Muhammad Majid Butt

On the Near-Far Gain in Opportunistic and Cooperative Multiuser Communications

Thesis for the degree of Philosophiae Doctor

Trondheim, January 2011

Norwegian University of Science and Technology
Faculty of Information Technology, Mathematics
and Electrical Engineering
Department of Electronics and Telecommunications

NTNU – Trondheim
Norwegian University of
Science and Technology

NTNU
Norwegian University of Science and Technology

Thesis for the degree of Philosophiae Doctor

Faculty of Information Technology, Mathematics and Electrical Engineering
Department of Electronics and Telecommunications

ISBN 978-82-471-2547-2 (printed ver.)
ISBN 978-82-471-2549-6 (electronic ver.)
ISSN 1503-8181

Doctoral theses at NTNU, 2011:16

Printed by NTNU-trykk

Abstract

In this dissertation, we explore the issues related to opportunistic and cooperative communications in a multiuser environment. In the first part of the dissertation, we consider opportunistic scheduling for delay limited systems. Multiuser communication over fading channels is a challenging problem due to fast varying channel conditions. On the other hand, it provides opportunities to exploit the varying nature of the channel and maximize the throughput by scheduling the user (or users) with good channel. This *gain* is termed as multiuser diversity. The larger the number of users, the greater is the multiuser diversity gain. However, there is an inherent scheduling delay in exploiting multiuser diversity. The objective of this work is to design the scheduling schemes which use multiuser diversity to minimize the system transmit energy. We analyze the schemes in large system limit and characterize the energy–delay tradeoff. We show that delay tolerance in data transmission helps us to exploit multiuser diversity and results in an energy efficient use of the system resources. We assume a general multiuser environment but the proposed scheduling schemes are specifically suitable for the wireless sensor network applications where saving of transmit energy at the cost of delay in transmission is extremely useful to increase the life of battery for the sensor node.

In the first part of the thesis, we propose scheduling schemes with the objective of minimizing transmit energy for a given fixed tolerable transmission delay. The fixed delay is termed as *hard deadline*. A group of users have channels better than a transmission threshold are scheduled for transmission simultaneously using superposition coding. The transmission thresholds depend on the fading statistics of the underlying channel and hard deadline of the data to be scheduled. As deadline is approached, the thresholds decrease monotonically to reflect the scheduling priority for the user.

We analyze the proposed schedulers in the large system limit. We compute the optimized transmission thresholds for the proposed scheduling schemes. We analyze the proposed schemes for practically relevant scenarios when the randomly arriving packets have individual, non–identical deadlines. We analyze the case when loss tolerance of the application is exploited to further decrease the system energy. The transmitted energy is not a convex function of transmission thresholds. Therefore, we propose heuristic optimization procedures to compute the transmission thresholds and evaluate the performance of the schemes. Finally, we study the effect of outer cell interference on the proposed scheduling schemes.

The second part of the thesis investigates the problem of cooperative communication between the nodes which relay the data of other sources multiplex with their own data towards a common destination, i.e. a relay node performs as a relay and data source at the same time. This problem setting is very useful in case of some wireless sensor network (WSN) applications where all the nodes relay sensed data towards a common destination sink node. The capacity region of a relay region is still an open problem. We use deterministic network model to study the problem. We characterize the capacity region for a cooperative deterministic network with single source, multiple relays and single destination. We also characterize the capacity region when communicating nodes have correlated information to be sent to the destination.

Preface

This dissertation is submitted in partial fulfillment of the requirements for the degree of Philosophiae Doctor (PhD) at the Department of Electronics and Telecommunications, Norwegian University of Science and Technology (NTNU). Prof. Ralf R. Müller and Prof. Geir E. Øien have been my main and co-supervisors, respectively. Both of them are with the Department of Electronics and Telecommunications (NTNU).

The studies have been carried out in the period from December 2006 to December 2010, including one semester of course work and approximately one year of teaching assistant duties in different graduate courses. I have been working at NTNU for most of this period. I visited Prof. Giuseppe Caire at University of Southern California from February 2009 to June 2009. I also got the opportunity to visit Prof. David Gesbert at Institute Eurecom, Sophia Antipolis for two months.

This work has been funded by The Research Council of Norway (NFR) via the project "Cross Layer Optimization of Wireless Sensor Networks" (CROPS) while teaching assistantship has been funded by the Department of Electronics and Telecommunications, NTNU.

Acknowledgements

This dissertation would not have been possible without the guidance and help of several individuals who in one way or the other contributed and extended their valuable assistance in the completion of this study. In words of William Arthur Ward, "Feeling gratitude and not expressing it is like wrapping a present and not giving it". I would take this opportunity to thank some of the key members of my personal and professional life.

First and foremost, my utmost gratitude to my supervisor Prof. Ralf R. Müller for his inspirational guidance and motivating advice throughout my work. I must admit that he is never short of ideas related to my scientific work. I truly believe that it is a privilege to get the opportunity to learn so many things about the scientific world from his vast experiences.

Special thanks to my colleague Dr. Kimmo Kansanen for his great assistance and endurance. He always managed to find time for discussion on technical issues related to my work. I would like to extend my feelings of gratitude to Prof. Geir Øien who co supervised me. As a project manager of my funding project, he was always more than helpful to resolve the technical and administrative issues.

I would like to mention deep feelings of gratitude for Prof. Giuseppe Caire who hosted me for four months at University of Southern California and introduced me to some new areas of research. I spent two months at Eurecom research institute with Prof. David Gesbert. Many thanks go to Prof. David Gesbert for sharing his valuable time and ideas with me, and constantly encouraging me to work on interesting problems.

It is the time to praise the informal, open and friendly environment at the signal processing group. Specially, I am thankful to our group secretary Kirsten Ekseth for all the administrative help during these four years. I cannot forget all the social gatherings and fun time we had with my fellow PhD students and other group members.

I managed to get married during my PhD. A lot of thanks to Hina, my wife, for her patience and understanding when I was doing my PhD stuff at home on weekends and she was looking after our baby Aayan, alone.

Last but not the least; I am greatly thankful to my parents and family who always have been very supportive and loving throughout my educational career. Special mention and thanks to the motivation and efforts of my father who taught me self belief and encouraged to look for the bright light in the darkest of the times.

Above all, I thank God Almighty, for giving me the intellect and intensity to complete this research, and strength to overcome all the hurdles in life with endurance.

M. Majid Butt
December 2010, Trondheim

Contents

List of Figures

List of Tables

Notation and Symbols

a_l	Power allocated to transmit antenna l
C	Spectral efficiency per channel
$\mathcal{C}$	Capacity region
D	Distance between two base stations
$\mathcal{E}$	Edge of the acyclic directed graph
$\mathbb{E}$	Expectation operator
E_k	Transmitted power by user k
E_k^R	Received power of user k
E_{sys}	Average system energy
ΔE_{sys}	Difference in system energies
$\left(E_b/N_0\right)_{\text{sys}}$	Average energy consumption of the system per transmitted bit
f_k	Short term fading for user k
g_k	Channel gain of user k
$\mathcal{G}$	Directed acyclic graph
$\mathbf{G}_{\mathcal{S}\mathcal{S}_c}$	Transfer Matrix for the cut with source $s \in \mathcal{S}$ and destination $d \in \mathcal{S}_c$
$I(\rho)$	Outer cell interference represented as Gaussian noise on cell ρ
K	Number of users in the multiuser system
$\mathcal{K}_m$	Set of users scheduled in same frequency band
$L(i,j)$	Number of packets scheduled as a result of transition from state i to state j

M	Number of Channels in a multichannel system
max	Maximum operator
min	Minimum operator
n	Hard deadline representation in a Markov process
$n_{i,j}$	Link capacity between nodes i and j
N	Number of nodes in a cooperative network
N^{const}	Constant Size of the arriving packet
N_0	Noise Spectral density
$\mathbf{P}$	Transition probability matrix
$\mathbf{P}^{\text{opt}}$	Optimal transition probability matrix
P_{arr}	Arrival probability
q	Deadline offset
R_k	Rate supported by the user k
R_{ψ_k}	Rate allocated to the user k in permutation ψ of the scheduled user indices
s_k	Path loss of user k
$\mathcal{S}_0$	Wideband Slope
$\mathbf{S}$	Down shift matrix
$\mathbf{S}^{\text{nid}}$	Offset Matrix for Non-identical deadline arrival case
T	Temperature in Simulated Annealing process
T_W	Transmission time for an encoded message into signal
T_0	Initial temperature in Simulated Annealing process
$\mathcal{T}$	Throughput measured over a certain time window
$\mathcal{T}_{\text{min}}$	Minimum throughput guarantee over a certain time window
t_c	Length of time window to measure the throughput
$T_{i\rightarrow j}$	Transition from backlog state i to backlog state j
$\mathcal{T}_{min}$	Minimum Throughput of the user

$\mathcal{V}$	Vertex of the acyclic directed graph
$\mathcal{W}_i$	Message set for user i
W	Waiting Time
$\mathbf{x}$	Vector of transmitted symbols
$\mathbf{y}$	Vector of received symbols
z	Zero mean circularly symmetric Gaussian random noise vector
α	Path loss exponent
α_{ij}	Transition probability from backlog state i to backlog state j
β	Constant to model outer cell interference
$\beta_{\text{in}}(\mathcal{S})$	Set of nodes in set $\mathcal{S}$ with a direct link to nodes in set $\mathcal{S}^c$
$\beta_{\text{out}}(\mathcal{S})$	Set of nodes in set $\mathcal{S}^c$ with a direct link from nodes in set $\mathcal{S}$
Γ	Spectral Efficiency of the system
δ	Radius of the forbidden region around the base station
δ_k	Violation probability for user k
ϵ	Rate of arrival for Bernoulli arrival process
θ_d	Per user Packet Dropping probability
θ_l	Phase shift applied at antenna l to the signal
θ_{in}	Per user dropping probability for the users in the inner part of the cell
θ_{out}	Per user dropping probability for the users in the outer part of the cell
θ_s	System wide dropping probability
ϑ	Rate of arrival for a Poisson arrival process
κ	Rician factor
$\kappa_{i\to j}$	Transmission threshold for the transition from backlog state i to j
κ_i	Transmission threshold for the transition from backlog state i to j when $j = 1$

κ_j	Transmission threshold for the transition from backlog state i to j when i is irrelevant
$\vec{\kappa}$	Vector of transmission thresholds
$\vec{\kappa}^{\text{opt}}$	Optimal Vector of transmission thresholds
λ	Weighting factor for LWDF scheduler
π_i	Limiting probability of state i in a Markov process
τ_{max}	Hard Deadline for transmission
τ_{max}^{vr}	Virtual hard deadline for transmission of data to fulfill throughput guarantee
χ	Deadline distance
$\psi_k^{(m)}$	Permutation of scheduled user indices in frequency band m

List of Acronyms

BA	Boltzmann Annealing
BS	Base Station
CDMA	Code Division Multiple Access
CSI	Channel State Information
DDOS	Deadline Depending Opportunistic Scheduling
DDPS	Deadline Dependent Partial buffer Scheduling
FA	Fast Annealing
FIFO	First in First out
FRPS	Future Rate Prediction based Scheduling
HOL	Head of Line
LOS	Line of Site
LSB	Least Significant Bit
LWDF	Largest Weighted Delay First
MC	Multiple Cell
MDDOS	Modified Deadline Dependent Opportunistic Scheduling
MLWDF	Modified Largest Weighted Delay First
MS	Mobile Station
MSB	Most Significant Bit
OSP	Opportunistic Superpositioning
PFS	Proportional Fair Scheduling

QoS	Quality of Service
RRM	Radio Resource Management
SA	Simulated Annealing
SC	Single Cell
SDDPS	Sequential Deadline Dependent Partial buffer Scheduling
SINR	Signal to Interference and Noise Ratio
SIR	Signal to Interference Ratio
SNR	Signal to Noise Ratio
SVU	Single Virtual User
TDMA	Time Division Multiple Access
WSN	Wireless Sensor Network

Chapter 1

Introduction

1.1 Motivation

Wireless communication has gone through a revolutionary phase in recent times. There is a lot of interest in exploiting the communication opportunities provided by the multiuser systems. Cooperative and opportunistic communication provides us new degrees of freedom. The multiuser diversity concept is already integrated into the downlink design of IS-856. With the introduction of new applications, demand for efficient use of resources has increased. These applications have widely contrasting requirements in terms of delay, throughput and loss tolerance. The task of efficient transmission and scheduling of data gets tougher if these requirements are to be fulfilled and practically, it becomes hard to meet all the requirements in a scheduling algorithm. A good communication scheme exploits every degree of freedom available to maximize the use of expensive resources.

The goal of this dissertation is to exploit delay tolerance and channel variations of the user channels to design the scheduling schemes which fulfill the hard deadline delay guarantees and minimize system energy resources. We consider a general multiuser system but the results are specifically useful for wireless sensor network applications where energy saving enhances the life time of sensor battery. We address the practically relevant scenarios and analyze the results for the proposed scheduling schemes. We integrate these practical consideration into our schemes such that the task of the scheduler remains simple and complexity is handled through preprocessing of data. We use large system analysis to investigate our schemes but numerical results provide evidence that the schemes work well for small number of users as well.

1.2 Outline of the Thesis

The rest of the dissertation is organized as follows. It comprises of two main parts. The first (main) part (Chapter 3 to Chapter 6) deals with deadline constrained opportunistic scheduling while the second part (Chapter 7) addresses cooperative relaying.

Chapter 2: We discuss background of this work and review the related literature briefly in Chapter 2.

Chapter 3: In Chapter 3, we propose an opportunistic scheduling scheme for a hard deadline constrained multiuser system. We assume constant arrival of a single packet in each time slot. The scheduler empties the buffer when a user is scheduled for transmission. The aim is to minimize the system energy while obeying hard deadline constraint. We analyze the scheme in asymptotically large user limit. As the energy function is a non-convex function, we use Simulated Annealing algorithm to optimize the transmission thresholds.

Chapter 4: In Chapter 4, we extend the results in Chapter 3 and propose a scheduling scheme which schedules data in small chunks depending on the instantaneous channel conditions of the user. The results are extended for random arrival case. Numerical results show the energy efficiency of the scheme as compared to emptying buffer scheme.

Chapter 5: In Chapter 5, we address the practical considerations for the proposed scheduling scheme in Chapter 4. We propose a scheduling scheme which gives comparable results to the scheme discussed in Chapter 4 at lower complexity. We propose a heuristic algorithm to compute the transmission thresholds and compare the numerical results with the optimal scheme. We extend our results to more practically relevant scenarios when all the arriving packets have individual non–identical deadlines. Also, we consider the system where all the users cannot be provided absolute deadline guarantees and a certain predefined allowed proportion of the packets are dropped to save the system energy.

Chapter 6: The work from Chapter 3 to Chapter 5 considers a single cell case. In Chapter 6, we consider the multicell case of the opportunistic schedulers proposed in previous work. We model the inter–cell interference and analyze the scheduling schemes in large user limit.

Chapter 7: Chapter 7 contains the second part of the thesis. We address the issue of cooperative communication between sensor nodes. We

consider a system where network nodes operate as source and relays at the same time and send data to a common node. We consider deterministic network model and characterize the capacity region for certain topologies of such networks. Then, we generalize our results and prove achievability for a general deterministic network. We also extend our results to the case when data from the sources is correlated.

Chapter 8: We conclude with the main contributions of this work in Chapter 8. We summarize the important results and their impact on the state of the art communication networks. Also, we discuss some of the open problems and their impact on the improvement of scheduling schemes proposed.

Appendix A: Appendix A introduces the channel model used in this work and we derive channel distributions for the proposed schemes and Rayleigh fading.

Appendix B: Appendix B contains proof of the properties for the opportunistic scheduling scheme proposed in Chapter 4.

Appendix C: Appendix C presents the relation between transmission thresholds and transition probabilities used to model the proposed scheduling schemes.

Appendix D: In Appendix D, we prove the equivalence of channel distributions of the scheduled packets for the case of non–identical deadlines (discussed in Chapter 5).

Appendix E: In Appendix E, we prove Theorem 7.1 and Theorem 7.2 introduced in Chapter 7 for the deterministic cooperative network.

Chapter 2

Background of the Work

In this chapter, we discuss the fundamental aspects of opportunistic and cooperative communication in wireless networks. First, we review the radio resource management (RRM) problem briefly and focus on the issues related to scheduling and its applications in wireless networks.

2.1 Radio Resource Management (RRM)

With rapid increase in demand for resources in wireless applications, it is important to use the available radio resources efficiently. The task of efficient spectrum management is termed as radio resource management (RRM) [Zander, 1997]. However, radio resource is not limited to radio spectrum. It includes management of access rights for the individual users, the time period a user is active, transmission power, admission control for the user and policies for user mobility, etc. [Zhang, Hu, and Fujise, 2007]. Reference [Berry and Yeh, 2004] reviews resource allocation problem for multi access channels in detail and discusses cross layer approaches by considering physical and networking layer together.

In the following, we discuss some of the radio resource allocation mechanisms briefly.

2.1.1 Power Control

Power control is the intelligent use of transmit power to effectively use a communication system. Power control is important mechanism for wireless systems. Without employing power control, *near far* effects may arise and affect the system performance. To achieve the same signal to noise ratio (SNR), the mobile stations (MS) near the base station (BS) transmit at

less power as compared to the MSs at the cell border. In a multicell environment, this kind of power control has adverse effects on the inter–cell interference [Goldsmith, 2005]. In literature, there are several methods of power control, termed as open-loop power control and closed loop power control [Rappaport, 2001]. Closed loop power control requires that the receiver compares the estimated signal to interference ratio (SIR) with target SIR and instructs the transmitter to increase or decrease transmit power accordingly. The target SIR is controlled by outer loop power control. The outer loop measures the link quality in terms of frame and bit error rate and adjusts the target SIR accordingly.

2.1.2 Admission Control

The system capacity evaluated in terms of available radio resources reduces after admission of a new user. Specifically, it increases the interference to other users in a heavily loaded system and may cause system instability [Zhang *et al.*, 2007]. It results in call dropping which is considered more annoying effect than denying a channel (call) by user point of view. For example, in a Code Division Multiple Access (CDMA) system , the cell coverage reduces as a result of increase in interference. Therefore, admission control is required to ensure the cell coverage. However, admission control can have different controlling parameters like service level, load, user profile etc. in order to optimize the system performance. In a cellular system, the admission controller keeps track of the radio resources for all the ongoing calls in the current cell as well as in the adjacent cells. This is essential to keep the call going in the case when the mobile user moves into an adjacent cell and requires resources for handoff. Therefore, it is important to keep some resources reserved for handoffs from the neighbouring cells.

2.1.3 Mobility Management

The task of keeping track of the location of MS in a large wireless network is termed as mobility management [Zander, 1997]. Mobility can further be classified as discrete and continuous. Discrete terminal mobility is the ability of a terminal to move to a new location, connect to the network and continue to access the service again [Chen and Zhang, 2004]. It is often referred to as portability. Continuous mobility is the ability of a terminal to remain connected with the network and service while on the move. There are certain trade offs involved in determining the whereabouts of the user. It involves the functions of location management, packet delivery to the mobile, handoff, roaming and network access control. The details of these

functions and their operations are omitted here. The interested reader is referred to [Chen and Zhang, 2004] for the details.

2.1.4 Scheduling

Scheduling is the fundamental function of wireless communication systems which divides and allocates available radio resources to different MSs. Based on different matrices, the scheduler decides to allocate resources to the users. These matrices include individual quality of service (QoS) requirements from the users. QoS can be defined in terms of delay requirement or throughput requirement. Furthermore, these parameters can be guaranteed on *long term* or *short term* basis depending on the requirement of the target application. One of the simplest schedulers used in Time Division Multiple Access (TDMA) systems is round robin (RR) scheduler which schedules users periodically. Every user is scheduled for a fixed number of time slots and then waits for her turn. The disadvantage with this type of schedulers is that users with different QoS requirements cannot be satisfied. Also, if a user has no data to schedule, still she gets the channel for fixed number of time slots and wastes system resources that can be used by other users.

First In First Out (FIFO) is another simple scheduler that schedules the users according to their waiting times. The user k whose head of line (HOL) packet has spent the largest time at the base station is selected for transmission.

$$j = \underset{k}{\arg\max} \quad W_k(t) \tag{2.1}$$

where $W_k(t)$ is the waiting time of the HOL packet. In [Stolyar and Ramanan, 2001] a similar scheduler, Largest Weighed Delay First (LWDF), is proposed which provides QoS guarantees in the form of a deadline n_k and an allowed violation probability δ_k for user k. For stationary delay W_k of the kth user, this scheduler always chooses for service the longest waiting customer for which the current weighted delay $\lambda_k W_k(t)$ is maximal.

$$j = \underset{k}{\arg\max} \quad \lambda_k W_k(t) \tag{2.2}$$

where λ is a weighting factor. If all the elements in the vector of weighting factors $\vec{\lambda}$ are unity, LWDF scheduler becomes FIFO scheduler. The authors in [Stolyar and Ramanan, 2001] show that for large delays and small allowed violation probabilities, the LWDF scheduler with weights

$\lambda_k = -log\delta_k/n_k$ is *nearly* optimal in order to satisfy the QoS constraints. These policies do not take channel conditions into account and performs poorly as compared to the scheduling algorithms which exploit channel conditions [Shakkottai and Stolyar, 2001].

On the other hand, we have an other class of schedulers which exploits the opportunities provided by the wireless channel to improve the system performance. This work specifically deals with such scheduling algorithms in a multiuser communication system. Highly unpredictable channel conditions of wireless system adds an other dimension to the scheduling problem. Before going into the details of scheduling problem, we review some important characteristics of the wireless channel in the next section.

2.2 Wireless Channel Characteristics

Strength of a typical wireless channel varies over frequency and time. We can characterize this variation as two effects.

2.2.1 Large scale fading

The channel varies in terms of path loss and shadowing from the large objects. Path loss is a function of the distance between a transmitter and the receiver and modeled as a path loss exponent. It is independent of the frequency.

2.2.2 Small scale fading

This effect arises due to constructive and destructive interference of the multiple signal paths between the transmitter and receiver and can further be divided into two parts.

Flat Fading

If the bandwidth of the input is considerably less than the coherence bandwidth, the channel is usually referred to as flat fading.

Frequency Selective Fading

When the bandwidth is much larger than the coherence bandwidth of the channel, the channel is said to be frequency selective.

2.3 Diversity

In the presence of a single signal path, the reliable communication depends heavily on signal to noise ratio (SNR) at that path. When the path is in deep fade, any communication scheme will suffer from errors. This is a natural motivation for using more than one signal paths for reliable communication between a source-destination pair. If all the paths fade independently, there is a high probability that at least one of them would be strong enough to make the reliable communication. The idea of providing multiple paths for the reliable communication between two nodes is termed as *diversity* [Tse and Viswanath, 2005].

Diversity can be provided by using degrees of freedom in time, frequency or space. The idea of transmitting a coded symbol on independently fading coherence periods is called time diversity. Interleaving can further exploit the temporal diversity to improve the performance. For a frequency selective channel, degrees of freedom in frequency domain can be exploited by transmitting coded symbol on multiple frequency channels. Similarly, for a system with sufficiently spaced multiple transmit and/or receive antennas, spatial diversity can also improve the performance. Specifically, in case of cellular network, macro-diversity exploits the fact that a signal from a mobile can be received from two base stations [Tse and Viswanath, 2005]. Any combination of these diversity schemes typically improves the performance of a communication system.

All of these schemes require use of certain system resources to combat the fading effect. As resources such as frequency, time, antennas are always expensive and using resources with effective communication schemes is complex, there is always a tradeoff involved in enhancing the diversity paths.

2.4 Multiuser Diversity and Maximum Rate Scheduler

In a large multiuser environment, every user experiences an independent channel. If we allow, only a single user to transmit at a given time, we can achieve another form of diversity. In [Knopp and Humblet, 1995], the authors propose a scheduling scheme which schedules a user with the best channel gain for transmission. This concept is known as *multiuser diversity* and the scheduler is known as Maximum-Rate Scheduler as it maximizes the rate supported by the channel. Multiuser diversity is inherent in the system and we do not need to create it in contrast to all the diversity

schemes mentioned in Section 2.3. The multiuser diversity gain arises due to improvement in the channel gain from $g[t]$ to $\max_k g_k[t]$ in a time slot t with K users in the system. The larger the number of users K in the system, the greater the multiuser gain would be. By allowing the user with the strongest channel, the shared channel resources are used most efficiently and throughput is maximized. To exploit the multiuser diversity, channel state information (CSI) at the transmitter side needs to be known and we make this assumption throughout this work.

It is interesting to note that conventional diversity techniques are designed to counteract fading while multiuser diversity gain exploits the variation in channel gain. More fluctuations in channel gains ensure that there is a high probability of finding a user with greater channel gain than the mean level. Contrast to the schedulers discussed in Section 2.1.4, by allocating all the system resources to the stronger user, the benefit of the stronger channel is fully exploited. This fact can also be verified from the fact that multiuser gain is significantly smaller in case of Rician fading as compared to Rayleigh fading [Tse and Viswanath, 2005]. In Rician fading, there is a strong line of sight (LOS) path between the transmitter and receiver in addition to multiple small reflected paths. The parameter Rician factor κ is defined as the ratio of energy in direct path to the energy in reflected paths. The presence of LOS component in Rician fading distribution makes the channel less random as compared to Rayleigh channel. Equivalently, for asymptotically large user limit, the mean of the equivalent channel gain approaches to $\frac{1}{1+\kappa}$ for Rician channel as compared to one for Rayleigh channel.

2.5 Variants of Maximum Rate Scheduler

In literature [Shakkottai and Stolyar, 2000], [Shakkottai and Stolyar, 2001], [Andrews, Kumaran, Ramanan, Stolyar, Vijayakumar, and Whiting, 2000], [Chaponniere, Black, Holtzman, and Tsc, 2002] different modified versions of [Knopp and Humblet, 1995] have been discussed. In [Andrews *et al.*, 2000] a modified version of LWDF is proposed which takes channel variations of the users into account. This is called Modified Largest Weighted Delay First (MLWDF) scheduler. The scheduler schedules a user which maximizes

$$j = \max_k \quad \gamma_k W_k(t) R_k(t) \tag{2.3}$$

where $\gamma_k > 0, k = 1, \ldots, K$ is an arbitrary set of constants and $R_k(t)$ is the actual rate supported by user k at time t. The following choice of γ_k has been shown to exhibit good QoS results.

$$\gamma_k = \frac{\lambda_k}{\overline{T_k}} \tag{2.4}$$

where $\overline{T_k}$ is measured short term average throughput for user k and λ_k is parameterized similarly as in LWDF in Section 2.1.4.

In [Shakkottai and Stolyar, 2000, 2001], the authors propose a scheduler based on the following *exponential* weighting factor.

$$j = \arg\max \gamma_k R_k(t) \exp\Big(\frac{\lambda_k W_k(t) - \overline{\lambda W}}{1 + \sqrt{\overline{\lambda W}}}\Big) \tag{2.5}$$

where γ and λ are same as in MLWDF scheme and

$$\overline{\lambda W} = \frac{1}{K} \sum_k \lambda_k W_k \tag{2.6}$$

This scheme tries to equalize the weighted delays of all the users when their differences are large. If one of the user has large weighted delay than others, the exponent term becomes very large and overrides the channel conditions. This makes the user get selected for transmission. However, if weighted delay difference is small, the exponential term becomes equal to one and no priority is given to the user [Andrews *et al.*, 2000], [Shakkottai and Stolyar, 2000].

For both of the M-LWDF and exponential rule schedulers, the authors prove that both of the schedulers are throughput optimal in the sense that they make the queues stable in any system for which stability is feasible at all with any other rule.

2.6 Proportional Fair Scheduling

The solution proposed in [Knopp and Humblet, 1995] has an inherent drawback for the real time or hard delay constrained systems. In the large multi-user environment, many users will never get the opportunity for transmission in the maximum allowed time slots and data has to be dropped. Similarly, due to different distances from the base station, the channel of the user closer to BS will always be better than the user at the edge of the cell. To address the issues of fairness and delay, Proportional Fair Scheduling (PFS) algorithm was proposed in [Chaponniere *et al.*, 2002],[Kelly, Maulloo,

and Tan, 1998]. This scheduler has been used in downlink standardization IS-856 as a base scheduler.

The task of the scheduler is to select a user for transmission in each time slot based on the requested rates received from the users. A simple round robin scheduler can do the job but in a dynamic wireless environment, it results in wastage of energy resources. PFS scheduler exploits the channel variations to schedule the users in an energy efficient manner. It keeps track of average throughput $\mathcal{T}_k[t]$ of each user in an exponentially weighed window of length t_c [Viswanath, Tse, and Laroia, 2002]. In a time slot t, the base station receives the requested rates $R_k(t)$ from all the users and schedules the user k^* with the largest $\frac{R_k[t]}{\mathcal{T}_k[t]}$. For each user k, the average throughput $\mathcal{T}_k[t]$ is updated with an exponentially weighted factor and given by

$$\mathcal{T}_k[t+1] = \begin{cases} (1-1/t_c)\mathcal{T}_k[t] + 1/t_c R_k[t] & k = k^* \\ (1-1/t_c)\mathcal{T}_k[t] & k \neq k^* \end{cases} \tag{2.7}$$

PFS schedules a user when its instantaneous channel quality is relatively higher than its own average channel quality over the time scale t_c. In other words, data is transmitted to a user when her channel is near to her own peak. The users compete for resources not directly based on the requested rates but based on the rates normalized by their average throughput. It has been observed that the total throughput increases with the number of users because greater the number of users, greater is the probability that a users will find a channel near her peak [Viswanath *et al.*, 2002].

2.6.1 The Role of Mobility

Total throughput is also affected by the mobility of the users. In a high mobility environment, the rate of channel variations is much greater than the low mobility case. This implies that over the latency time scale, the peaks of channel fluctuations are likely to be much higher in the mobile environment and inherent multiuser diversity is more pronounced in high mobility environment. It seems as if multiuser diversity will increase the throughput but there is a certain limitation on that. At very high speed, the users have troubles in measuring and tracking channel variations and the predicted channel is a low pass smoothed version of the actual fading process [Tse and Viswanath, 2005]. Therefore, in spite of highly dynamic environment, opportunistic communication can only be used when channel estimation can be done with a reasonable accuracy.

2.6.2 Opportunistic Beamforming

In [Viswanath *et al.*, 2002] a scheme has been proposed that induces a random fading when the environment has little scattering or fading is slow. For a downlink, multiple antennas are used at the base station to transmit the same signal from each antenna modulated by a gain whose phase and magnitude is changing in time in a pseudo random fashion. The gain in different antennas are independent and channel variation is created due to constructive and destructive addition of the signals from the different transmit antennas. The overall signal to interference plus noise ratio (SINR) is tracked by each user and is fed back to the base station [Viswanath *et al.*, 2002].

Consider a system with L transmit antennas at the base station. Let $g_{lk}(t)$ be the complex channel gain from antenna l to the kth user in time slot t. In time slot t, the same block of symbols $x(t)$ is transmitted from all the antennas except that it is multiplied by a complex number $\sqrt{a_l(t)}e^{j\theta_n(t)}$ at antenna l such that $\sum_{l=1}^{L} a_l(t) = 1$. The received signal at user k is given by

$$y_k(t) = \Big(\sum_{l=1}^{L} \sqrt{a_l(t)}e^{j\theta_l(t)} g_{lk}(t)\Big) x(t) + z_k(t) \tag{2.8}$$

where $x(t)$ is the vector of T_s transmitted symbols in time slot t, $y_k(t)$ is the vector of T_s received symbols of user k and $z_k(t)$ is an independent and identically distributed (i.i.d.) sequence of zero mean circularly symmetric Gaussian random vectors. The symbols $a_l(t)$ and $\theta_l(t)$ denote the fractions of power allocated to each of the transmit antennas and phase shift applied at each antenna to the signal, respectively. Variation of these parameters introduce fluctuations in the physical channel even if the original channel $g_{lk}(t)$ have very little fluctuations [Viswanath *et al.*, 2002]. Each receiver k feedbacks the overall SNR of its channel to the base station (and not the individual channel gain $g_{lk}(t)$) and therefore requires a single pilot signal for the channel measurement.

We consider the case of slow fading and fast fading separately.

Slow Fading

We consider the case when the channel gain of each user $g_{lk}(t) = g_{lk}$ remains constant during the latency time t. Without using additional antennas, the received SNR for this user would have remained constant. If all the users in the system experience a slow fading, multiuser diversity cannot be exploited. However, using multiple antennas on the transmit side, fluctuations in the channel gain $g_k(t)$ are introduced and opportunistic scheduling

can now exploit multiuser diversity. It should be noticed that to be able to beamform a user, amplitude and phase of all the antennas must be known at the base station. As the proposed scheme keeps track of SNR only, at any time, the transmission is scheduled to a user which is closest to its beamforming configuration. This type of beamforming is called opportunistic beamforming [Viswanath *et al.*, 2002].

Fast Fading

In a fast fading environment, rate of fluctuations is already fast enough. Here, the impact of opportunistic beamforming depends on the fact that how the stationary distributions of the over all gains can be modified by a power and phase randomization. In case of independent Rayleigh fading $g_{lk}(t)$ has the same distributions as $g_k(t)$ and overall gains are independent across the users. Therefore, in a fast fading environment, opportunistic beamforming does not provide any performance gain. However, in case of Rician fading, a non time varying line of sight (LOS) component is also present. When the ratio of energy in the LOS component to the diffused components is large, opportunistic beamforming can largely increase the dynamic range of fluctuations and thus multiuser diversity can be exploited by the opportunistic scheduling [Viswanath *et al.*, 2002].

2.7 Delay-Energy Tradeoff

In [Tse and Hanly, 1998; Hanly and Tse, 1998] the notion of throughput capacity region and delay limited capacity region has been introduced. The first term refers to the application of Shanon capacity on fading channels. The channel statistics are assumed to be fixed and the code word length can be chosen arbitrarily large to average over all the fading of the channel. To achieve these rates, the users will experience long delays as long term rates are achieved averaged over the fading process [Tse and Hanly, 1998].

On the other hand, we have some situations where users cannot be allowed to wait for long time to achieve the rates. Delay limited capacity refers to the situation where the time requirement of the application is shorter than the time scale of the channel variations. Therefore, it is not possible to average over all the fading states and it is necessary to maintain the desired rate at all the fading states. It has been shown that successive decoding is the optimal solution and optimal decoding order and resource allocation can be found explicitly as a function of fading states [Hanly and Tse, 1998].

2.7.1 Scheduling For Hard Deadline Delay Constrained Applications

A lot of applications require data to be transmitted before a fixed deadline called hard deadline [Hajek and Seri, 1998; Agarwal and Puri, 2002]. For example, many applications in wireless sensor networks (WSN) and multimedia belong to this class of applications. At the same time, it is important to use resources like energy and spectrum efficiently. The idea behind deadline constrained energy efficient scheduling is to schedule transmission of data such that both of the requirements are fulfilled.

