\lim_{\xi_i \to \infty} \hat{G}_{|\theta=0} = \frac{L + 1}{N} \leq 1.$$

Thus, if the channels are perfectly orthogonal and $\tilde{\xi}_i$ is large, regular CDMA transmission is favored. That is, for high $\tilde{\xi}_i$, reducing the interference by scheduling results in an average throughput loss. This is due to that the effective transmission rate increases very little from the intracell interference reduction but the channel utilization reduces to $(L + 1)/N$. For $\theta > 0$ and $L = 0$, the gain
Figure 7.10: The scheduling gain as function of channel quality \( \tilde{\xi} \) for different orthogonality factors \( \theta = 0, 0.1, \ldots, 1 \). The number of scheduled users equals 3 \( (L = 2) \) and the total is \( N = 10 \).

This is expected since \( R_t \) is for \( L = 0 \) independent of \( \theta \) and an increasing function of \( \tilde{\xi} \). \( R_t(N) \) on the other hand, can be bounded by Jensen's inequality to

\[
R_t(N) \leq W \log_2 \left( 1 + \frac{1}{\theta(N - 1)} \right),
\]

which gives an increasing scheduling gain. Therefore, the curves in Fig. 7.8 for \( \theta > 0 \), will eventually exceed unity as \( \tilde{\xi} \) increases. As the channel quality degrades, we find that the limit value takes the following form:

\[
\lim_{\tilde{\xi} \to 0} \tilde{G} = \sum_{j=1}^{N+1} \sum_{k=1}^{N+1-j} \binom{N}{j} \binom{N+1-j}{k} \frac{(-1)^{N+1-j-k}}{L+1} \times \frac{j^k}{(N+1-j)(N+1-k)^2} \tag{7.44}
\]
In particular for $L = 0$, it can after some algebra be shown that

$$
\lim_{\xi \to 0} \hat{G}_{L=0} = \sum_{k=1}^{N} \binom{N}{k} (-1)^{N-k} \frac{k}{(N + 1 - k)^2}
= \sum_{k=1}^{N} \frac{1}{k}
$$

(7.45)

where the last equality is shown in Appendix C. As opposed to the case of large $\xi$, here, the scheduler gains over regular CDMA when $N$ increases. It is noteworthy that $G$ is in this case exactly equal to the scheduling gain that was found in (7.12) when the throughput was compared to that of a round robin scheduler, under a linear rate relation. This result is expected since $\log(1 + \xi) = \xi + O(\xi^2)$, i.e., a linear rate is well approximated in this region. In Fig. 7.9 and 7.10, we plot $G$ when two and three users are scheduled, $L = 1, 2$. The region where the scheduling gain is larger than unity, is here broadened compared to the case of $L = 0$ but the peak value at $\xi = 0$ is lower. Here the gain always decreases with increasing $\xi_i$.

### 7.6.3 System Multiuser Scheduling Gain

As was found, the scheduling gain depends on the actual value of $\tilde{\xi}_i$. To determine the gain of a system, the variations in $\tilde{\xi}_i$ must be accounted for. By regarding the throughput $R_i$ as a conditional result, the system gain can be obtained by averaging over $\tilde{\xi}_i$. If we let $S_i$ be lognormally distributed with zero mean and standard deviation $\sigma_s$ dB, then we can define $\tilde{\xi}_i$ as a random variable

$$
\tilde{\xi}_i = S_i \cdot \frac{K_0}{r_i^\alpha} \cdot \frac{P}{I_i + \nu}, \quad K_0 > 0
$$

(7.46)

where $r_i$ is the distance to the transmitter, also being as a random variable. If we consider a circular cell with uniformly dispersed mobiles within the radius $[r_{min}, r_{max}]$, the density function becomes for $x \geq 0$

$$
\int_{\xi} (x) = \frac{1}{\sqrt{2\pi} \sigma_s x \log_e 10} \int_{\log_e (10 \log_{10} K_0 P_{r_{max}^\alpha} / (I_i + \nu))^2}^{\log_e (10 \log_{10} K_0 P_{r_{max}^\alpha} / (I_i + \nu))^2} e^{-\frac{(10 \log_{10} K_0 P_{r_{max}^\alpha} / (I_i + \nu))^2}{2\sigma_s^2}} \frac{2r}{1 - \frac{r^2}{r_{max}^2}} dr
$$

(7.47)

where $C = 10 \log_{10} (K_0 P_{r_{max}^\alpha} (I_i + \nu))$ is the expected SIR in dB at the cell border. The system scheduling gain can be obtained by averaging (7.40) and (7.42) over (7.47) and taking their ratio. In Fig. 7.11, we plot results obtained from numerical integration for different values of $L$ and $\theta$. Here $r_{max} = 1000 \cdot r_{min}$, $\alpha = 4$ and $\sigma_s = 6$ dB. It can be verified that, except for the case $L = 0$ and $\theta = 1$, a higher average quality $C$ provides less system gain from scheduling. If however, the orthogonality factor $\theta$ is known to take large values, single-user transmissions are preferable. It is observed that the larger $L$ becomes, the less the span in system gain becomes between $\theta = 0$ and 1.
Figure 7.11: The average scheduling gain as function of channel quality $C$ for different orthogonality parameters. The three upper curves show $\theta = 1$ and the three lower $\theta = 0$. The number of scheduled users equals 1-3 ($L = 0, 1, 2$).

### 7.7 Concluding Remarks

In this chapter we have suggested and evaluated channel adaptive schedulers for various environments. The fast scheduler base its decision on average and instantaneous channel quality and provides different individual asymptotic time access to the channel. However, the main fairness criteria was here to give the users access to the channel the same fraction of time, or alternatively the same amount of energy. The scheduling gain over a round robin scheduler was shown to reduce to that of selection diversity and our closed form expressions are in accordance with the principal behavior of several simulation studies. The gain of a fast scheduler, should however, be seen in the light of a higher complexity and the need for a feedback loop. Also, it is heavily dependent on the achievable rate for a given channel quality. If there is a diminishing return in the SIR to rate relation, the scheduling gain becomes higher for channels with low average SIRs. The performance loss due to time delays or noisy estimates was also quantified. In practice, uncertainty of the estimate is due to hardware-, propagation- and scheduling delays as well as measurement errors. Motivated by Proposition 7.1, only one user per cell was scheduled at a time. This may not necessarily be the most efficient way if the rate is a nonlinear function of the SIR. For such
a scenario, the scheduler could be generalized to select several users. In that case, scheduling of several users simultaneously was found to have merits when the average channel quality was good. If the orthogonality among the codes is small ($\theta$ large), one user should be allocated to the channel at a time. In comparison, the PF scheduler suggested in [48, 49, 114], normalizes with a filtered throughput value. Small values of the time constant $t_c$ result in more emphasis on controlling delays whereas large values utilize channel variations, which will give higher throughput. Our scheduler normalizes with the channel mean, assumed to be estimated accurately. The consequence thereof is that short-term fairness will not be controlled. Hence, suitable applications are limited to best-effort data with high elasticity. Since the throughput will be lower for small values of $t_c$, one would expect that the larger the $t_c$, the closer the throughput of the PF scheduler will approach the results presented in this chapter. Some of the analysis performed here, considered a Shannon type of rate mapping. This implicitly assumes the interference to be Gaussian, which may be less valid when few users are scheduled. Additional work is required to validate this assumption more rigorously. Further extensions of this work include how scheduling in higher layers could be intertwined with the given channel schedulers. Another issue is also to incorporate power control, which could be used to provide intercell scheduling. By time multiplexing transmissions from different base stations, the variance of the fading could be increased, which could lead to even higher multiuser diversity effects.
Chapter 8

Linear Successive
Interference Cancellation

POWER control is an effective and necessary instrument for controlling the multiple access interference between cells and, in particular for CDMA, alleviating near-far effects within the cells. In addition to interference reduction by power control, the full potential of the DS-CDMA technique may be obtained by exploiting useful information of the nonwhite interference, which is colored by the structure of the spreading waveforms. Since the spreading waveforms are known at the receiver, at least in the uplink direction, multiuser detection is a viable option. Successive interference cancellation is a sub-optimum but simple scheme, which successively subtracts off the decoded signal from the composite received signal. Subsequently decoded users will thereby experience reduced interference. If done in conjunction with proper power control, this effect will translate into higher system capacity. In this chapter, we consider detection and power control for a linear SIC receiver. A linear receiver is characterized by that only linear operations may be performed on the received signal vector. Hence it can be built around simple processing elements of the matched filter concept. Therefore, for this receiver, the required number of cancellation steps is linear in the number of users.

8.1 Introduction

In order to implement ideal SIC detection, the receiver would typically need to have knowledge of all spreading waveforms in the network. Usually, the mobile has knowledge only about its own spreading waveform, whereas the base station has access to all spreading waveforms within the cell. For that reason, a significant thrust of work has focused on the uplink, where the canceled users are restricted reside in the same cell. For SIC, signal parameters such as am-
Amplitude, phase and timing are sometimes needed in the cancellation process to regenerate the detected signal. The accuracy in estimating these parameters has influence on the detection of each user's signal and resulting decoding errors can potentially propagate, yielding imperfect cancellation and an unreliable decoding process. To maintain acceptable decoding performance, the power control must take into account and mitigate the effects of incomplete cancellations. As the remaining interference becomes less with progressing cancellation, power control for multiuser detection will differ from the one of the single user spread spectrum receiver. Integrated power control and SIC for a single cell has been considered for both single-rate systems [29,30,46,70,89,117] and multi-rate systems [99,115]. However, a practical power control algorithm should be also able to perform in a multicellular network, preferably in a distributed fashion. Therefore, it should be fast and agile to varying channel and traffic conditions, to assure that the powers are rapidly directed to levels at which the users meet their QoS requirements.