In [Tarello, Sun, Zafar, and Modiano, 2008], the authors consider energy minimization problem for the deadline constrained applications. They consider a system with a single transmitter and K users. The channel for each user is discretized and can be in one of the finite sets of states. In each scheduling operation, a single user is scheduled for transmission. They consider two cases of rate-power curves. For both the cases, they obtain dynamic programming based optimal solutions. When the rate power relation is linear, they obtain a threshold based scheduler which follows the optimal stopping theory formulation in [Bertsekas, 2007]. For the convex case of rate-power curve, a heuristic algorithm is proposed which gives a solution quite close to the optimal. Similar approach is used in [Fu, Modiano, and Tsitsiklis, 2006] to maximize expected data throughput for a single transmitter and single receiver in presence of deadline delay constrained.

A similar approach is applied in [Lee and Jindal, 2009] where the authors consider the same problem for a point to point network. They consider a packet of B bits which has to be transmitted within hard deadline of n time slots. During the transmission of the packet, no other packets are scheduled. The authors obtain close form expressions for the optimal policy only for the case $n = 2$ using dynamic programming solution. For $n > 2$, the optimal policy is numerically determined. Based on optimal policy for the case when $n = 2$, heuristic sub optimal scheduling policies are considered. As a variant of the problem they consider the scheduling of the complete packet in one time slot and propose optimal channel threshold based scheduling scheme. It should be noted that optimal solution is obtained only when either the rate-power curve is linear [Tarello *et al.*, 2008] or scheduling of a single packet is considered following the frame work of optimal stopping theory.

Following the work in [Tarello *et al.*, 2008; Lee and Jindal, 2009], we consider the problem of scheduling data before a hard deadline n. However, we do not limit ourselves to the constraint of transmission of single packet during the whole transmission period as it is too impractical assumption.

The focus of this work is to generalize the problem. We analyze the problem asymptotically for large number of users. Then, we propose heuristic scheduling schemes which provide comparable results to the optimal solution.

2.8 Literature Survey

In this section we review some further work relating opportunistic scheduling. In literature, a lot of work deals with average delay parameterizations and opportunistic scheduling. For example, reference [Berry and Gallager, 2002] deals with the tradeoffs between average delay and average power for a single user case. Average delay-energy relation is analyzed asymptotically when delay approaches to infinity. Similar work in [Rajan, Sabharwal, and Aazhang, 2004] addresses average delay-energy tradeoff for the case of bursty traffic. A suboptimal scheduler called log-linear scheduler is proposed which provides performance close to the optimal scheduler.

In [Liu, Chong, and Shroff, 2003], the authors discuss the performance of opportunistic scheduling under different fairness and minimum performance requirement constraints. Optimal solutions, implementation considerations and parameter estimation procedures are discussed as well. In reference [Wu and Negi, 2005], the authors use multiuser diversity to provide statistical quality of service (QoS) in terms of data rate, delay bound, and delay bound violation probability. In [Phan and Kim, 2007], threshold–dependent opportunistic transmission is discussed for WSNs using the IEEE 802.11 standard. Significant improvement is reported as compared to the non–opportunistic version.

In [Coleman and Medard, 2004] a similar work to [Tarello *et al.*, 2008] considers a scheduling policy without a centralized scheduler and discusses the energy delay trade–off with full and partially shared information about the queue lengths of all the users. In [Chan, Neely, and Mitra, 2007], an exact solution for the average packet delay under the optimal offline scheduler is presented when an asymmetry property of packet inter–arrival times and packet inter–transmission times holds. Online scheduling algorithms that assume no future packet arrival information are discussed as well. Their performances are comparable to those of the offline schedulers which assume identically and independently distributed inter-arrival times. The results of [Berry and Gallager, 2002] have been extended to the multiuser context in [Neely, 2007]. It is found that to achieve an average power within the $\mathcal{O}(1/n)$ of the minimum power required for network stability, there must be an average queuing delay greater or equal to $\Omega(\sqrt{n})$. Recently, ref-

erence [Neely, 2009] shows that fractional packet dropping changes energy-delay square root law to a logarithmic law. In [Neely, 2010], an algorithm is provided that ensures a worst case delay guarantee and maximizes a throughput utility. Reference [Biyikoglu and Gamal, 2004] proposes an iterative algorithm for finding the optimal offline scheduler. Moreover, an energy saving online scheduler is discussed which adapts to both the channel fading and the backlog.

In [Kabamba, Meerkov, and Tang, 2005], the authors show that for any modulation and coding schemes and under general assumptions on channel model, the optimal policy is always of threshold nature. They discuss the tradeoffs between long and short term power efficiency, long term average throughput and short term performance. To improve the fairness among the users at the cell border, they propose an adaptive threshold policy and prove that it is both power efficient and location fair. Reference [Kim and Hwang, 2009] discusses a threshold based scheduling scheme for downlink which fulfills the ergodic rate requirements of the users. The authors discuss feasibility of the given rate requirements and propose feasible thresholds that maximize the ergodic sum rate while guaranteeing the ergodic rate requirements. The work in [Chen and Jordan, 2009] addresses the problem of scheduling multiple transmissions on the downlink with performance guarantees in terms of probabilities that the short term throughput exceeds user specified throughput. The authors consider the case when the channel does not vary during the time slot and provide optimality conditions. An online scheduling algorithm is proposed which aims to maximize the time when short term throughput of the user exceeds a target throughput. Reference [Hassel, Øien, and Gesbert, 2007] provides approximate expressions for the minimum throughput guarantee violation probability. Closed form expressions for the violation probability are evaluated for some of the existing schedulers and the result show that approximations are quite useful for the practical networks with correlated channels for the users and realistic throughput guarantees.

The role of opportunistic scheduling in a network has been a well investigated topic. Opportunistic scheduling has been addressed in different type of networks and problem settings. For example, the ideas of opportunistic scheduling for adhoc networks have been discussed in [Zheng, Pun, Ge, Zhang, and Poor, 2008]. The authors propose a threshold based scheduling scheme for the case when the transmitter has perfect knowledge of channel. In [Zheng, Ge, and Zhang, 2009], the proposed scheme is generalized for imperfect channel state information and shown that optimal scheme is still threshold based but thresholds depend on the variance of estimation error. [Urgaonkar and Neely, 2009], [Rashid, Hossain, Hossain,

and Bhargava, 2009] discuss opportunistic scheduling in cognitive radio networks. [Cui, Chen, and Ho, 2008] and [Chen, Letaief, and Cao, 2007] address network coding issues for opportunistic networks.

2.9 Background For the Deterministic Cooperative Network

In the second part of the dissertation, we discuss deterministic cooperative network. In this section, we review the main concepts and the related work briefly.

Wireless channels differ from their wired line counterpart in two fundamental aspects [Avestimehr, Diggavi, and Tse, 2007a]. First, the wireless channel is a broadcast (shared) medium and the signal from any transmitter is received by potentially many receivers. This is called *broadcast* constraint. On the other hand, any receiver observes the superposition (linear combination) of signals from possibly many transmitters. This is called *interference* constraint. The simultaneous presence of these two constraints makes a general wireless network quite difficult to analyze.

The multiuser Gaussian channel that models a relay network, unfortunately, has so far escaped a sharp general characterization, even in the simplest case of a Gaussian relay network with a single source, single destination and a single relay [Meulen, 1977]. The capacities of Gaussian relay channel and certain discrete relay channels are evaluated in [Cover and Gamal, 1979] and a lower bound to the capacity of general relay channel is presented. In [Gastpar and Vetterli, 2005], capacity is determined for a Gaussian relay network when the number of relays is asymptotically large. Reference [Gamal and Zahedi, 2005] shows that capacity of a class of discrete memoryless relay channels with orthogonal channels from sender to relay and from the sender and relay to the receiver is equal to max-flow min-cut. The authors of [Bölcski, Nabar, Özgür Oyman, and Paulraj, 2006] discuss a setup in which a single source-destination pair equipped with M antennas communicates with the help of K intermediate nodes having single or multiple antennas. For an asymptotically large K, it is shown that with perfect channel state information at the relays, the capacity of the network scales as $(M/2)\log(K) + \mathcal{O}(1)$. In [Kramer, Gastpar, and Gupta, 2005], decode-and-forward and compress and forward strategies are studied for the relay network and it is shown that decode-and-forward achieves the ergodic capacity with phase fading if phase information is available locally and the relays are near the source node.

Recently a simple deterministic channel is proposed which is still able to capture the key aspects, broadcast and interference, of the wireless channel [Avestimehr *et al.*, 2007a; Avestimehr, Diggavi, and Tse, 2007b]. This model, referred to as the linear finite-field deterministic model, determines the capacity for a general relay network with one source and one destination, as well as the multicast capacity with one source, multiple destinations and common information only. This model is used in this work and will be presented briefly in Section 7.1. In [Aref, 1980], a similar deterministic network has been used to model the broadcast effect but it does not model the interference effect. The authors of [Gamal and Aref, 1982] determined the capacity of a semi-deterministic network with a single relay. Similarly, an achievable rate region is presented for a general deterministic relay network in [Avestimehr, Diggavi, and Tse, 2007c; Avestimehr *et al.*, 2007b]. It is shown that capacity for such a network follows the same interpretation as max-flow min-cut solution for a wire line network.

In this work, we extend the ideas of [Avestimehr *et al.*, 2007a,c,b] and discuss the achievable rate region of a relay network in a setup that is very relevant for WSN. We consider the "sensor reachback problem" [Barros and Servetto, 2006] for a linear finite-field deterministic network with arbitrary topology, a single destination node and independent information at the source nodes. We show that the capacity region for this network is given by the cut-set bound and takes on a very simple and appealing closed-form expression. Also, for a specific sources correlation model, we find necessary and sufficient conditions for the sources transmissibility. This result reminds closely Theorem 1 of [Barros and Servetto, 2006], with the following main differences: on one hand, the result of [Barros and Servetto, 2006] is more general since it applies to general correlated discrete sources observed at the sensor nodes and general noisy channels. On the other hand, our result applies to networks with broadcast and interference constraints while the result of [Barros and Servetto, 2006] requires "orthogonal" channels, i.e., with neither broadcast nor interference constraints.

Part I

Opportunistic Multiuser Communications

Chapter 3

Deadline Dependent Opportunistic Scheduling

Energy saving and delay constraints are one of the most demanding requirements for the state of the art wireless communication networks. Specifically, wireless sensor networks (WSN) put an emphasis on the energy saving aspect of the system. WSN consists of a large number of nodes with sensing, computation and communication capabilities merged together. One of the fundamental and most important tasks in designing protocols for WSN is minimization of energy expenditure to increase the life time of the communicating nodes. Sensor nodes can save the measured data locally for some duration and wait in *sleep mode* before transmitting it to the data collecting node, called fusion node. When they find good channel conditions, they *wake up* and empty the buffer by transmitting all of the data. Some applications explicitly require the transmission of sensed data before a hard deadline and therefore, often an upper delay bound for each node needs to be provided. Reference [Miao, Himayat, Li, and Swami, 2009] is an excellent survey on the topic of energy minimization techniques in wireless communications.

This work deals with the dual task of minimizing the energy of the system while providing an upper delay bound for each node. Similar task has been discussed in [Yao and Giannakis, 2005] where the authors address the problem of minimizing the energy in a sensor network by varying the transmission times assigned to different sensor nodes and propose a scheduling scheme that achieves near optimal energy efficiency.

In this work, multi-user scheduling is performed by extending the idea of a single user scheduling in the large system limit. We use the new modalities provided by the physical layer for opportunistic communication and

the scheduling parameters at MAC layer are tuned according to the channel distributions at the physical layer. This tuning of parameters across different layers can be regarded as vertical calibration in a cross layer design approach [Srivastava and Motani, 2005].

3.1 System Model

We consider a multiple–access system with K users randomly placed within a certain geographical area. Each user is provided a certain fraction of the total data rate available to the system. The required average rate R for each user is $\frac{\Gamma}{K}$ where Γ denotes the spectral efficiency of the system. We consider a time–slotted system. The arrival rate is constant for all users. Arrivals are queued in a finite buffer of size $\tau_{\max}$ before transmission. In each time slot $\frac{\Gamma}{K}$ bits arrive in the buffer of each user. We consider an uplink (reverse link) scenario but the results can be generalized to a downlink (forward link) scenario in a straightforward manner using the multiple–access broadcast duality of the Gaussian channel [Jindal, Vishwanath, and Goldsmisth, 2004].

The fading environment of the multi-band multi-access system is described as follows. Fading is termed as *short–term fading* if the coherence time of the channel is much shorter than the delay requirement of the application. If the coherence time is greater than the delay requirements, it is called *long–term fading* [Tse and Viswanath, 2005]. Therefore, the nature of the fading is associated with the variability of the channel as well as with the delay requirements of the application. Each user k experiences a channel gain[1] $g_k(t)$ in slot t. The channel gain $g_k(t)$ is the product of path gain s_k and short–term fading $f_k(t)$ i.e. $g_k(t) = s_k f_k(t)$. Path loss and short–term fading are assumed to be independent. The path gain is a function of the distance between the transmitter and the receiver and we assume it not to change within the time-scales considered in this work. Short–term fading depends on the scattering environment. It changes from slot to slot for every user and is independent and identically distributed across both users and slots but remains constant within each single transmission. This model is often referred to as block fading. For a multi-band system of M channels, short–term fading over the best channel is represented by, $f_k(t) = \max(f_k^{(1)}(t), f_k^{(2)}(t), \ldots, f_k^{(M)}(t))$.

[1]For mathematical convenience the propagation loss is introduced as a gain which, of course, is a number smaller than 1 and, therefore, actually a loss.

$E_k^R(t)$ and $E_k(t)$ represent the received and the transmitted energy for each user k such that

$$E_k^R(t) = g_k(t)E_k(t). \tag{3.1}$$

Note that the distribution of $g_k(t)$ differs from user to user. Let N_0 denotes the noise power spectral density. The channel state information is assumed to be known at the both transmitter and the receiver side. This can be accomplished by channel measurement on opposite link (downlink) in time-division duplex systems or obtaining explicit information from the receiver within the same time slot.

We allow multiple users to be scheduled simultaneously in the same frequency band. The scheme follows the results for the asymptotic user case analysis and therefore, there is no limit on the number of users scheduled simultaneously. Those scheduled users are separated by superposition coding. Let $\mathcal{K}_m$ be the set of users to be scheduled in frequency band m. Let $\psi_k^{(m)}$ be the permutation of the scheduled user indices for frequency band m that sorts the channel gains in increasing order, i.e. $g_{\psi_1}^{(m)} \leq \cdots \leq g_{\psi_k}^{(m)} \leq \cdots \leq g_{\psi_{|\mathcal{K}_m|}}^{(m)}$. Then, the energy of the user $\psi_k^{(m)}$ with rate $R_{\psi_k}^{(m)}$, as scheduled by the scheduler to guarantee an error free communication, is given by [Tse and Hanly, 1998; Caire, Müller, and Knopp, 2007]

$$E_{\psi_k}^{(m)} = \frac{N_0}{g_{\psi_k}^{(m)}} \left[2^{\sum_{i \leq k} R_{\psi_i}^{(m)}} - 2^{\sum_{i<k} R_{\psi_i}^{(m)}} \right]. \tag{3.2}$$

This energy assignment results in the minimum total transmit energy for the scheduled users. Collisions between simultaneous transmissions are avoided because in a multiuser environment, superposition coding and successive decoding ensure that data from multiple users are decoded successfully without error on the receiver side[2].

3.2 Background: Opportunistic Superpositioning(OSP)

For delay limited systems, it is necessary to provide QoS in terms of average or deadline delay. In [Chaporkar, Kansanen, and Müller, 2009] an opportunistic superpositioning (OSP) scheduling scheme is proposed for an asymptotically large multiuser system. A block fading model is considered where channel gain is a product of distance dependent path loss and

[2]The problem of error propagation in successive decoding can easily be overcome by means of iterative (soft) multiuser decoding [Caire, Müller, and Tanaka, 2004].

environment dependent short term fading. Contrast to single user scheduling in PFS, OSP schedules a group of users having instantaneous short term fading better than an opportunistic threshold. The scheduling is performed on short term fading to maintain the fairness among the users[3]. When the users are uniformly distributed in a geographical area, the users far from the base station have greater path loss as compared to the users located near the base station. If scheduling is performed on the basis of channel gain, better path loss gives an advantage to the users located near the base station. To eliminate this advantage, scheduling decisions are made on short term fading and fairness is guaranteed among all the users.

The opportunistic threshold depends on the average delay of the users. If the users can tolerate more average delay, the opportunistic threshold will be higher and less number of users will be transmitted in every time slot. OSP eliminates the users with bad channels and makes the system more energy efficient. At the same time, all the users are provided with the average delay guarantees. In OSP, when a user is scheduled for transmission, she empties her buffer by transmitting all the packets queued in the buffer. The buffer size is assumed to be infinite.

The scheme has been analyzed in large system limit but behaves nicely in finite user case as well. In [Butt, Kansanen, and Müller, 2007], the finite user convergence results of OSP have been discussed as shown in Fig. 3.1. The results show that at small spectral efficiencies, a finite user system behaves similar to asymptotic case for number of users $K = 500$ and $K = 1000$. For larger spectral efficiency values, it requires more number of users to converge to the asymptotic results. Similarly, the convergence is faster at small average delay values as compared to large values. For finite number of users, the number of simultaneously scheduled users and their buffer sizes vary from slot to slot. This variation increases at larger values of spectral efficiency. Similarly, scheduled users' buffer size grows for larger average delay constraint. To minimize these slot to slot energy variations, the system size should be large enough so that a large number of users are scheduled in every time slot and amount of data scheduled becomes nearly constant in each time slot.

3.3 Deadline Dependent Opportunistic Scheduler

Opportunistic Superpositioning (OSP) has been proposed to exploit the channel diversities of the users. We have generalized the framework of

[3]For the same reason, we perform all the scheduling decisions based on short term fading of the users in the rest of this work.

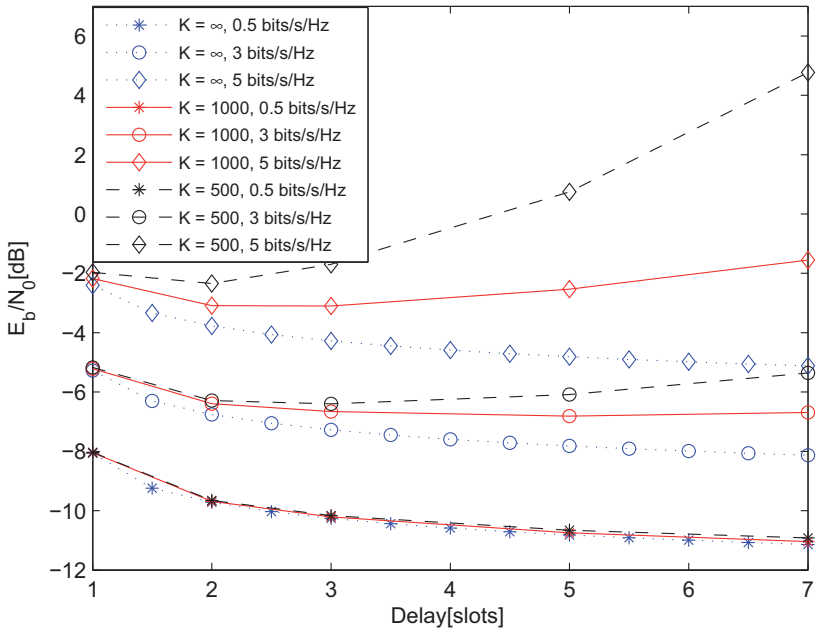


FIGURE 3.1: Convergence results for OSP for a multichannel system with number of channels M equals 10.

OSP and proposed Deadline Dependent Opportunistic Scheduling (DDOS) in this work [Butt, Kansanen, and Müller, 2008c, b].

It should be noted that computation of transmission thresholds for all the users depending on their respective backlog–states is usually not feasible in a multiuser environment because of a large state space. We consider an asymptotically large user system in this work. Therefore, the backlog–states of the users decouple and we can formulate the problem of energy efficient transmission in a multiuser system as an equivalent single user scheduling problem [Guo and Verdu, 2005]. A similar decoupling principle is applied in [Benaim and Le Boudec, 2008] and called a *mean field limit* principle.

In DDOS, individual queues of the users are observed in addition to the short term fading $f_k(t)$ for the scheduling purpose. The proposed scheme schedules a set of users experiencing high short term fading gains. If backlog of a user is equal to the maximum delay parameter, called deadline, the user is scheduled and channel is assigned regardless of its instantaneous fading state. We are using the approach of emptying the queue to keep the

scheduling operation simple.

DDOS provides the energy efficient solution by scheduling the users experiencing high short term fading gains as compared to an opportunistic threshold which depends on the backlog state of the user. As long as the backlog of the user is less than the deadline, the scheduler attempts to exploit the multiuser gain. When the deadline is reached for the *oldest* packet in the buffer, it provides the required data rate to the user regardless of its fading state. The users are scheduled in an energy efficient way as long as they do not reach the deadline. Opportunistic threshold is set to zero to ensure the scheduling of the user reaching the deadline.

Pseudo code for DDOS has been shown below.

Algorithm 3.3.1: DDOS($Backlog, Deadline$)

comment: User k knows Backlog and Deadline

$i \leftarrow Backlog$

$n \leftarrow Deadline$

comment: Current Buffer contains $i\frac{\Gamma}{K}$ data

$Buffer \leftarrow i\frac{\Gamma}{K}$

comment: Rate R is provided by the scheduler

if $(f_k > \kappa_i)$

$$\textbf{then} \begin{cases} R \leftarrow Buffer \\ i \leftarrow 1 \end{cases}$$

else

$$\textbf{then} \begin{cases} R \leftarrow 0 \\ i \leftarrow i+1 \end{cases}$$

3.3.1 Modified Deadline Dependent Opportunistic Scheduling (MDDOS)

Instead of emptying the buffer on reaching maximum buffer length, a more suitable approach would be to transmit only the data that has reached the

TABLE 3.1: Comparison of the Scheduling schemes

Scheme	Threshhold in state $i \neq n$	Threshhold in state $i = n$	Deadline Rate Allocation
OSP	κ	NA	NA
DDOS	κ_i	0	Full Buffer
MDDOS	κ_i	κ_{n-1}	Full Buffer if $f_k > \kappa_{n-1}$ $\frac{\Gamma}{K}$ if $f_k \leq \kappa_{n-1}$

deadline and this scheme is referred to as Modified Deadline Dependent Opportunistic Scheduling (MDDOS). The queue is emptied at the deadline only if short term fading gain f_k of the user is greater than the opportunistic threshold in the time slot before the deadline i.e. κ_{n-1} for the case $\tau_{max} = n$. This scheme schedules the user opportunistically even in deadline state and transmits *the oldest data unit* $\frac{\Gamma}{K}$ if deadline is reached but $f_k < \kappa_{n-1}$, resulting in further energy saving. The comparison of the characteristics of OSP, DDOS and MDDOS has been summarized in table 3.1.

3.4 Asymptotic Analysis of DDOS

We define some terms used in this work.

Definition 3.1 (Waiting Time) The waiting time W of a packet is defined as the number of time slots a packet has spent in the buffer waiting to be scheduled.

We use a Markov chain description to model the scheduling process.

Definition 3.2 (Backlog State) The backlog state in a Markov chain is the *maximum* of waiting times of all of the packets in the buffer.

$$i = \max(W_1, W_2 ... W_j) \tag{3.3}$$

For a constant arrival and identical deadline case, backlog state represents the number of packets buffered in the queue.[4] □

[4]However, this is not true when arriving packets have individual non-identical deadlines.

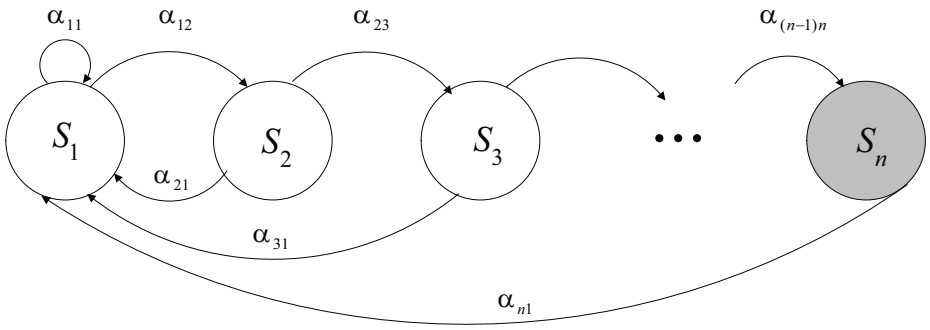


FIGURE 3.2: State diagram for the transition states of a single user for DDOS. Due to constant arrival rate, user moves into next adjacent state if not scheduled.

Definition 3.3 (Opportunistic Threshold For DDOS) An opportunistic threshold κ_i for DDOS is defined as the minimum short–term fading value allowing for the transition $T_{i\to 1}$ from backlog state i to backlog state 1. □

In the Markov state description of DDOS, the deadline constraint τ_{max} is reflected by the maximum number of states[5] n. Forward state transition $T_{i\to j}$ from a state i to the next higher state j occurs if no data is transmitted. Due to constant arrival of one packet in a time slot and identical deadline for all the arrived packets, j is always equal to $i+1$ for a forward transition. Due to *emptying buffer* property, the backward state transition occurs always from state i to state 1 and the scheduler schedules i packets for transmission.

In a Markov process if a user is in state i, then the next state j is determined according to transition probabilities α_{ij} [Rose, 2003]. Let S_t be the state of the process at time t. In order to ease notation, introduce a dummy state $i=0$ and make it impossible to be reached by the definition $\kappa_{i\to 0}=\infty \forall i$. Then, the transition probabilities are

$$\alpha_{ij} = \Pr\{S_{t+1}=j | S_t = i\} \tag{3.4}$$

$$= \begin{cases} \Pr(f > \kappa_i) & j=1 \\ \Pr(f \le \kappa_i) & j = i+1 \\ 0 & else \end{cases} \tag{3.5}$$

Transition probabilities for the next state S_{t+1} depend on the current state S_t and are independent of all the past states $S_0, \ldots, S_{t-1}$. The current state

[5] In this work, we use the term state and backlog state interchangeably.

i represents the backlog of the user of interest and the next state j depends on the short–term fading. The fading randomizes the state transitions.

Let the fading be an ergodic process. Then the limiting probability that the process will be in state j is denoted by

$$\pi_j = \lim_{t\to\infty} \Pr\{S_t = j\} \quad \forall j. \tag{3.6}$$

The limiting probability π_j is independent of state i and can be expressed as

$$\pi_j = \sum_i \pi_i \alpha_{ij} \quad \forall j. \tag{3.7}$$

The sum of the limiting probabilities of all the states must obey

$$\sum_j \pi_j = 1. \tag{3.8}$$

These $n+1$ linear equations determine the limiting probabilities for all n states.
The state transition diagram is shown in Fig. 3.2 and the transition probability matrix is given by

$$\mathbf{P}_{\text{DDOS}} = \begin{pmatrix} \alpha_{11} & \alpha_{12} & 0 & \cdots & 0 \\ \alpha_{21} & 0 & \alpha_{23} & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ \alpha_{(n-1)1} & 0 & 0 & \cdots & \alpha_{(n-1)(n)} \\ \alpha_{n1} & 0 & 0 & \cdots & \alpha_{nn} \end{pmatrix}. \tag{3.9}$$

where the state transitions represented by zero implies impossible state transition for DDOS scheduling scheme. The limiting probabilities for state i are given by

$$\pi_i = \sum_{b=i-1}^{n} \alpha_{bi} \pi_b. \tag{3.10}$$

Let

$$P_i^j = \lim_{t\to\infty} \Pr(S_t = i, S_{t+1} = j) \tag{3.11}$$

denote the joint probability that the buffer is in state i and moves to state j. Then, we have asymptotically

$$P_i^j = \lim_{t\to\infty} \Pr(S_{t+1} = j | S_t = i) \Pr(S_t = i) \tag{3.12}$$

$$= \alpha_{ij}\pi_i \tag{3.13}$$

$$= \alpha_{ij} \sum_{b=i-1}^{n} \alpha_{bi}\pi_b. \tag{3.14}$$

We assume that the users exhibit independent fading processes. Furthermore, the proposed scheduler is independent of the other users' fading. Therefore, the law of large numbers drives the proportion of users in state i at time t to be identical to $\Pr(S_t = i)$ in the large user limit. If a user is scheduled in a small queue state, it utilizes a good channel and allocates comparatively more rate. If the channel is not good, the user may decide to wait for the next slot due to the following dilemma: On the one hand, the user is hoping that the next channel will be better than the current channel and energy will be saved. But on the other hand, the user is coming closer to the deadline for transmission. If the deadline will be reached and the channel has gotten worse, energy will be wasted by transmitting data on a bad channel. The task of the optimizer is to find the optimal threshold values such that the transmitted energy in the system is minimized by smart scheduling decisions.

3.4.1 Optimization of Thresholds

Next we would like to optimize the opportunistic thresholds. Our objective is to minimize the average transmitted energy. We would like to apply the large-system results of [Caire *et al.*, 2007], in particular Theorem 2. However, Theorem 2 is restricted to cases where the rate is independent of the channel realization which does not hold for our scheduler. In order to circumvent this limitation, we model a user that sends L packets at a time as L virtual users with identical fading that send single packets. The average energy consumption of the system per transmitted information bit at the large system limit $K \to \infty$ is then given by [Caire *et al.*, 2007]

$$\left(\frac{E_\mathrm{b}}{N_0}\right)_{\mathrm{sys}} = \log(2) \int_0^\infty \frac{2^{R\,\mathrm{P}_{g,\mathrm{SVU}}(x)}}{x}\, \mathrm{d}P_{g,\mathrm{SVU}}(x) \tag{3.15}$$

where $\mathrm{P}_{g,\mathrm{SVU}}(\cdot)$ denotes the cumulative distribution function (cdf) of the fading of the scheduled virtual users (SVU). It is composed of the short-term fading of the SVUs and the long-term fading of the SVUs. Note that

in the large system limit, the long-term fading of the SVUs follows the same distribution as the long-term fading of all users because long-term and short-term fading are mutually statistically independent and state transitions depend only on the short-term fading.

The probability density function (pdf) of the short-term fading of the SVUs is given by

$$\mathrm{p}_{f,\mathrm{SVU}}(y) = \sum_{i=1}^{n} \pi_i \, \mathrm{p}_{f,\mathrm{SVU}}(y|i) \tag{3.16}$$

where the channel distribution of the users in state i is given by

$$\mathrm{p}_{f,\mathrm{SVU}}(y|i) = c_i i \, \mathrm{p}_{\max\{f\}}(y) \tag{3.17}$$

with $\mathrm{p}_{\max\{f\}}(y)$ and c_i denoting the short–term fading of the best of the channels for a multi-channel system and a constant to normalize the conditional pdf. The cumulative distribution function (cdf) of the SVUs is given by

$$\mathrm{P}_{f,\mathrm{SVU}}(y) = \sum_{i=1}^{n} \pi_i \int_{\kappa_i}^{y} \mathrm{p}_{f,\mathrm{SVU}}(\xi|i)\mathrm{d}\xi, \tag{3.18}$$

since for $y < \kappa_i$ no users are scheduled. Using Eq. (3.17), Eq. (3.18) can be written as sum of integrals

$$\mathrm{P}_{f,\mathrm{SVU}}(y) = \sum_{i=1}^{n} c_i \pi_i i \int_{\kappa_i}^{y} \mathrm{p}_{\max\{f\}}(\xi)\mathrm{d}\xi \tag{3.19}$$

$$= \sum_{i=1}^{n} c_i \pi_i i \Big(\mathrm{P}_{\max\{f\}}(y) - \mathrm{P}_{\max\{f\}}(\kappa_i) \Big). \tag{3.20}$$

Using standard methods for calculating the distribution of the product of two independent random variables, $\mathrm{P}_{g,\mathrm{SVU}}(y)$ is calculated out of Eq. (3.20) and the CDF of the path loss in Appendix A.

3.4.2 Optimization by Simulated Annealing

The energy in Eq. (3.15) is not a convex function of the transmission thresholds. Therefore, we choose to use the Simulated Annealing (SA) algorithm to optimize the energy function for the transmission thresholds that result in a minimum energy for a given maximum delay parameter. The simulated annealing algorithm was proposed in [Kirkpatrick, Gelatt, and Vecchi, 1983] and [Cerny, 1985] separately. It uses the ideas from statistical

mechanics to solve combinational problems. It is believed to provide near optimal solutions (even optimal) in many combinatorial problems.

The main components of the simulated annealing algorithm are described briefly here.

1. Objective Function
   In this work, objective function is the system energy as given in Eq. (3.15).

2. Description of the configuration of the system
   It is essential to provide a clear description of the configuration of the system. In our case, the transmission thresholds are the parameters which represent the configuration of the system at a particular instant. The transmission thresholds are related with the transition probabilities for a given deadline and short–term fading. Therefore, determining an optimal vector of thresholds $\vec{\kappa}^{\text{opt}}$ is equivalent to determine an optimal state transition probability matrix $\mathbf{P}^{\text{opt}}$.

3. A random generator for the new configuration
   At the start of the algorithm, any configuration can be provided. In the next step, there must be a suitable method to provide a random change in the configuration. In this work, transition probabilities are varied in each step to provide a new configuration to evaluate Eq. (3.15).

4. A cooling temperature schedule
   The system is "heated" at high temperature T at the start of the algorithm. Afterwards, the temperature is decreased slowly up to the point where the system "freezes". The terms heating and cooling come from statistical thermodynamics where freezing of the system represents a situation where the system reaches a near optimal solution and no more state[6] transitions occur for further lowering of the temperature parameter. The cooling schedule depends on the specific problem and can be developed after certain experiments. In our simulations, we tested both Boltzmann annealing (BA) and Fast annealing (FA) temperature cooling schedules which have been proven to provide global minimum solutions for a wide range of problems [Geman and Geman, 1984; H.Szu and Hartley, 1987]. In FA, it is sufficient to decrease the temperature linearly in each step b such that,

[6]The state in SA refers to the configuration of the system, i.e. the current transmission thresholds. It has no relation with the state of the Markov process given by the backlog of the user.

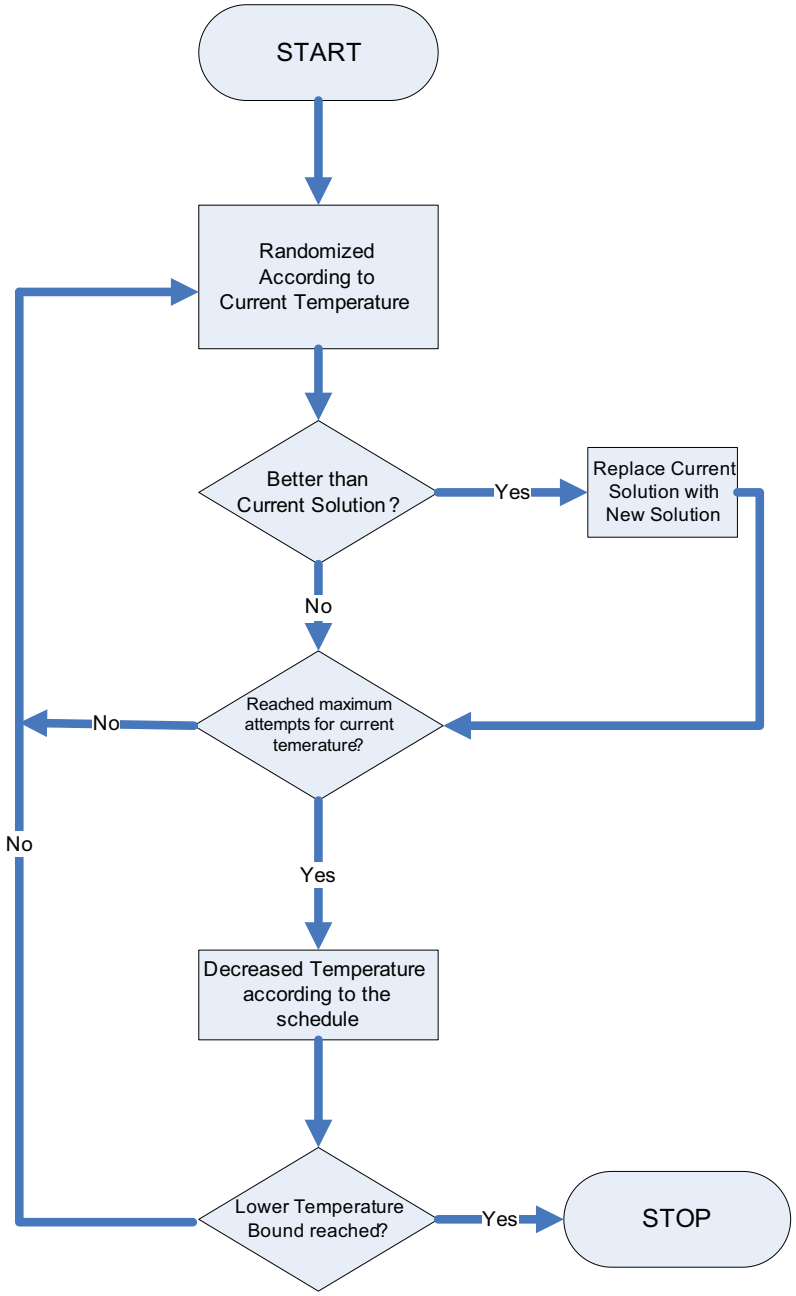


FIGURE 3.3: Flow chart for SA algorithm.

$$T_b = \frac{T_0}{b+1} \tag{3.21}$$

where T_0 is a suitable starting temperature. Similarly in BA, global minima can be found sufficiently (in many problems) if temperature decreases logarithmically such that,

$$T_b = \frac{T_0}{\ln(b+1)} \tag{3.22}$$

5. Acceptance Probability
   Any new configuration in SA is accepted if it results in a lower system energy with probability 1. A change in energy in each step is denoted by ΔE. Any new state is accepted with probability $\Delta E/T$ if it results in a higher energy state and it is referred to as *muting*. Muting occurs frequently at the start of the algorithm and vanishes to happen as the temperature T approaches zero.

Block diagram for SA algorithm has been shown in Fig. 3.3.

Using the SA algorithm, an optimal set of transmission thresholds is obtained for a given deadline τ_{max}. The muting step makes it likely that local minima are avoided in the optimization process by moving into higher energy solutions with some temperature dependent probability. Numerical results relating the optimization process using the SA are discussed in Section 3.6.

3.5 Minimum Throughput Guarantees

We formulate the problem in terms of providing throughput guarantees to the individual users instead of deadline delay guarantees. Throughput for a user k is defined over a rectangular window of length t_c such that

$$\mathcal{T}_k = \frac{1}{t_c} \sum_{b=1}^{t_c} \delta(t-b) R_k^{sch}(t-b) \tag{3.23}$$

where $R_k^{sch}(t)$ is the rate requirement of user k in time slot t to empty the buffer and

$$\delta(t) = \begin{cases} 1 & \text{if user scheduled} \\ 0 & \text{otherwise} \end{cases} \tag{3.24}$$

Each user is provided a minimum throughput guarantee and the task is to achieve the throughput at minimum transmit energy. We adapt DDOS

algorithm such that it achieves minimum throughput guarantee for each user.

We assume infinite buffer size for this case. The rate allocation scheme follows emptying buffer property as in DDOS. Instead of a fixed deadline τ_{max}, we have a *virtual deadline* that depends on the difference between instantaneous throughput $\mathcal{T}(t)$ in slot t and minimum throughput $\mathcal{T}_{min}$. A virtual deadline is defined as the number of time slots an unscheduled user tolerates before her instantaneous throughput $\mathcal{T}(t)$ falls below $\mathcal{T}_{min}$. Virtual deadline needs to be computed at the start of time slot before the scheduling operation.

It should be noted that τ_{max}^{vr} needs not to be computed in every time slot. The routine of computation of τ_{max}^{vr} is invoked only when a user is scheduled to empty the buffer.

Virtual deadline is computed in the following steps.

1. Initialize virtual deadline τ_{max}^{vr} by one.
2. Replace rate $R_k^{sch}(t - t_c - \tau_{max}^{\text{vr}} + 1)$ with zero and compute $\mathcal{T}$.
3. If $\mathcal{T} > \mathcal{T}_{min}$, increment τ_{max}^{vr} by one.
4. Repeat step 2 and step 3 until $\mathcal{T} \leq \mathcal{T}_{min}$.
   τ_{max}^{vr} is virtual deadline at the end of the routine.

Once τ_{max}^{vr} is computed, DDOS scheduler computes transmission thresholds by treating the system as deadline constrained system with deadline τ_{max}^{vr} .

3.6 Numerical Results

We have considered a multi-access channel with M bands and it is assumed that fading on these channels is statically independent. It implies that every user senses M channels instead of a single channel and selects its best channel as a candidate channel for the transmission scheduling. This is the optimal multi-band allocation for the asymptotic case [Caire *et al.*, 2007]. We consider a system where users are placed uniformly at random in a cell except for a forbidden region around the access point of radius $\delta = 0.01$. The path loss is exponential with exponent 2. All users experience fast fading with exponential distribution with mean one on each of the M channels. The details of path loss model can be found in Appendix A. The spectral efficiency values used in the results are divided by M to get spectral efficiency/channel C.

TABLE 3.2: Threshold Computation by SA

τ_{max}	κ_1	κ_2	κ_3	κ_4	(E_b/N_o)
2	0.16	0	NA	NA	-0.76dB
3	0.36	0.17	0	NA	-2.24dB
4	0.55	0.38	0.19	0	-3.23dB

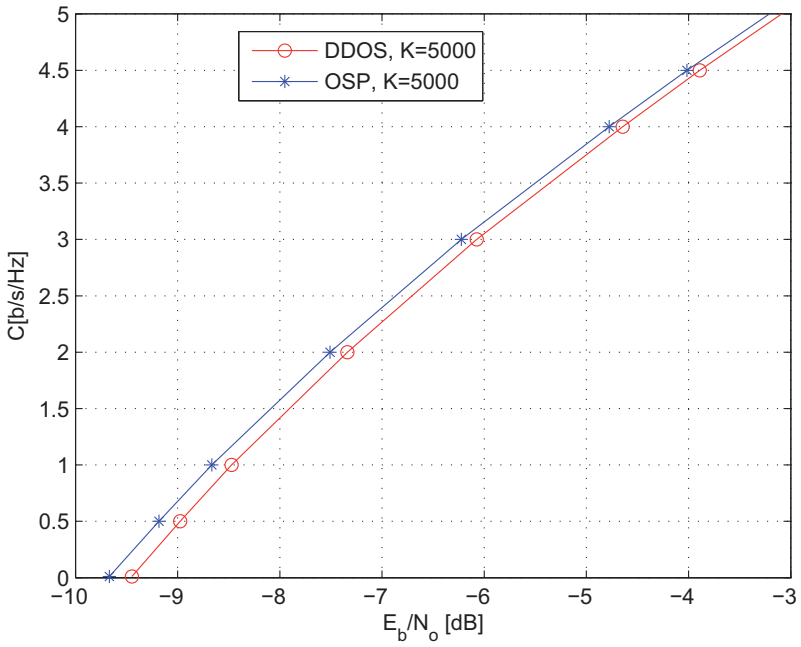


FIGURE 3.4: Comparison of OSP and DDOS for $\tau_{max} = 3$, M=10 and same average delay.

We compute the vector of optimal thresholds using Simulated Annealing algorithm as explained in Section 3.4.2. We set the value of spectral efficiency per channel equals to 0.5 and $M = 1$. We use FA cooling schedule with 50 temperature iterations. For each temperature value, 10 random configuration (transition probabilities) are produced. Table 3.2 shows the vector of transmission thresholds for $n = 2$, $n = 3$ and $n = 4$, and the

corresponding average system energy.

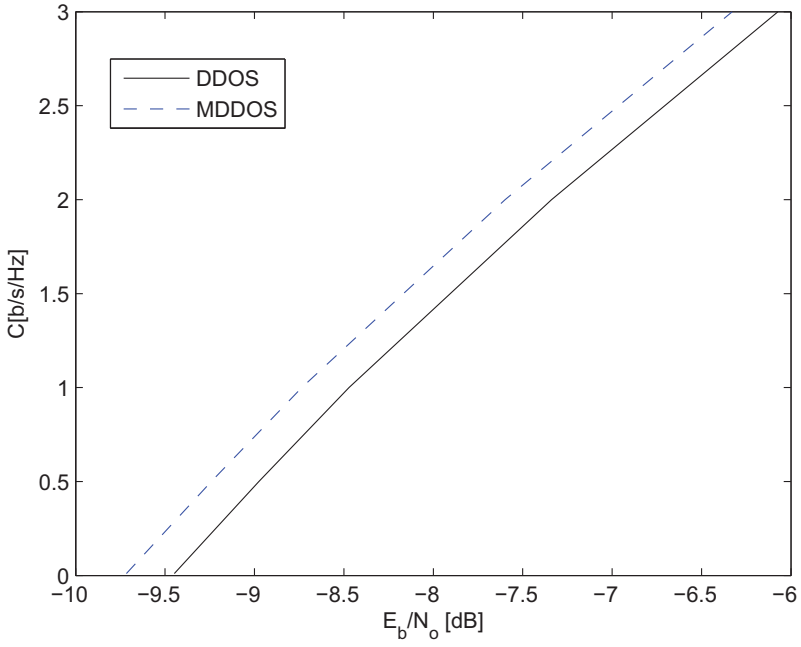


FIGURE 3.5: Energy efficiency of DDOS and MDDOS for $\tau_{max} = 3$, M=10.

We evaluate the results for the finite user case. All the numerical results have been obtained by simulating a multiuser environment where 5000 users have simultaneous access to 10 channels. For each operation, 100 path loss environments have been simulated to remove the effect of variation in path loss on the system energy. For a single path loss environment, 200 scheduling operations have been performed for the convergence of the sum energy of the system. We consider the constant arrivals for all the users.

Fig. 3.4 demonstrates the comparison of OSP with DDOS for $\tau_{max} = 3$ and $K = 5000$. For a fair comparison with OSP, average delay has been kept same for the evaluation of both of the scheduling schemes. A small loss in system energy is observed for DDOS as compared to OSP but it provides the deadline delay guarantees to the users as well.

Fig. 3.5 shows the comparison of energy efficiency between DDOS and MDDOS schemes for $\tau_{max} = 3$. As described in Section 3.3.1, MDDOS behaves similarly when scheduling the users opportunistically but opera-

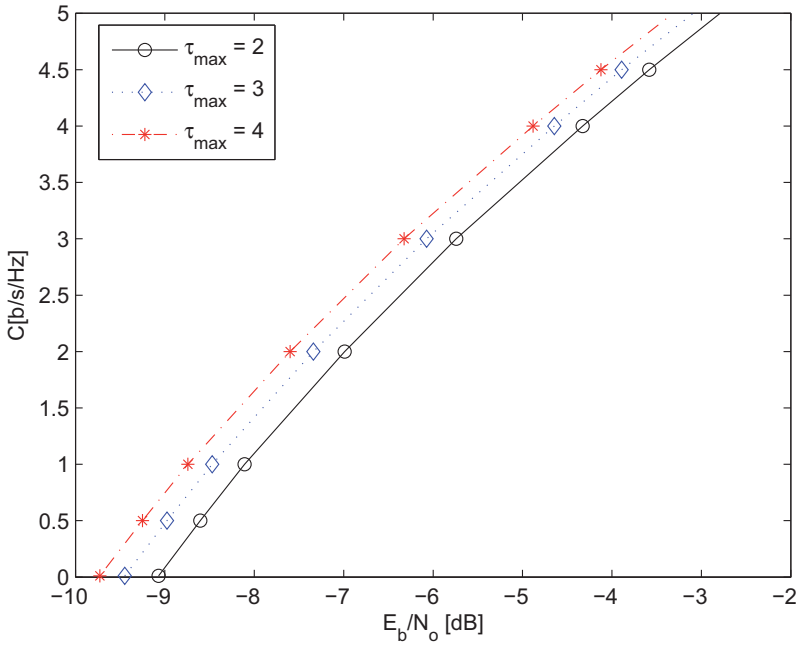


FIGURE 3.6: Comparison of DDOS scheme for different values of τ_{max} and M=10.

tion is modified in deadline state only. This modification results in further improvement in energy efficiency while fulfilling the same deadline delay constraint as shown in Fig. 3.5.

Fig. 3.6 shows the numerical example of energy-delay tradeoff exhibited by DDOS. It is observed that an increase in deadline delay constraint results in an energy efficient system. If application is delay tolerant, delay provides another degree of freedom to achieve energy efficiency. However, gain in system energy diminishes for large deadline delays.

Chapter 4

Opportunistic Partial Buffer Scheduling

In this chapter, we extend the work in Chapter 3 and propose a more generic scheduler as compared to the emptying buffer scheduler in DDOS. We use the system model introduced in Section 3.1 for DDOS.

In the proposed scheme, the problem of optimization of the transmission thresholds for energy efficient data transmission is formulated analogously to the dynamic programming concept where a complicated problem is broken into a set of smaller, less complicated problems. Our goal is to compute transmission thresholds for all the backlog states such that system energy is minimized. We break the problem of energy optimal transmission of data packets (before the deadline) into a less complicated problem of transmitting the data in the current time slot or waiting for the next slot. For the current backlog state, transmission threshold is computed in such a way that outcome of the decision (waiting or transmission) in current time slot results in a minimum energy. Based upon the decision in the current time slot, backlog–state is modified and we ask the same question again in the next time slot. In this way, by solving the less complicated problem of determination of optimal threshold for the current time slot (with a given backlog), we compute the energy optimal thresholds for all the backlog–states (Markov process).

We consider an asymptotically large user system as in Chapter 3 and formulate the problem of energy efficient transmission in a multiuser system as an equivalent single user scheduling problem.

4.1 Deadline Dependent Partial Buffer Scheduling (DDPS)

In this work, we discuss a scheduling scheme in which data can be transmitted in discrete steps (packets) [Butt, Kansanen, and Müller, 2008b, a]. We define some basic terms for the DDPS scheduler.

Definition 4.1 (Transmission threshold For DDPS:) A transmission threshold $\kappa_{i\to j}$ is defined as the minimum short–term fading value allowing for the transition $T_{i\to j}$ from state i to state j.

The state represents the size of current buffer occupancy. If we assume that p packets arrive, $i - j + p$ packets are scheduled when the state transition $T_{i\to j}$ takes place.

In the sequel, we use the following proposition.

Proposition 4.1 Given a fixed backlog state, transmit more for the better fading channel [Viswanath, Tse, and Anantharam, 2001]. This is known as water–filling principle . □

We impose a few restrictions on the transmission thresholds called *properties* of the transmission thresholds.

Property 4.1 $\kappa_{i\to j} = 0 \quad \forall i < j$. The transition from a lower to the higher state requires no minimum channel quality. Note that no packets are scheduled for this event to happen.

Property 4.2 $\kappa_{i\to j} \leq \kappa_{i\to j-1} \quad \forall i, j$. Scheduling more packets requires better channel conditions. Equality is only allowed for sake of consistency with Property 4.1.

Proof: See Appendix B. □

Property 4.3 $\kappa_{i\to j} \leq \kappa_{i-1\to j-1} \quad \forall i, j$. Both transitions schedule the same number of packets. The inequality ensures that scheduling from longer queues is preferred over scheduling from shorter queues.

Proof: See Appendix B. □

Property 4.4 $\kappa_{n\to n} = 0$ for the highest state n. This ensures that the hard deadline is kept and, in fact, that the number of states is finite. If the buffer is full, a packet is scheduled regardless of the channel quality. □

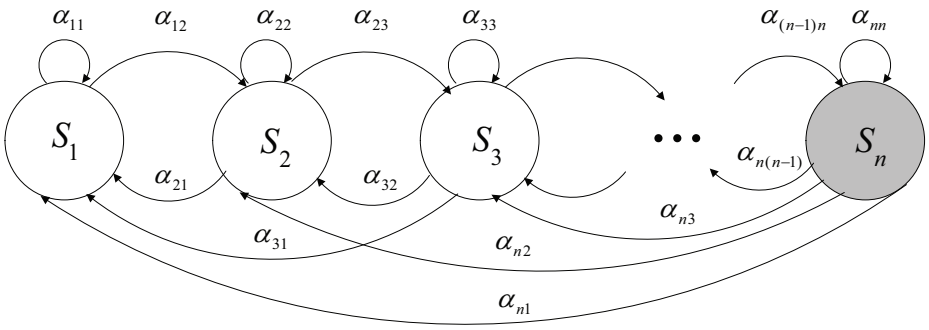


FIGURE 4.1: State diagram for the transition states of DDPS scheduler.

In addition to the short–term fading $f_k(t)$, the backlog states of the users are considered to make a decision that which of the users are to be scheduled. Similar to DDOS, DDPS schedules a set of users experiencing favorable short–term fading. The short–term fading of the users is compared against *a set of transmission thresholds*. The rate R_k is, then, allocated in discrete steps according to the number of transmission thresholds that are exceeded. If the backlog of a user is equal to the hard deadline, the user is scheduled for at least a single packet regardless of its instantaneous fading state. The scheduler performs an energy efficient opportunistic scheduling as long as the backlog of the user is less than the hard deadline. When the hard deadline is reached, a deadline guarantee is provided via Property 4.4 by scheduling the user for at least a single packet. This may be costly, but the deadline is to be respected by all means. To avoid these potentially costly events happening too frequently, the algorithm gives preference to older packets as compared to newer packets by proper choice of the transmission thresholds. This is the idea behind Property 4.3 of the transmission thresholds.

In each time slot only integer multiples of rate $\frac{\Gamma}{K}$ (packet size) are transmitted. We consider a constant arrival rate of one packet per time slot. For this special case, the delay of the oldest packet in a buffer is represented by the user's state and the hard deadline is equal to his buffer size. The assumption of constant arrival makes the analysis simpler and readily understandable. Later, in the next section, we will consider random arrivals. The deadline constraint τ_{max} is reflected by the maximum number of states n. State transition $T_{i\rightarrow j}$ from state i to the next higher state j occurs if no data is transmitted (Property 4.1). Similarly, state transition $T_{i\rightarrow j}$ from a state i to a lower or equal state j occurs by transmitting $i-j+1$ packets of size $\frac{\Gamma}{K}$ depending upon the transmission thresholds.

We can represent the state transition model for each user as a Markov chain as shown in Fig. 4.1 and the corresponding transition probability matrix is given by

$$\mathbf{P}_{\text{DDPS}} = \begin{pmatrix} \alpha_{11} & \alpha_{12} & 0 & \cdots & 0 \\ \alpha_{21} & \alpha_{22} & \alpha_{23} & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ \alpha_{(n-1)1} & \alpha_{(n-1)2} & \alpha_{(n-1)3} & \cdots & \alpha_{(n-1)(n)} \\ \alpha_{n1} & \alpha_{n2} & \alpha_{n3} & \cdots & \alpha_{nn} \end{pmatrix}. \tag{4.1}$$

where the state transitions represented by zero implies impossible state transition for DDPS scheduling scheme. If a Markov process is in state i, then the next state j is determined according to transition probabilities α_{ij}. Similar to DDOS, forward transition is still limited to adjacent higher state due to identical deadline for all the arriving packets. However, the backward transition probabilities for DDPS result from the scheduling operation and differ from DDOS. The relation between transition probabilities and transmission thresholds have been shown in Appendix C.

Pseudo code of DDPS is shown in Algorithm 4.1.1.

Algorithm 4.1.1: DDPS($Backlog, Threshold$)

comment: User k knows backlog and transmission thresholds

$i \leftarrow Backlog$

$\vec{\kappa} \leftarrow Threshold$

comment: Buffer contains $i\frac{\Gamma}{K}$ and rate R is provided by the scheduler

$R \leftarrow 0$

comment: Find maximum allowed rate

for $j \leftarrow 1$ **to** $Backlog$

do { **if** $f_k > \kappa_{i \rightarrow j}$

then { $R \leftarrow (i - j + 1)\frac{\Gamma}{K}$
$i \leftarrow j$
break

if $(R = 0)$

$i \leftarrow i + 1$

$Backlog \leftarrow i$

4.1.1 Modeling Random Arrivals

We consider a random arrival process such that the arrival of a random number of packets in each time slot is modeled by the arrival of a single packet with random content size N [Butt, Kansanen, and Müller, c].

Lemma 4.1 Random arrivals with fixed packet size are identical to constant arrivals with random packet size.

In our arrival model, we denote the number of packets arriving within a single time slot by the random variable X and the size of the arriving packets by the random variable N. Denoting $\Pr(X = x)$ shortly by $p_X(x)$, constant arrival of a single packet per time slot with random packets size N^{rand} is then represented by the

probability distributions $p_{X^{\text{const}}}(1) = 1$ and $p_{N^{\text{rand}}}(\mu)$ where $\mu \in \{[0, 1, 2 \ldots] \frac{\Gamma}{K}\}$. Thus, for a fixed packet size N^{const}, we have

$$p_{X^{\text{rand}}}(x) = p_{N^{\text{rand}}}(xN^{\text{const}}). \tag{4.2}$$

□

Following Lemma 4.1, we consider the random arrivals as constant arrivals with random size in this work. This will be useful in the sequel as in the large system limit, the packet size distribution has no effect onto the spectral efficiency of the system [Caire *et al.*, 2007]. In Section 4.2, we analyze the effect of Lemma 4.1 to the DDPS scheduler.

4.1.2 Asymptotic Analysis of DDPS

We perform asymptotic analysis of DDPS analogous to the analysis of DDOS in Section 3.4. The distribution of short term fading for SVUs differs for DDPS as compared to DDOS.

In case of DDPS, the probability density function (pdf) of the short-term fading of the SVUs is given by

$$\mathrm{p}_{f,SVU}(y) = \sum_{i=1}^{n} \pi_i \, \mathrm{p}_{f,\mathrm{SVU}}(y|i) \tag{4.3}$$

where the channel distribution of the users in state i is given by

$$\mathrm{p}_{f,\mathrm{SVU}}(y|i) = c_i(i - j(y,i) + 1) \, \mathrm{p}_{\max\{f\}}(y) \tag{4.4}$$

with $\mathrm{p}_{\max\{f\}}(y)$, $j(y,i)$ and c_i denoting the short–term fading of the best of the channels for a multi-channel system, the end state of the system after scheduling and a constant to normalize the conditional pdf. Note that the function $j(y,i)$ is uniquely defined by Definition 4.1 given all transmission thresholds $\{\kappa_{i\to j}\}$. The cumulative distribution function (cdf) of the SVUs is given by

$$\mathrm{P}_{f,\mathrm{SVU}}(y) = \sum_{i=1}^{n} \pi_i \int_{\kappa_{i\to i}}^{y} \mathrm{p}_{f,\mathrm{SVU}}(\xi|i) \mathrm{d}\xi, \tag{4.5}$$

since for $y < \kappa_{i\to i}$ no users are scheduled. Using Eq. (4.4), Eq. (4.5) can be

written as sum of integrals

$$\begin{aligned} \mathrm{P}_{f,\mathrm{SVU}}(y) &= \sum_{i=1}^{n} c_i \pi_i \Big[(i - j(y,i) + 1) \int_{\kappa_{i \to j(y,i)}}^{y} \mathrm{p}_{\max\{f\}}(\xi) \mathrm{d}\xi \\ &+ \sum_{b=1}^{i-j(y,i)} b \int_{\kappa_{i \to i-b+1}}^{\kappa_{i \to i-b}} \mathrm{p}_{\max\{f\}}(\xi) \mathrm{d}\xi \Big] \qquad (4.6) \\ &= \sum_{i=1}^{n} c_i \pi_i \Big[(i - j(y,i) + 1) \mathrm{P}_{\max\{f\}}(y) - \sum_{b=0}^{i-j(y,i)} \mathrm{P}_{\max\{f\}}(\kappa_{i \to i-b}) \Big] \\ &. \qquad (4.7) \end{aligned}$$

$\mathrm{P}_{g,\mathrm{SVU}}(y)$ is calculated in Appendix A out of Eq. (4.7) and the CDF of the path loss.

Note that $\mathrm{P}_{f,\mathrm{SVU}}(y)$ is piecewise constant due to the constraining of the transmitted rate to a discrete set of rates. The case where real valued allocation is allowed corresponds to the current case with infinitely granularity of allocation and infinitely many thresholds being applied. In this asymptotic limit, $\mathrm{P}_{f,\mathrm{SVU}}(y)$ would be continuous, and the backlog state described by a continuous state Markov chain. Since increasing the number of possible backlog states and transmit rates corresponds to relaxing the constraints on allocated rate, increasing the number of allowed states and admissible transmit rates implies a decreasing optimum system energy, with the minimum system energy achievable by the asymptotic system with infinitely many backlog states.

4.2 Equivalence of User Distribution For Constant and Random Arrivals

We claim in Lemma 4.1 that a random arrival distribution can be expressed as a constant arrival distribution with random packet size. In this section, we explore the consequences of Lemma 4.1 to the DDPS scheduler. We evaluate Eq. (4.4) for random and constant arrivals and prove equivalence of the channel distributions of the scheduled users for both cases in the limit when $K \to \infty$.

We define a subset of users having backlog state i_o as

$$A_{\{i=i_o\}} = \{k : i_k = i_o\}. \qquad (4.8)$$

Then,

$$\lim_{K\to\infty} \frac{|A_{\{i=i_o\}}|}{K} \quad = \quad \Pr(i = i_o) \quad \forall \quad i, i_o. \tag{4.9}$$

Thus, at the large user limit, the probability of a user to be in backlog state i represents the proportion of the users in that state. Also, the buffered data with backlog state i_o can be expressed as

$$\lim_{K\to\infty} c_a \frac{\Gamma}{K} |A_{\{i=i_o\}}| = c_a \Gamma \Pr(i = i_o). \tag{4.10}$$

where the constant c_a represents the content size and depends on the distribution of the arrival process.

Similarly, we denote the number of arrivals in time t by $X(t)$ and define a set of users as

$$A_{\{X(t)=x\}} \quad = \quad \{k : X(t) = x\}. \tag{4.11}$$

Then, we have

$$\lim_{K\to\infty} \frac{|A_{\{X(t)=\mu\}}|}{K} \quad = \quad \Pr(X(t) = x) \quad \forall \quad X(t), x \tag{4.12}$$

Asymptotically, the probability that a user transmits x arrived packets is equal to the proportion of the users transmitting x packets in a time slot.

Eq. (4.8) to Eq. (4.12) show that we can analyze an asymptotically large user system with the help of the state space representation in Eq. (4.1). The scheduling function depends upon the number of backlog states $L(y, i)$ that have to be scheduled for the current fading f_k and the number of packets having arrived in these scheduled states. These numbers are independent of each other. The arrival process is ergodic and $X(t)$ packets are assumed to arrive in every time slot. Following Lemma 4.1, we express random arrival distribution as random size distribution. For any packet size distribution $p_{N(t)}(\mu)$, the channel distribution of virtual users can be written as

$$\mathrm{p}_{f,\mathrm{SVU}}(y|i) = c_i^a \mathrm{p}_{\max\{f\}}(y) L(y, i) \sum_{\mu} \mu p_{N(t)}(\mu) \tag{4.13}$$

where the summation represents the actual number of packets to be transmitted due to the random arrival process. c_i^a is a constant normalizing the conditional probability distribution and the content size of the random arrival. The summation in Eq. (4.13) can be expressed as the expectation $\mathbb{E}(\mu_t)$ of the arrival process.

$$\mathrm{p}_{f,\mathrm{SVU}}(y|i) = c_i^a \mathbb{E}(\mu_t) L(y, i) \mathrm{p}_{\max\{f\}}(y) \tag{4.14}$$

Corresponding channel distribution for constant arrivals is given by

$$\mathrm{p}_{f,\mathrm{SVU}}(y|i) = c_{\mathrm{const}} L(y,i)\, \mathrm{p}_{\max\{f\}}(y) \tag{4.15}$$

where c_{const} is a normalizing constant. A comparison of the channel distributions of the scheduled virtual users in Eq. (4.14) and Eq. (4.15) for random and constant packet arrivals show that they differ only in the constant. The two channel distributions are therefore equivalent. Furthermore, the thresholds optimized for the constant arrivals are equally valid for the systems with randomly arriving packets.

4.2.1 Special Cases

We consider examples of two commonly used arrival processes.

Bernoulli Process

In a Bernoulli process, a single data packet arrives with probability ϵ in a time slot and has a Bernoulli distribution. Therefore, Eq. (4.14) can be written as

$$\mathrm{p}_{f,\mathrm{SVU}}(y|i) = c_i^a \epsilon L(y,i) \mathrm{p}_{\max\{f\}}(y) \tag{4.16}$$

where ϵ is the probability of arrival in a time slot.

Poisson Process

We discuss Poisson arrival process which is commonly used in network analysis. For Poisson arrival process with arrival rate ϑ , Eq. (4.14) can be written as

$$\mathrm{p}_{f,\mathrm{SVU}}(y|i) = c_i^a \vartheta L(y,i) \mathrm{p}_{\max\{f\}}(y) \tag{4.17}$$

4.3 Implementation Considerations

The proposed scheme solves the optimization problem offline as a function of the channel statistics and backlog of the user. The offline optimization task can be performed locally by the users and needs no centralized control, since it only involves the fading statistics, but not the fading realizations. However, a centralized optimization would save complexity, since the outcome of the optimization is identical for all users. Similarly, the scheduling decisions can fully be taken by each user individually. However, the powers required by the users to transmit their packets depend on the ordering

TABLE 4.1: DDPS with Thresholds Computed via SA For $M = 1, C = 0.5$ bits/s/Hz

n	$\kappa_{1\to1}$	$\kappa_{2\to1}$	$\kappa_{3\to1}$	$\kappa_{2\to2}$	$\kappa_{3\to2}$	$\kappa_{3\to3}$	$(E_b/N_0)_{sys}$
2	0.24	0.24	-	0	-	-	-1.42 dB
3	0.52	0.51	0.52	0.24	0.23	0	-3.06 dB

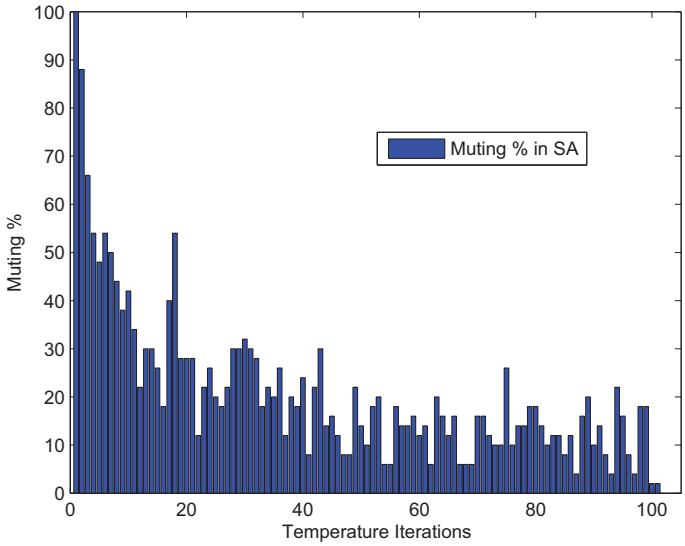
of the successive decoding. For a finite user system, it is not possible for the users to get the exact knowledge of the required transmit power to provide the rate. Therefore, the users need to transmit at slightly extra power. The excess power from all the users should vanish at the large system limit for the system to follow the result of Eq.(3.15) which does not happen when successive decoding is used. However, joint decoding does not suffer from this problem as all the users are decoded at the same time (without a specific order). Thus, for successive decoding, there is a need for a centralized assignment of the transmit powers. If the number of scheduled users, however, is very large and joint decoding is employed, the users can calculate their transmit powers individually by closely approximating the empirical fading and rate distributions of the other scheduled users by their statistical averages following the ideas of [Viswanath *et al.*, 2001]. With the application of joint decoding, the proposed schemes have the potential to be implemented in a distributed manner.

The simple nature of DDPS makes our scheduling scheme well-suited for wireless sensor networks. By using the parameter τ_{max}, we can control the energy-delay tradeoff. A higher value of τ_{max} implies that the application data is more delay tolerant and the energy consumption will be closer to the energy consumption of the schemes without deadline delay guarantees.

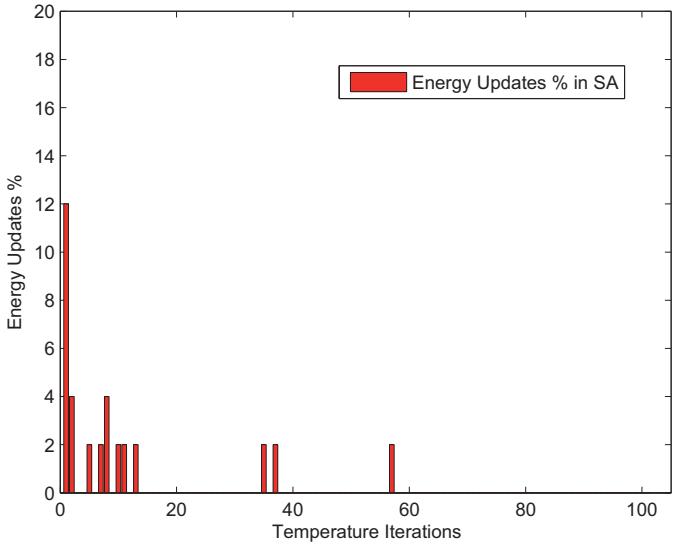
4.4 Numerical Results

To evaluate numerical results for DDPS, we consider the same simulation model and parameters as used in Section 3.6 for DDOS.