In this chapter, we will consider SIC for a DS-CDMA system taking into account effects of imperfect-, partial- and limited cancellation. We assume a receiver based on the same principle as in [29,89], which uses soft feedback cancellation by subtracting off signals based on spreading waveform correlations. This method eliminates the need for direct amplitude estimation and the risk of error propagation due to hard decisions when regenerating the signal to be canceled. A drawback though is that, due to cross-correlation between the spreading waveforms, each cancellation stage introduces extra noise. Further, imperfect cancellation may occur when the signals undergo channel fading. Extension of the principle includes partial- and limited cancellation. When performing partial cancellation, purposely only a fraction of the signal is canceled. A positive consequence thereof is reduced cancellation noise. Limited cancellation is a way to reduce the complexity of the receiver by canceling only a subset of all signals. To minimize the transmitter powers, the problem of optimal power allocation for this concept is considered. First, we determine a general closed form power solution for a single cell. It is shown that the same type of closed form expression applies to many other receiver models as well, which serves as a succinct basis to compare capacity between them. Then, we investigate the decoding order of the SIC and find that under certain conditions, users should be decoded by the order of their path gains in order to minimize the total transmitted power. For the multicellular cell case, two iterative distributed power control algorithms, which converge to the optimal solution, are suggested.

8.2 Receiver Analysis

The main principle of SIC is that, for a given decoding order, the received signal is passed through a first correlator. The soft output of the correlator is then used to detect the first signal and also used to cancel the detected signal by subtraction
from the total input signal. From the remaining signal, subsequently the next signal is detected, and so on.

### 8.2.1 Single-rate System

Let us for analytical simplicity consider a flat fading channel. The received lowpass equivalent signal can be expressed as

\[
r(t) = \sum_{i=1}^{N} h_i(t) \sqrt{g_i} P_i a_i(t - \tau_i) b_i(t - \tau_i) + n(t)
\]

(8.1)

where \( h_i(t) \) is a stationary complex Gaussian process. Moreover, let us redefine the cross-correlation from (2.7),

\[
\hat{\beta}_{ik} = \frac{1}{\sqrt{T}} \int_{0}^{T} h_i(t + \tau_k) a_i(t - \tau_i) a_k(t) b_i(t - \tau_i + \tau_k) dt,
\]

now including the channel fading. The output sample of the correlator, for the first symbol, of user 1 becomes

\[
y_1 = \frac{1}{\sqrt{T}} \int_{\tau_1}^{T+\tau_1} r(t) a_1(t - \tau_1) dt
\]

\[
= \sum_{i=1}^{N} \sqrt{g_i} P_i \hat{\rho}_{i1} + n_1 \overset{\text{def}}{=} \sqrt{g_i} P_i \hat{\rho}_{i1} + \eta_1
\]

where \( n_1 = \frac{1}{\sqrt{T}} \int_{\tau_1}^{T+\tau_1} n(t) a_1(t - \tau_1) dt \). Now, this soft output is modulated and fed back such that the input signal to the correlator, for the first symbol of user 2 becomes

\[
y_2(t) = r(t) - y_1 \sqrt{R} a_1(t - \tau_1) \cap_T (t - \tau_1)
\]

which gives its output sample\(^1\)

\[
y_2 = \frac{1}{\sqrt{T}} \int_{0}^{T} y_2(t + \tau_2) a_2(t) dt
\]

\[
= \epsilon_{12} \sqrt{g_1} P_1 \hat{\rho}_{12} + \sum_{i=2}^{N} \sqrt{g_i} P_i \hat{\rho}_{i2} - \eta_1 \sqrt{R} \hat{\rho}_{12} + n_2 \overset{\text{def}}{=} \sqrt{g_1} P_1 \hat{\rho}_{12} + \eta_2
\]

---

\(^1\) Here the effect of uncanceled future bits is ignored and it is assumed that \( \int_{0}^{T+\tau_1-\tau_2} a_1(t - \tau_2) a_2(t) dt = \rho_{12} \). This is reasonable if the system is (close to) synchronous or if data can be buffered for simultaneous cancellations.
where
\[
\epsilon_{12} = \sqrt{W} (\hat{\rho}_{12} - \sqrt{R}\hat{\rho}_{11}\rho_{12}) \\
= \sqrt{\frac{W}{T}} \int_0^T \left( h_1(t + \tau_2)h_1(t - \tau_1 + \tau_2) - \frac{1}{T} \int_0^T h_1(t + \tau_1)h_1(t)dt \right) a_1(t - \tau_1 + \tau_2)a_2(t)dt, \tag{8.3}
\]
and \(\rho_{ik}\) is defined as in (2.8). If \(h_i(t)\) is slowly varying, it can be noted from (8.3), that \(\epsilon_{12} \approx 0\). Hence, the faster the channel varies the more imperfect cancellation. Generalizing, the input to correlator \(k\) takes the form
\[
y_k(t) = r(t) - \sum_{i=1}^{k-1} y_i \sqrt{W} a_i(t - \tau_i) \cap_T (t - \tau_i) \tag{8.4}
\]
with the output sample
\[
y_k = \frac{1}{\sqrt{T}} \int_0^T y_k(t + \tau_k) a_k(t) dt \\
= \sqrt{\frac{W}{T}} \sum_{i=1}^{k-1} \epsilon_{ik} \sqrt{\frac{g_{ik}}{W}} \leq \sum_{i=k}^N \sqrt{\frac{g_{ik}}{W}} \rho_{ik} - \sum_{i=1}^{k-1} \sqrt{R} \tilde{\rho}_{ik} + n_k \tag{8.5}
\]
defined as \(\sqrt{W} (\hat{\rho}_{ik} - \sqrt{R}\hat{\rho}_{ii}\rho_{ik})\). The receiver is depicted in Fig. 8.1. It will be assumed that for \(i \neq k\), var[\(\rho_{ik}\)] = \(\varphi/W\) and var[\(\rho_{ik}\)] = \(\tilde{\varphi}/W\) where \(0 < \{\varphi, \tilde{\varphi}\} \leq 1\). With the assumption of the terms in (8.5) being Gaussian and mutually independent, the variance of the interference becomes
\[
\sigma_k^2 = \text{var}[\eta_k] = \sum_{i=1}^{k-1} \theta_{ik} \frac{g_{ik}}{W} + \sum_{i=k+1}^N \frac{g_{ik}}{W} + \sum_{i=1}^{k-1} \frac{R}{W} \sigma_i^2 + N_0 \tag{8.6}
\]
where \(\theta_{ik} = \text{var}[\epsilon_{ik}]\). Averaged over time, following the steps in Chapter 2, the post-detection quality equals
\[
\left( \frac{E_b}{I_0} \right)_k = \frac{W}{R} \sum_{i=1}^{k-1} \theta_{ik} g_{ik} p_i + \sum_{i=k+1}^N \frac{g_{ik}}{W} + \sum_{i=1}^{k-1} \frac{R\sigma_i^2}{W} + \nu \tag{8.7}
\]
with \(\nu = N_0 W\). As can be observed from the third term of (8.6), every stage of the SIC detector experiences extra noise accumulated from the previous stages even when \(\theta_{ik} = 0\). This agrees with the results in [29,89]. It can be noted that this receiver, due to the accumulated cancellation noise, can potentially give worse performance than conventional detection.

\(^2\)For mathematical convenience, we assume that \(\eta_k\) is zero-mean and uncorrelated of the data symbol of user \(k\). It is further assumed that the fading process is ergodic and has zero mean and unit variance.
Figure 8.1: Simplified block diagram of a limited linear SIC detector for DS-CDMA signals. The first $M + 1$ users experience canceled interference whereas the others deploy conventional single-user detection.
8.2.2 Multirate System

To generalize to a multirate system, the analysis can be carried out with different symbol times $T_i$. If the users are ordered such that $T_1 \geq T_2 \geq \ldots \geq T_N$, the receiver output of each user's symbol can be found by small changes in notation.

$$y_i = \frac{1}{\sqrt{T_i}} \int_0^{T_i} y_i(t + \tau_i) a_i(t) \, dt$$

$$y_k(t) = r(t) - \sum_{i=1}^{k-1} y_i \sqrt{R_i} a_i(t - \tau_i) \cap T_i (t - \tau_i)$$

If there is no specified decoding order, the structure becomes more complex.

$$\hat{y}_{i,j} = \frac{1}{\sqrt{T_i}} \int_{T_i}^{(t+1)T_i} y_i(t + \tau_i) a_i(t) \, dt$$

$$y_k(t) = r(t) - \sum_{i=1}^{k-1} \sum_{j=0}^{L_{k,i}-1} \hat{y}_{i,j} \sqrt{R_i} a_i(t - \tau_i) \cap T_i (t - iT_i)$$

$$L_{k,i} = \max \left\{ 1, \left\lfloor \frac{T_k}{T_i} \right\rfloor \right\}$$