The thresholds optimized by SA algorithm for DDPS are shown in Table 4.1. For all the numerical results, the SA algorithm used 50 random configurations per temperature iteration. For $n = 2$ and $n = 3$, we observe insignificant difference in the thresholds corresponding to the transitions



(a) muting



(b) Energy updates

FIGURE 4.2: Muting and Energy updates statistics for BA cooling schedule.

$T_{r \to j}$ where $r \geq j$. Comparing system energy resulting from DDPS scheduler with the system energy of DDOS for the same hard deadline τ_{max}, we observe that DDPS scheduler is significantly energy efficient as compared to DDOS.

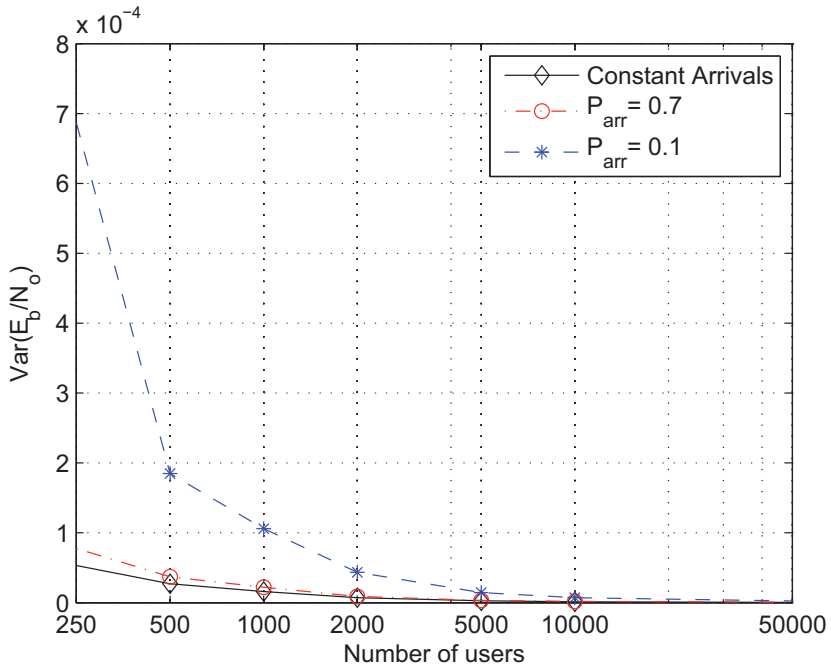


FIGURE 4.3: Variance of system energy for random and constant arrival distributions for $M = 10$, $C = 0.5$ and $n = 2$.

Fig. 4.2 shows the statistics for the SA algorithm for DDPS with FA temperature cooling schedule. As explained in section 3.4.2, mutations are 100% at the start and then decrease with every iteration. Similarly, energy updates are more frequent at the start. Once the system finds the minimum energy solution, no more updates occur in spite of the occurrence of the muting. It should be noted that statistics can differ a bit for different cooling schedules like BA and different configuration schedules but the final results remain identical.

Fig. 4.3 shows the convergence behaviour of system energy for both constant and random packet arrival distributions. For the numerical results, 250 simulations with different fading values have been performed for a single path loss. For a fixed number of users and iterations, we compute and compare the variance of the system energy for the cases of constant and random arrivals.

For Bernoulli process with arrival probability $P_{arr} = 0.7$, the variance of the system energy for both constant and random arrivals converges to same value for 5000 users. However, as arrival probability decreases, variance of system energy with random arrivals requires more users to converge to the variance of system energy with constant arrivals.

Fig. 4.4 shows effect of the number of users on DDPS for a system with $M = 10$ frequency bands and $n = 3$. The results are obtained by varying the number of users in Eq. (3.2) which is a finite user approximation of the asymptotic expression in Eq. (3.15). For each simulation point, 100 path loss environments have been averaged and for each path loss environment, 200 scheduling operations have been performed to find the average system energy. The performance of DDPS is invariant to the number of users for small spectral efficiencies, but degrades for small numbers of users at higher spectral efficiencies. A lot of applications like WSNs do not operate on high spectral efficiencies and this allows the scheme to operate effectively for small number of users as well. Furthermore, we find, though we do not show in an explicit figure, that with fewer bands, the dependency of the performance on the number of users is less pronounced.

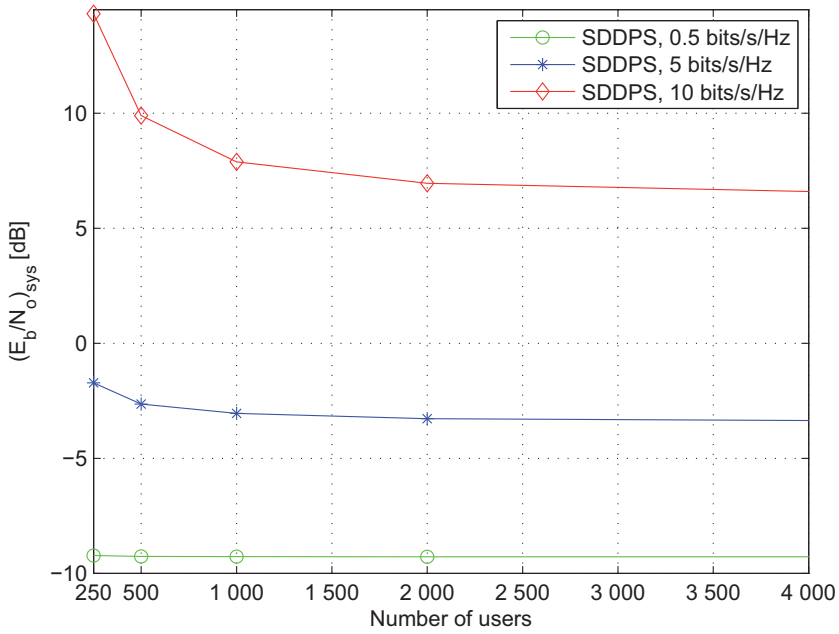


FIGURE 4.4: Effect of decrease in number of users for the DDPS scheme with M=10.

Chapter 5

Partial Buffer Scheduling: Practical Approach

In Chapter 4, we discussed and analyzed an energy efficient multiuser scheduling scheme for asymptotically large multiuser system. The simulation results show that DDPS scheduler converges for a few hundred users and works well for the realistic scenarios.

In this chapter, we discuss the practical aspects of DDPS scheduler and address the issues which are practically relevant for modern networks. One of the drawbacks of DDPS scheduler is the computation of $\mathcal{O}(n^2)$ transmission thresholds for the deadline of n time slots. For large n, it is computationally expensive to compute transmission thresholds using Simulated Annealing algorithm. We propose a scheduler which gives comparable results to DDPS and its computational complexity is smaller than DDPS.

Another practically important situation is the concept of outage. In practical systems, it is very difficult to provide strict deadline guarantee for all the packets. Some applications are loss tolerant and it is advantageous to drop certain proportion of the arriving packets if the quality of the application remains acceptable, e.g. multimedia applications have hard deadline limitation but can afford some data loss. It this case, it is important to identify the packets which require large energy for transmission and drop them. The tradeoff between system energy efficiency and packet drop rate essentially depends on the loss and delay tolerance of the application. Reference [Striegel and Manimaran, 2002] discusses a scheduler which differentiates traffic based on loss and delay tolerance of the user in traditional internet. In [Chen, Mitra, and Neely, 2006], the authors address the problem of optimal dropping of packets. They obtain optimal dropping scheme when size of the packet grows asymptotically large. In this chapter, we con-

sider scheduling with outage to provide energy-outage tradeoff in addition to energy–delay tradeoff for a multiuser system.

Similarly, it is too ideal to consider identical deadlines for all the arriving packets. In a practical network, there is a huge difference in the delay tolerance of the arriving packets depending on the corresponding applications. One way of solving problem is to assign different classes of QoS to individual users and schedule them according to the delay tolerance assigned to the specific class. This solution has traditionally been discussed for static and mobile ad-hoc networks [Blake, Black, Carlson, Davies, Wang, and Weiss, 1998],[Xiao, Seah, Lo, and Chua, 2000]. At the boundary of the network, traffic entering the network is differentiated and assigned a different behaviour. However, in this case, all the traffic from a specific user should have identical QoS requirements for all the packets. In this chapter, we propose a generic solution which treats individual packets according to their own deadlines.

5.1 Sequential Deadline Dependent Partial Buffer Scheduling (SDDPS)

In DDPS, the number of transmission thresholds to be computed grows quadratically as the deadline increases. For a deadline of n time slots, it becomes computationally costly to compute the $\frac{n}{2}(n+1)-1$ thresholds. In the following, we propose a sequential version of DDPS which reduces the number of thresholds from $\mathcal{O}(n^2)$ to $\mathcal{O}(n)$ at a negligible energy penalty [Butt, Kansanen, and Müller, d].

In order to understand SDDPS, assume that we were only allowed to decide whether one packet is scheduled or not, but not allowed to ask for having multiple packets from the same user scheduled. Then, we could base all scheduling decisions solely upon the $n-1$ non-zero components of the threshold vector $\vec{\kappa} = [\kappa_{1\to1}, \kappa_{2\to2}, \ldots, \kappa_{n-1\to n-1}, 0]$. In order to decide whether more than one packet of a given user should be scheduled, we repeat the threshold decisions sequentially within each time slot until no packet is supposed to be scheduled anymore.

The sequential scheduling decisions are equivalent to setting

$$\kappa_{i\to j} = \max\{\kappa_{i\to i}, \kappa_{i-1\to i-1}, \ldots, \kappa_{j\to j}\} \quad \forall i > j, \tag{5.1}$$

since with any decision to schedule one packet the queue goes into the next lower state. Due to Property 4.3, Eq. (5.1) is equivalent to

$$\kappa_{i\to j} = \kappa_{j\to j} \quad \forall i > j. \tag{5.2}$$

The transmission thresholds for entering into a state j from all the higher or equal states are identical, irrespective of i. In order to simplify notation, we simply write κ_j to denote $\kappa_{j\to j}$ in the following.

Sequential deadline dependent partial buffer scheduling is not optimal, in general. Note that scheduling multiple packets increases the aggregate data rate. Therefore, we have

Property 5.1 $\kappa_{i\to j} \geq \kappa_{i-1\to j} \quad \forall i, j$ since the required energy grows super-linearly with aggregate data rate, state transitions that result in higher data rates must be penalized compared to state transitions that result in lower data rates even if the final states are identical.

5.1.1 Recursive Optimization

In the following, the threshold vector $\vec{\kappa} = [\kappa_1, \kappa_2, \ldots, \kappa_{n-1}, 0]$ is optimized to minimize the power consumption for the SDDPS scheme. We use a simple heuristic optimization approach which suits to SDDPS due to the constraint applied in Eq. (5.2). This technique provides similar performance as SA at reduced complexity. The optimized threshold vector is found using a recursive procedure explained in the following:

1. Start the optimization procedure for a buffer length of $n = 2$ such that the optimization is a scalar problem and we only need to find the threshold κ_1 since $\kappa_n = 0$ by Property 4.4.

2. Given the optimized threshold vector for a deadline n, i.e. $\vec{\kappa}^{\,\mathrm{opt}}(n) = [\kappa_1^{\mathrm{opt}}(n), \kappa_2^{\mathrm{opt}}(n), \ldots, \kappa_{n-1}^{\mathrm{opt}}(n), 0]$, we find the threshold vector for buffer length $n+1$ by the heuristic postulate $\vec{\kappa}(n+1) = [\kappa_1(n+1), \vec{\kappa}^{\mathrm{opt}}(n)]$ and optimize over $\kappa_1(n+1)$. Again, this is a scalar optimization problem.

This heuristic optimization procedure is given here without theoretical justification. However, Table 5.1 gives numerical evidence that thresholds optimized for the smaller state space remain close to optimal for the higher state space though. Recursively computed threshold values for SDDPS are shown in Table 5.1. For $n = 2$ and $n = 3$, we compare the energy expenditure of SDDPS scheme with the energy expenditure of optimal DDPS as shown in Table 4.1. We observe insignificant difference in the thresholds corresponding to the transitions $T_{i\to j} \forall i \geq j$. This is easily understood due to the minor (and no) differences in the threshold assignments resulting from DDPS and SDDPS. Although marginal energy saving is provided by DDPS for $n = 3$, yet extra energy cost in computation of $\mathcal{O}(n^2)$ thresholds as compared to $\mathcal{O}(n)$ thresholds in SDDPS makes it practically useless.

Though, our scheduling schemes have been designed at low spectral efficiency due to linear energy-rate relationship, we observe negligible difference between energy efficiency of DDPS and SDDPS at high spectral efficiency values as well. Based upon our observations, we conjecture that

Conjecture 5.1 In large user limit, DDPS and SDDPS are equivalent scheduling schemes at all spectral efficiencies. □

System energy is not a linear function of spectral efficiency at high spectral efficiency. However, we believe that presence of large (infinite) number of users in the system makes per user rate linear and SDDPS behaves equivalent to DDPS.

TABLE 5.1: SDDPS with Recursively Computed Thresholds

n	κ_1	κ_2	κ_3	κ_4	$(E_b/N_0)_{sys}$
2	0.24	0	-	-	-1.44 dB
3	0.52	0.24	0	-	-3.05 dB
4	0.75	0.52	0.24	0	-4.08 dB

We optimize transmission thresholds at low spectral efficiencies and continue to use them at high spectral efficiencies. In practice, this is not true and transmission thresholds need to be optimized at each spectral efficiency. However, numerical evidence shows that this assumption results in a small energy loss. Table 5.2 shows exact optimized threshold and resulting energy for each spectral efficiency for a single channel system with $n = 2$.

In Fig. 5.1, the system energy is plotted versus the maximum tolerated delay for a single channel system when scheduling is performed by SDDPS scheduler. Obviously, a trade-off between delay tolerance and energy consumption occurs which is more noticeable at smaller spectral efficiencies. Moreover, savings in system energy are more pronounced when n varies from 1 to 2 as compared to the case when n varies from 4 to 5. This effect is similar to time diversity where performance improvement is more pronounced at the addition of a few initial degrees of diversity.

Fig. 5.2 demonstrates the effect of frequency diversity on both DDPS and SDDPS for $n = 2$. A unique set of thresholds need to be optimized for

TABLE 5.2: Approximate and Exact Thresholds for every Spectral Efficiency

	Approximate $\vec{\kappa}$ For n=2		Exact $\vec{\kappa}$ For n=2	
C	κ_1	$(E_b/N_0)_{sys}$	κ_1	$(E_b/N_0)_{sys}$
0.05	0.24	-1.42 dB	0.24	-1.42 dB
2	0.24	-0.35 dB	0.27	-0.36 dB
4	0.24	1.56 dB	0.34	1.49 dB
6	0.24	4.15 dB	0.46	3.96 dB
10	0.24	11.28 dB	0.78	10.8 dB

a specific number of channels as different number of channels have different sets of optimal thresholds. DDPS and SDDPS show indistinguishable performance when there is frequency diversity because marginal improvement in DDPS is hard to observe in presence of frequency diversity. Without frequency diversity, DDPS slightly outperforms SDDPS.

5.2 Future Rate Prediction Based Scheduling

To compare our results for the SDDPS scheduler, we consider another scheduling scheme which uses same algorithm for scheduling of users but computes transmission thresholds differently.

We consider Future Rate Prediction based Scheduling (FRPS) for the deadline constrained multiuser wireless systems [Butt, Kansanen, and Müller, 2008a]. Like SDDPS, fading of each user is compared with a set of transmission thresholds. These thresholds are computed using the fading statistics and deadline delay constraint. For a deadline constraint of n time slots, if a packet waits for W slots in the buffer, it still has $n - W$ opportunities to transmit the data before the deadline.

For a multi-band system with M channels, a packet with deadline of n time slots is transmitted if,

$$f_k(t-W) > \mathbb{E}[\max(X_k(t-W+1), X_k(t-W+2), ...X_k(t-W+n-W))] \tag{5.3}$$

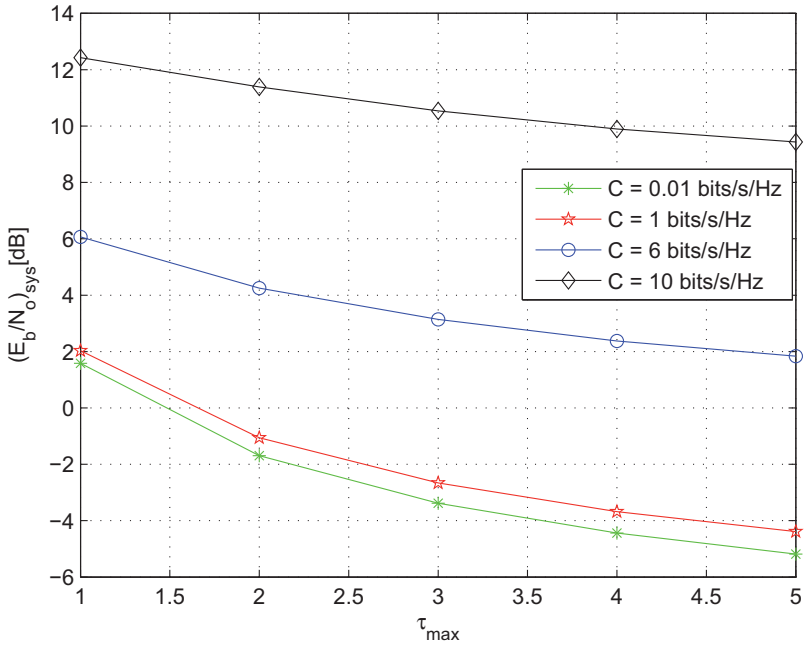


FIGURE 5.1: Energy-Delay tradeoff exhibited by SDDPS scheme.

where $f_k(t-W)$ is the fading in slot W while $X_k(t-W+j)$ is the random variable representing estimate of fading in the jth time slot in future. A packet is transmitted only if channel gain in the current slot is greater than the expectation of *all* the channel gains in the remaining slots before the deadline. This idea resembles the car fueling problem where a car is fueled if fuel price today is less than the expected cost of fueling in the coming days unless the car completely runs out of the fuel. This idea is used to compute the transmission thresholds on the basis of deadlines of individual packets.

5.2.1 Threshold Computation in FRPS

From order statistics, the cumulative distribution function $F_n(y)$ of the maximum of iid distributed n random variables $X_1, X_2, ...X_n$ is given by [David and Nagaraja, 2003],

$$\begin{aligned} F_n(y) &= P[(X_1 \leq y) \bigcap (X_2 \leq y) \bigcap ... \bigcap (X_n \leq y)] \\ &= [F_X(y)]^n \end{aligned} \tag{5.4}$$

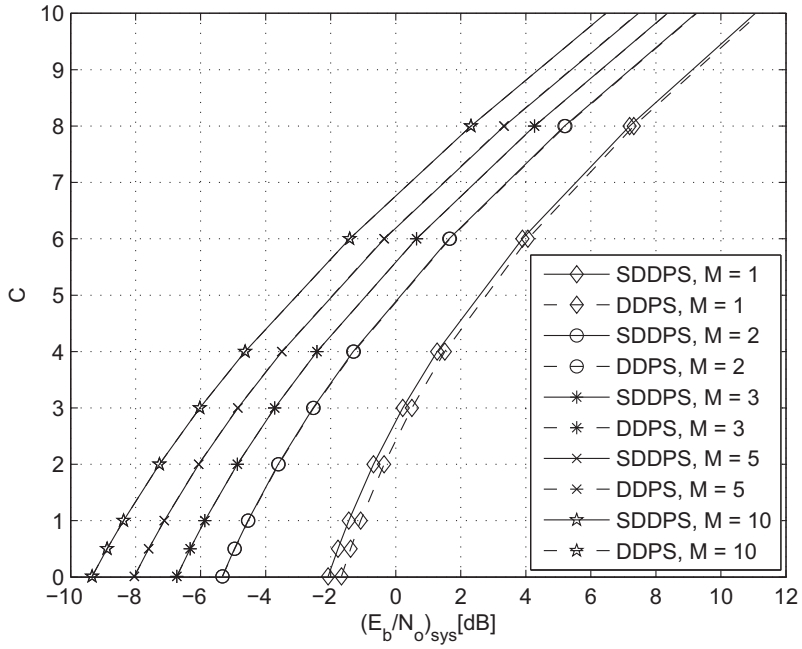


FIGURE 5.2: Comparison of DDPS and SDDPS schemes for different number of channels M.

Corresponding probability density function (pdf) is obtained by differentiating Eq. (5.4) and is given by,

$$f_n(y) \quad = \quad n f_x(y)(F_x(y))^{n-1} \tag{5.5}$$

We have assumed exponential short–term fading in this work. For the exponential short term fading distribution with mean 1, the threshold κ_W of a packet with waiting time W is computed by finding the expectance of maximum of $M(n-W)$ exponential random variables with means $\vec{\nu} = (\nu_1 ... \nu_{M(n-W)})$.

The *bth* moment $L_n(\frac{1}{\vec{\nu}}, b)$ of the maximum of $n \geq 1$ independent, negative exponential random variables with means $\vec{\nu} = (\nu_1 ... \nu_n)$ can be computed recursively using the following equation [Harrison and Zertal, 2007],

$$L_n(\frac{1}{\vec{\nu}}, b) = \frac{b}{\sum_{j=1}^{n} \frac{1}{\nu_j}} L_n(\frac{1}{\vec{\nu}}, b-1) + \frac{\sum_{j=1}^{n} \frac{1}{\nu_j} L_{n-1}(\frac{1}{\vec{\nu}}, b)}{\sum_{j=1}^{n} \frac{1}{\nu_j}} \tag{5.6}$$

for $n \geq 1$, $L_0(\vec{\gamma}, b) = 0$ for all $b \geq 1$ and $L_n(\frac{1}{\vec{\nu}}, 0) = 1$ for all $n \geq 0$. $\vec{\gamma}$ represents a null vector of zero components.

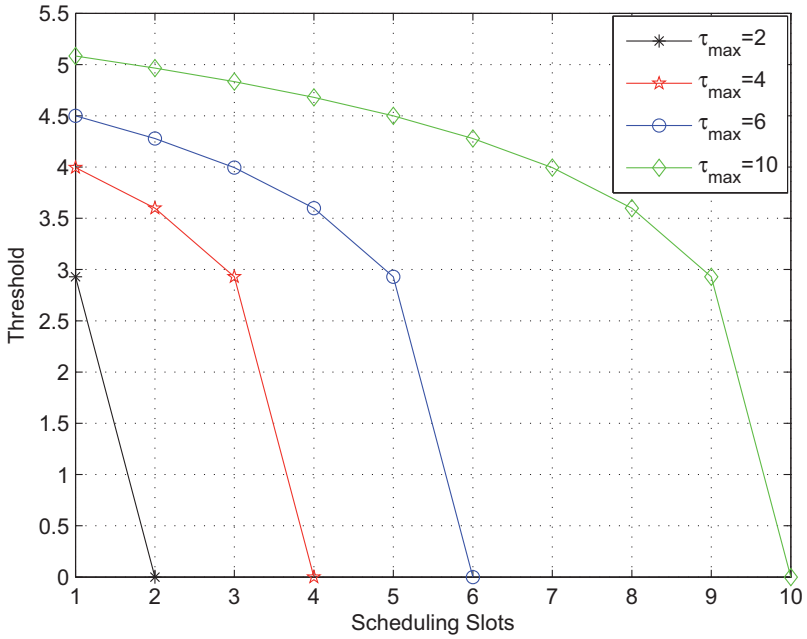


FIGURE 5.3: Threshold function for the FRPS scheduler in each time slot for $M = 10$ and different deadline constraints.

For the special case when the mean ν of all the distributions is equal, the result can be simplified. We assume this case here for computing the expectance of maximum of $M(n-W)$ random variables for $n-W$ slots left before the deadline. The threshold κ_W of a packet with backlog W is given by,

$$\kappa_W = \nu \sum_{j=1}^{M(n-W)} \frac{1}{j} \tag{5.7}$$

Fig. 5.3 shows the threshold function for FRPS for different deadlines. Note, the function behaves similar to SDDPS but it is deterministic and thresholds can be computed in closed form using order statistics.

In Fig. 5.4, we compare the energy delay tradeoff for SDDPS and FRPS. We observe that SDDPS outperforms FRPS significantly. Note that the only difference between two schemes is the computation of thresholds.

For SDDPS, the cost of transmission $C(t-W)$ for a packet with waiting time W in a time slot $t-W$ is the minimum of transmit energy $E(t-W)$ in current time slot and the expectation of cost $E(t-W+1)$ in time slot

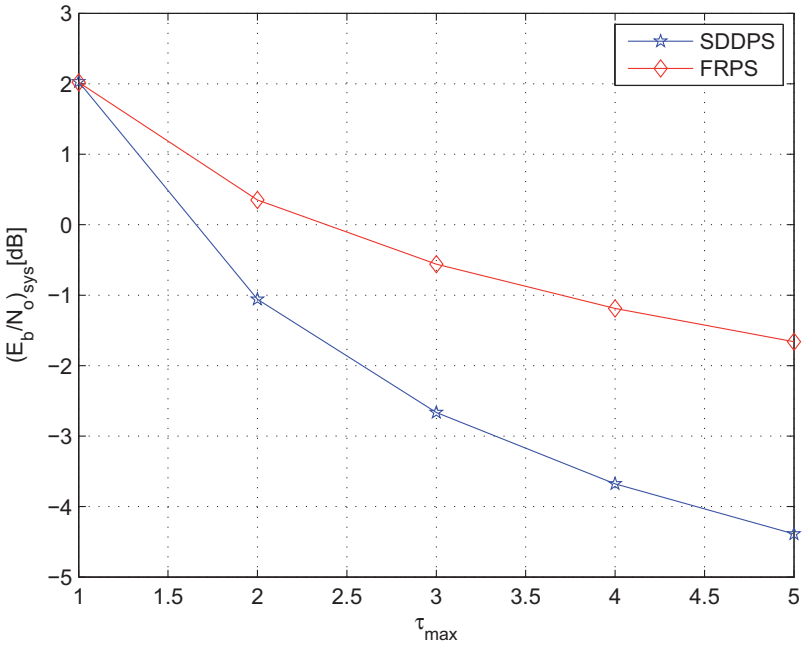


FIGURE 5.4: Comparison of energy-delay tradeoff for a single channel system at C=1.

$t - W + 1$.

$$C(t-W) = \min\Big[E(t-W), \mathbb{E}\big(C(t-W+1)\big)\Big] \tag{5.8}$$

For FRPS, the cost of transmission $C(t-W)$ in a time slot $t-W$ is the minimum of current transmit energy and the expectation of minimum of transmit energies in remaining $n-W$ time slots.

$$C(t-W) = \min\Big[E(t-W), \mathbb{E}\big(\min(E(t-W+1), \ldots E(t-W+n-W)\big)\Big] \tag{5.9}$$

The expectation terms in Eq. (5.8) and Eq. (5.9) represent the transmission thresholds for the respective schemes. Eq. (5.9) can be computed in closed form while Eq. (5.8) requires dynamic programming solution [Bertsekas, 2007].

Remark: Though, SDDPS outperforms FRPS for Rayleigh fading and our problem settings, we do not have mathematical analysis to claim that SDDPS outperforms FRPS in all conditions. However, based on our results, we use SDDPS in rest of the work. □

5.3 Packet Dropping

In this section, we discuss the idea of scheduling with outage. In practical delay constrained systems, it is impossible to fulfill the hard deadline delay requirements without some kind of outage. We consider the case when dropping of a certain fraction θ_d of packets is allowed. An individual dropping probability θ_d is provided to each user and there is no system wide dropping fraction. We specify an SDDPS scheduler is by a tuple $(\vec{\kappa}, \theta_d, n)$.

Packet dropping is a quite practical scenario in modern wireless networks as a lot of multimedia applications are loss tolerant and dropping of certain fractions of packets still keeps the quality of the application within acceptable limits. Moreover, providing rate R to schedule all the packets in a bad channel state requires a lot of energy and is not always practical.

Without packet drop, the transmission threshold $\kappa_{n\to n}$ is set to zero to transmit the packet in the deadline state n unconditionally. To model the effect of packet drop, let the lowest transmission threshold in the deadline state, i.e. $\kappa_{n\to n}$ be non-zero. When state transition $T_{n\to n}$ takes place, two different things may happen from the scheduling perspective: Either a single packet is scheduled or a packet is dropped. The state transition probability α_{nn} is a sum of the two probabilities $\hat{\alpha}_{nn}$ and $\tilde{\alpha}_{nn}$ where we define $\hat{\alpha}_{nn}$ and $\tilde{\alpha}_{nn}$ as

$$\tilde{\alpha}_{nn} = \Pr(\kappa_{n\to n} < f \leq \kappa_{n\to n-1}) \tag{5.10}$$

$$\hat{\alpha}_{nn} = \Pr(f \leq \kappa_{n\to n}) \tag{5.11}$$

The probabilities $\hat{\alpha}_{nn}$ and $\tilde{\alpha}_{nn}$ correspond to the events when the scheduler drops a packet and schedules a single packet for transmission, respectively.

The transmission threshold $\kappa_{n\to n}$ in the deadline state determines the probabilities $\hat{\alpha}_{nn}$ and $\tilde{\alpha}_{nn}$. For a dropping probability of θ_d and a given possible state transition matrix $\mathbf{P}$, the probability $\tilde{\alpha}_{nn}$ is computed such that the following equality holds.

$$\theta_d = \prod_{i=1}^{n}(1 - P_i) \tag{5.12}$$

$$= \prod_{i=1}^{n}\Big(1 - \sum_{l=1}^{i}\alpha_{il}\Big) \tag{5.13}$$

where P_i denotes the probability of scheduling a packet in state i. The equation states that a fraction θ of the packets enter the system and are not able

to be scheduled in n time slots. From Eq. (5.13)

$$\tilde{\alpha}_{nn} = 1 - \sum_{l=1}^{n-1} \alpha_{nl} - \frac{\theta_d}{\prod_{i=1}^{n-1}\left(1 - \sum_{h=1}^{i} \alpha_{ih}\right)} \tag{5.14}$$

Due to additional constraint in terms of dropping probability, the heuristic algorithm for optimization of transmission thresholds does not work well. Therefore, we use SA when we have non-zero dropping probability. For the SDDPS scheduler state space description in Section 5.1 with packet drop capability and identical deadline, we use the following simplified state transition matrix $\mathbf{P}_{\mathrm{d}}$ to vary the configuration in each step in SA.

$$\mathbf{P}_{\mathrm{d}} = \begin{pmatrix} \alpha_1 & 1-\alpha_1 & 0 & \cdots & 0 \\ \alpha_1 & \alpha_2 & 1-\alpha_1-\alpha_2 & \cdots & 0 \\ \ddots & \ddots & \ddots & \ddots & \ddots \\ \alpha_1 & \alpha_2 & \cdots & \cdots & \tilde{\alpha}_n + \hat{\alpha}_n \end{pmatrix}. \tag{5.15}$$

Note that transition probabilities for entering a state from any of the higher backlog states are equal and therefore, the scheduler requires the same threshold to enter a specific backlog state (Property 5.1). This is due to the reason that the scheduler (virtually) visits all the intermediate states between the backlog states i and j and moves into the adjacent lower state sequentially. Thus, the scheduler always enters into state j from state $j+1$ and requires same transmission threshold.

Pseudo code of SDDPS scheme with packet dropping and random ar-

rivals is shown in Algorithm 5.3.1.

Algorithm 5.3.1: SDDPS($Backlog, \vec{\kappa}, n$)

comment: User k knows backlog, thresholds and deadline

$i \leftarrow Backlog$

comment: Rate R is provided by the scheduler

$R \leftarrow 0$

$L \leftarrow 0$

comment: Find rate R to be provided

for $b \leftarrow Backlog$ **to** 1

comment: $N(t-b+1)$ packets arrived at time $(t-b+1)$

do
- **if** $f_k > \kappa_b$
  - **then**
    - $L \leftarrow L+1$
    - $R \leftarrow R + \frac{\Gamma}{K} N(t-b+1)$
  - **else**
    - **if** $(b = n)$
      - **comment:** Packet Dropping in deadline state
      - **then** $i \leftarrow i-1$
    - **break**

$Rate \leftarrow R$

$Backlog \leftarrow i - L + 1$

Table 5.3 shows the thresholds values for the SDDPS scheduler when

we allow different dropping probabilities. As dropping probability increases, transmission thresholds increase to demonstrate the fact that a packet has more freedom to wait for the better channel as compared to lossless delay constrained systems.

Fig. 5.5 shows the effect of dropping of packets in the deadline state for a system with $n = 2$. At small values of C, the effect of packet dropping is more pronounced and energy saving is not as large at higher values of spectral efficiency. At $C = 1$, dropping probability of 0.1 saves more than 3 dB of system energy. Therefore, SDDPS scheduler allows the users to have tradeoff between energy efficiency and quality of the application in addition to providing energy-delay tradeoff.

TABLE 5.3: Thresholds Computation for SDDPS with $M = 1$

	$\theta_d = 0\%$				$\theta_d = 5\%$			
n	κ_1	κ_2	κ_3	E_b/N_0	κ_1	κ_2	κ_3	E_b/N_0
2	0.24	0	-	-1.4 dB	0.44	0.15	-	-3.4 dB
3	0.52	0.25	0	-3.0 dB	0.78	0.53	0.26	-4.7 dB

5.3.1 Fair Scheduling Unfair Dropping

In Section 3.2, we mentioned that the purpose of making scheduling decisions based on fast fading (and not channel gain) is to provide fair chance of scheduling to all the users. Throughout this work, we have followed this practice. However, there are certain applications where introducing some kind of unfairness does not hurt much and helps in saving the overall system energy. For example, some of the applications in wireless sensor networks require location based transmission of data towards the fusion node. The fusion node can be placed near the location where lossless (or small loss tolerant) data transmission is the requirement.

We provide a solution where a cell can be divided into multiple circular areas around the fusion node. We still make the scheduling decisions based on fast fading and keep delay requirements of the users identical. However, we introduce a small unfairness by having non–uniform dropping in these circular regions. For example, we divide the cellular region of area A into two small equal regions of areas $A/2$. As users (sensors) are

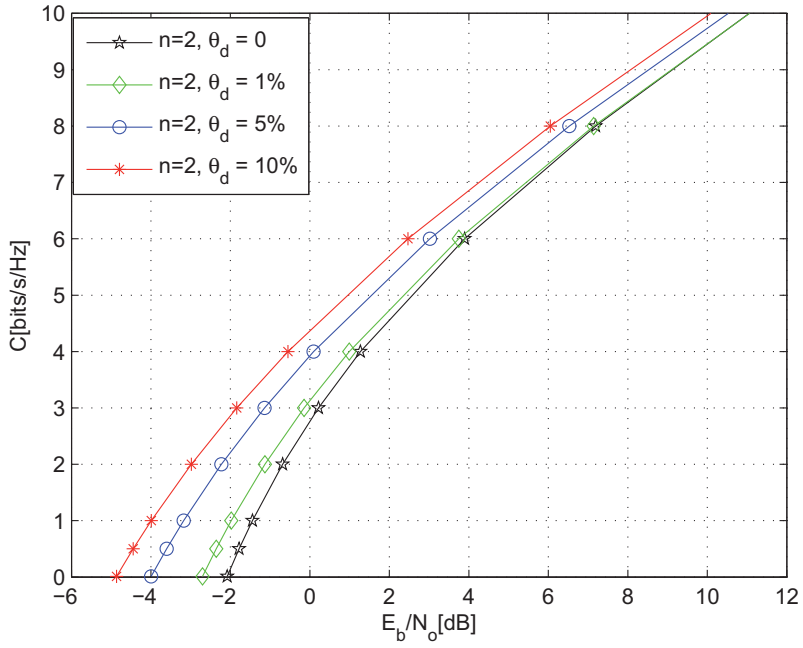


FIGURE 5.5: System energy as a function of packet dropping probabilities.

assumed to be equally distributed, both of the regions have equal number of users statistically. A schematic diagram for such a WSN has been shown in Fig. 5.6.

We introduce non-uniform dropping probabilities. In the inner area (close to the fusion node), we allow packet drop with probability θ_{in}. In the outer area, we allow packet dropping with probability θ_{out} where $\theta_{\text{out}} > \theta_{\text{in}}$ as fusion node is placed close to the sensor nodes having strict requirement on data loss. It should be noted that dropping rate is not user based anymore. We have a system wide drop rate θ_s that is weighted (by area) sum of θ_{out} and θ_{in} such that

$$A\theta_s = A_{\text{in}}\theta_{\text{in}} + A_{\text{out}}\theta_{\text{out}} \tag{5.16}$$

Due to mobility of the users, we do not know which user will be in the outer cell when $i = n$. Therefore, we allow each user in the system to use thresholds from the same tuple $(\vec{\kappa}, \theta_{\text{out}}, n)$ for $n-1$ time slots. If a user is in state n, only then she needs to figure out about her location. If she finds herself in the inner cell, she simply uses κ_n corresponding to tuple $(\vec{\kappa}, \theta_{\text{in}}, n)$ otherwise keeps on using κ_n corresponding to her original tuple.

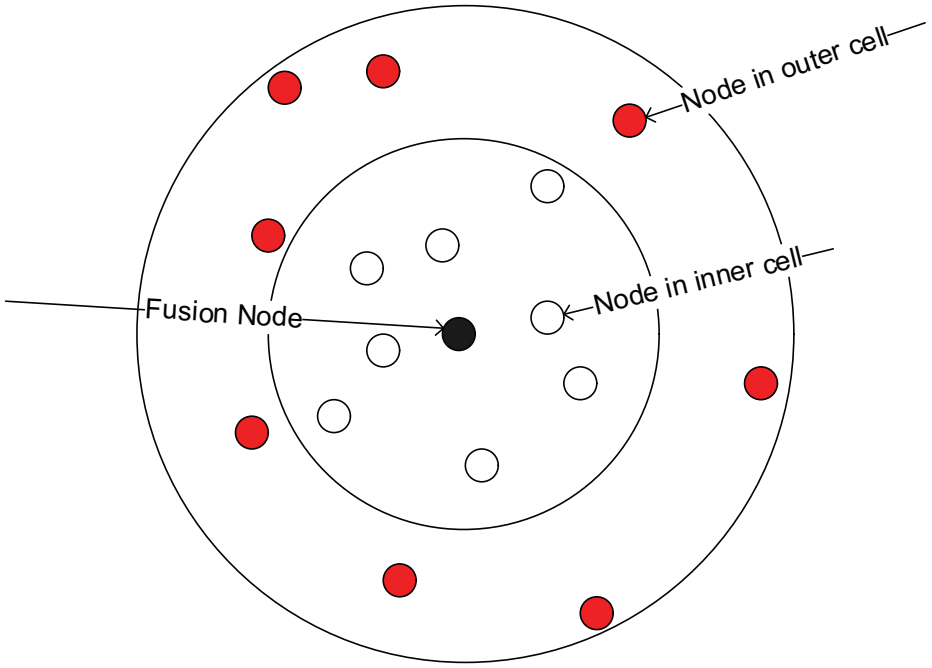


FIGURE 5.6: Schematic diagram for a WSN with inner and outer cells defined with equal area.

Implementation of the scheme is explained with the help of an example in the following.

Example 5.1 Assume that we want to have drop probability $\theta_{\text{out}} = 0.1$ for the users in the outer cell and lossless data transmission for the users in the inner cell i.e. $\theta_{\text{in}} = 0$ for the same deadline requirement n.

All the users follow the same set of thresholds which correspond to deadline n and user based drop probability 0.1. In the last time slot, if a user finds herself in the inner cell, she uses the thresholds corresponding to θ_{in} and deadline n (which is zero in this case). If she finds herself into outer cell, she uses the threshold corresponding to θ_{out} and deadline n. As users are equally distributed in the whole cell, half of the users drop the packets with rate 0.1 and half of them do not drop any packet at all which results into system wide drop probability of 0.05. Extending this approach, we can divide geographical area into multiple areas and allow different dropping probabilities according to the system requirement. □

Fig 5.7 shows the effect of unfair dropping on system energy. We plot two curves with uniform dropping probability θ_d of 0.05 and 0.1. We divide the cell into two areas where inner area requires lossless data, i.e. $\theta_{\text{in}} = 0$. The areas of outer regions are $A/2$ and $A/4$ for the simulations. We observe an energy saving when we employ unfair dropping. For example,

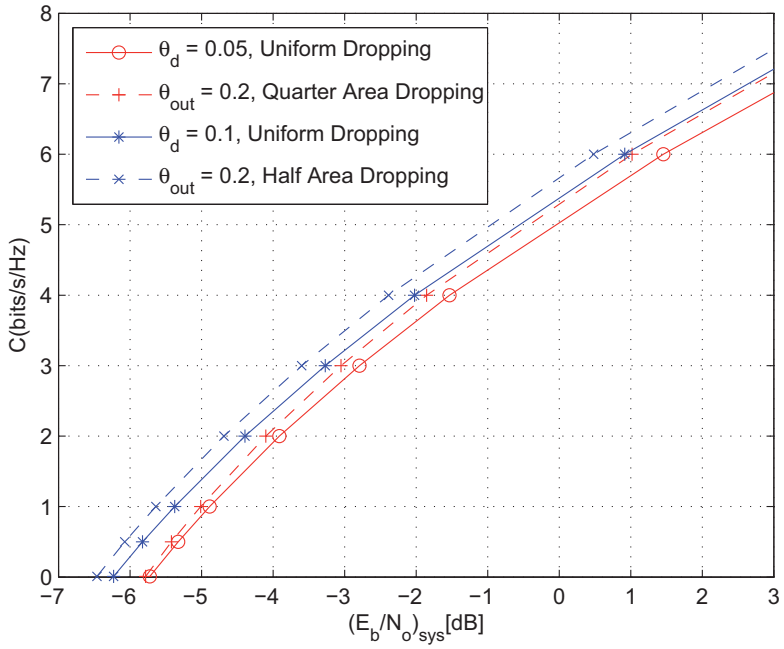


FIGURE 5.7: System energy as a result of region based dropping probabilities for a system with $n = 2, M = 2$.

unfair dropping with probability 0.2 and inner area $A/2$, is more energy efficient to fair dropping with probability 0.1 though both of them result in same aggregate data loss. Similarly, with the inner region area $A/4$, unfair dropping with probability 0.2 is same as fair dropping with probability 0.05 but it provides a more energy efficient system.

5.4 SDDPS: Non–identical Packet Deadlines

We model this case as a special case of SDDPS scheduler [Butt, Kansanen, and Müller, e]. We use Markov process as before. The *maximum* possible deadline of an arriving packet in the buffer is denoted by n. Then, we define the deadline of an arriving packet $q < n$ with respect to n. i.e. the deadline of a packet with deadline of $0 \leq a < n$ time slots less than n is represented by $q = n - a$ where a is termed as *deadline offset*. We denote probability of arrival of a packet with deadline q by p_q. We assume infinite size buffer for simplicity of analysis.

We model the arrival of a packet with deadline $q = n - a$ by a packet

that already has spent a time slots in the buffer. By this initial offsetting, we model a system with non-identical packet deadlines equivalent to a system with identical packet deadlines.

We define some terms used in the work.

Definition 5.1 (Deadline Distance) The deadline distance $\chi \in \{1 \dots n\}$ for a packet is defined as the number of time slots remaining before it reaches its hard deadline. □

In case of individual packet deadline, we redefine backlog state in terms of deadline distance.

Definition 5.2 (Backlog State) Backlog state i in a Markov chain is defined as the *minimum* of the deadline distance for the packets waiting to be scheduled in the buffer.

$$i = \min\Big(\chi_1, \chi_2 \cdots \chi_j\Big) \tag{5.17}$$

where χ_j represents the deadline distance of the jth packet in the buffer. □

For better understanding, we enumerate the backlog states in descending order such that a packet with deadline distance χ_j enters the buffer in state j as shown in Fig. 5.8. As deadline distance of the HOL packet decreases, backlog state i *decreases*[1] as well.

The definition for transmission threshold holds for this case as well. SDDPS scheduler works in the similar way as explained in Section 4.1 and 5.1. Due to change in the numbering scheme for the backlog states, we mention the main changes briefly. As before, each user compares the current short-term fading f_k with the threshold $\kappa_{i \to j}$ for every backlog state $j \geq i$ sequentially. If $f_k > \kappa_{i \to i}$, all the packet with deadline distance χ_j are scheduled for transmission and the threshold of the next *higher* backlog state is compared with f_k. Similarly, the thresholds of all the backlog states $j \geq i$ are compared *sequentially* with f_k until f_k is less than the threshold $\kappa_{i \to j+1}$ of a state. The last backlog state in which a packet is scheduled is termed as *ending* state $j(t)$. The scheduler moves to backlog state $j(t)$ from a backlog state $i(t)$ by scheduling packets in the L intermediate states and therefore, ending state $j(t)$ is given by

$$j(t) = i(t) + L(y, i) - 1 \tag{5.18}$$

The backlog state $i(t+1)$ is determined after the arrival of a new packet at time $t+1$. Depending on the deadline of the next arriving packet, the

[1]Note that contrast to the models discussed in previous chapters, states are numbered in decreasing order now. Therefore, backlog state decreases for the unscheduled packets.

scheduler stays in backlog state $j(t)$ or moves into backlog state $q(t+1)$ such that

$$i(t+1) = \begin{cases} j(t) & if \quad q(t+1) \geq j(t) \\ q(t+1) & if \quad q(t+1) < j(t) \end{cases} \tag{5.19}$$

where $q(t+1)$ denotes the deadline of the arriving packet at time $t+1$. If f_k is less than the thresholds of all the states, no packet is scheduled. Then, the scheduler moves into backlog state $i(t+1)$ according to the Eq. (5.19) where $j(t) = i(t) - 1$. The deadline distance χ of all the remaining un–scheduled packets in the buffer is decremented by one in each time slot.

We model the special case when no packet arrives in a time slot by considering arrival of a packet with zero size and deadline n. This assumption keeps our state space description consistent for no arrival case as well. When a packet with zero size is scheduled for transmission, no rate is allocated.

State transition is a two step process. The scheduler schedules packets in $L(y,i)$ backlog states, allocates rate for the actual data and then moves into backlog state $i(t+1)$ according to Eq. (5.19). Note that the scheduler schedules packets in $L(y,i)$ backlog states without the knowledge that arriving packets have identical or non-identical deadlines. It is a packet based scheduling algorithm and the transmission thresholds depend only on the fading distribution and number of time slots left before the deadline.

State transition diagram of the SDDPS scheduler with packets having non-identical deadlines is shown in Fig. 5.8 and corresponding state transition matrix (STM) is given by

$$\mathbf{P}^{\text{nid}}_{\text{SDDPS}} = \begin{pmatrix} \acute{\alpha}_{11} & \acute{\alpha}_{12} & \acute{\alpha}_{13} & \cdots & \acute{\alpha}_{1n} \\ \acute{\alpha}_{21} & \acute{\alpha}_{22} & \acute{\alpha}_{23} & \cdots & \acute{\alpha}_{2n} \\ \ddots & \ddots & \ddots & \ddots & \ddots \\ \acute{\alpha}_{n1} & \acute{\alpha}_{n2} & \acute{\alpha}_{n3} & \cdots & \acute{\alpha}_{nn} \end{pmatrix}. \tag{5.20}$$

With our enumeration scheme for states, STM for identical deadline case is given by [Butt *et al.*, 2008b]

$$\mathbf{P}^{\text{id}}_{\text{SDDPS}} = \begin{pmatrix} \alpha_{11} & \alpha_{12} & \cdots & \alpha_{1n} \\ \ddots & \ddots & \ddots & \ddots \\ 0 & \cdots & \alpha_{(n-1)(n-1)} & \alpha_{(n-1)n} \\ 0 & \cdots & \alpha_{n(n-1)} & \alpha_{nn} \end{pmatrix}. \tag{5.21}$$

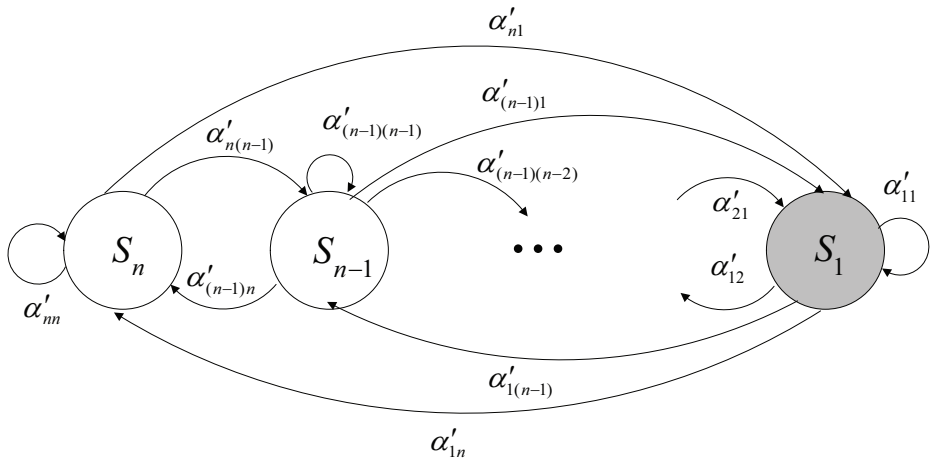


FIGURE 5.8: State diagram for the transition states for the SDDPS scheduler. Due to non-identical deadlines of arriving packets, the scheduler moves into any of the next state if no packet is scheduled.

where α_{ij} is defined as

$$\alpha_{ij} = \Pr(\kappa_j < f \leq \kappa_{j+1}). \tag{5.22}$$

The termination condition $\kappa_{i\to n+1}$ is defined as ∞. We set $\kappa_{1\to 1}$ to zero to ensure the transmission of the packet reaching its deadline.

Note that STM for identical deadline case is a triangular matrix while STM for non-identical deadline case is a sparse matrix. To compute STM for the non–identical deadline case, we evaluate the effect of non-identical deadlines on state space representation. In the first step, we compute the value of j as described in Eq. (5.22) for the identical deadline case and remains the same for non–identical deadline case. However, STM for non-identical case is modified by the deadline distribution as well. In the second step, for a given pair of i, j, we compute the offset produced by this distribution by evaluating the *offset matrix* $\mathbf{S}^{\text{nid}}$ given by

$$\mathbf{S}^{\text{nid}} = \begin{pmatrix} \sum_{\mu=1}^{n} p_\mu & 0 & \cdots & 0 \\ p_1 & \sum_{\mu=2}^{n} p_\mu & \cdots & 0 \\ \ddots & \ddots & \ddots & \ddots \\ p_1 & \cdots & p_{n-1} & \sum_{\mu=n}^{n} p_\mu \end{pmatrix}. \tag{5.23}$$

where probability $S^{\text{nid}}(q, j)$ is defined as

$$S^{\text{nid}}(q,j) = \begin{cases} 0 & if \quad q > j \\ \sum_{\mu=j}^{n} p_\mu & if \quad q = j \\ p_q & if \quad q < j \end{cases} \tag{5.24}$$

Diagonal elements in Eq. (5.23) represent the sum of probabilities of deadline distribution which keeps $i(t+1) = j(t)$ while non zero non diagonal elements represent the probability when $i(t+1) = q(t+1)$ as explained in Eq. (5.19). We can represent STM for non identical case as a product of Eq. (5.21) and Eq. (5.23) such that

$$\mathbf{P}^{\text{nid}}_{\text{SDDPS}} = \mathbf{P}^{\text{id}}_{\text{SDDPS}} \mathbf{S}^{\text{nid}} \tag{5.25}$$

One obvious difference between the STMs for the systems having arrived packets with identical deadlines and the systems having arrived packets with non–identical deadlines is in the transition probabilities when no packet is scheduled. If all the arriving packets have identical deadline of n time slots, the only possible state is $i-1$. However, if the deadlines are non identical, depending on the deadline of the arriving packet, the scheduler moves into any of the states $j < i$.

Example 5.2 We explain it with the help of an example when $n = 2$. If all the packets have identical deadline, $\mathbf{P}^{\text{id}}_{\text{SDDPS}}$ is given by

$$\mathbf{P}^{\text{id}}_{\text{SDDPS}} = \begin{pmatrix} \alpha_{11} & \alpha_{12} \\ \alpha_{21} & \alpha_{22} \end{pmatrix}. \tag{5.26}$$

If the packets have non-identical deadlines of $n_1 = 1$ and $n_2 = 2$, n equals 2 and resulting $\mathbf{P}^{\text{nid}}_{\text{SDDPS}}$ is given by

$$\begin{aligned} \mathbf{P}^{\text{nid}}_{\text{SDDPS}} &= \begin{pmatrix} (p_n + p_{n-1})\alpha_{11} + p_{n-1}\alpha_{12} & p_n\alpha_{12} \\ (p_n + p_{n-1})\alpha_{21} + p_{n-1}\alpha_{22} & p_n\alpha_{22} \end{pmatrix} \\ &= \begin{pmatrix} (p_2 + p_1)\alpha_{11} + p_1\alpha_{12} & p_2\alpha_{12} \\ (p_2 + p_1)\alpha_{21} + p_1\alpha_{22} & p_2\alpha_{22} \end{pmatrix} \\ &= \begin{pmatrix} \alpha_{11} + p_1\alpha_{12} & p_2\alpha_{12} \\ \alpha_{21} + p_1\alpha_{22} & p_2\alpha_{22} \end{pmatrix}. \end{aligned} \tag{5.27}$$

where p_n and and p_{n-1} represent the probabilities of arrival of a packet with deadline $n = 2$ and $n = 1$, respectively. Note that Eq. (5.27) is reduced to Eq. (5.26) if all the arriving packets have identical deadline of 2. □

5.5 Asymptotic Analysis of SDDPS For Non–Identical Deadline Case

We analyze SDDPS in the large system limit. For the case of individual packet deadlines, the channel distribution of the users in state i is given by

$$\mathrm{p}_{f,\mathrm{SVU}}(y|i) = c_i^{nid} L(y,i)\, \mathrm{p}_{\max\{f\}}(y) \tag{5.28}$$

with $\mathrm{p}_{\max\{f\}}(y)$ and c_i^{nid} denote the short–term fading of the best of the channels for a multi-channel system and a constant to normalize the conditional pdf.

In the following, we derive channel distributions of SVUs with two different methods and prove their equivalence.

First, we model two step state transition process as a single step stochastic process and weight every possible state transition probability α_{ij} according to Eq. (5.24). We represent the channel distribution of SVUs in terms of weighted sum of rate allocation for probabilistic state transitions. Note that scheduled rate is only a function of fading and state as shown in Eq. (5.22), but deadline distribution of the arriving packets randomizes the state transition and corresponding rate allocation function. For a given fading f and state i, rate is probabilistically allocated to the states $i \leq r \leq j$ and weighted by the probability of ending up in state r. The probability of ending in state r can directly be computed from Eq. (5.24) for given i, j. We can write $\mathrm{p}_{f,\mathrm{SVU}}(y|i)$ as

$$\begin{aligned} \mathrm{p}_{f,\mathrm{SVU}}(y|i) &= c_i^{nid} \mathrm{E}(\mu_t) \Big((j-i+1) \sum_{l=j}^{n} p_l \\ &+ \sum_{r=i}^{j-1} (r-i+1) p_r \Big)\, \mathrm{p}_{\max\{f\}}(y) \end{aligned} \tag{5.29}$$

where j is uniquely defined by Eq. (5.22) and rate of arrival in each scheduled state is $\mathrm{E}(\mu_t)$.

In the second method of modeling, we give a more intuitive description of the effect of non-identical deadlines on system behaviour. For a given fading and state, we consider a fixed state transition (and rate allocation function). However, the contents (number of packets in each state) of the scheduled states depend on the deadline distribution. This method follows directly from the two step process explained earlier.

Note that arrival of a new packet always takes place in state n for the case when all the packets have identical deadlines. For a two state system

with $n = 2$, if a packet is not scheduled at time t, it moves into state 1 with probability one. In non–identical deadline case, the probability of arrival in state two is p_2. If a packet is not scheduled in state 2, it moves into state 1 with probability p_2, making rate of arrival p_2 for state 1. However, there is an additional *direct* source of arrival in state 1 with probability p_1 at time $t + 1$ (due to deadline offset). Therefore, cumulative buffer content of state 1 is summation of p_1 and p_2 while buffer content of state 2 is p_2. Recall that arrival with non–identical deadlines results in insertion of the arriving packet with deadline less than n in state $n - a$. The resulting content size for the states with small deadline distances is greater as compared to the states with large deadline distances. For a state r, this effect is modeled in Eq. (5.30) by considering an equivalent arrival process whose arrival rate is a summation of the arrival probability p_l over $r \leq l \leq n$. For a given i, j pair, due to non-identical deadline of the arriving packets, the random content size in $L(y, i)$ scheduled states is given by

$$L(y, i) = \sum_{r=i}^{j} \sum_{l=r}^{n} p_l \tag{5.30}$$

This situation has been depicted in Fig. 5.9. In a state r, p_{n-a} is the rate of arrival for the packets with deadline *offset* equal to a where $r = n - a$. The second source of arrival originates from the previously un–scheduled packets in the buffer who are left with deadline distance $\chi_r = n - a$.

Consequently, using Eq. (5.30) in Eq. (5.28), the channel distribution of the scheduled users for the non–identical deadline case can be written in terms of rate allocation function for a fixed state transition and random content, and given by

$$\mathrm{p}_{f,\mathrm{SVU}}(y|i) = c_i^{nid} \mathbb{E}(\mu_t) \sum_{r=i}^{j} \sum_{l=r}^{n} p_l \, \mathrm{p}_{\max\{f\}}(y) \tag{5.31}$$

where c_i^{nid} is a normalization constant.

Lemma 5.2 The representations of distribution of scheduled virtual users in Eq. (5.29) and Eq. (5.31) are equivalent. □

Proof: See Appendix D. □

Note that there is an important difference between modeling of random arrival as constant arrival with random size in Section 4.2 and modeling packet arriving with non identical deadlines. For random arrival modeling, in the limiting case, the content of every scheduled state converges to the

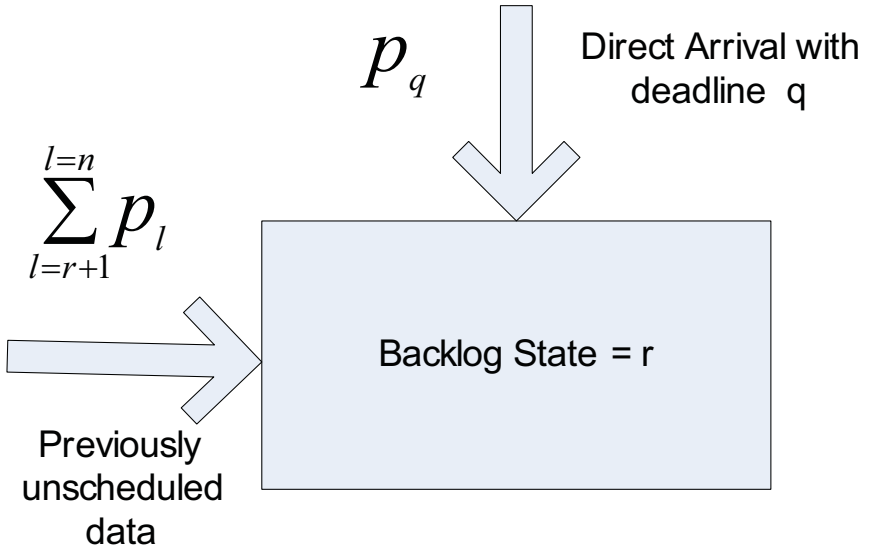


FIGURE 5.9: Arrival process demonstration for a backlog state r.

rate of arrival process $\mathbb{E}(\mu_t)$. However, contents of states are non uniform for non-identical packet deadline case and increase for the states closer to deadline. Random content size description helps in understanding the energy behaviour of the system when deadline constraints are non-identical for individual packets.

Corresponding channel distribution of the SVUs for the identical deadline case is given by

$$\mathrm{p}_{f,\mathrm{SVU}}(y|i) = c_i^{id}\mathbb{E}(\mu_t)\Big(j-i+1\Big)\,\mathrm{p}_{\max\{f\}}(y) \tag{5.32}$$

where c_i^{id} is a normalization constant.

Although, non–identical deadline case has been modeled as an extension of identical deadline case, state space description and channel distributions are not equivalent due to different STMs. However, as SDDPS scheduler treats every packet individually based on its deadline distance, transmission thresholds optimized for the systems with identical deadlines remain valid for the systems with non-identical deadlines as well.

We show the convergence result of non-identical arrival case for the finite user population in Fig. 5.10. The figure shows the convergence of variance of the system energy for identical and non–identical deadline cases. We use Eq. (3.2) to evaluate system energy for finite number of users K. To compute the system energy for a specific number of users, 250 fading val-

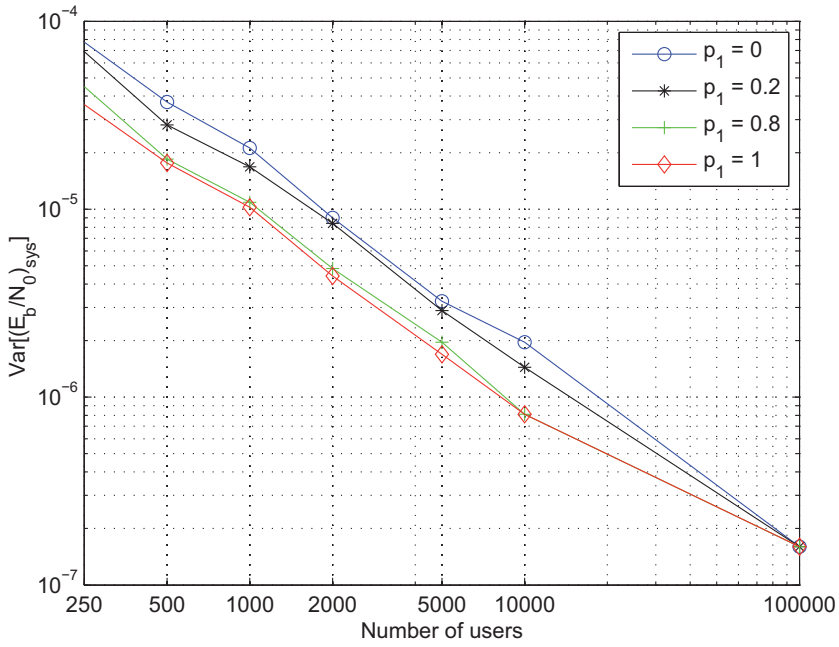


FIGURE 5.10: Convergence of System energy of scheduled users for identical and non–identical deadline cases.

ues have been used for a single path loss. The arrival process is Bernoulli with $p_{arr} = 0.7$. The curves with $p_1 = 0$ and $p_1 = 1$ represent the identical deadline case when all the arriving packets have deadline 2 and 1, respectively. $p_1 = 0.2$ represents the case when 20% of the packets arrive with $n = 1$ and 80% with $n = 2$. For all the cases, variance of the computed average system energy decreases as the number of users increases. The system energy for the system with a smaller deadline delay constraint converges faster as compared the distribution with a larger deadline delay. Hence, as ratio of p_1 decreases, it requires more number of users to converge.

Fig. 5.11 demonstrates the delay–energy trade off for a single channel system when the arriving packets have non–identical deadlines. We evaluate system performance at different spectral efficiencies. As the proportion of the packets with tight deadline constraint increases, average system energy increases correspondingly. However, this effect is more pronounced at small spectral efficiencies.

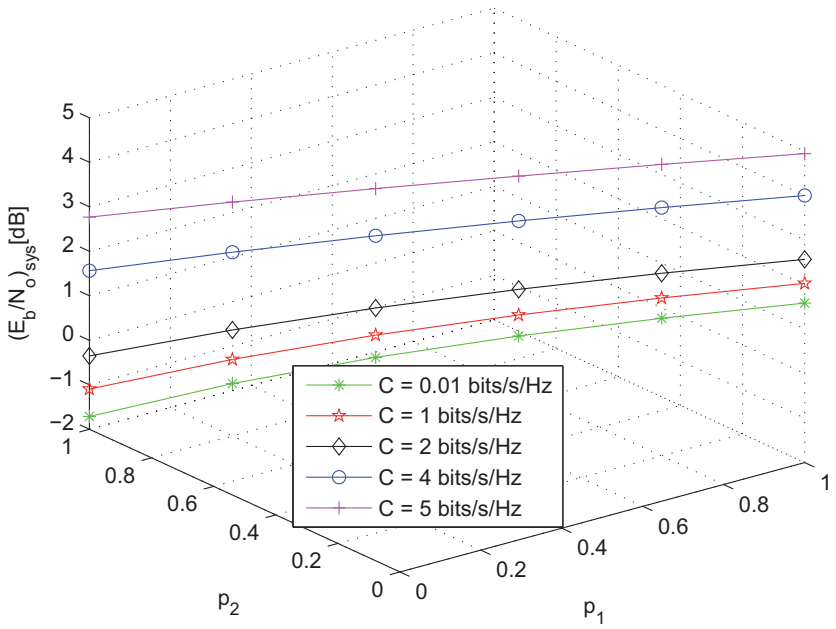


FIGURE 5.11: For $M = 1$ and different spectral efficiencies, system energy is plotted when p_1 and p_2 varies from 0 to 1.

5.5.1 Discussion on Implementation

In this work, we have derived the distribution of the scheduled users analytically and discussed the practical implementation of the SDDPS scheme for random arrivals, non–identical deadline and fractional packet drop. These characteristics are integral part of any practical scheduler and therefore, it is important to give practical implementation along with the theoretical details. In all the implementations, we try to keep the task of the scheduler as simple as possible. We focus on pre-processing of the packet in such a way that complexity arising due to random arrivals and non–identical deadlines is handled at higher layers and the scheduler does simple scheduling decisions. The details of the packet deadlines, arrival statistics and the packet drop ratio are provided by the higher layers and the media access layer uses this information to keep the scheduling operation simple on the physical layer. This approach is quite in line with the cross layer modeling approaches proposed in modern wireless networks.

Chapter 6

Opportunistic Multiuser-Multicell Scheduling

We have discussed deadline constrained opportunistic scheduling schemes in previous chapters for a single cell case. In this chapter, we consider the schemes in a multicell environment. For multicell analysis of the schemes, we use the system model and results in [Park and Caire, 2008],[Wyner, 1994]. In so called Wyner's model, all users share the same bandwidth and Base stations (BS) are arranged at a uniform distance on a line. The path gain to the closest BS is 1, the path gain to the adjacent base stations is α and zero elsewhere as shown in Fig. 6.1.

6.1 System Model

We consider a multiple–access system with K users randomly placed within a multicell system. System model is similar to the model presented in Section 3.1. For ease of understanding in the rest of the chapter, we describe it briefly for a multicell system. The minimum distance between two base stations is denoted by $D = 2r$ where cell radius is r. Each user is provided a certain fraction of the resources available to the system. The required average rate R for each user is $\frac{\Gamma}{K}$ where Γ denotes the spectral efficiency of the system. We consider a multiband system with M channels and spectral efficiency is normalized by M to get spectral efficiency per channel. We consider random arrivals in each time slot for all the users and model them as constant arrivals with different content size. Arrivals are queued in a finite buffer of length n before transmission. We consider an uplink

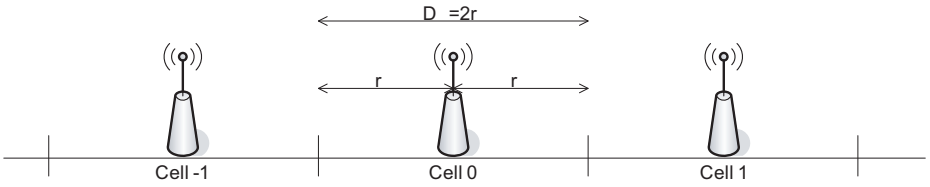


FIGURE 6.1: Infinite linear array cellular model [Park and Caire, 2008].

scenario but the results can be generalized to a downlink scenario as well. We assume perfect channel state information (CSI) on both transmitter and receiver sides. However, a user has no information of her channels to the base stations in the neighboring cells.

The fading environment of the multi-band multi-access system is same as for the single cell case. Each user k experiences a channel gain $g_k(\rho, j)[t]$ at cell j to cell ρ in slot t. For a multi-band system of M channels, short–term fading over the best channel is represented by,

$$f_k(t) = \max(f_k^{(1)}(t), f_k^{(2)}(t), \ldots, f_k^{(M)}(t)). \tag{6.1}$$

The channel gain is the product of path loss $s_k(\rho, j)$ and short–term fading $f_k(t)$, i.e. $g_k(\rho, j)[t] = s_k(\rho, j) f_k(t)$. The path loss is a function of the distance between the transmitter and the receiver and we assume it to not change within the time-scales considered in this work. In a multiple–cell system, path loss equals to $d_k(\rho, j)^{-\alpha}$ where depending on the location of user k in cell j, distance $d_k(\rho, j), \rho \neq j$ is given by

$$d_k(\rho, j) = |\rho - j| D \pm d_k(j, j) \tag{6.2}$$

Path gain and short–term fading are assumed to be independent. Short–term fading changes from slot to slot for every user and is independent and identically distributed across both users and slots but remains constant within each single transmission.