The value of $L_{k,i}$ gives integration of interferers during the whole symbol interval of the desired signal. With the same assumptions as in the previous case, the post-detection quality of the first detected symbol of user $k$ equals

$$\left( \frac{E_b}{I_0} \right)_k = \frac{W}{R_k} \cdot \frac{g_k p_k}{\sum_{i=1}^{k-1} \theta_{ik} g_i p_i + \sum_{i=k+1}^{N} \hat{g}_i p_i + \sum_{i=1}^{k-1} \hat{g}_i R_i \sigma_i^2 + \nu}$$

8.3 Power Control and User Capacity

Let us assume again that each user $k$ requires a maximum tolerable bit error rate which can be mapped into an equivalent minimum $E_b/I_0$ value, denoted $\Gamma_k$. Thus, as described in Chapter 3, a power control problem arises to determine the individual powers $p_k$ so that $(E_b/I_0)_k \geq \Gamma_k$. The problem of optimal power allocation, for any user $1 \leq k \leq N$, reduces to an equation of the form

$$\Gamma_k = \frac{W}{R_k} \sum_{i=1}^{k-1} \theta_{ik} g_i p_i + \sum_{i=k+1}^{N} \hat{g}_i p_i + \sum_{i=1}^{k-1} \hat{g}_i R_i \sigma_i^2 + \nu$$

where $\theta_{ik}$ for simplicity is assumed the same for all $i$ and independent of $k$. This can be seen as a worst case scenario when choosing $\theta$ as the maximum of all $\theta_{ik}$. The SIC efficiency is captured by $\theta$, where $\theta = 0$ is referred to as perfect SIC. This type of problem formulation, results in a linear equation system like (3.2) and the resulting power solution will be Pareto optimal.
8.3.1 Minimum Power Allocation

To probe further into issues on achievable capacity, i.e., if the selected $R_k$ and $\Gamma_k$ guarantee a feasible power solution, we start by obtaining the solution to the equation system (8.14). After some tedious algebra, we find that the solution can be written on this compact form

$$p_k = \frac{\nu \ w_k R_k \Gamma_k}{g_k \ \frac{1}{1 + \hat{\theta} R_k \Gamma_k} \left(1 - \sum_{j=1}^{N} \frac{w_j R_j \Gamma_j}{1 + \hat{\theta} R_j \Gamma_j} \right),}$$

(8.15)

where we have defined

$$w_k = \begin{cases} 1, & k = 1 \\ \prod_{i=1}^{k-1} q_i, & k > 1 \end{cases}$$

(8.16)

with

$$q_i = \frac{1 + \hat{\theta} R_i \Gamma_i \left(1 + \frac{R_i \Gamma_i}{\hat{\theta} \Gamma_i} \right)}{1 + \hat{\theta} R_i \Gamma_i}.$$\hspace{1cm} (8.17)

**Remark 8.1** It can be shown that the more general problem obtained when $\theta$ is replaced by $\theta_i$ in (8.14), is covered by replacing $\theta$ with $\theta_i$ in (8.17).

Further analysis shows that the structure of (8.15) applies to other receivers as well. The receiver models in [99,115] can be covered by excluding the accumulated cancellation noise. That is the recursive term of $\sigma_k^2$ is excluded. In the case of $\hat{\theta} = \theta = 1$, we obtain after some algebra that the power solution will be given with

$$q_i = \frac{1 + \theta R_i \Gamma_i}{1 + \hat{\theta} R_i \Gamma_i}.$$\hspace{1cm} (8.18)

For example, with $\theta = 1$ and no accumulated cancellation noise we obtain a conventional single-user receiver and it follows that $q_i = 1$. Such a power solution was used as the building block of GRP in Chapter 5. In general, a small $q_i$ value is desirable and would render in high user capacity. We observe in (8.17) that if $\theta < \hat{\theta} - \theta / \Gamma_i$ then $q_i < 1$, which means that the reduced interference from the SIC is captured in (8.15) as the $w_k$s are decreasing with increasing decoding number $k$. For $\theta > \hat{\theta} - \theta / \Gamma_i$, the power solution (8.15) can be greater than the one in the case without interference cancellation. This means that the cancellation efficiency is too low and the accumulated cancellation noise would not be compensated for. Using the $q_i$s of (8.18), assuming perfect SIC and rearranging, (8.15) reduces to

$$p_k = \frac{\nu \ R_k \Gamma_k \ W}{g_k \ \frac{1}{1 + \hat{\theta} R_k \Gamma_k} \left(1 + \prod_{j=k+1}^{N} \frac{R_j \Gamma_j}{W} \right)}.$$\hspace{1cm} (8.19)

which appeared in [99] and partly also in [29,30,115,117].
8.3.2 Partial Successive Interference Cancellation

The penalty of soft feedback is the extra generated noise. A possible approach to reduce the accumulated cancellation noise, is to cancel only a fraction of the signal. This can be represented as

\[ y_k(t) = r(t) - \sum_{i=1}^{k-1} \alpha_i y_i \sqrt{R_i a_i(t - \tau_i)} \cap r_i(t - \tau_i) \tag{8.20} \]

where \( 0 < \alpha_i \leq 1 \) is a control parameter. Repeating previous analysis, the variance of the interference becomes

\[ \sigma_k^2 = \sum_{i=1}^{k-1} \theta_{ik} \frac{g_i p_i}{W} + \sum_{i=k+1}^{N} \theta_i \frac{g_i p_i}{W} + \sum_{i=1}^{k-1} \var{\tilde{\theta}_i} \frac{R_i}{W} \sigma_i^2 + \nu_0. \tag{8.21} \]

where \( \var{\tilde{\theta}_i} = \var{\sqrt{W}(\tilde{\theta}_i - \alpha_i \sqrt{R_i} \tilde{\rho}_i \tilde{\rho}_i)} \). It can be shown that for \( \theta_i = \theta \), the minimum power solution is also in this case given by (8.15), but with

\[ q_i = \frac{1 + \frac{g_i r_i}{W} (\alpha_i^2 + \frac{\Gamma_i}{\theta})}{1 + \frac{g_i r_i}{W}}. \tag{8.22} \]

For comparison, consider the case of a slowly varying channel, where for \( i \neq k \), \( \var{\tilde{\theta}_i} = \var{\tilde{\theta}_k} = \var{\rho_i} = 1/R_i \). Then, we find that

\[ \theta_i(\alpha_i) = \var{1 - \alpha_i}^2. \tag{8.23} \]

It can be seen that, as we decrease \( \alpha_i \), \( \theta_i \) increases and there will be more interference due partial cancellation. On the other hand, the accumulated cancellation noise decreases at the same time. If we focus on the numerator of (8.22), we find that it is minimized by \( \alpha_i^* = \Gamma_i/(1 + \Gamma_i) \). It follows that with \( \theta_i = \var{1 - \alpha_i}^2 \)

\[ q_i|_{\alpha_i^*} = \frac{1 + \frac{g_i r_i}{W} \Gamma_i}{1 + \frac{g_i r_i}{W} + \frac{\Gamma_i}{\theta}} < \frac{1 + \frac{g_i r_i}{W}}{1 + \frac{g_i r_i}{W}} = q_i|_{\alpha_i=1}. \tag{8.24} \]

Therefore, for a slowly varying channel, a lower \( q_i \) value can be obtained by not canceling the whole signal. Further, the lower the values of \( \Gamma_i \), the less of the signal should be canceled.

8.3.3 Limited Successive Interference Cancellation

A drawback with SIC is that additional delay is introduced by each cancellation stage. The receiver complexity can to some degree be reduced, to the price of performance, by canceling only a subgroup of all users. If we assume that \( M \)
users are being canceled, there will be one set of users experiencing canceled interference and another employing conventional single-user detection.

\[ \begin{align*}
1, 2, \ldots, M, & \\
M + 1, M + 2, \ldots, N, & \\
\text{Canceled signals} & \\
\text{Uncanceled signals} & 
\end{align*} \]

The previous case of full cancellation is covered by letting \( M = N - 1 \). The variance of the interference becomes:

\[ \sigma_k^2 = \begin{cases} 
\sum_{i=1}^{k-1} \theta_{ik} \frac{R_i}{W} + \sum_{i=k+1}^{N} \frac{R_i}{W} + \sum_{i=1}^{k-1} \theta_{ik} \frac{R_i}{W} \sigma_i^2 + N_0, & 1 \leq k \leq M + 1 \\
\sum_{i=1}^{M} \theta_{ik} \frac{R_i}{W} + \sum_{i=M+1}^{N} \frac{R_i}{W} + \sum_{i=1}^{M} \theta_{ik} \frac{R_i}{W} \sigma_i^2 + N_0, & M + 2 \leq k \leq N
\end{cases} \]

It can be shown that for \( \theta_{ik} = \theta \), the minimum power solution is also in this case given by (8.15), but with

\[ q_i = \begin{cases} 
1 + \frac{\theta_i}{\theta} \left( 1 + \frac{\theta_i}{\theta} \right), & 1 \leq i \leq M \\
1, & M + 1 \leq i \leq N.
\end{cases} \tag{8.25} \]

### 8.3.4 Effective Bandwidth

From a networking perspective, we can use (8.15) to characterize the achievable user capacity of this receiver within a single cell. That is, given \( N \), if the desired levels of \( R_k \) and \( \Gamma_k \) can be accommodated within the feasible dynamic power range of the transmitters or given \( R_k \) and \( \Gamma_k \), the maximum \( N \). Hence, we do not consider an information-theoretic capacity definition. Several comparisons between different linear multuser receivers based on the same approach were carried out in [109]. The characterization of the user capacity was described in terms of a decoupled quantity, effective bandwidth, defined for each user; the SIR requirements could be met if the sum of all the effective bandwidths was less than a certain value. Let us here first observe the capacity characterization when the powers are only restricted to be positive. From (8.19), it is observed that for any assignment with \( R_k \cdot \Gamma_k < \infty \), there exists a positive and limited power solution. Clearly that will not be the case for a receiver with \( q_i \) given according to (8.17) and (8.18), which requires the denominator in (8.15) being positive. Therefore, we may regard (8.19) with \( \theta = 0 \) as the solution to a noise-limited system, whereas it otherwise becomes an interference-limited system.

#### Example 8.3.1
Consider the case with uniform parameters \( R \) and \( \Gamma \) and \( \bar{\theta} = \theta = 1 \).

The feasibility is governed by \( \sum_{j=1}^{N} \frac{W_j}{RT} \leq 1 \). With \( \gamma = RT/W \) and

\[ q = (1 + (\Gamma^{-1} + \theta) \gamma) / (1 + \gamma) \neq 1, \]

the maximum number of users \( N \) must fulfill the feasibility condition

\[ N < 1 + 1 + q^{-M} \left( 1 + \frac{1}{\gamma} - \frac{q^{M+1} - 1}{q - 1} \right), \quad M < N. \]
Figure 8.2: Relative user capacity compared to single-user detection for different number of canceled signals $M$ as function of the SIC efficiency. Parameters are set to $W = 1.2288$ MHz and $\Gamma = 8$ dB. For $\theta > 1 - 1/\Gamma \approx 0.84$, SIC is not beneficial and a single-user detector would perform better.

For a conventional single-user receiver, $M = 0$, the above capacity reduces to

$$N < 1 + \frac{1}{\gamma}$$  \quad (8.26)$$

which was also given in Example 3.1.1. For $M = 1$,

$$N < 1 + \frac{1}{\gamma q}$$  \quad (8.27)$$

and for $M = N - 1$

$$N < \frac{\log \left( \left( 1 + \frac{1}{\gamma} \right) (q - 1) + 1 \right)}{\log q}$$  \quad (8.28)$$

In Fig. 8.2 and 8.3, we plot the upper bounds (8.27) and (8.28) and normalize with the limit in (8.26). In both Figs., it can be seen that for certain $\theta$ (when $q > 1$), SIC offers less user capacity than single-user detection. Fig. 8.2 contains curves for different data rates. We note that, when many cancellations are performed, the accumulated noise, which in each step is proportional to the data rate, makes the relative user capacity larger for low data rates. On the other hand, when few cancellation steps are performed, the opposite behavior is shown. In Fig. 8.3, $R$ is fixed but the threshold $\Gamma$ is varied. The relative user capacity is here found to decrease when $\Gamma$ decreases. Also,
Figure 8.3: Relative user capacity compared to single-user detection for different number of canceled signals $M$ as function of the SIC efficiency. Parameters are set to $R = 9.6$ kbps and $W = 1.2288$ MHz. For $\theta > 1 - 1/\Gamma$, SIC is not beneficial and a single-user detector would perform better.

the range in $\theta$ at which the SIC receiver may outperform single-user detection shrinks. It can be seen that for $q < 1$, $M = N - 1$, i.e., full cancellation maximizes the relative user capacity, whereas for $q > 1$ it becomes $M = 0$, i.e., single-user detection. For $q = 1$, the maximum user capacity $N$ is independent of the number of cancellations. For the chosen parameters, we conclude that the gain from canceling just one signal was moderate.

**Example 8.3.2** Using the same parameters as in previous example and assuming a slowly fading channel, the system capacity for the case of partial successive interference cancellation can be upper bounded by the expression given in (8.28) with

$$q = \frac{1 + \frac{\alpha^2}{\Gamma} + (1 - \alpha)^2 \Gamma}{1 + \frac{\alpha^2}{\Gamma}}$$

where $0 < \alpha \leq 1$ represents the fraction of the signal canceled. Fig. 8.4 illustrates the capacity upper bound relative to the single-user detection case ($\alpha = 0$) as a function of the parameter $\alpha$ and for different values of the threshold $\Gamma$. We notice that full signal cancellation does not achieve the maximum relative user capacity, which confirms what was stated in Section 8.3.2. In fact, as seen in Fig. 8.4, the maximum user capacity is very much dependent on the kind of modulation scheme used. With a
robust modulation scheme (small \( \Gamma \)), SIC loses in efficiency and in this case less part of the signal should be canceled in order to maximize the capacity. As seen in Fig. 8.4, with a proper selection of the parameter \( \alpha \), the user capacity can in some cases be increased by as much as 25% when compared to the full cancellation case.

Now, let \( \gamma_k = R_k \Gamma_k / W \) be the SIR-target of user \( k \) and let the maximum achievable SIR for user \( k \) be denoted as \( \tilde{\gamma}_k = g_k \tilde{p}_k / \nu \). Then we can conclude the capacity region for feasible power assignment in a single cell as

\[
\sum_{j=1}^{N} \frac{w_j \gamma_j}{1 + \gamma_j} \leq 1 - \max_{1 \leq k \leq N} \frac{1}{\gamma_k} \frac{w_k \gamma_k}{1 + \gamma_k}.
\]

(8.29)

The effective bandwidth of this receiver could correspondingly be defined as \( w_j \gamma_j / (1 + \gamma_j) \). An important difference from [109], is that the effective bandwidths are coupled due to the factors \( w_j \), which is inherent from the successive decoding procedure. With respect to this, instead the decoupled \( q_i \) could be thought of as an intermediate effective bandwidth that would characterize the receiver efficiency. In the special case where \( q_i \) is given by (8.18), \( \tilde{\sigma} = \sigma = 1 \) and
\( \theta = 0 \), (8.29) reduces to
\[
\prod_{j=1}^{N} (1 + \gamma_j) \leq \min_{1 \leq k \leq N} \frac{\gamma_k}{\gamma_k} \prod_{j=1}^{k} (1 + \gamma_j). \tag{8.30}
\]

Taking the logarithm and rearranging, we see that a proper definition of effective bandwidth in this case is \( \log(1 + \gamma_j) \), which is proportional to the information-theoretic rate.
\[
\sum_{j=1}^{N} \log(1 + \gamma_j) \leq 1 - \max_{1 \leq k \leq N} \left\{ 1 + \log \frac{\gamma_k}{\gamma_k} - \frac{k}{\gamma_k} \sum_{j=1}^{k} \log(1 + \gamma_j) \right\} \tag{8.31}
\]

The effective bandwidths are in this case decoupled. If \( \gamma_k \) and \( \theta \) can be measured, the feasibility of the single cell problem can easily be determined from the conditions above.

### 8.4 Decoding Order

To the SIC scheme, a decoding order has to be specified. For any given decoding order, the power allocation in (8.15) makes the \( E_k/I_0 \) for user \( k \) meet its target \( \Gamma_k \) if it is feasible. In terms of multicellular network capacity, it is advantageous to minimize the interference and therefore the transmit powers. To determine the decoding order that would minimize the transmit powers of a given cell, we first find the following result\(^3\).

**Lemma 8.1** Under the optimal power control (8.15)-(8.17), the total received power is minimized by decoding by the order of the targets, \( \Gamma_1 \geq \Gamma_2 \geq \ldots \geq \Gamma_N \).

**Proof:** Since the total received power is given by
\[
\sum_{k=1}^{N} g_k p_k = \frac{\nu \sum_{k=1}^{N} \frac{w_k g_k \Gamma_k}{1 + \frac{w_k g_k \Gamma_k}{\nu}}}{1 - \sum_{k=1}^{N} \frac{w_k g_k \Gamma_k}{\nu}},
\]
we only need to show that the sum in the numerator is minimized by the given decoding order. Let us compare two such decoding orders, where under the first decoding order, user \( A \) is decoded as number \( 1 \leq i < N \) and user \( B \) as number \( i + 1 \) and in the second order vice versa. Then, with the assignments \( R_i, \Gamma_i \) in