$E_k^R[t]$ and $E_k[t]$ represent the received and the transmitted energy for each user k such that

$$E_k^R[t] = g_k[t] E_k[t]. \tag{6.3}$$

Let N_0 denote the noise power spectral density. The scheduled users are separated by superposition coding. Let $\mathcal{K}_m$ be the set of users to be scheduled in frequency band m. Let $\psi_k^{(m)}$ be the permutation of the scheduled user indices for frequency band m that sorts the channel gains in increasing order, i.e. $g_{\psi_1}^{(m)} \leq \cdots \leq g_{\psi_k}^{(m)} \leq \cdots \leq g_{\psi_{|\mathcal{K}_m|}}^{(m)}$. Then, the energy of the user

$\psi_k^{(m)}$ with rate $R_{\psi_k}^{(m)}$, as scheduled by the scheduler to guarantee an error free communication, is given by [Park and Caire, 2008]

$$E_{\psi_k} = \frac{N_0 + I(\rho)}{g_{\psi_k}} \left[2^{\sum_{i \leq k} R_{\psi_i}} - 2^{\sum_{i<k} R_{\psi_i}} \right]. \tag{6.4}$$

where $I(\rho)$ denotes the outer cell interference represented as Gaussian noise on cell ρ and is given by

$$I(\rho) = \sum_{j \neq \rho} \sum_{k=1}^{K} g_k(\rho, j) E_k(j) \tag{6.5}$$

This energy assignment results in the minimum total transmit energy for the scheduled users.

6.2 Multicell Analysis

For the multicell system analysis, we use the large system results from [Park and Caire, 2008]. The average energy consumption of the system per transmitted information bit at the large system limit $K \to \infty$ is then given by

$$\left(\frac{E_\mathrm{b}}{N_0}\right)_{\mathrm{sys}}^{\mathrm{MC}} = \frac{\log(2) \int_0^\infty \frac{2^{C\mathrm{P}_{g,\mathrm{SVU}}(x)}}{x} \, \mathrm{d}P_{g,\mathrm{SVU}}(x)}{1 - \beta C \log(2) \int_0^\infty \frac{2^{C\mathrm{P}_{g,SVU}(x)}}{x} \, \mathrm{d}P_{g,\mathrm{SVU}}(x)} \tag{6.6}$$

where $\mathrm{P}_{g,\mathrm{SVU}}(\cdot)$ is the the cumulative distribution function (cdf) of the fading of the scheduled virtual users (SVU) and β is a constant commonly used to model the effect of interference in multicell analysis [Park and Caire, 2008] and bounded by

$$2D^{-\alpha}\zeta(\alpha, 1) \leq \beta \leq D^{-\alpha}\left(\zeta(\alpha, \frac{1}{2}) + \zeta(\alpha, \frac{3}{2})\right) \tag{6.7}$$

where $\zeta(.)$ denotes the standard zeta function. It has been observed that β does not change much with spectral efficiency and any β within the bound can be used effectively to model the inter–cell interference [Park and Caire, 2008]. We can rewrite average energy consumption of multicell case in Eq. (6.6) in terms of average energy consumption of single cell as

$$\left(\frac{E_\mathrm{b}}{N_0}\right)_{\mathrm{sys}}^{\mathrm{MC}} = \frac{\left(\frac{E_\mathrm{b}}{N_0}\right)_{\mathrm{sys}}^{\mathrm{SC}}}{1 - \beta C \left(\frac{E_\mathrm{b}}{N_0}\right)_{\mathrm{sys}}^{\mathrm{SC}}} \tag{6.8}$$

As compared to single cell case, performance of multicell case behaves differently at large spectral efficiency. Performance degradation in multicell case is dependent on the behaviour of multiple cell interference. Depending on on the multiple cell interference, we can distinguish the following operating regions [Park and Caire, 2008].

$$\begin{cases} \left(\frac{E_b}{N_0}\right)^{\text{SC}}_{\text{sys}} \leq \frac{1}{2\beta C} & \text{noise dominated region} \\ \frac{1}{2\beta C} < \left(\frac{E_b}{N_0}\right)^{\text{SC}}_{\text{sys}} < \frac{1}{\beta C} & \text{interference dominated region} \\ \left(\frac{E_b}{N_0}\right)^{\text{SC}}_{\text{sys}} \geq \frac{1}{\beta C} & \text{forbidden region} \end{cases} \tag{6.9}$$

Performance degradation as compared to single cell case is up to 3 dB in the noise dominated region. When interference dominates the noises, performance degrading is more than 3dB. Contrast to single cell cases, there is a certain limit on spectral efficiency C beyond which it is not allowed to operate and this region is termed as forbidden region.

6.2.1 Multiple-cell Wideband Slope

As our scheduling schemes are more suitable for operation at low spectral efficiencies, we investigate low spectral efficiency behaviour of multicell systems. It has been shown that systems with same $(E_b/N_0)_{\min}$ may have very different behaviours in the wideband regime and this behaviour can be expressed by evaluating *wideband slope* [Verdu, 2002]. We denote the spectral efficiency as a function of E_b/N_0 by $C(E_b/N_0)$. Then, derivative of C with respect to E_b/N_0 expressed in decibels, evaluated at $(E_b/N_0)_{\min}$ and normalized to 3dB is called a wideband slope and denoted by $\mathcal{S}_0$[Caire *et al.*, 2007]. The low spectral efficiency behaviour is characterized by the minimum system $(E_b/N_0)_{\min}$ and the wideband slope $\mathcal{S}_0$, such that [Verdu, 2002], [Caire *et al.*, 2007]

$$\left(\frac{E_b}{N_0}\right)_{\text{sys}}\Big|_{\text{dB}} = \left(\frac{E_b}{N_0}\right)_{\min}\Big|_{\text{dB}} + \frac{C}{\mathcal{S}_0} 10\log_{10}(2) + \mathcal{O}(C) \tag{6.10}$$

We evaluate $\mathcal{S}_0$ for the multiple cell case in this section. The reader is referred to [Caire *et al.*, 2007] for the detailed derivation of $\mathcal{S}_0$ for the single cell case where S_0 is derived in terms of $\frac{E_b}{N_0}(C)$.

For the multicell case, $\mathcal{S}_0$ is given by

$$\mathcal{S}_0 = \log 2 \frac{f(0)}{f'(0)} \tag{6.11}$$

where $f(C)$ denotes Eq. (6.6) as a function of spectral efficiency C. Evaluating Eq. (6.6) at $C \to 0$

$$f(0) = \log(2) \int_0^\infty \frac{\mathrm{dP}_{g,\mathrm{SVU}}(x)}{x} \tag{6.12}$$

where $\mathrm{P}_{g,\mathrm{SVU}}(x)$ is the cdf of channel gain distribution.

Differentiating Eq. (6.6) as a function of C yields

$$f'(C) = \log(2) \frac{\Big[\int_0^\infty 2^{C\mathrm{P}_{g,\mathrm{SVU}}(x)} \frac{\mathrm{P}_{g,\mathrm{SVU}}(x)}{x} \mathrm{dP}_{g,\mathrm{SVU}}(x) \Big(d(c)\Big) + \int_0^\infty \frac{2^{C\mathrm{P}_{g,\mathrm{SVU}}(x)} \mathrm{dP}_{g,\mathrm{SVU}}(x)}{x} d'(c)\Big]}{\Big[d(c)\Big]^2} \tag{6.13}$$

where the term $d(c)$ is given by

$$d(c) = 1 - \beta C \log(2) \int_0^\infty \frac{2^{C\mathrm{P}_{g,\mathrm{SVU}}(x)}}{x} \mathrm{dP}_{g,\mathrm{SVU}}(x) \tag{6.14}$$

Differentiating $d(c)$, we get

$$d'(c) = -\beta \log(2) \Big(C \frac{2^{C\mathrm{P}_{g,\mathrm{SVU}}(x)} \mathrm{P}_{g,\mathrm{SVU}}(x)}{x} \mathrm{dP}_{g,\mathrm{SVU}}(x) + \int_0^\infty \frac{2^{C\mathrm{P}_{g,\mathrm{SVU}}(x)}}{x} \mathrm{dP}_{g,\mathrm{SVU}}(x) \Big) \tag{6.15}$$

Now evaluating $f'(C)$ at $C \to 0$

$$f'(0) = \log(2) \Big[\int_0^\infty \frac{\mathrm{P}_{g,\mathrm{SVU}}(x)}{x} \mathrm{dP}_{g,\mathrm{SVU}}(x) + \beta \Big(\int_0^\infty \frac{\mathrm{dP}_{g,\mathrm{SVU}}(x)}{x} \Big)^2 \Big] \tag{6.16}$$

Using Eq. (6.12) and Eq. (6.16) in Eq. (6.11) wideband slope is given by

$$\mathcal{S}_0 = \frac{\int_0^\infty \frac{\mathrm{dP}_{g,\mathrm{SVU}}(x)}{x}}{\int_0^\infty \frac{P_{g,\mathrm{SVU}}(x)}{x} \mathrm{dP}_{g,\mathrm{SVU}}(x) + \beta \Big(\int_0^\infty \frac{\mathrm{dP}_{g,\mathrm{SVU}}(x)}{x} \Big)^2} \tag{6.17}$$

Wideband slope of multicell system directly gives us information about the performance loss as compared to a single cell case. A cell size dependent term in denominator results in decrease in slope of the multicell system which is an indicator that inter–cell interference reduces the energy efficiency of the system as compared to a single cell case in low spectral efficiency regime.

6.2.2 SDDPS: Multicell Case

In this section we evaluate performance of SDDPS in a multiple cell environment and compare with the results with the single cell case. Fig. 6.2 compares the delay–energy trade off for the single and multicell cases when all the arriving packets have identical deadlines. As the deadline of transmission increases, average system energy decreases for both the cases. The loss in system energy efficiency due to multicell interference is similar for different cell sizes.

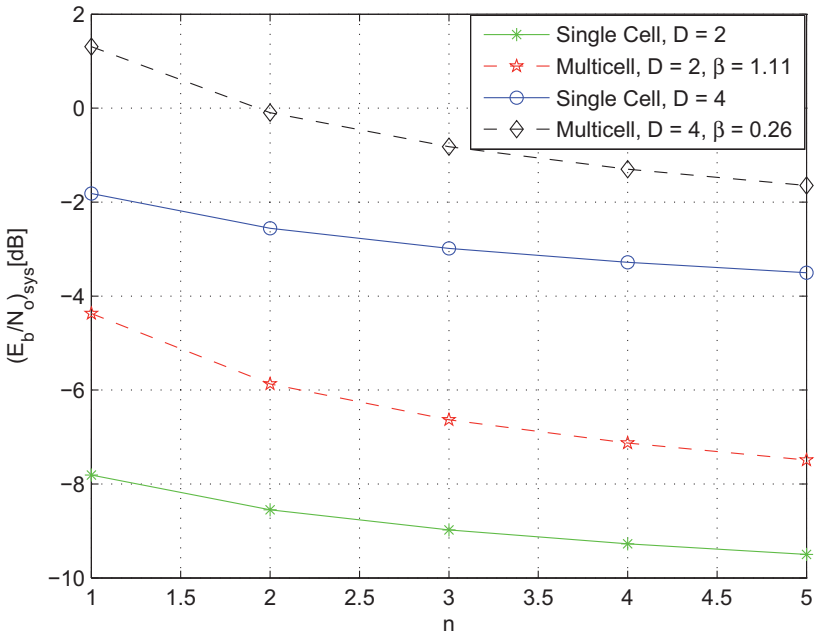


FIGURE 6.2: Delay–energy tradeoff comparison for the single and multiple cell cases at $C = 3$ bits/s/Hz for different cell size D where β is chosen as in Eq. (6.7).

Fig 6.3 shows the behaviour of SDDPS scheme in a multicell system for different spectral efficiencies. It is important to note that contrast to single cell system, multicell system operates in the region where single cell energy $(E_b/N_0)^{\text{SC}}_{\text{sys}} < \frac{1}{\beta C}$ as explained in Section 6.2. The region beyond that is termed as forbidden region for multicell system whereas there is no such limitation for a single cell system. For a delay limited system with $n = 1$ and $\beta = 1.1$, $C > 4.3$ is a forbidden region. However, for a delay tolerant system with $n = 2$, $(E_b/N_0)^{\text{SC}}_{\text{sys}}$ decreases and it allows system to operate

at higher spectral efficiencies. This effect is quite evident in Fig. 6.3 as the curve belonging to $n = 1$ saturates at lower spectral efficiency as compared to the curve with $n = 2$. Consequently, for a multicell system, scheduling with delay tolerance not only makes system more energy efficient, it allows system to operate at higher spectral efficiencies as compared to single cell case as well.

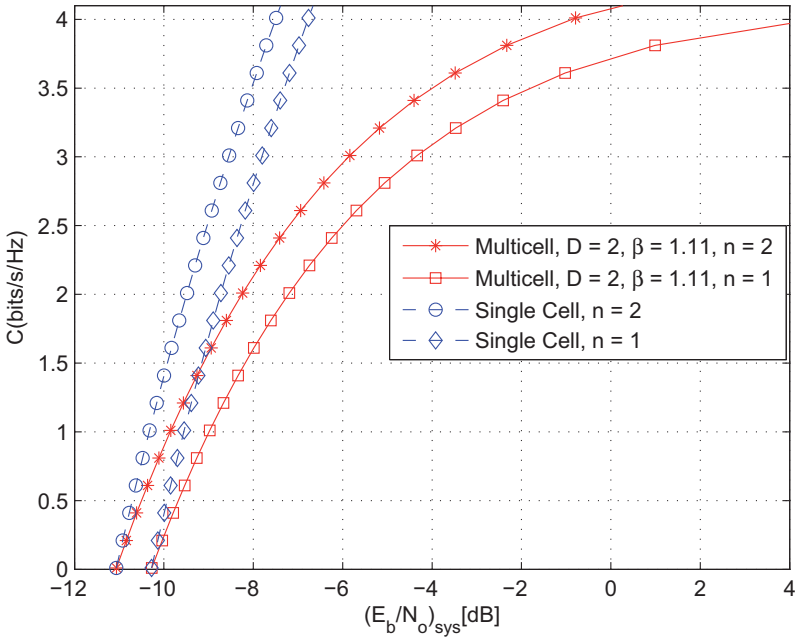


FIGURE 6.3: Comparison of system energy for single and multicell systems for SDDPS scheme with M=1.

Part II

Cooperative Communications

Chapter 7

Cooperative Communication in Deterministic Networks

This chapter deals with the second part of the dissertation. We consider a cooperative network where we have multiple sources, multiple relays and a single destination. The motivation behind the work is the need of cooperation among the sensor nodes in a typical WSN. In a typical WSN configuration, a large number of nodes measure the relevant data and transmit to a single node for further processing. In many application of the WSN, the amount of data to be transmitted at each node is very small but the physical distance between a sensing node and a common sink (destination) makes the transmission difficult or (energy wise) expensive. We investigate such a scenario where communicating nodes cooperate with each other to transport their own data along with the data from the other sensing nodes to reduce the cost of communication.

7.1 Review of Deterministic Network Model

In this section we briefly review the deterministic channel model proposed in [Avestimehr *et al.*, 2007a] and used in this work. The received signal at each node is a deterministic function of the transmitted signal. This model focuses on the signals interaction rather than on the channel noise. In a Gaussian (real) network, a single link from node i to node j with SNR $\mathrm{snr}_{i,j}$ has capacity $C_{i,j} = \frac{1}{2}\log(1+\mathrm{snr}_{i,j}) \approx \log\sqrt{\mathrm{snr}}_{i,j}$. Therefore, approximately, $n_{i,j} = \left\lceil \log\sqrt{\mathrm{snr}_{i,j}} \right\rceil$ bits per channel use can be sent reliably. In [Avestimehr *et al.*, 2007a] (see also references therein), the Gaussian channel is replaced by a finite-field deterministic model that reflects the above behav-

ior. Namely, the transmitted signal amplitude is represented through its binary[1] expansion $X = \sum_{\ell=1}^{\infty} B_\ell 2^{-\ell}$ where $B_\ell \in \mathbb{F}_2$. At the receiver, all the input bits such that $\sqrt{\text{snr}_{i,j}} 2^{-\ell} > 1$ (i.e., received "above the noise level") are perfectly decoded, while all those such that $\sqrt{\text{snr}_{i,j}} 2^{-\ell} \leq 1$ (i.e., received "below the noise level") are completely lost. It follows that only the most significant bits (MSBs) can be reliably decoded, such that the capacity of the deterministic channel is given exactly by $n_{i,j}$ and it is achieved by letting $B_1, \ldots, B_{n_{i,j}}$ i.i.d. Bernoulli-1/2.

A linear finite-field deterministic relay network is defined as a directed acyclic graph $\mathcal{G} = \{\mathcal{V}, \mathcal{E}\}$ such that the received signal at any node $j \in \mathcal{V}$ is given by

$$\mathbf{y}_j = \sum_{i \in \mathcal{V}:(i,j) \in \mathcal{E}} \mathbf{S}^{q-n_{i,j}} \mathbf{x}_i \tag{7.1}$$

where $\mathbf{y}_j, \mathbf{x}_i \in \mathbb{F}_2^q$, sum and products are defined over the vector space $\mathbb{F}_2^q$, and where

$$\mathbf{S} = \begin{bmatrix} 0 & 0 & 0 & \cdots & 0 \\ 1 & 0 & 0 & \cdots & 0 \\ 0 & 1 & 0 & \cdots & 0 \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ 0 & \cdots & 0 & 1 & 0 \end{bmatrix}$$

is a "down-shift" matrix. Notice that $n_{i,j} \leq q$ indicates the deterministic channel capacity for the link (i, j) as described before. Without loss of generality, the integer q can be set equal to the maximum of all $\{n_{i,j} : (i, j) \in \mathcal{E}\}$. The broadcast constraint is captured by the fact that the input $\mathbf{x}_i$ for each node i is common to all channels $(i, j) \in \mathcal{E}$.

In the case of single source (denoted by s) single destination (denoted by d), Theorem 4.3 of [Avestimehr *et al.*, 2007a] yields the capacity of linear finite-field deterministic relay networks in the form

$$C = \min_{(\mathcal{S}, \mathcal{S}^c) \in \Lambda_d} \text{rank}\{\mathbf{G}_{\mathcal{S},\mathcal{S}^c}\} \tag{7.2}$$

where Λ_d is the set of cuts $\mathcal{S} \subset \mathcal{V}$, $\mathcal{S}^c = \mathcal{V} - \mathcal{S}$ such that $s \in \mathcal{S}$ and $d \in \mathcal{S}^c$, and where $\mathbf{G}_{\mathcal{S},\mathcal{S}^c}$ is the transfer matrix for the cut $(\mathcal{S}, \mathcal{S}^c)$, formally defined as follows. Let $\mathcal{N}(i)$ denote the set of nodes j for which $(i, j) \in \mathcal{E}$ (this is the "fan-out" of node i) and let $\mathcal{P}(j)$ denote the set of nodes i for which

[1] The generalization to p-ary expansion is trivial. Here we focus on the binary expansion as in [Avestimehr *et al.*, 2007a].

$(i, j) \in \mathcal{E}$ (this is the "fan-in" of node j). The transfer matrix $\mathbf{G}_{\mathcal{S},\mathcal{S}^c}$ is defined as the matrix of the linear transformation between the transmitted vectors (channel inputs) of nodes $\beta_{\text{in}}(\mathcal{S})$ and the received vectors (channel outputs) of nodes $\beta_{\text{out}}(\mathcal{S})$, where the inner and outer boundaries $\beta_{\text{in}}(\mathcal{S})$ and $\beta_{\text{out}}(\mathcal{S})$ of $\mathcal{S}$ are defined as [Kramer, 2009]:

$$\beta_{in}(\mathcal{S}) = \{i \in \mathcal{S} : \mathcal{N}(i) \cap \mathcal{S}^c \neq \emptyset\}$$

and

$$\beta_{\text{out}}(\mathcal{S}) = \{j \in \mathcal{S}^c : \mathcal{P}(j) \cap \mathcal{S} \neq \emptyset\}$$

In words: $\beta_{\text{in}}(\mathcal{S})$ is the set of nodes of $\mathcal{S}$ with a direct link to nodes in $\mathcal{S}^c$, and $\beta_{\text{out}}(\mathcal{S})$ is the set of nodes in $\mathcal{S}^c$ with a direct link from nodes in $\mathcal{S}$.

Going through the proof of Theorem 4.3 in [Avestimehr *et al.*, 2007a] we notice that the "down-shift" structure for the individual channels is irrelevant. In fact, this structure is useful in making the connection between the linear finite-field model and the corresponding Gaussian case. As a matter of fact, if the channel matrices $\mathbf{S}^{q-n_{i,j}}$ in the above model are replaced by general matrices $\mathbf{S}_{i,j} \in \mathbb{F}_2^{q\times q}$, the result in Eq. (7.2) still holds.

7.2 Relay with one source

To motivate for the result in Section 7.4, we consider an example of a simple network and show achievability of the capacity region. We use a deterministic channel model. The signal received at the receiver is a deterministic function of the signal transmitted. This model helps us to concentrate on the interaction between the signals transmitted from different sources rather than the noise.

Fig. 7.1 shows a single relay network when a relay is acting as a data source as well. We focus here on the constraints due to the broadcast nature of the wireless network and interference of the signals coming from different nodes. The relay R injects its own data in addition to the data originated at the source S and *relays* the cumulative data to the common destination D. n_{SD}, n_{SR}, n_{RD} represent the number of bits successfully transmitted on the source-destination, source-relay and relay-destination links, respectively and termed as link capacities. The rate region constrained by the source rate R_S and relay rate R_R inequalities can be described as,

$$\begin{aligned} R_S &\leq \max(n_{SR}, n_{SD}) && (7.3)\\ R_R &\leq n_{RD} && (7.4)\\ R_S + R_R &\leq \max(n_{SD}, n_{RD}) && (7.5) \end{aligned}$$

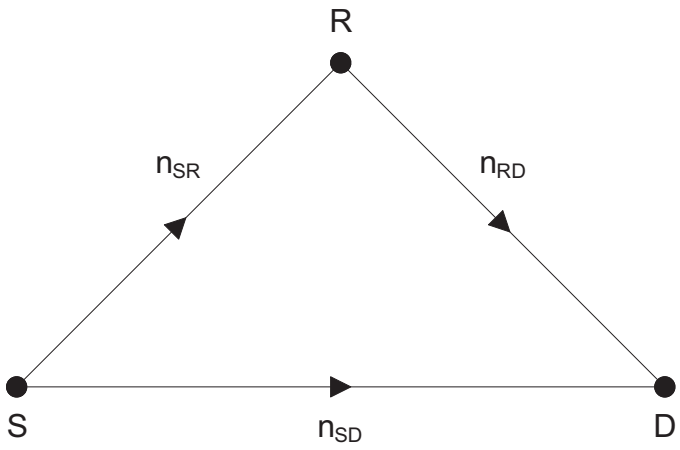


FIGURE 7.1: Single Source, single Relay and single Destination.

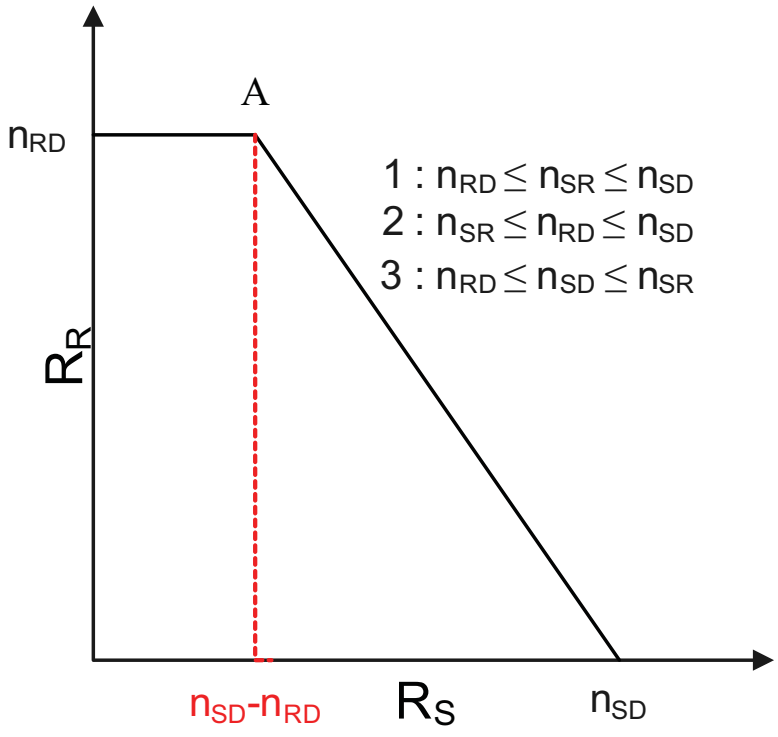


FIGURE 7.2: Capacity region for the hypotheses when $n_{SD} \geq n_{RD}$

.

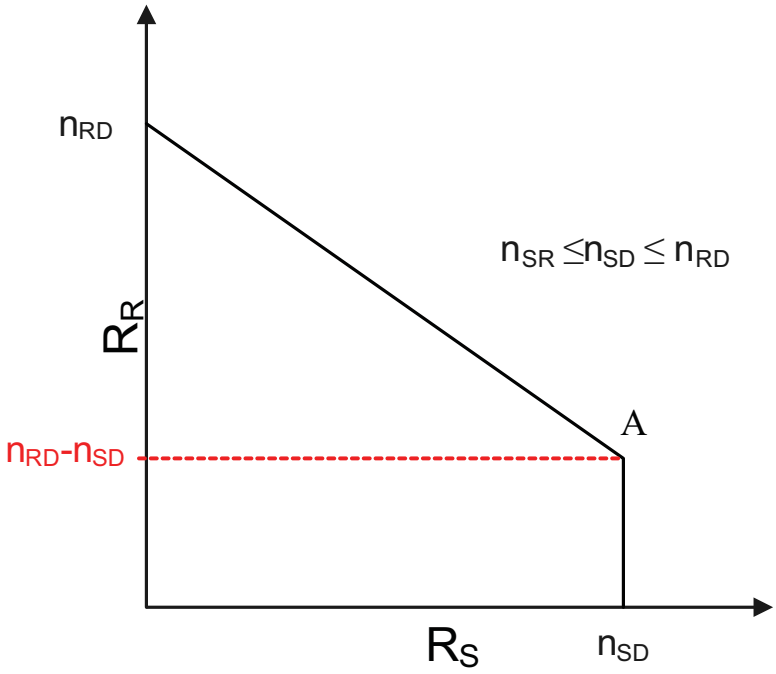


FIGURE 7.3: Capacity region for the hypothesis $n_{SD} \leq n_{SR} \leq n_{RD}$.

We prove that the capacity region bounded by these constraints can be achieved. To show achievability of capacity region $\mathcal{C}$, we show achievability for every possible hypothesis of *link capacity ordering*.

The capacity region can be achieved by the following scheme. The source transmits on the direct link SD only if $n_{SD} \geq \min(n_{SR}, n_{RD})$, otherwise it relays its data through the relay R. Depending on the link capacities n_{SD}, n_{RD}, n_{SR}, six different capacity regions can be defined and all of them are achievable with the given scheme as shown in Fig. 7.2–7.5.

The capacity region has been plotted for 3 possible hypotheses[2] in Fig. 7.2 when $n_{SD} \geq n_{RD}$. To achieve the corner vertex A, the source sends n_{SD} bits directly to destination. The source sends $n_{SD} - n_{RD}$ bits at MSB positions without interference from the relay while n_{RD} bits are shared between source and relay.

Fig. 7.3 represents the capacity region for the hypothesis $n_{SR} \leq n_{SD} \leq n_{RD}$. Here, we achieve the corner vertex A by sending n_{SD} bits on direct link. Relay sends $n_{RD} - n_{SD}$ bits without interference on MSB positions

[2]For three of the hypotheses, the resulting capacity region is identical, as shown in Fig. 7.2.

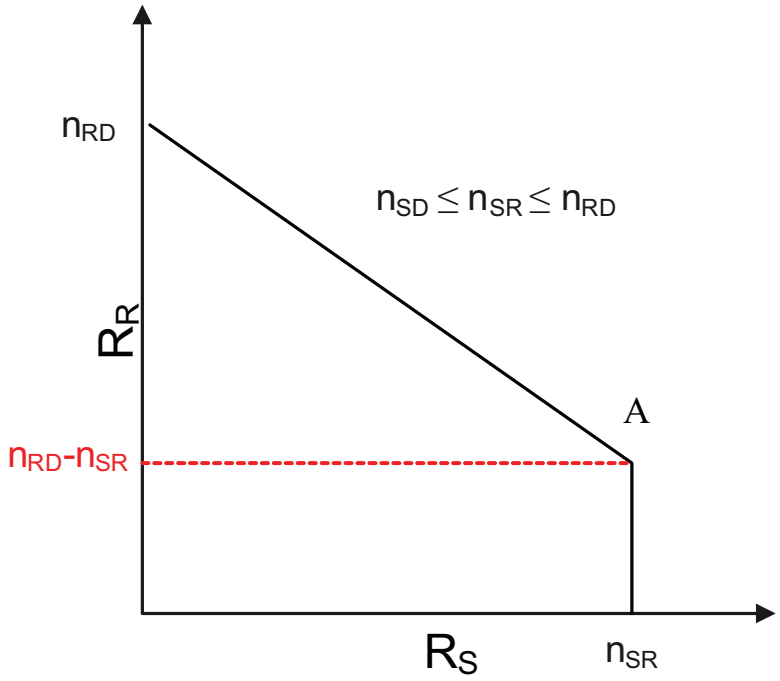


FIGURE 7.4: Capacity region for the hypothesis $n_{SR} \leq n_{SD} \leq n_{RD}$

while n_{SD} bits are shared between relay and the source.

Fig. 7.4 shows the capacity region for the hypothesis $n_{SD} \leq n_{SR} \leq n_{RD}$. The corner vertex A is achieved by letting source transmit n_{SR} bits to the relay. The relay decodes and forwards n_{SR} to the destination. The relay sends its own $n_{RD} - n_{SR}$ bits to the relay at MSB positions without interference from the source. However, n_{SR} bits are shared between the source and the relay.

Fig. 7.5 shows the capacity region for the hypothesis $n_{SD} \leq n_{RD} \leq n_{SR}$. This is the only hypothesis where *sharing* of rates R_S and R_R is the only possible solution and no data can be transmitted from either of the source or relay without interference from the other data source. the source sends n_{SR} bits to the relay. The relay decodes n_{SR} bits and after multiplexing its own $n_{RD} - n_{SR}$ bits, forwards them to the destination node.

7.3 A specific example: diamond network

In this section we work out an other example and provide an explicit achievability strategy. Consider the "diamond" network shown in Fig. 7.6, with

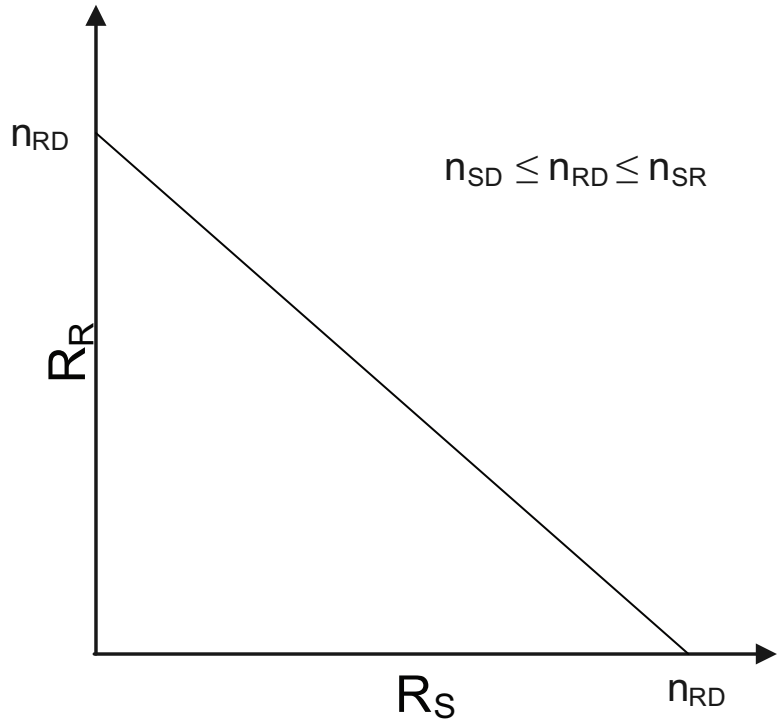


FIGURE 7.5: Capacity region for the hypothesis $n_{SD} \leq n_{RD} \leq n_{SR}$.

nodes $\{1, 2, 3, d\}$ and links of capacity $n_{1,2}, n_{1,3}, n_{2,d}$ and $n_{3,d}$. In this case, capacity region $\mathcal{C}$ is constrained by the following inequalities.

$$\begin{aligned} R_1 + R_2 + R_3 &\leq \max(n_{2,d}, n_{3,d}) && (7.6)\\ R_1 + R_2 &\leq n_{2,d} + n_{1,3} && (7.7)\\ R_1 + R_3 &\leq n_{3,d} + n_{1,2} && (7.8)\\ R_1 &\leq \max(n_{1,2}, n_{1,3}) && (7.9)\\ R_2 &\leq n_{2,d} && (7.10)\\ R_3 &\leq n_{3,d} && (7.11) \end{aligned}$$

Next, we provide simple coding strategies that achieve all relevant vertices of $\mathcal{C}$. Any point $\mathbf{R} \in \mathcal{C}$ can be obtained by suitable time-sharing of the vertices-achieving strategies. There are 24 possible orderings of the individual link capacities $n_{1,2}, n_{1,3}, n_{2,d}$ and $n_{3,d}$. Due to symmetry, the regions for the case $n_{3,d} > n_{2,d}$ will be the mirror image of the regions for the case $n_{2,d} \geq n_{3,d}$. Therefore, we shall consider only the cases where $n_{2,d} \geq n_{3,d}$.