\(^3\)In the following we assume \( M = N - 1 \) and \( \tilde{\gamma} = \gamma = 1. \)
the first decoding order and $R_i^*, \Gamma_i^*$ in the second,

$$
\sum_{k=1}^{N} \frac{w_k R_k \Gamma_k}{1 + R_k \Gamma_k} \sum_{k=1}^{N} \frac{w_k' R_k' \Gamma_k'}{1 + R_k' \Gamma_k'} = 1 + \frac{R_i}{1 + R_i \Gamma_i} + \frac{R_{i+1} \Gamma_{i+1}}{1 + R_{i+1} \Gamma_{i+1}} \times \frac{R_{i+2} \Gamma_{i+2}}{1 + R_{i+2} \Gamma_{i+2}} \times \ldots \times \frac{R_N \Gamma_N}{1 + R_N \Gamma_N} \times \left( \frac{\frac{R_i \Gamma_i}{1 + R_i \Gamma_i} + 1 + \frac{R_{i+1} \Gamma_{i+1}}{1 + R_{i+1} \Gamma_{i+1}} \times \frac{R_{i+2} \Gamma_{i+2}}{1 + R_{i+2} \Gamma_{i+2}} - 1 + \frac{R_i \Gamma_i}{1 + R_i \Gamma_i} + \frac{R_{i+1} \Gamma_{i+1}}{1 + R_{i+1} \Gamma_{i+1}}}{1 + \frac{R_i \Gamma_i}{1 + R_i \Gamma_i} + \frac{R_{i+1} \Gamma_{i+1}}{1 + R_{i+1} \Gamma_{i+1}} + \frac{R_{i+2} \Gamma_{i+2}}{1 + R_{i+2} \Gamma_{i+2}} \times (\frac{R_B}{1 + R_B \Gamma_B} - \frac{R_A}{1 + R_A \Gamma_A})}{1 + \frac{R_i \Gamma_i}{1 + R_i \Gamma_i} + \frac{R_{i+1} \Gamma_{i+1}}{1 + R_{i+1} \Gamma_{i+1}} + \frac{R_{i+2} \Gamma_{i+2}}{1 + R_{i+2} \Gamma_{i+2}} + \ldots} \right)
$$

since $R_k = R_k'$ and $\Gamma_k = \Gamma_k'$ for $k \neq i, i + 1$. Thus, the sum and therefore the received power of the first decoding order is smaller if $\Gamma_A \geq \Gamma_B$. Therefore, we can apply this argument recursively for any pair of users and it follows that the received power is minimized by decoding in the order of decreasing targets. If $\Gamma_k = \Gamma$, the difference becomes zero and the total received power is constant for any decoding order. 

This result says that for a homogeneous case $\Gamma_k = \Gamma$, independent of the decoding order within the cell, the total received power is constant for a given SIC efficiency. Considering the case $\Gamma_k = \Gamma$ and using Lemma 8.1, we find the following condition for minimizing the transmit powers.

**Proposition 8.1** Under the optimal power control (8.15)-(8.17) with $\Gamma_k = \Gamma$ and $\theta < 1 - \Gamma^{-1}$, the total transmitted power is minimized by decoding by the order of the path gains, $g_1 \geq g_2 \geq \ldots \geq g_N$.

**Proof:** Let under the first decoding order, user $A$ be decoded as number $1 \leq i < N$ and user $B$ as number $i + 1$ and in the second decoding order vice versa. Then, applying Lemma 8.1 with the power assignment $p_k$ in the first decoding
order and \( p'_k \) in the second,

\[
\left( 1 - \sum_{k=1}^{N} \frac{w_k R_k \Gamma}{1 + R_k \Gamma} \right) \frac{1}{\nu} \sum_{k=1}^{N} p_k = \left( 1 - \sum_{k=1}^{N} \frac{w_k R_k \Gamma}{1 + R_k \Gamma} \right) \frac{1}{\nu} \sum_{k=1}^{N} p'_k =
\]

\[
\frac{1}{\nu} \sum_{k=1}^{N} p_k - \frac{1}{\nu} \sum_{j=k}^{N} p'_k =
\]

\[
\frac{1 + \frac{R_i \Gamma}{W} + \theta \frac{R_i \Gamma}{W}}{1 + \frac{R_i \Gamma}{W}} \times \cdots \times \frac{1 + \frac{R_{i-1} \Gamma}{W} + \theta \frac{R_{i-1} \Gamma}{W}}{1 + \frac{R_{i-1} \Gamma}{W}} \times
\]

\[
\frac{1}{g_A} \cdot \frac{1 + \frac{R_k \Gamma}{W} + \theta \frac{R_k \Gamma}{W}}{1 + \frac{R_k \Gamma}{W}} \times \frac{R_k \Gamma (g_B - g_A) \Gamma (1 - \theta) - 1}{g_A g_B (1 + \frac{R_k \Gamma}{W}) (1 + \frac{R_k \Gamma}{W})}
\]

which is \( < 0 \) when \( g_A > g_B \) and \( \geq 0 \) when \( g_A \leq g_B \), since \( R_k = R'_k \) for \( k \neq i, i + 1 \). Thus, the user with the highest path gain should be the first decoded user. Therefore, we can apply this argument recursively for any pair of users and it follows that the total transmit power is minimized by decoding by the order of the path gains. \( \square \)

The interpretation hereof, is that links with poor path gains should be allocated low interference. When \( \theta = 1 - \Gamma^{-1} \), we have \( q_i = 1 \) and the total power becomes independent of the decoding order. However, as was noted previously, if \( \theta > 1 - \Gamma^{-1} \), \( q_i > 1 \) and each “cancellation” step introduces more interference. In this case, to provide the lowest interference to links with poor path gains, the decoding order need to be reversed. If instead of (8.17), the value (8.18) is used, Lemma 8.1 and Proposition 8.1 need to be reformulated. In that case, it can be found that the received power is always constant for any decoding order and the transmitted power is always minimized by decoding by the order of the path gains. It should be pointed out that Proposition 8.1 holds for a single cell and is not necessarily an optimal decoding order for multicellular systems. It is also noteworthy that, due to the variable rates and targets, the strongest received signal may not have the highest path gain. In particular, the received signal power from user \( k + 1 \) is stronger than that from user \( k \) if \( R_{k+1} \Gamma_{k+1} > WR_k \Gamma_k / (W + R_k (1 - \Gamma_k (1 - \theta))) \). This means that the strongest received signal is not necessarily the one of the first decoded user.
8.5 Iterative Power Control

Now consider an extension of the model to a multicellular system and the objective to find the optimal power allocation given a certain decoding order. We will assume a given decoding order in every cell such that all users can be supported with their requirements $R_k$ and $\Gamma_k$ within the feasible dynamic power range. For a power control scheme to be practical, it should allow for distributed implementation. For this we suggest the following iterative power control algorithms,

(PC1)

$$p_k(n + 1) = \min \left\{ \frac{\Gamma_k R_k}{W} \frac{1}{g_k} \left( \sum_{i=1}^{k-1} \theta g_i p_i(n) + \sum_{i=k+1}^{N} \tilde{g}_i p_i(n) + \sum_{i=1}^{k-1} \tilde{\theta}_i \sigma_i^2(n) + I_k(n) + \nu \right) + \tilde{\beta}_k, \right\}$$

(PC2)

$$p_k(n + 1) = \min \left\{ \frac{I_k(n) + \nu}{g_k} \frac{w_k R_k \Gamma_k}{1 + \tilde{\theta} R_k \Gamma_k} \frac{1}{1 - \sum_{j=1}^{N} \frac{w_j \tilde{\theta}_j}{1 + \tilde{\theta}_j \nu_j}} \tilde{\beta}_k, \right\}$$

where $p_k(n)$ is the power value at iteration $n$ and $I_k(n)$ is the inter-cell interference power at iteration $n$ at user $k$'s receiver. For PC1, the power value for user $k$ can be computed based on measurements of the total interference after each cancellation stage and the path gain. Alternatively PC1 can be rewritten as a DCPC mapping with the SIR measured after each cancellation step for each user respectively. PC2 makes direct use of the SIC efficiency parameter $\theta$ through the $w_k$'s. PC1 is fully distributed compared to PC2, which requires some coordination at the base station to distinguish the inter-cell interference from the intracellular interference. It is assumed that the $p_k(n)$ values can be correctly delivered to the mobile terminal for all users $k$ and iterations $n$. The gain of updating the powers according to PC2, can be understood from that the intracellular interference is directly compensated for in each iteration, increasing the rate of convergence to the optimal fixed point. The same principle was applied for the GSPC algorithm in Chapter 5. This we state below, omitting a rather lengthy proof.

**Proposition 8.2** The algorithms PC1 and PC2 converge to the fixed point of a feasible system and PC2 is asymptotically faster than PC1.

**Sketch of Proof:** A proof follows the same principle as was used in Chapter 5,
1.1 

Relative power reduction

\[ \theta = 0 \]

\[ \theta = 0.2 \]

\[ \theta = 0.4 \]

\[ \theta = 0.6 \]

\[ \theta = 0.8 \]

Figure 8.5: Relative power reduction compared to \( \theta = 1 \) as function of decoding number for different degrees of SIC efficiency. Parameters are set to \( R = 9.6 \) kbps, \( W = 1.2288 \) MHz, \( \Gamma = 8 \) dB and \( N = 10 \) users.

Comparing spectral radii of iteration matrices. Since (8.14) is a linear equation system in the powers, an \( \mathbf{H} \) matrix can be determined and the problem can be written in matrix form like (3.2). It can be verified that the steady state solutions of PC1 and PC2 are consistent with (8.14). From Proposition 5.2 in [56], it follows that if \( N > 1 \), the matrix splitting of \( \mathbf{H} \) reducing to PC2 is asymptotically faster than the one of PC1.

Hence, if the SIC efficiency \( \theta \) can be found, PC2 can be used and the convergence rate of the power control can be increased.

8.6 Numerical Results

Inherent from the SIC, for small \( \theta \) interference is reduced as cancellation proceeds, which implies less transmitter power is needed than for a conventional single-user receiver. For the single cell homogeneous case \( \gamma = \mathbf{R}^\top / W \) for all users, \( \hat{\theta} = \theta = 1 \) and \( M = N - 1 \), we plot reduction in transmitter powers as function of the SIC efficiency parameter \( \theta \). In Fig. 8.5, for every user, we plot its power when \( \theta = 1 \), normalized with the power, for different cases of \( \theta \). The parameters are set to \( \Gamma = 8 \) dB, \( R = 9.6 \) kbps, \( W = 1.2288 \) MHz and \( N = 10 \)
Figure 8.6: Relative received power reduction as function of SIC efficiency. Parameters are set to $R = 9.6$ kbps and $W = 1.2288$ MHz, $\Gamma = 8$ dB and $N = 10$ users. The curves show that the results are quite sensitive to the value of $\theta$ and that the relative gain is more pronounced for low $\theta$. Naturally, the later on decoded users experience a larger gain. Note that the curves of Fig. 8.5 serve as an upper bound to a comparison with a single-user receiver, due to the accumulated cancellation noise. The relative reduction in intracell interference is related to the total received power at the base station. The plot in Fig. 8.6 shows the relative reduction of total received power as function of the SIC efficiency for the same parameters. A significant reduction can be seen over a big range of $\theta$ and this will correspond to increased multicellular capacity as well. The iterative algorithms are evaluated by simulations in a system consisting of 19 hexagonal cells with centrally located base stations using omni-directional antennas. The cell radius is set to 1 km and a wrap-around technique is used. Evaluation is performed in the DS-CDMA system described in Chapter 2 and illustrated in Fig. 1.2. The cell radius is set to 1 km and a wrap-around technique is used along with omni-directional antennas. For a given instance, 10 mobiles/cell are generated, the locations of which are uniformly distributed over the cells. At any given instance, the link gain is modeled by $g_{ij} = S_{ij} \cdot r_{ij}^4$, where $S_{ij}$ is the shadow fading factor and $r_{ij}$ is the distance between base $i$ and mobile $j$. The shadow fading factor is generated from a lognormal distribution.
Figure 8.7: The average power for different degree of SIC efficiency. Parameters are set to $R = 9.6$ kbps, $\Gamma = 8$ dB and $W = 1.2288$ MHz and $N = 10$ users/cell.

with $\mu_s = 0$ dB, and $\sigma_s = 8$ dB, where the $S_i$’s are mutually independent. The base receiver noise is taken to be $\nu = 10^{-12}$ W and the maximum mobile power is set to unity. The initial power for each mobile is randomly chosen from the interval $[0,1]$. Each user is assigned to the base station that provides the lowest signal attenuation. We consider the uplink of a system that has a spreading bandwidth $W = 1.2288$ MHz, $R = 9.6$ kbps and a required bit-energy-to-noise-spectral-density ratio $\Gamma_k = 8$ dB for all users. Fig. 8.7 contains the result of one thousand independently generated snapshots. The decoding order is in each cell set according to Proposition 8.1. We let the algorithms iterate for thirty iterations for $\theta = 0, 0.5$ and 1 respectively and plot the average power per user in each iteration. We see that when $\theta = 0$, the gain in rate of convergence for PC2 is rather limited, mainly due to low intracell interference. For larger $\theta$, the gain from PC2 becomes more distinct since it reaches the steady level much faster. It is noteworthy that the average power level decreases monotonously for PC2 for all values of $\theta$. This follows from that the intracell interference is inherently compensated for by PC2 in a more rapid sense. Further, the rate of convergence becomes slower with increasing $\theta$. This can be explained by writing PC1 and PC2 on matrix forms and observing that the magnitude of the eigenvalues of the resulting iteration matrices increases with increasing $\theta$. It is also noticeable that the magnitude of the steady state power levels is highly dependent on the
8.7 Concluding Remarks

Power controlled multiuser detection is an area that has not been that well examined. Since power control and resource allocation have significant impact on system performance, it is important to explore such features as well. This chapter shows that achievable performance of a SIC scheme is dependent on suitable power control. We have derived the optimal single cell power control law for imperfect cancellation with diverse QoS requirements. The analysis was extended to include several other receiver models. From this, a condition for single-cell user capacity could be determined. The results indicate that the gain from SIC is significant when the channel fading is slow and the $E_b/I_0$ requirements are high. Our results show that, among our schemes, partial SIC is the most suited detection scheme as it achieves the maximum capacity when its parameter $\alpha$ is adjusted accordingly. It is noteworthy that $\alpha^* < 1$, i.e., the whole signal should never be canceled. The single cell result was extended to an iterative convergent algorithm for the multicellular case. We point out that if the SIC efficiency parameter can be estimated, it can be directly incorporated into the power control through an update form like PC2. Using this information, a gain in rate of convergence could be found, making the powers reach desirable levels faster. In the expression (8.14), the fast fading is averaged out and hence the power control does not explicitly track these fast variations. The power control algorithms can however be straightforwardly modified to take this into account.

We have used the terminology of effective bandwidth to characterize the user capacity. In [109], several linear receivers were characterized by means of this quantity. However, their analysis included random spreading sequences and very large systems. Large here means that both the number of users and the processing gain go to infinity but their ratio is fixed, e.g., $\beta = N/(W/R)$. To relate to their results, an asymptotic interpretation of the conventional receiver feasibility condition is

$$\sum_{i=1}^{N} \frac{\Gamma}{1 + \frac{W/R}{W/R}} = \frac{N\Gamma}{1 + \frac{W/R}{W/R}} \rightarrow \beta \Gamma < 1.$$  

In [109], they give $\Gamma$ as the effective bandwidth of a matched filter receiver.
Chapter 9

Conclusions

Radio resource management plays an important role in the design of cellular radio systems. To allocate resources in real time for maximizing system capacity, is an important but enormous task. Power control is one essential issue in this problem, in particular for DS-CDMA. The cellular systems deployed so far, mainly have provided services characterized by being connection oriented, delay sensitive and offering a rather low transmission rate. Appropriate network control for delay insensitive and multirate applications is not as straightforward to specify. In this thesis the main focus has been on heterogeneous resource allocation. The main objective has been to suggest power control based RRM methods, suitable for more complex QoS support.

9.1 Summary

The thesis started with the focus on the detrimental effect an overloaded system could have on system performance. In a highly congested system, powers may drift toward their maximum and users cannot reach sufficient quality. The consequence is that an over-sized set of non-supported connections appears. This undesirable event must be rapidly detected and congestion be avoided. Congestion control is an adopted terminology, which in our first sub-problem corresponds to removing radio connections until a feasible power control solution can be found. Unfortunately, the connection removal problem is NP-complete, which motivates heuristic approaches. For keeping the practical relevance, low complexity algorithms that lend themselves into distributed implementation are sought. The approach taken here, is to incorporate connection removals during power control execution, a natural way of handling congestion. The suggested GDCPC algorithm, which is distributed, incorporates so called temporary removals. This happens if the requested power is too high. The algorithm then prevents the user for making an attempt transmitting with maximum power. This was shown to quickly direct the powers, so that outage probability became
low. The benefit from this is that the bandwidth can be utilized more efficiently and terminals do not waste energy on bad channels. This may be interpreted as a distributed autonomous channel access mechanism integrated with the power control. The interference levels can here be said to play the role of inducing the removal/admission decision. Our numerical results demonstrated that this simple algorithm could provide a suitable bridge to energy-efficient multirate RRM.

With multiple rates, there are several combinations of rate allocations that would yield the same total throughput. Due to the different propagation conditions among users, these combinations will result in different energy consumption. We formulated a throughput maximization problem, where the combination requiring the least power is sought. Again we encounter a NP-complete problem for which we suggest a heuristic greedy admission control algorithm, GRP, that allocates high data rates to channels with good quality. GRP has linear complexity and is executed in a cell-wise manner. For a single cell and downlink direction or unconstrained uplink direction, we found that the GRP scheme finds the optimal solution, which motivates its use under more general conditions too. This algorithm constitutes a practical interpretation of the information theoretic water filling concept. Inevitably, this creates some unfairness among the users, so practically some mobility will be necessary if the resulting solution should also be fair in the long run. Fairness versus maximum throughput though, was not specifically part of the problem definition for this best effort service. However, a more conservative rate allocation algorithm, GRP2, could deal with issues of a minimum required data rate. The GRP uses knowledge of link gains and assigned rates within the cell, i.e., slightly more information than a fully distributed scheme. This model is realistic when the resource allocation is done at each base station but distributed between them. The gain from semi-distributed allocation was shown to be large in terms of faster convergence and more energy-efficient resource allocation. In principle, the transmission schemes applied for throughput maximization, allocate more rate and sometimes more power to good channels. That is, quite a different behavior from the usual SIR-balancing power control, which has its cause in the new application.

Another type of nonreal time data service is, where users require an average data rate. That is a certain amount of data should be received over a given time period. It is easy to conclude that to use the minimum possible energy for delivering the data, users should be scheduled to transmit in a one-by-one fashion within the cells. An important issue is whether such time division will cause a throughput loss. We found that if a throughput could be provided with simultaneous transmission, it could also be achieved as an average rate with time division. It should be remarked though, that this claim is based on that the rate adaptation can be made in linear fashion to the channel quality. Our approach for the QoS support encompasses a close relation between power control and multiple access, as time division is introduced on the CDMA channel by purpose. The impact of this is lower intracell interference, which reduces the energy
consumption, and leads to higher system capacity. Also, the rate of convergence of the power control can be improved. Differing from previous chapters, where the power control problem can rather easily be discerned, the power control here attempts to allocate the minimum cell site power level that guarantees the average rate. As was found, it has notationally quite a big resemblance with the normal power control problem introduced in Chapter 3 and the suggested power control algorithm can be analyzed similarly.