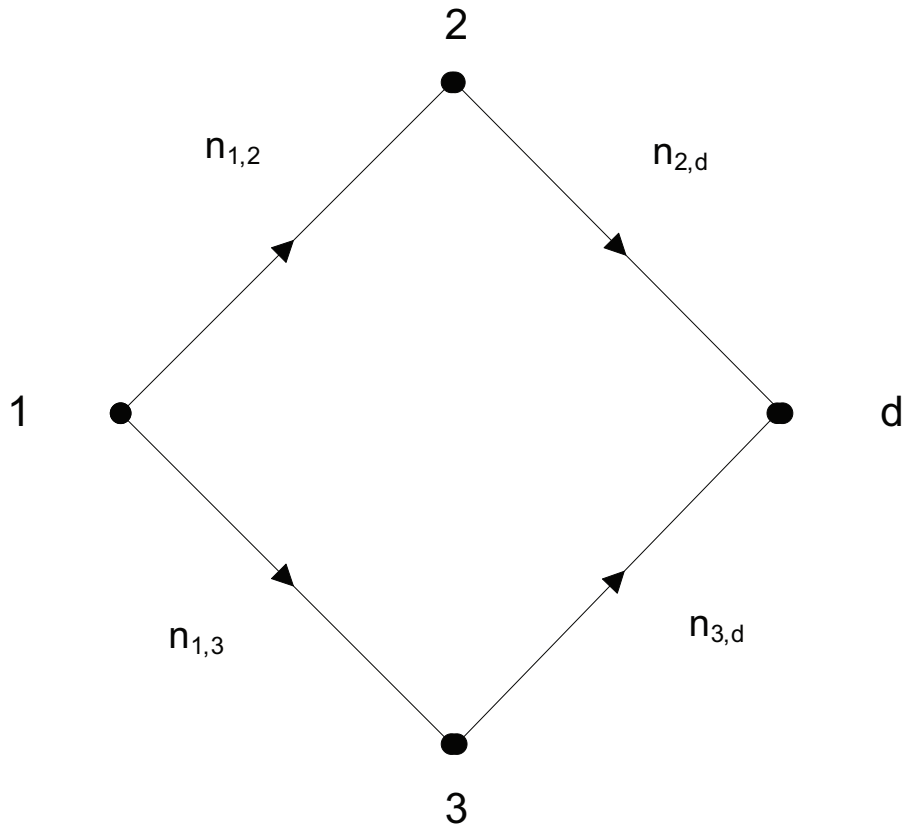


FIGURE 7.6: A diamond network with a source node 1, two relay nodes 2 and 3 and a common destination d.

The remaining 12 cases have to be discussed individually. For example, let's focus on the case $n_{3,d} \leq n_{1,2} \leq n_{1,3} \leq n_{2,d}$. An example of the network for the choice of the link capacities $n_{3,d} = 1, n_{1,2} = 2, n_{1,3} = 3, n_{2,d} = 4$ is given in Fig. 7.7. Fig. 7.8 shows qualitatively the shape of the capacity region in the three possible sub-cases of the link-capacity ordering $n_{3,d} \leq n_{1,2} \leq n_{1,3} \leq n_{2,d}$: case 1) for $n_{1,2} + n_{3,d} < n_{1,3}$; case 2) for $n_{1,2} + n_{3,d} \geq n_{1,3}$, and case 3) for $n_{1,2} + n_{3,d} \geq n_{2,d}$. In all cases, the achievability of the vertices B and C of the region of Fig. 7.8 is trivial, since these correspond to vertices of the multi-access channel with node 2 and 3 as transmitters and node d as receiver.

Case 1): Vertex A has coordinates $(R_1 = n_{1,2}, R_2 = n_{2,d} - n_{1,2} - n_{3,d}, R_3 = n_{3,d})$ and can be achieved by letting node 1 send $n_{1,2}$ to node 2. Node 2 decodes and forwards these bits after multiplexing its own $n_{2,d} - n_{1,2} - n_{3,d} > 0$ bits in the MSB positions, such that node 3 can send $n_{3,d}$ bits without inter-

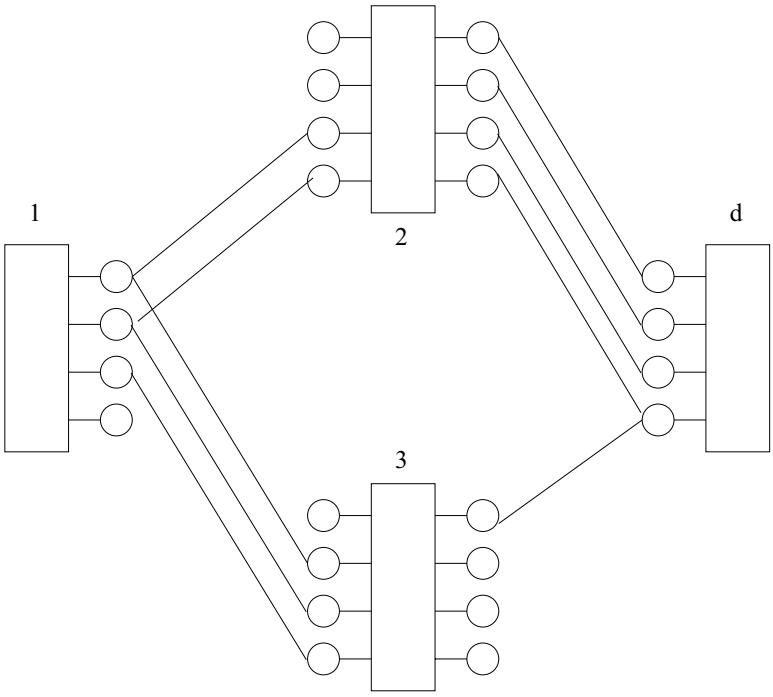


FIGURE 7.7: The configuration of the diamond network in the example (Case (1) in Fig.7.8)

.

ference from node 2. Vertex D has coordinates $(R_1 = n_{1,2} + n_{3,d}, R_2 = n_{2,d} - n_{1,2} - n_{3,d}, R_3 = 0)$ and can be achieved by letting node 1 send $n_{1,2} + n_{3,d}$ bits. These can be all decoded by node 3, then node 3 can forward the bottom (least-significant) $n_{3,d}$ bits of node 1 to node d. Node 2 decodes the top (most-significant) $n_{1,2}$ bits from node 1, and forwards them after multiplexing its own bits.

Case 2): Vertices A, D and E have coordinates $(R_1 = n_{1,2}, R_2 = n_{2,d} - n_{1,2} - n_{3,d}, R_3 = n_{3,d})$, $(R_1 = n_{1,3}, R_2 = n_{2,d} - n_{1,3}, R_3 = 0)$ and $(R_1 = n_{1,3}, R_2 = n_{2,d} - n_{1,2} - n_{3,d}, R_3 = n_{1,2} + n_{3,d} - n_{1,3})$, respectively. Vertex A can be achieved in the same way as in Case 1). Vertex D can be achieved by letting node 1 send $n_{1,3}$ bits to node 3. Node 3 decodes and forwards the bottom $n_{3,d}$. Since in this case $n_{1,2} \geq n_{1,3} - n_{3,d}$, node 2 can decode the top $n_{1,3} - n_{3,d}$ bits of node 1, and forwards them to node d after multiplexing its own $n_{2,d} - n_{1,3}$ bits, using its $n_{2,d} - n_{3,d}$ MSBs. Vertex E can be achieved by letting node 1 transmit $n_{1,3}$ bits, where the top $n_{1,2}$ of which are received by node 2. Node 3 forwards the bottom $n_{1,3} - n_{1,2}$ bits of node 1, and multiplex its own $n_{3,d} + n_{1,2} - n_{1,3}$ bits. Node 2 forwards the top $n_{1,2}$ bits from node

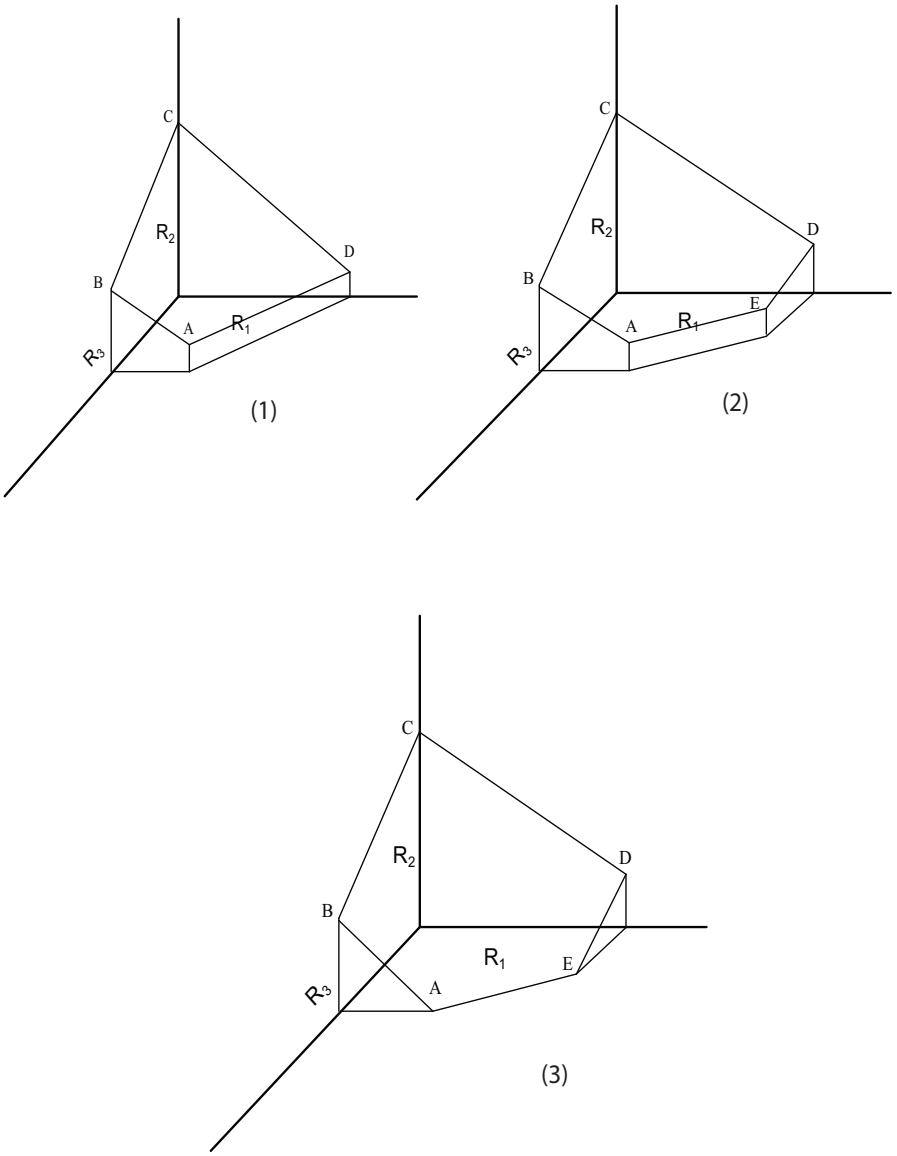


FIGURE 7.8: The capacity region of the diamond network for the hypothesis $n_{3,d} \leq n_{1,2} \leq n_{1,3} \leq n_{2,d}$.

1, by multiplexing its own $n_{2,d} - n_{1,2} - n_{3,d}$ bits, transmitting over its $n_{2,d} - n_{3,d}$ MSBs.

Case 3): Vertices A, D and E have coordinates $(R_1 = n_{2,d} - n_{3,d}, R_2 = 0, R_3 = n_{3,d})$, $(R_1 = n_{1,3}, R_2 = n_{2,d} - n_{1,3}, R_3 = 0)$ and $(R_1 = n_{1,3}, R_2 = 0, R_3 = n_{2,d} - n_{1,3})$, respectively. Vertex A can be achieved by letting node

1 send $n_{2,d} - n_{3,d}$ bits to node 2. Since $n_{2,d} - n_{3,d} \leq n_{1,2}$ these can be decoded and forwarded to node d in the MSB positions. Node 3 simply sends $n_{3,d}$ bits to node d without interfering with node 2. Vertex D is achieved by letting node 1 send $n_{1,3}$ bits. The top $n_{1,3} - n_{3,d}$ of these are decoded by node 2 and forwarded together with $n_{2,d} - n_{1,3}$ own bits. The bottom $n_{3,d}$ bits of node 1 are decoded and forwarded by node 3. Finally, vertex E is achieved by letting node 1 sent $n_{1,3}$ bits. The bottom $n_{3,d} - n_{2,d} + n_{1,3}$ of these are forwarded by node 3, after multiplexing its own $n_{2,d} - n_{1,3}$ bits. Since $n_{2,d} - n_{3,d} \leq n_{1,2}$, node 2 can decode the top $n_{2,d} - n_{3,d}$ bits from node 1 and forward them to node d using its MSB positions.

We consider another hypothesis $n_{3,d} \leq n_{1,2} \leq n_{2,d} \leq n_{1,3}$ and show achievability of the capacity region in Fig. 7.9. This hypothesis has two cases as well : case 1) for $n_{1,2} + n_{3,d} \geq n_{2,d}$ and case 2) for $n_{1,2} + n_{3,d} < n_{2,d}$.

Case 1): Vertices A and B have coordinates $(0, n_{2,d} - n_{3,d}, n_{3,d})$ and $(n_{2,d} - n_{3,d}, 0, n_{3d})$, respectively. Vertex A is achieved by letting node node 3 transmit $n_{3,d}$ bits to the destination at the MSB positions without interference from node 2. Node 2 transmits $n_{2,d} - n_{3,d}$ bits to the destination in the remaining positions. Similarly, vertex B is achieved by letting node 3 transmit $n_{3,d}$ bits to the destination without interference from node 2. Node 1 sends $n_{2,d} - n_{3,d}$ bits to node 2. Node 2 decodes $n_{2,d} - n_{3,d}$ bits and forwards them to the destination to the top $n_{2,d} - n_{3,d}$ positions.

Case 2): Vertices A, B and C have vertices $(0, n_{2,d} - n_{3,d}, n_{3,d})$, $(n_{1,2}, n_{2,d} - n_{1,2} - n_{3,d}, n_{3,d})$ and $(n_{1,2} + n_{3,d}, n_{2,d} - n_{1,2} - n_{3,d}, 0)$, respectively. Vertex A has same coordinates as in case 1) and is achieved in the same way. To achieve vertex B, node 1 sends $n_{1,2}$ bits to node 2. Node 3 sends $n_{3,d}$ bits to the destination at the MSB positions. As $n_{2,d} > n_{1,2} + n_{3,d}$, node 2 multiplexes its own $n_{2,d} - n_{1,2} - n_{3,d}$ with $n_{1,2}$ bits of node 1 and sends to the destination at the top $n_{2,d} - n_{3,d}$ positions. Vertex C is achieved by letting node 1 transmit top $n_{1,2}$ bits to node 1 and bottom $n_{3,d}$ bits to node 3. Node 3 decodes and forwards $n_{3,d}$ bits to the destination. Node 2 decodes $n_{1,2}$ bits from the source. As $n_{2,d} > n_{1,2} + n_{3,d}$, node 2 multiplexes its own $n_{2,d} - n_{1,2} - n_{3,d}$ bits and sends to the destination at the top positions.

Other hypotheses follow similarly and the whole capacity region is achieved by decode and forward.

7.4 Generalization: Main result

In a linear finite-field deterministic network defined as above, let $\mathcal{V} = \{1, \ldots, N, d\}$, where node d denotes the common destination and all other nodes $\{1, \ldots, N\}$ have independent information to send to node d. For any

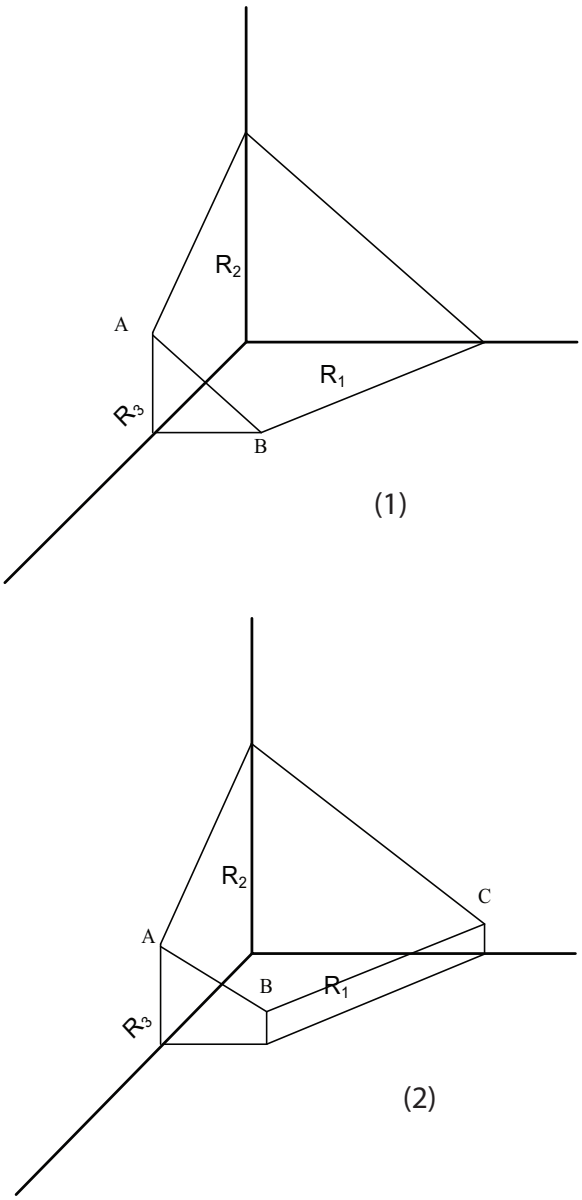


FIGURE 7.9: The capacity region of the diamond network for the hypothesis $n_{3,d} \leq n_{1,2} \leq n_{2,d} \leq n_{1,3}$.

integer $T_W = 1, 2, \ldots$ we let $\mathcal{W}_i = \{1, \ldots, \lceil 2^{T_W R_i} \rceil\}$ denote the message set of node $i = 1, \ldots, N$. A $(T_W, R_1, \ldots, R_N)$ code for the network is defined by a sequence of *strictly causal* encoding functions $f_i^{[t]} : \mathcal{W}_i \times \mathbb{F}_2^{q(t-1)} \to \mathbb{F}_2^q$,

for $t = 1, \ldots, T_W$ and $i = 1, \ldots, N$, such that the transmitted signal of node i at (discrete) time t is given by $\mathbf{x}_i[t] = f_i^{[t]}(w_i, \mathbf{y}_i[1], \ldots, \mathbf{y}_i[t-1])$, and by a decoding function $g : \mathbb{F}_2^{T_W q} \to \mathcal{W}_1 \times \cdots \mathcal{W}_N$, such that the set of decoded messages is given by $(\widehat{w}_1, \ldots, \widehat{w}_N) = g(\mathbf{y}_d[1], \ldots, \mathbf{y}_d[T])$.

The average probability of error for such code is defined as $P_n(e) = \mathbb{P}((W_1, \ldots, W_N) \neq (\widehat{W}_1, \ldots, \widehat{W}_N)$, where the random variables W_i are independent and uniformly distributed on the corresponding message sets $\mathcal{W}_i$. The rate N-tuple $(R_1, \ldots, R_N)$ is *achievable* if there exists a sequence of $(T_W, R_1, \ldots, R_N)$-codes with $P_n(e) \to 0$ as $T_W \to \infty$. The capacity region $\mathcal{C}$ of the network is the closure of the set of all achievable rates. With these definitions, we have:

Theorem 7.1 The capacity region $\mathcal{C}$ of a linear finite-field deterministic network $(\mathcal{V}, \mathcal{E})$ with independent information at the nodes $\{1, \ldots, N\}$ and a single destination d is given by

$$\sum_{i \in \mathcal{S}} R_i \leq \text{rank}\left\{\mathbf{G}_{\mathcal{S},\mathcal{S}^c}\right\}, \quad \forall\, \mathcal{S} \subseteq \{1, \ldots, N\}. \tag{7.12}$$

□

Proof: The proof of Theorem 7.1 is given in Appendix E. □

7.5 Transmissibility for correlated sources

Consider the case of a sensor network where the nodes $\{1, \ldots, N\}$ observe samples from a spatially-correlated, i.i.d. in time, discrete vector source $\mathbf{U} = (U_1, \ldots, U_N)$ (see the source model in [Barros and Servetto, 2006]). The goal is to reproduce the source blocks $\mathrm{u}[1], \ldots, \mathrm{u}[T_W]$ at the common destination node d. If the source blocks can be recovered at the destination with vanishing probability of error as $T_W \to \infty$, the vector source is said to be *transmissible*. In the case of a network of orthogonal links with capacities $C_{i,j}$, this problem was solved in [Barros and Servetto, 2006] and yields the necessary and sufficient transmissibility condition[3]

$$H(U_{\mathcal{S}}|U_{\mathcal{S}^c}) \leq \sum_{(i,j) \in \mathcal{S} \times \mathcal{S}^c} C_{i,j}, \quad \forall\, \mathcal{S} \subseteq \{1, \ldots, N\}. \tag{7.13}$$

From the system design viewpoint, the above result yields the optimality of the "separation" approach consisting of the concatenation of Slepian-Wolf coding for the source with routing and single-user channel coding for the network [Barros and Servetto, 2006].

[3]The notation $U_{\mathcal{S}} = \{U_i : i \in \mathcal{S}\}$ is standard.

With the same assumptions and linear finite-field deterministic network defined before, we consider a specific model for the vector source as defined in [Mohammad Ali Maddah-Ali and David N. C. Tse, 2009]. Let n_0 be a non-negative integer, and let $\mathbf{V} \in \mathbf{F}_2^{n_0}$ be a random vector of uniform i.i.d. bits. For all $i = 1, \ldots, N$, let $\mathcal{U}_i \subseteq \{1, \ldots, n_0\}$ and define $U_i \in \mathbb{F}_2^{|\mathcal{U}_i|}$ as the restriction of $\mathbf{V}$ to the components $\{V_\ell : \ell \in \mathcal{U}_i\}$ of $\mathbf{V}$. Then, the correlation model for the source $(U_1, \ldots, U_N)$ is reduced to the following "common bits" case: sources U_i and U_j have common part $\{V_\ell : \ell \in \mathcal{U}_i \cap \mathcal{U}_j\}$ while the bits V_ℓ in $\mathcal{U}_i - \mathcal{U}_j$ and in $\mathcal{U}_j - \mathcal{U}_i$ are mutually independent. It follows that $H(U_i|U_j) = |\mathcal{U}_i| - |\mathcal{U}_i \cap \mathcal{U}_j|$.

This source model is somehow "matched" to a correlated source defined over the reals in the following intuitive sense. Consider $N = 2$ and let U_1 and U_2 denote the binary quantization indices resulting from quantizing two correlated random variables $A_1 \in \mathbb{R}$ and $A_2 \in \mathbb{R}$ using "embedded" scalar uniform quantizers with n bits, such that their first m MSBs are identical and their last $n - m$ least significant bits (LSBs) are mutually independent. If A_1, A_2 are marginally uniform and symmetric, U_1 and U_2 are *exactly* obtained by defining $\mathbf{V}$ as above, with $n_0 = 2n - m$ independent bits, and letting U_1 include the m MSBs and the first set of $n - m$ LSBs of $\mathbf{V}$, and U_2 include the same m MSBs and the second set of $n - m$ LSBs of $\mathbf{V}$. This model trivially generalizes to the case of N correlated sources and is related to the Gaussian sources with "tree" dependency considered in [Mohammad Ali Maddah-Ali and David N. C. Tse, 2009]. For the source model defined above we have the following simple result[4]:

Theorem 7.2 The vector source $\mathbf{U} = (U_1, \ldots, U_N)$ is transmissible over the linear finite-field deterministic network $(\mathcal{V}, \mathcal{E})$ if and only if

$$H(U_{\mathcal{S}}|U_{\mathcal{S}^c}) \leq \text{rank}\left\{\mathbf{G}_{\mathcal{S},\mathcal{S}^c}\right\}, \quad \forall\, \mathcal{S} \subseteq \{1, \ldots, N\}. \tag{7.14}$$

□

Proof: The proof of Theorem 7.2 is given in Appendix E. □

[4]The results presented in Section 7.4 and Section 7.5 have been produced independent of the recent similar parallel work in [Perron, Diggavi, and Telatar, 2009] and [Mohajer, Tian, and Diggavi, 2010].

Chapter 8

Conclusions

This dissertation comprises of two parts. In the first part, we discussed opportunistic scheduling for large multiuser systems. Efficient use of system resources available in terms of spectral efficiency and transmit energy is the need of hour in modern communication networks. On the other hand, QoS constraints in terms of throughput, delay and loss tolerance requirements have a large variation depending on the application requirements. For example, multimedia applications are loss tolerant but have certain strict data and delay requirement. Many applications in WSNs do not have high requirements for throughput but they do not tolerate data loss and delay. Similarly, by operator point of view, considerations like fairness and minimal use of available energy resources is more important.

This work deals with energy efficient radio resource allocation for heterogenous network where all the applications have different types of constraints. Main task of the work is to develop scheduling algorithms that maximize information capacity in a fading environment for a hard deadline constrained system. Then, we accommodate other practically relevant features in our scheduler and analyze the effects on our scheduling scheme. The motivation behind the work is to model these effects such that core complexity of the scheduler does not increase and new tasks are accommodated in existing framework.

The second part of the dissertation addresses the information theoretic aspect of a cooperative relay network. We characterize the capacity region for some simple network topologies of wireless networks and prove the general results.

8.1 Main Contributions of the Work

8.1.1 Part I

Chapter 3: In chapter 3, we propose a scheduling scheme, called DDOS, which makes use of good channel and schedules a group of users having better channel than a transmission threshold. Our objective is to minimize transmit energy while each user has a deadline delay constrained for each packet. Each backlog state has a corresponding state dependent transmission threshold. It is an emptying buffer policy which implies that a scheduled user transmits all the packets buffered. The scheduling scheme is analyzed in large system limit. We provide a framework to adapt the proposed scheme with the objective of minimizing transit energy constrained by minimum throughput guarantee for each user. We use Simulated Annealing algorithm for optimization of transmission thresholds because energy function is not a convex function of transmission thresholds. Numerical results show the energy delay tradeoff exhibited by the scheme. This scheme is specially suited to applications like WSN where nodes enter into a sleep mode if they sense a bad channel. In this way, only sensing part of the circuitry is invoked in each time slot and rest of the parts consume no power. On sensing a good channel, the sensor nodes wake up and empty the buffer.

Chapter 4: Chapter 4 discusses a more generalized scheduler as compared to DDOS. We propose DDPS scheduling scheme which allows transmission of data in discrete steps. In DDOS, we have a single backlog state dependent transmission threshold. In DDPS, for each backlog state i, we have i buffered packets and corresponding i transmission thresholds. This certainly increase complexity of the scheme but numerical results provide evidence that it is more energy efficient as compared to DDOS. One of the main contribution of this chapter is proof of equivalence of channel distributions for random and constant arrivals in large system limit. This results allows us to use same analysis and transmission threshold for DDPS as for more idealistic case of constant arrivals.

Chapter 5: In this chapter, we analyze DDPS for more practically relevant scenarios. The main drawback of DDPS is its large computational complexity. We assume that transmission thresholds ending in the same backlog state can be coupled and call this scheme SDDPS. This assumption reduces complexity of DDPS from $\mathcal{O}(n)^2$ to $\mathcal{O}(n)$. Nu-

merical results show that there is negligible energy loss as a result of this assumption. For SDDPS, we propose a recursive algorithm to compute transmission thresholds which is much simpler than Simulated Annealing algorithm.

We model SDDPS scheme for the case of individual, non-identical packet deadlines. This case is practically very relevant as real systems can have such users who would like to transmit data with different deadlines. In literature, this problem is usually solved by assigning a different priority class to different deadline packets but we model this scenario without making such class differentiation. We model state transition matrix for non–identical case as a product of state transition matrix for identical case and a deadline offset matrix. In the large system limit, we prove that the packets with identical and non–identical deadlines have same transmission thresholds but resulting energy depends on the deadline distribution of the packets.

We extend our results to the case when we allow outage in packet transmission. In practice, it is costly (impossible sometimes) to provide deadline guarantee to all the users. Therefore, we allow a predefined proportion of packets to be dropped if they do not meet deadline. We model the system with packet outage probabilities. The transmission thresholds needs to be re–optimize for every deadline and dropping probability. We also consider the system when we allow packet dropping from the users at the outer edge of the cell at higher rate as compared to the users in the inner cell. Though, it results in unfair data loss but results in saving of transmit energy.

Chapter 6: This chapter addresses the effect of inter–cell interference in a multicell environment on SDDPS scheduler. In the large system limit, we show that energy efficiency reduces by a intercell interference limited term β as compared to a single cell system. Therefore, multicell systems cannot operate beyond a certain spectral efficiency. For delay tolerant systems, we show that reduction in system energy due to delay tolerance allows us to operate at higher spectral efficiencies as compared to delay limited systems.

8.1.2 Part II

Chapter 7: This chapter addresses a different concept as compared to rest of the dissertation. We consider a deterministic cooperative network where relay nodes relay the data from the source but multiplex their own data to a common destination. We characterize capacity region

of single source, single relay, single destination and single source, two relays, single destination networks, and provide achievability results.

We have generalized our results and characterized the capacity region for a linear finite-field deterministic network with independent information at all nodes and a single destination node. In our setup, all nodes may relay information from other nodes as well as inject their own information into the network. This may serve as a simplified model for a large WSN where sensing nodes cooperate with each other to send the collective data towards a single collector node. For a specific model of discrete binary source correlation at the nodes, we have also found necessary and sufficient conditions for the source transmissibility. Albeit restrictive, this correlation model may be useful (e.g., see [Mohammad Ali Maddah-Ali and David N. C. Tse, 2009]) as a simple discrete "equivalent" (up to some bounded mean-square distortion penalty) for a spatially-correlated real sources whose components are observed and encoded separately at the network nodes.

8.2 Further Research Directions

In this section we discuss some of the problems demanding further research.

- Throughout this work, we have assumed perfect channel state information (CSI) on transmitter and receiver sides. Practically, it is very hard to have perfect CSI on transmitter side in a fast fading environment. It would be challenging to investigate the proposed schedulers when perfect CSI is not available on transmitter side. Reference [Bertsekas, 2007] provides a few techniques to reduce a problem with imperfect CSI to a problem with perfect CSI using dynamic programming approach. In general, the solution gets more complicated because perfect CSI solution prompts us to specify a rule for transmission for each state i at time t while solution with imperfect CSI requires us to compute a vector of controls applied for every sequence of observations received and controls applied by time t.

- We have characterized energy-delay tradeoff in this work. The results can easily be extended to throughput-delay tradeoff.

- Our objective in this work is to minimize transmit energy for a given deadline. A trivial extension is to maximize system throughput while maintaining a minimum user throughput. A dual of this problem is to

minimize system transmit energy for a given minimum user throughput. We have considered this topic briefly in Section 3.5 for large system limit. However, this problem gets complicated in real scenario and providing strict guarantee without outage is impossible. Work in [Min, Kim, Woo, and Kim, 2005], [Zorba, Pérez-Neira, Foglar, and Verikoukis, 2009], [Chen and Jordan, 2009], [Zorba and Verikoukis, 2010], [Pitic and Capone, 2008] discuss the problem in different settings and propose some scheduling schemes but none of them provides any guarantee. Up to best of our knowledge, characterization of a scheduler which maximizes system throughput for a given outage probability (in terms of number of slots, a user's throughput falls below minimum throughput) is still an open problem.

- We have assumed uncorrelated channels from slot to slot. An interesting extension is the characterization of the schemes when channels are correlated.

- In the multicell case, we assumed that a user has information of her channel to the base station in her own cell but no information to the channels in the neighbouring cells. This assumption helps us to model interference from other cells as Gaussian noise. It is a challenging problem to investigate the problem when a user has information of her channels to other base stations as well. In this case, it would not be straight forward to schedule a user who has good channel to her base station. If she has good channel to other base stations as well, it means she will cause a lot of intercell interference. Minimum energy solution requires that a user is scheduled when she has a good channel to her base station and bad channels to neighboring base stations. As a result, task of scheduling becomes complicated.

- We mainly have discussed problem for a single hop system. The work can be extended to multihop systems where end to end delay depends on multiple transmission and single deadline. It will be interesting to investigate how energy-delay tradeoff can be characterized in this scenario.

Part III

Appendices

Appendix A

Channel Characteristics:

In this work the channel model of [Caire *et al.*, 2007] is used. Signal propagation is characterized by a distance dependent path loss factor and a frequency-selective short-term fading that depend on the scattering environment around the user terminal. As described in Section 3.1, these two effects are taken into account by letting $g_k^m = s_k f_k^m$ where s denotes the path loss of user k and f_k^m is the short term fading of user k in channel m.

As in [Caire *et al.*, 2007], we assume that users are uniformly distributed in a geographical area but for a forbidden circular region of radius δ centered around the base station where $0 < \delta \leq 1$ is a fixed system constant. Using this model, the cdf of path loss is given by

$$\mathrm{F}_s(x) = \begin{cases} 0 & x < 1 \\ 1 - \frac{x^{-2/\alpha} - \delta^2}{1-\delta^2} & 1 \leq x < \delta^{-\alpha} \\ 1 & x \geq \delta^{-\alpha} \end{cases} . \tag{A.1}$$

where the path loss at the cell border is normalized to one.

Frequency selective short–term block fading is modeled by M parallel channel which are i.i.d. For a Rayleigh channel, the distribution of $\max\{f_k^1, \ldots, f_k^M\}$ is given by

$$\mathrm{P}_{\max\{f\}}(y) = (1 - \exp(-y))^M \tag{A.2}$$

$\mathrm{P}_{\max\{g\}}(x)$ is defined as the cdf of the random variable $\max\{g_k^1, \ldots, g_k^M\} = s_k \max\{f_k^1, \ldots, f_k^M\}$. Recall from Eq. (4.7), the cdf $\mathrm{P}_{f,\mathrm{SVU}}(y)$ of SVUs for DDPS is a weighted function of the cdf of actual fading $\mathrm{P}_{\max\{f\}}(y)$ given by Eq. (A.2). Using Eq. (4.7) and Eq. (A.1), we compute a convenient expression for the cdf $\mathrm{P}_{g,\mathrm{SVU}}(x)$ of the SVUs for this product channel. As

path loss and Rayleigh fading occur simultaneously and independently, the cdf of the channel gain is given by

$$\mathrm{P}_{g,\mathrm{SVU}}(x) = \int F_s(x/y)\mathrm{dP}_{f,\mathrm{SVU}}(y). \tag{A.3}$$

A.1 Case I: DDPS

Using Property 4.4 and the path loss distribution in Eq. (A.1), Eq. (A.3) is computed as follows

$$\begin{aligned}\mathrm{P}_{g,\mathrm{SVU}}(x) &= \int_0^{x\delta^\alpha} \mathrm{p}_{f,\mathrm{SVU}}(y)\mathrm{d}y + \int_{x\delta^\alpha}^{x} F_s(x/y)\mathrm{dP}_{f,\mathrm{SVU}}(y) &\text{(A.4)}\\ &= \mathrm{P}_{f,\mathrm{SVU}}(x\delta^\alpha) + \int_{x\delta^\alpha}^{x}\Big(1-\frac{(y/x)^{2/\alpha}-\delta^2}{1-\delta^2}\Big)\mathrm{dP}_{f,\mathrm{SVU}}(y) &\text{(A.5)}\end{aligned}$$

Changing variables and integrating by parts yields,

$$\mathrm{P}_{g,\mathrm{SVU}}(x) = \frac{1}{x^{2/\alpha}(1-\delta^2)}\int_{x^{2/\alpha}\delta^2}^{x^{2/\alpha}} \mathrm{P}_{f,\mathrm{SVU}}(y^{\alpha/2})\mathrm{d}y. \tag{A.6}$$

For $\alpha = 2$, Eq. (A.6) can be written in closed form.