In the above approach, the transmission schedule is arbitrary. We refer to T/CDMA as a static gain that could be achieved by a simple round robin scheduler. A large static gain means that simpler schedulers or less channel estimates can be considered. For a fading channel, additional gains can also be obtained by selectively choosing the users with favorable channels. Therefore, there are two gains possible compared to pure CDMA: a static gain from the time division property and a dynamic gain from channel fading. To maximize throughput in a single cell, the maximum power level should be used, which could relax the requirements on sophisticated power control. Different fairness criteria can be considered and we suggested a scheduler that will provide all users with the same long-run channel access. This may alternatively be regarded as all users get asymptotically the same amount of energy. This is achieved by scheduling the user with the best unbiased channel quality normalized with its average quality. The analysis shows that the throughput gain is equal to the one of a common selection diversity scheme. The gain was further shown to decrease when the rates were constrained or discrete. Our results agree with previous work in [48, 49, 114], where the gain increases with the variance of the channel fading. If the relation between effective throughput and channel quality is nonlinear, e.g., a logarithmic Shannon type of growth, several users could be scheduled simultaneously to utilize the whole channel capacity. The results show that the gain from multiuser scheduling is to be found for channels with large average quality.

In the final chapter, the receiver model was modified slightly so that already detected signals could be fed back for cancellation. Such a receiver could still be built around the matched filter concept. Here we considered soft feedback in the cancellation step. This avoids error propagation, which can occur with hard decision feedback but instead accumulated cancellation noise is introduced. This extra noise could cause worse performance than conventional single-user detection. Hence, suitable power control and achievable user capacity needs to be found for this receiver. We determined the achievable user capacity on closed-form and found that it applies to a larger class of linear SIC receivers. The given solution is general and provides results for imperfect-, partial- and limited cancellation. It was exemplified that by canceling only part of the signal, higher user capacity can be achieved. In fact, one should opt for partial cancellation for maximizing user capacity. Although imperfect cancellation can cause worse performance than a single-user detector, for the case of a slowly varying channel, the user capacity can be significantly increased with proper power control.
9.2 Discussion

The work presented in this thesis has aimed to provide insight on how to improve system performance by means of suitable radio resource functions. We have focused on suggesting and evaluating transmission schemes based on power control, transmission rate control and transmission scheduling. For this, we have applied referred system models which have allowed a close resemblance between the analytical work and the numerical evaluation. The results are basic in nature and should be interpreted as such, providing insights of how to design real implementable schemes and principal solution behavior.

A main focus has been on global stability and performance. Since the radio environment is highly changing, quick schemes are likely to allocate the resources better. Therefore, a common theme has been to focus on designing fast algorithms. From different matrix splittings, algorithms with different asymptotic average rate of convergence could be derived. In the studied schemes, faster rate of convergence comes from semi-distributiveness, that is some more knowledge, like measurements, could be required than a fully distributed algorithm. For practical systems, time delays and information aging may be disastrous to the stability of the local loops. Therefore, there should be some critical point at which there is practically no gain of designing any theoretically faster algorithm, as the overhead signaling burden is too large or slow. In this thesis we have limited the information exchange to be within the cell.

Temporary removals were suggested as a useful means toward better channel utilization. One drawback that temporary removals may face, is that they practically may not be feasible on such a short term notice as was assumed. There may exist a positive minimum power level and some synchronization issues may arise when shutting the power totally down. However, this is dependent on the transceiver implementation. For example in WCDMA, during silent periods, no data is transmitted on the in-phase branch, whereas pilot bits and power control commands are still transmitted through the quadrature branch [32]. From theoretical point of view, the results in Chapter 4-7, could be interpreted in a wider perspective as PCMA, where transmission is executed depending on the current interference level. Thus some TDMA behavior occurs and we can classify this as a form of induced hybrid multiple access. In this thesis we suggest that time division should be induced when interference is high, cf., the GDCPC algorithm. The results on the multirate part touches upon the question if single transmissions within the cells are always favorable. There exists a great deal of information-theoretic results for this problem. Practically though, if throughput is not logarithmic, we may arrive at other conclusions. Therefore, general claims should be cautiously used. We find that the assumed model for the transmission rates plays a major part of how many users will transmit simultaneously. If there is a maximum rate, maximum power or discrete rates, the work presented herein confirms that not only one single user can take the whole capacity. That is, under more practical conditions it may not be optimal to just let one user
transmit at a time.

In several of the chapters, a time independent model has been assumed. Hence, the algorithm analysis becomes deterministic. Reasonably, this can not capture all effects of the channel and traffic characteristics. In practice, power control algorithms are adapted to this by the outer loop. However, to assure reliable performance over a time varying channel, over-provisioning of radio resources is common. Therefore, a practically achievable capacity could be lower than the figures presented in this work.

The analysis in Chapter 8 concerned a linear SIC detector, i.e., the soft output was fed back. Another, perhaps more common approach of SIC, is to feed back the hard decision output, resulting in a nonlinear detector. The main reason for adopting linear SIC is implementation aspects. For nonlinear SIC, first the data symbol needs to be detected. An erroneous decision could propagate, causing unreliable decisions later on. Naturally, the first decision is very crucial. Then the wave form (amplitude, phase and timing) needs to be estimated separately. Poor estimation could therefore also be disastrous. Since most radio systems of today use strong FEC, the symbol error probability could be made quite small. On the other hand, estimation is still important and if the FEC allows for operating at low SNRs, the estimation procedure could become more difficult, which would affect the signal regeneration. For linear SIC, a joint estimate of bit and amplitude can be used. It has been claimed that, if the estimation process is accurate, nonlinear SIC generally outperforms linear SIC [85]. With the opposite, gains can be significantly reduced, or even reversed. Soft feedback is often regarded as a way of improving the intermediate decisions. Another method, which we also studied, is to rely less on the detected signals and use only partial cancellation. Another important issue in SIC is the decoding order. For single-rate systems, it is intuitive that the strongest received user should be the first to be decoded. This because its acquisition is better and it is more reliable. Canceling the strongest user also has a larger positive effect on the other users. A motivation for the linear SIC in [89] was that the ordering of users could be performed directly on the correlation outputs, using no knowledge on the received amplitudes. It has also been argued that it is simpler to implement noncoherent detection with linear SIC [89].

We conclude this thesis by emphasizing proper RRM as a necessary mainstay for QoS provisioning. This thesis has suggested transmission schemes that offer high capacity, energy conservation and fast convergence. We find that careful resource allocation is essential and that significant improvements can be obtained to reasonable complexity.

### 9.3 Further Studies

The schemes proposed in this thesis have been evaluated with rather simple channel and traffic models. To fully assess performance of these principles, full
evaluation with more relevant traffic models would be needed. Further challenges include the introduction of mixed services, which adds another dimension to the resource allocation problem. Different service classes can be associated with services of different elasticity. Radio resources should then also be allocated such that relative service levels can be maintained between classes. For example in Chapter 5, congestion was treated by that users back off their transmission rates. In a more general setting, this should also be determined by the elasticity of the class to which the service belongs. Depending on which spreading codes that are available, there may be other constraints on feasible rate combinations. Even if a preferred combination would not violate the channel capacity, code blocking could occur.

With random data arrival, a scheduler needs to establish queue stability. If the buffers are of finite length, risk of overflow or packet dropping must be taken into account. For the current scheduler, queuing aspects are still an open issue. In its current form, there is a risk that the schedulers in this work could try to allocate resources to users with empty buffers. This could give rise to a situation where users with poor channels always have full buffers, whereas they are empty for favorable users. A related issue occurs when the analysis should encompass a slotted time scale. Then, there could be occasions where the amount of data in the buffer is less than what can be transmitted during the slot. Scheduling such a user could result in an underutilization of the channel. With finite amount of data for each user, a maximum SIR scheduler could be relevant. By scheduling the best user, according to the minimum time-span solution [54], the total delay could be minimized. Moreover, an open issue is to design scheduling algorithms that adapt also the number of allocated users to the channel conditions. Based on our results of the multiuser scheduling, such a decision would be dependent on orthogonality factors and average channel qualities.

Some of our schemes encompass time division properties. The general validity of the gain of TDMA, e.g., stated in Chapter 6 is of course dependent on the system model. Other nonlinear rate versus SIR relations need to be investigated. The conclusions could also change if there is a finite amount of data in each user’s buffer and the time scale is slotted.

Joint resource allocation and adaptive antennas is also an interesting topic. The results in [114] proposed scheduling combined with a rather different beamforming technique, aiming to increase the channel variance.
Appendix A

Multicellular Outage Probability

Consider a desired signal and \( B - 1 \) independent interfering intercell signals \( C_i \), exponentially distributed, each with a mean \( \bar{C}_i \). The complementary outage probability of user 1 can be obtained as:

\[
\int_0^\infty \cdots \int_0^\infty \Pr \left[ C_1 > x \sum_{j \neq 1} c_j + x \nu \right] \prod_{j=2}^B \frac{1}{C_j} e^{-\frac{c_j}{C_j}} dc_2 \cdots dc_B =
\]

\[
\int_0^\infty \cdots \int_0^\infty e^{-x \sum_{j=1}^B \frac{c_j}{C_j}} e^{\nu x} \prod_{j=2}^B \frac{1}{C_j} e^{-\frac{c_j}{C_j}} dc_2 \cdots dc_B =
\]

\[
e^{-\frac{\nu}{C_1}} \prod_{j=2}^B \int_0^\infty \frac{1}{c_j} e^{-\left(\frac{x}{C_1} + \frac{1}{c_j}\right) c_j} dc_j =
\]

\[
e^{-\frac{\nu}{C_1}} \prod_{j=2}^B \frac{1}{C_j} \left(\frac{x}{C_1} + \frac{1}{c_j}\right) =
\]

\[
e^{-\frac{\nu}{C_1}} \prod_{j=2}^B \frac{1}{\frac{x}{C_1} + 1}
\]
Appendix B

Average Throughput with Time Delays

From (7.33) and (7.34), it follows that

\[
\xi_i(t + \delta) = \rho^2 \xi_i(t) + (1 - \rho^2)(N_i^{(1)}(t)^2 + N_i^{(2)}(t)^2) + 2\rho \sqrt{1 - \rho^2} (Z_i^{(1)}(t)N_i^{(1)}(t) + Z_i^{(2)}(t)N_i^{(2)}(t)).
\]

Previously, this results was used.

\[
\int_0^\infty \xi_i(t) \prod_{j \neq i}^N (1 - e^{-\xi_i(t)/\xi_j(t)}) j_{\xi_i(t)}(\xi_i(t)) d\xi_i(t) = \frac{\bar{\xi}_i}{N} \sum_{k=1}^N \frac{1}{k}
\]

Now, consider the following integrals.

\[
\int_0^\infty \int_{-\infty}^\infty \int_{-\infty}^\infty (N_i^{(1)}(t)^2 + N_i^{(2)}(t)^2) \prod_{j \neq i}^N (1 - e^{-\xi_i(t)/\xi_j(t)}) \times \]

\[
\int_{-\infty}^\infty f_{N_i^{(1)}(t)}(N_i^{(1)}(t)) f_{N_i^{(2)}(t)}(N_i^{(2)}(t)) j_{\xi_i(t)}(\xi_i(t))
\]

\[
dN_i^{(1)}(t) dN_i^{(2)}(t) d\xi_i(t) = \bar{\xi}_i \int_0^\infty \prod_{j \neq i}^N (1 - e^{-\xi_i(t)/\xi_j(t)}) j_{\xi_i(t)}(\xi_i(t)) d\xi_i(t) = \frac{\bar{\xi}_i}{N}
\]

157
The first equality is due to that \( N_i^{(1)}(t) \), \( N_i^{(2)}(t) \) and \( \xi_i(t) \) are by definition mutually independent.

\[
\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} (Z_i^{(1)}(t) N_i^{(1)}(t) + Z_i^{(2)}(t) N_i^{(2)}(t)) \prod_{j \neq i}^{N} (1 - e^{-\frac{\xi_j(t)}{\xi_i}}) \times \\
\frac{f_{N_i^{(1)}(t)}(N_i^{(1)}(t)) f_{N_i^{(2)}(t)}(N_i^{(2)}(t)) f_{Z_i^{(1)}(t)}(Z_i^{(1)}(t)) f_{Z_i^{(2)}(t)}(Z_i^{(2)}(t)) f_{\xi_i}(t)(\xi_i(t))}{dN_i^{(1)}(t)dN_i^{(2)}(t)dZ_i^{(1)}(t)dZ_i^{(2)}(t)d\xi_i(t)} = 0
\]

The integral equals zero since \( N_i^{(1)}(t) \) and \( N_i^{(2)}(t) \) are zero mean Gaussian random variables and by definition mutually independent from \( \xi_i(t) \). Thus the integral (7.36) then equals (7.35).
Appendix C

Scheduling Gain Expressions

C.1 Equivalent Representations

The following well-known relations will be used

\[
\sum_{k=1}^{N} \binom{N}{k} (-1)^{k+1} = 1 \tag{C.1}
\]

\[
\binom{N + 1}{k} = \binom{N}{k-1} + \binom{N}{k} \tag{C.2}
\]

\[
\frac{k}{N} \binom{N}{k} = \frac{N-1}{k-1} \tag{C.3}
\]

\[
\binom{N}{k} = \binom{N}{N-k} \tag{C.4}
\]

C.1.1 Equation 7.12

Assume that \( \sum_{k=1}^{N} \binom{N}{k} (-1)^{k+1} \frac{1}{k} = \sum_{k=1}^{N} \frac{1}{k} \). Since

\[
\sum_{k=1}^{N+1} \binom{N+1}{k} (-1)^{k+1} \frac{1}{k} = \sum_{k=1}^{N+1} \binom{N}{k-1} (-1)^{k+1} \frac{1}{k} + \sum_{k=1}^{N} \binom{N}{k} (-1)^{k+1} \frac{1}{k}
\]

\[
= \sum_{k=1}^{N+1} \binom{N+1}{k} (-1)^{k+1} \frac{1}{N+1} + \sum_{k=1}^{N} \frac{1}{k}
\]

\[
= \sum_{k=1}^{N+1} \frac{1}{k}
\]

159
it follows by induction that it holds for any \( N \geq 1 \).

**C.1.2 Equation 7.26**

The throughput is obtained from:

\[
\int_0^{\infty} \log_2(1 + \xi_i) \left(1 - e^{-\xi_i} \xi_i \right)^{N-1} \frac{1}{\xi_i} e^{-\xi_i} d\xi_i =
\]

\[
\sum_{k=0}^{N-1} \binom{N-1}{k} (-1)^k \int_0^{\infty} \log_2(1 + \xi_i) \frac{1}{\xi_i} e^{-\xi_i} d\xi_i =
\]

\[
\sum_{k=1}^{N} \frac{k}{N} \binom{N}{k} (-1)^k H(k, \xi_i) \tag{C.5}
\]

The scheduling gain (7.26) is found by normalizing (C.5) with

\[
\frac{1}{N} \int_0^{\infty} \log_2(1 + \xi_i) \frac{1}{\xi_i} e^{-\xi_i} d\xi_i = \frac{1}{N} H(1, \xi_i). \tag{C.6}
\]

**C.1.3 Equation 7.43**

We find that

\[
\sum_{j=1}^{L+1} \sum_{k=1}^{N+1-j} \binom{N}{j} \frac{(-1)^{N+1-j-k}}{N} \frac{j}{(N + 1 - j) (N + 1 - k)} =
\]

\[
\sum_{j=0}^{L} \sum_{k=0}^{N-1-j} \binom{N-j}{k+1} \frac{(-1)^{N-1-j-k}}{N} \frac{j+1}{(N-j) (N-k)} =
\]

\[
\sum_{j=0}^{L} \sum_{k=0}^{N-1-j} \binom{N-1-j}{k} \frac{(-1)^{N-1-j-k}}{N} \frac{1}{(N-k)}
\]

From the mathematical software package Mathematica, an alternative representation is given by

\[
\sum_{k=0}^{N-1-j} \binom{N-1-j}{k} \frac{(-1)^{N-1-j-k}}{N-k} \frac{1}{(N-k)} = (-1)^{N-j} \hat{\Gamma}(-N) \hat{\Gamma}(N-j) \hat{\Gamma}(-j)
\]

where \( \hat{\Gamma}(x) = \int_0^{\infty} t^{x-1} e^{-t} dt \) is the Gamma function. With the following known representation,

\[
\hat{\Gamma}(z) \propto \frac{(-1)^n(1 + O(z + n))}{n!(z + n)}, z \to -n, n \in \mathbb{N}^+
\]
we find that
\[
\sum_{k=0}^{N-1-j} \binom{N-1}{j} \binom{N-1-j}{k} (-1)^{N-1-j-k} \frac{1}{(N-k)} = \frac{1}{N}.
\]

Hence, the following relation can be established
\[
\sum_{j=1}^{L+1} \sum_{k=1}^{N+1-j} \binom{N}{j} \binom{N+1-j}{k} (-1)^{N+1-j-k} \frac{j}{N} = \frac{L+1}{N}.
\]

### C.1.4 Equation 7.45

We find by simple inspection that
\[
\sum_{k=1}^{N} \binom{N}{k} (-1)^{N-k} \frac{k}{(N+1-k)^2} = \sum_{k=1}^{N} \binom{N}{k} \frac{(N+1-k)}{(N+1-k)^2} = \sum_{k=1}^{N} \frac{(N+1-k)}{k^2} = \sum_{k=1}^{N} \frac{1}{k^2}.
\]

### C.2 Limit Values

The following relations can be found, which are useful in determining the scheduling gain.
\[
\lim_{x \to 0} V(k/x) = 0
\]
\[
\lim_{x \to \infty} V(k/x) = \infty
\]

For large \( x \),
\[
V(k/x) = -\gamma - \log x - \frac{1}{x} + \frac{k + k(-\gamma - \log x + \log x)}{x} + O(x^{-2}).
\]
\[
\Gamma(x) = \int_{x}^{\infty} t^{-1}e^{-t} dt \propto x^{-1}e^{-x}(1 + O(1/x)), \quad x \to \infty
\]
References