For DDPS scheduler, using Eq. (4.7) and the Rayleigh fading model, Eq. (A.6) is given by

$$\begin{aligned}\mathrm{P}_{g,\mathrm{SVU}}(x) &= \frac{1}{x(1-\delta^2)}\int_{x\delta^2}^{x}\sum_{i=1}^{n} c_i\pi_i\Big[(i-j(y,i)+1)(1-\exp(-y))^M\\ &- \sum_{b=0}^{i-j(y,i)}(1-\exp(-\kappa_{i\to i-b}))^M\Big]\mathrm{d}y. &\text{(A.7)}\end{aligned}$$

Using geometric series expansion, the closed form expression is given by

$$\begin{aligned}\mathrm{P}_{g,\mathrm{SVU}}(x) &= \sum_{i=1}^{n} c_i\pi_i\Big[(i-j(y,i)+1)\Big(1-\frac{1}{x(1-\delta^2)}\sum_{m=1}^{M}\frac{1}{m}\big[(1-\exp(-x))^m\\ &- \big(1-\exp(-x\delta^2)\big)^m\big]\Big) - \sum_{b=0}^{i-j(y,i)}\big(1-\exp(-\kappa_{i\to i-b})\big)^M\Big]. &\text{(A.8)}\end{aligned}$$

where $j(y,i)$ is defined by Definition 4.1.

A.2 Case II: DDOS

For DDOS scheduler, using Eq. (3.20), $\alpha = 2$ and the Rayleigh fading model, Eq. (A.6) is given by

$$\begin{aligned} \mathrm{P}_{g,\mathrm{SVU}}(x) &= \frac{1}{x(1-\delta^2)} \int_{x\delta^2}^{x} \sum_{i=1}^{n} c_i \pi_i \Big[i(1-\exp(-y))^M \\ &- (1-\exp(-\kappa_i))^M \Big] \mathrm{dy}. \end{aligned} \tag{A.9}$$

Again, using geometric series expansion, the closed form expression is given by

$$\begin{aligned} \mathrm{P}_{g,\mathrm{SVU}}(x) &= \sum_{i=1}^{n} c_i \pi_i i \Big[\Big(1 - \frac{1}{x(1-\delta^2)} \sum_{m=1}^{M} \frac{1}{m} \big[(1-\exp(-x))^m \\ &- \big(1-\exp(-x\delta^2)\big)^m \big] \Big) - \big(1-\exp(-\kappa_i)\big)^M \Big]. \end{aligned} \tag{A.10}$$

Appendix B

Proof of Properties For DDPS Scheduling Scheme

B.1 Proof of Property 4.2

Proof follows directly from Proposition 4.1.

The system energy $E_{\text{sys}}(\vec{R}, \vec{y})$ is a function of a allocated rate vector $\vec{R}$ and fading vector $\vec{y}$. We denote the required rate for transmission of a single packet by $R' = \frac{\Gamma}{K}$. To schedule a packet for transmission, we have a discrete (quantized) vector of transmission thresholds $\vec{\kappa}_i$ over the fading state y. From Definition 4.1, it is clear that $\vec{\kappa}$ should necessarily be ordered (increasing or decreasing). Proposition 4.1 states that more rate should be allocated at good channels. Thus, given a backlog state i, $\vec{\kappa}_i$ should be ordered in such a way that the lowest assigned rate $R = R'$ and the highest rate $R_i = iR'$ correspond to the lowest threshold $\kappa_{i\to i}$ and the highest threshold $\kappa_{i\to 1}$, respectively.

Following Proposition 4.1, for the buffer state i, we can order the vector $\vec{\kappa}_i$ only in increasing order such that

$$\vec{\kappa}_i = [\kappa_{i\to i}, \kappa_{i\to i-1} \dots \kappa_{i\to 1}] \tag{B.1}$$

which implies

$$\kappa_{i\to j} \leq \kappa_{i\to j-1} \forall i, j \tag{B.2}$$

and proves Property 4.2

B.2 Proof of Property 4.3

The allocated rate R is equal in both of the state transitions $T_{i\to j}$ and $T_{i-1\to j-1}$. However, in state i, the user is closer to the deadline as compared to state

$i-1$. The user's decision to transmit is analogous to dynamic programming formulation where data is transmitted if cost of transmission in state i is less than the expected future cost. Equivalently, the cost of transmission in a state $i-1$ can be represented as

$$E_{i-1} = \min\Big(E(R, y_{i-1}), \mathbb{E}\big[E_i\big]\Big) \tag{B.3}$$

where $\mathbb{E}[E_i]$ is the expected future cost in state i. The future expected cost of transmission is only the function of quantized fading vector ($\vec{\kappa}$). The problem belongs to a class of monotone optimal stopping problems where it has been shown that if the *one step look ahead* optimal stopping rule prompts to wait, then it is optimal to wait. In a monotone optimal stopping problem, there is always a certain ordering (decreasing or increasing) of costs (energy) in every state i. As future expected cost of transmission is a function of quantized fading vector (transmission thresholds), to prove $\kappa_{i\to j} \leq \kappa_{i-1\to j-1}$, it is sufficient to prove that vector of future expected costs $\vec{C}$ is ordered in increasing order.

$$\vec{C} = \Big(\mathbb{E}[E_1] \leq \mathbb{E}[E_2] \leq \cdots \leq \mathbb{E}[E_i] \leq \cdots \leq \mathbb{E}[E_n]\Big) \tag{B.4}$$

We need to prove,

$$\mathbb{E}[E_{i-1}] \leq \mathbb{E}[E_i] \quad \forall i \tag{B.5}$$

By monotonicity property of dynamic programming, it is sufficient to show that

$$\mathbb{E}[E_{n-1}] \leq \mathbb{E}[E_n] \tag{B.6}$$

By Property 4.4, $\kappa_{n\to n} = 0$. In a dynamic programming problem, this condition is represented by allowing an infinity energy expenditure for a future termination state $n+1$. For the state n, Eq. (B.3) can be written as,

$$E_n = \min\Big(E(R, y_n), \infty\Big) = E(R, y_n) \tag{B.7}$$

In state $n-1$, Eq. B.3 can be written as

$$E_{n-1} = \min\Big(E(R, y_{n-1}), \mathbb{E}\big[E_n\big]\Big) \tag{B.8}$$

However, it always holds that

$$\mathbb{E}[E_n] \quad \geq \quad \min\big(E(R, y_{n-1}), \mathbb{E}[E_n]\big) \tag{B.9}$$

Taking expectation on both sides and using Eq. (B.8)

$$\mathbb{E}\big[\mathbb{E}[E_n]\big] \geq \mathbb{E}\Big[\min\big(E(R, y_{n-1}), \mathbb{E}[E_n]\big)\Big] \tag{B.10}$$

$$\mathbb{E}[E_n] \geq \mathbb{E}[E_{n-1}] \tag{B.11}$$

which confirms that $\vec{C}$ has been ordered in an increasing order of future cost. Equivalently, transmission threshold vector $\vec{\kappa}$ is ordered in decreasing order and therefore $\kappa_{i\to j} \leq \kappa_{i-1\to j-1}$.

Appendix C

Relationship between Transmission Thresholds and Transition Probabilities

For a given deadline delay and short–term fading distribution $\mathrm{P}_{\max\{f\}}(y)$, the transmission thresholds are a function of the transition probabilities. Therefore, the procedure of computing a set of optimal thresholds is equivalent to the computation of a set of optimal transition probabilities. For a deadline of n time slots, the transition probability matrix for DDPS, $\mathbf{P}_{\mathrm{DDPS}}$ in Eq. (3.6) is expressed as,

$$\mathbf{P}_{\mathrm{DDPS}} = \begin{pmatrix} \Pr(y \geq \kappa_{1\to 1}) & \Pr(y < \kappa_{1\to 1}) & \cdots & 0 \\ \Pr(y \geq \kappa_{2\to 1}) & \Pr(\kappa_{2\to 2} \leq y < \kappa_{2\to 1}) & \cdots & 0 \\ \cdots & \cdots & \cdots & \cdots \\ \Pr(y \geq \kappa_{n\to 1}) & \Pr(\kappa_{n\to 2} \leq y < \kappa_{n\to 1}) & \cdots & \Pr(y < \kappa_{n\to n-1}) \end{pmatrix}.$$

Corresponding state transition matrix for DDOS, $\mathbf{P}_{\mathrm{DDOS}}$ is given by

$$\mathbf{P}_{\mathrm{DDOS}} = \begin{pmatrix} \Pr(y \geq \kappa_1) & \Pr(y < \kappa_1) & 0 & \cdots & 0 \\ \Pr(y \geq \kappa_2) & 0 & \Pr(y < \kappa_2) & \cdots & 0 \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ \Pr(y \geq \kappa_n) & 0 & 0 & \cdots & \Pr(y < \kappa_n) \end{pmatrix}.$$

where zero transition probability represents the impossible transition.

For Rayleigh fading the transition probability matrix $\mathbf{P}$ can be written in terms of transmission thresholds as,

$$\mathbf{P}_{\text{DDPS}} = \begin{pmatrix} 1-(1-e^{-\kappa_{1\to1}})^M & (1-e^{-\kappa_{1\to1}})^M & \dots & 0 \\ 1-(1-e^{-\kappa_{2\to1}})^M & (1-e^{-\kappa_{2\to1}})^M-(1-e^{-\kappa_{2\to2}})^M & \cdots & 0 \\ \cdots & \cdots & \cdots & \cdots \\ 1-(1-e^{-\kappa_{n\to1}})^M & (1-e^{-\kappa_{n\to1}})^M-(1-e^{-\kappa_{n\to2}})^M & \cdots & (1-e^{-\kappa_{n\to n-1}})^M \end{pmatrix} \quad \text{(C.1)}$$

$$\mathbf{P}_{\text{DDOS}} = \begin{pmatrix} 1-(1-e^{-\kappa_1})^M & (1-e^{-\kappa_{1\to1}})^M & \dots & 0 \\ 1-(1-e^{-\kappa_2})^M & 0 & \cdots & 0 \\ \cdots & \cdots & \cdots & \cdots \\ 1-(1-e^{-\kappa_n})^M & 0 & \cdots & (1-e^{-\kappa_n})^M \end{pmatrix} \quad \text{(C.2)}$$

It is convenient to vary the configuration in the SA algorithm by using transition probabilities as compared to varying transmission thresholds because the range of variation (from 0 to 1) is known. Using these relations, corresponding optimal thresholds can always be computed.

Appendix D

Equivalence of Channel Distribution of SVUs For SDDPS: Non–identical Deadline Case

We prove equivalence of $\mathrm{p}_{f,SVU}(y|i)$ for two two descriptions represented in Eq. (5.29) and Eq. (5.31) for individual non–identical deadline case. We evaluate equation Eq. (5.31) and prove equals to Eq. (5.29).

$$\begin{aligned}
\mathrm{p}_{f,SVU}(y|i) &= c_i^{nid}\mathrm{E}(\mu_t)\sum_{r=i}^{j}\sum_{l=r}^{n} p_l\,\mathrm{p}_{\max\{f\}}(y) && \text{(D.1)}\\
&= c_i^{nid}\mathrm{E}(\mu_t)\Big[\sum_{l=i}^{n} p_l + \sum_{l=i+1}^{n} p_l + \sum_{l=i+2}^{n} p_l + \cdots \sum_{l=j}^{n} p_l\Big]\,\mathrm{p}_{\max\{f\}}(y)\\
&= c_i^{nid}\mathrm{E}(\mu_t)\Big[p_i + 2\sum_{l=i+1}^{n} p_l + \sum_{l=i+2}^{n} p_l + \cdots \sum_{l=j}^{n} p_l\Big]\,\mathrm{p}_{\max\{f\}}(y)\\
&= c_i^{nid}\mathrm{E}(\mu_t)\Big[p_i + 2p_{i+1} + 3\sum_{l=i+2}^{n} p_l + \cdots \sum_{l=j}^{n} p_l\Big]\,\mathrm{p}_{\max\{f\}}(y)\\
&= c_i^{nid}\mathrm{E}(\mu_t)\Big[p_i + \cdots + (j-i)p_{j-1} + (j-i+1)\sum_{l=j}^{n} p_l\Big]\,\mathrm{p}_{\max\{f\}}(y)\\
&= c_i^{nid}\mathrm{E}(\mu_t)\Big[\sum_{r=i}^{j-1}(r-i+1)p_r + (j-i+1)\sum_{l=j}^{n} p_l\Big]\,\mathrm{p}_{\max\{f\}}(y) && \text{(D.2)}
\end{aligned}$$

Appendix E

Proof of Capacity Region in a Cooperative Relay Network

E.1 Proof of Theorem 7.1

The converse of Eq. (7.12) follows directly from the general cut-set bound and by the fact that, for the linear deterministic network model, uniform i.i.d. inputs maximize all cut-set values at once [Cover and Thomas, 2006; Avestimehr *et al.*, 2007a; Kramer, 2009].

For the direct part, we build an augmented network by introducing a virtual source node 0 and by expanding the channel output alphabet of each node $i = \{1, \ldots, N\}$. Let $\{n_{0,i} : i = 1, \ldots, N\}$ be arbitrary non-negative integers. The channel output alphabet of node i in the augmented network is given by $\mathbb{F}_2^{q+n_{0,i}}$. The virtual source node 0 has $n_0 = \sum_{i=1}^{N} n_{0,i}$ input bits, partitioned into N disjoint sets $\mathcal{U}_i$ of cardinality $n_{0,i}$ for $i = 1, \ldots, N$, respectively, such that the bits of subset $\mathcal{U}_i$ are sent directly to node i and are received at the top $n_{0,i}$ MSB positions of the expanded channel output alphabet. Fig. E.1 shows an example of such network augmentation for a "diamond" network [Avestimehr *et al.*, 2007a].

After introducing the virtual source node, the augmented linear finite-field deterministic network belongs to the class studied in [Avestimehr *et al.*, 2007a] with the minor difference that the channel linear transformations are not necessarily limited to "down-shifts". Nevertheless, as we observed before, Theorem 4.3 of [Avestimehr *et al.*, 2007a] still applies. Letting R_0 denote the rate from the virtual source node 0 to the destination node d, we have that all rates R_0 satisfying

$$R_0 \leq \min_{(\Omega_0, \Omega_0^c) \in \Lambda_d} \operatorname{rank} \left\{ \mathbf{G}_{\Omega_0, \Omega_0^c} \right\} \tag{E.1}$$

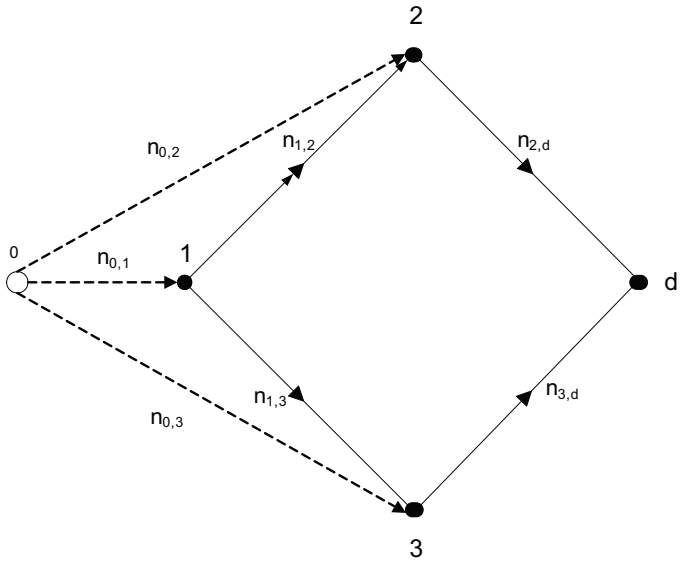


FIGURE E.1: A diamond network with a source node 1, two relay nodes 2 and 3 and a common destination d is augmented by adding node 0 and virtual links to nodes 1, 2 and 3.

are achievable, where Λ_d is the set of all cuts (Ω_0, Ω_0^c) of the augmented network such that $0 \in \Omega_0$ and $d \in \Omega_0^c$.

For any such set Ω_0 we have that $\Omega_0 = \mathcal{S} \cup \{0\}$, for some $\mathcal{S} \subseteq \{1, \ldots, N\}$. Consequently, we have that $\Omega_0^c = \mathcal{S}^c$, where $\mathcal{S}, \mathcal{S}^c$ are subsets as defined in the statement of Theorem 7.1. Since the links from 0 to any nodes $i \in \{1, \ldots, N\}$ are orthogonal by construction (not subject to any broadcast or interference constraint), we have that $\mathbf{G}_{\Omega_0, \Omega_0^c}$ has a block-diagonal form where a block is given by $\mathbf{G}_{\mathcal{S}, \mathcal{S}^c}$ (the links of the original network, corresponding to the cut (Ω_0, Ω_0^c) via the correspondence $\Omega_0 \leftrightarrow \mathcal{S}$ defined above) and other blocks, denoted by $\mathbf{G}_{0,j}$ for all $j \in \mathcal{S}^c$, have rank $n_{0,j}$, respectively. By construction, there is no direct link between 0 and d so, without loss of generality, we can assume $n_{0,d} = 0$. The general form for $\mathbf{G}_{\Omega_0, \Omega_0^c}$ is

$$\mathbf{G}_{\Omega_0, \Omega_0^c} = \begin{bmatrix} \mathbf{G}_{\mathcal{S}, \mathcal{S}^c} & 0 & \cdots & 0 \\ 0 & \mathbf{G}_{0,i_1} & & \vdots \\ \vdots & & \ddots & 0 \\ 0 & \cdots & 0 & \mathbf{G}_{0,i_{|\mathcal{S}^c|}} \end{bmatrix}$$

where we have indicated $\mathcal{S}^c = \{i_1, \ldots, i_{|\mathcal{S}^c|}\}$. Therefore, we have

$$\text{rank}\left\{\mathbf{G}_{\Omega_0,\Omega_0^c}\right\} = \text{rank}\left\{\mathbf{G}_{\mathcal{S},\mathcal{S}^c}\right\} + \sum_{j\in\mathcal{S}^c} n_{0,j} \tag{E.2}$$

In particular, the cut $\Omega_0 = \{0\}$ yields

$$R_0 \leq \sum_{j=1}^{N} n_{0,j} \tag{E.3}$$

By letting this inequality hold with equality, and by replacing this into all other inequalities, we obtain the set of inequalities

$$\sum_{i\in\mathcal{S}} n_{0,i} \leq \text{rank}\left\{\mathbf{G}_{\mathcal{S},\mathcal{S}^c}\right\}, \quad \forall\ \mathcal{S} \subseteq \{1,\ldots,N\} \tag{E.4}$$

where we used the fact that $\sum_{j=1}^{N} n_{0,j} - \sum_{j\in\mathcal{S}^c} n_{0,j} = \sum_{i\in\mathcal{S}} n_{0,i}$.

Consider now the ensemble of augmented networks for which there exist integers $\{n_{0,i} : i = 1,\ldots,N\}$ that satisfy (E.4). For such networks, the rate $R_0 = \sum_{j=1}^{N} n_{0,j}$ is achievable (by [Avestimehr *et al.*, 2007a]) and therefore the individual rates $R_i = n_{0,i}$ are achievable by the argument above. Finally, the closure of the convex hull of all individual rate vectors $\mathbf{R} = (n_{0,1},\ldots,n_{0,N})$ of such networks is achievable by time-sharing. It is immediate to see that this convex hull is provided by the inequalities in Eq. (7.12).[1]

E.2 Proof For Theorem 7.2

Again, we consider an augmented network with a single source node denoted by 0, with n_0 output bits that we denote by $\mathbf{V}$. As before, subsets $\mathcal{U}_i$ of cardinalities $n_{0,i}$ of these bits are sent to nodes i, respectively. However, differently from before we choose the subsets $\mathcal{U}_i$ to overlap in accordance with the vector source model. For the augmented network, the rate R_0 from the virtual source to the destination d must satisfy (E.1). In particular, choosing $\Omega_0 = \{0\}$ we get $R_0 \leq n_0$. Generalizing the proof of Theorem 7.1 to the case of overlapping sets $\{\mathcal{U}_i\}$, we find that for any cut (Ω_0, Ω_0^c) of the augmented network such that $\Omega_0 = \mathcal{S} \cup \{0\}$ and $\Omega_0^c = \mathcal{S}^c$, with $\mathcal{S} \subseteq \{1,\ldots,N\}$ we have

$$\text{rank}\left\{\mathbf{G}_{\Omega_0,\Omega_0^c}\right\} = \text{rank}\left\{\mathbf{G}_{\mathcal{S},\mathcal{S}^c}\right\} + \text{rank}\left\{\mathbf{G}_{0,\mathcal{S}^c}\right\}$$

[1] Indeed, the inequalities Eq. (7.12) represent the convex relaxation of the integer constraints (E.3).

where $\mathbf{G}_{0,\mathcal{S}^c}$ is the linear transformation between the inputs $\mathbf{V}$ and the (augmented) channel outputs of nodes $j \in \mathcal{S}^c$. By construction, the matrix $\mathbf{G}_{0,\mathcal{S}^c}$ is formed by linear independent columns for all bits V_ℓ with $\ell \in \bigcup_{j\in\mathcal{S}^c} \mathcal{U}_j$. Therefore,

$$\text{rank}\,\{\mathbf{G}_{0,\mathcal{S}^c}\} = \left|\bigcup_{j\in\mathcal{S}^c} \mathcal{U}_j\right| = H(U_{\mathcal{S}^c})$$

Since $\mathbf{V}$ is uniform i.i.d., we have $R_0 = n_0 = H(\mathbf{V}) = H(\mathbf{U})$. Replacing these equalities into the set of inequalities (E.1) and using the chain rule of entropy $H(\mathbf{U}) = H(U_{\mathcal{S}}|U_{\mathcal{S}^c}) + H(U_{\mathcal{S}^c})$ we obtain that the conditions (7.14) are sufficient for transmissibility. On the other hand, if a source as defined in our model was transmissible, then the set of conditions (7.14) must hold, otherwise the rate R_0 of the corresponding single-source single destination augmented network would violate Eq. (E.1). Hence, necessity also holds.

Bibliography

Agarwal, M. and A. Puri (2002, November).
Base station scheduling of requests with fixed deadlines.
In *Twenty-First Annual Joint Conference of the IEEE Computer and Communications Societies*, New York, USA.

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

Aref, M. R. (1980).
Information flow in relay networks.
Stanford University, Stanford, CA.
Ph.D. dissertation.

Avestimehr, A. S., S. N. Diggavi, and D. N. Tse (2007a, September).
A deterministic approach to wireless relay networks.
In *Allerton Conference on Communication, Control and Computing*, Illinois.

Avestimehr, A. S., S. N. Diggavi, and D. N. Tse (2007b, July).
A deterministic model for wireless relay networks and its capacity.
In *IEEE Information Theory Workshop*, Bergen, Norway.

Avestimehr, A. S., S. N. Diggavi, and D. N. Tse (2007c, September).
Wireless network information flow.
In *Allerton Conference on Communication, Control and Computing*, Illinois.

Barros, J. and S. Servetto (2006, January).
Network Information Flow With Correlated Sources.
IEEE Trans. Inform. Theory 52(1), 155–170.

Benaim, M. and J.-Y. Le Boudec (2008).

A class of mean field interaction models for computer and communication systems.
Performance Evaluation 65(11-12), 823–838.

Berry, R. and E. Yeh (2004, sep.).
Cross-layer wireless resource allocation.
IEEE Signal Processing Magazine 21(5), 59 – 68.

Berry, R. A. and R. G. Gallager (2002, May).
Communication over fading channels with delay constraints.
IEEE Trans. Inform. Theory 48(5), 1135–1149.

Bertsekas, D. P. (2007).
Dynamic Programming and Optimal Control, Vol. 1 (3 ed.).
Athena Scientific.

Biyikoglu, E. U. and A. E. Gamal (2004, December).
On adaptive transmission for energy efficeincy in wireless data networks.
IEEE Trans. Inform. Theory 50(12), 3081–3094.

Blake, S., D. Black, M. Carlson, E. Davies, Z. Wang, and W. Weiss (1998).
An Architecture for Differentiated Service.
United States: RFC Editor.

Bölcski, H., R. U. Nabar, Özgür Oyman, and A. J. Paulraj (2006, June).
Capacity scaling laws in MIMO relay networks.
IEEE Tansactions on Wireless Communications 5(6), 1433–1444.

Butt, M. M., K. Kansanen, and R. R. Müller.
Deadline constrained opportunistic scheduling for large multiuser systems.
in preparation for journal publication.

Butt, M. M., K. Kansanen, and R. R. Müller.
Guaranteeing QoS for large wireless sensor networks: A simplified framework.
in preparation for journal publication.

Butt, M. M., K. Kansanen, and R. R. Müller.
Hard deadline constrained multiuser scheduling for random arrivals.
accepted for publication in IEEE WCNC, 2011.

Butt, M. M., K. Kansanen, and R. R. Müller.
Individual packet deadline constrained scheduling for a multiuser system.
accepted for publication in IEEE VTC–Spring, 2011.

Butt, M. M., K. Kansanen, and R. R. Müller.
Individual packet deadline dependent multiuser scheduling: A practical perspective.
in preparation for journal publication.

Butt, M. M., K. Kansanen, and R. R. Müller (2007, October).
Opportunistic scheduling and delay limited systems.
In *VERDIKT Conference, Trondheim, Norway*.

Butt, M. M., K. Kansanen, and R. R. Müller (2008a, October).
Future rate estimation based opportunistic scheduling for a multiuser system.
In *VERDIKT Conference, Bergen, Norway*.

Butt, M. M., K. Kansanen, and R. R. Müller (2008b, April).
Multiuser opportunistic scheduling for hard delay constrained systems.
In *IEEE Sarnoff Symposium 2008, Princeton, NJ, USA*.

Butt, M. M., K. Kansanen, and R. R. Müller (2008c, August).
Provision of maximum delay guarantee at low energy in a multiuser system.
In *10th International Symposium on Spread Spectrum Techniques and Applications (ISSSTA), Bologna, Italy*.

Caire, G., R. Müller, and R. Knopp (2007, April).
Hard fairness versus proportional fairness in wireless communications: The single-cell case.
IEEE Trans. Inform. Theory 53(4), 1366–1385.

Caire, G., R. Müller, and T. Tanaka (2004, September).
Iterative multiuser joint decoding: Optimal power allocation and low-complexity implementation.
IEEE Trans. Inform. Theory 50(8), 1950–1973.

Cerny, V. (1985, January).
Thermodynamical approach to the travelling salesman problem: An efficient simulation algorithm.
Journal of Optimization Theory and Applications 45(1), 41–52.

Chan, W., M. J. Neely, and U. Mitra (2007, june).
Energy efficient scheduling with individual packet delay constraints: offline and online results.
In *IEEE Infocom*.

Chaponniere, E., P. Black, J. M. Holtzman, and D. Tsc (2002, September).
Transmitter directed, multiple receiver system using path diversity to equitably maximize throughput.
U. S. Patent No. 6449490.

Chaporkar, P., K. Kansanen, and R. R. Müller (2009, November).
On the delay-energy tradeoff in multiuser fading channels.
EURASIP Journal on Wireless Communications and Networking 2009.

Chen, J.-C. and T. Zhang (2004).
IP-Based Next Generation Wireless Networks: Systems, Architectures and Protocols (1 ed.).
John Wiley and Sons.

Chen, N. and S. Jordan (2009).
Downlink scheduling with guarantees on the probability of short-term throughput.
IEEE Trans. Wireless. Comm. 8(2), 593–598.

Chen, W., K. Letaief, and Z. Cao (2007, jun.).
Opportunistic network coding for wireless networks.
In *IEEE International Conference on Communications*.

Chen, W., U. Mitra, and M. Neely (2006, jul.).
Packet dropping algorithms for energy savings.
In *IEEE International Symposium on Information Theory (ISIT)*, Nice, France, pp. 227 –231.

Coleman, T. P. and M. Medard (2004, june).
A distributed scheme for achieving energy-delay tradeoffs with multiple service classes over a dynamically varying channel.
IEEE journal on selected areas in communications 22(5), 929–941.

Cover, T. M. and A. E. Gamal (1979, September).
Capacity theorems for the relay channel.
IEEE Trans. Inform. Theory 25(5), 572–584.

Cover, T. M. and J. Thomas (2006).
Elements of Information Theory (2 ed.).
Wiley and Sons.

Cui, T., L. Chen, and T. Ho (2008, apr.).
Energy efficient opportunistic network coding for wireless networks.
In *The 27th Conference on Computer Communications*.

David, H. and H. Nagaraja (2003).
Order Statistics (3 ed.).
Wiley and Sons.

Fu, A., E. Modiano, and J. N. Tsitsiklis (2006, March).
Optimal transmission scheduling over a fading channel with energy and deadline constraints.
IEEE Transactions on Wireless Communications, 630–641.

Gamal, A. E. and M. Aref (1982, May).
The capacity of the semi-deterministic relay channel.
IEEE Trans. Inform. Theory 28(3), 536.

Gamal, A. E. and S. Zahedi (2005, May).
Capacity of a class of relay channels with orthogonal components.
IEEE Trans. Inform. Theory 51(5), 1815–1817.

Gastpar, M. and M. Vetterli (2005, March).
On the capacity of large Gaussian relay networks.
IEEE Trans. Inform. Theory 51(3), 765–779.

Geman, S. and D. Geman (1984, November).
Stochastic relaxation, gibbs distribution and the baysian restoration in images.
IEEE transactions on pattern analysis and machine intelligence 6(6), 721–741.

Goldsmith, A. (2005).
Wireless Communications (1 ed.).
Cambridge University Press.

Guo, D. and S. Verdu (2005, June).
Randomly spread CDMA: Asymptotics via statistical physics.
IEEE Trans. Inform. Theory 51(6), 1983–2010.

Hajek, B. and P. Seri (1998, aug.).
On causal scheduling of multiclass traffic with deadlines.
In *Proceedings IEEE International Symposium on Information Theory*, Cambridge, MA, USA.

Hanly, S. and D. Tse (1998, November).
Multi-access fading channels-part II: Delay limited capacities.
IEEE Trans. Inform. Theory 44(7), 2816–2831.

Harrison, P. and S. Zertal (2007, August).
Queuing models of RAID systems with maxima of waiting times.
Performance Evaluation 64(7), 664–689.

Hassel, V., G. E. Øien, and D. Gesbert (2007).
Throughput guarantees for wireless networks with opportunistic scheduling: a comparative study.
IEEE Transactions on Wireless Communications 6(12), 4215–4220.

H.Szu and R. Hartley (1987).
Fast simulated annealing.
Physics Letters A 122(3).

Jindal, N., S. Vishwanath, and A. Goldsmisth (2004, May).
On duality of gaussian multiple access and broadcast channels.
IEEE Trans. Inform. Theory 50(5), 768–783.

Kabamba, P. T., S. M. Meerkov, and C. Y. Tang (2005, April).
Optimal, suboptimal and adaptive threshold policies for power efficiency of wireless networks.
IEEE Trans. Inform. Theory 51(4), 1359–1376.

Kelly, F. P., A. K. Maulloo, and D. K. H. Tan (1998, March).
Rate control for communication networks: Shadow prices, proportional fairness and stability.
The Journal of the Operational Research Society 49(3), 237–252.

Kim, Y. and G. U. Hwang (2009).
Threshold-based opportunistic scheduling for ergodic rate guarantees in wireless networks.
IEEE Trans. Wireless. Comm. 8(1), 84–89.

Kirkpatrick, S., C. Gelatt, and M. Vecchi (1983, May).
Optimization by simulated annealing.
Science 220(4598), 671–680.

Knopp, R. and P. Humblet (1995, June).
Information capacity and power control in single cell multiuser communications.
In *IEEE, Int. Computer conf, Seattle, WA*.

Kramer, G. (2009, June).
Pipelined encoding for deterministic and noisy relay networks.
In *2009 Workshop on Network Coding, Theory and Applications, NETCOD09*, Lausanne, Swtizerland.

Kramer, G., M. Gastpar, and P. Gupta (2005, September).
Cooperative strategies and capacity theorems for relay networks.
IEEE Trans. Inform. Theory 51(9), 3037–3063.

Lee, J. and N. Jindal (2009, April).
Energy-efficient scheduling of delay constrained traffic over fading channels.
IEEE Trans. Wireless Communications 8(4), 1866–1875.

Liu, X., E. K. P. Chong, and N. B. Shroff (2003).
A framework for opportunistic scheduling in wireless networks.
Computer Networks 41, 451–474.

Meulen, E. C. V. d. (1977, January).
A survey of multi-way channels in information theory.
IEEE Trans. Inform. Theory 23(1), 1–37.

Miao, G., N. Himayat, Y. Li, and A. Swami (2009, September).
Cross-layer optimization for energy-efficient wireless communications: a survey.
Wireless Communications and Mobile Computing 9(4), 529–542.

Min, K. Y., E. S. Kim, S. I. Woo, and D. K. Kim (2005).
Downlink packet scheduling with minimum throughput guarantee in TDD-OFDMA cellular network.
In *NETWORKING 2005*, Volume 3462 of *Lecture Notes in Computer Science*, pp. 623–633. Springer Berlin / Heidelberg.

Mohajer, S., C. Tian, and S. Diggavi (2010, jan.).
On source transmission over deterministic relay networks.
In *IEEE Information Theory Workshop (ITW)*, Cairo.

Mohammad Ali Maddah-Ali and David N. C. Tse (2009, July).
Approximating the Rate-Distortion Region of the Distributed Source Coding for Three Jointly Gaussian Tree-Structured Sources.
In *IEEE Int. Symp. on Inform. Theory, ISIT 09*, Seoul, Korea.

Neely, M. J. (2007, September).
Optimal energy and delay tradeoffs for multiuser wireless downlinks.
IEEE Trans. Inform. Theory 53(9), 3095–3113.

Neely, M. J. (2009, March).
Intelligent packet dropping for optimal energy-delay tradeoffs in wireless downlinks.
IEEE Trans. on Automatic Control 54(3), 565–579.

Neely, M. J. (2010, March).
Delay-based network utility maximization.
In *Proceedings IEEE INFOCOM*.

Park, D. and G. Caire (2008, jul.).
Hard fairness versus proportional fairness in wireless communications: The multiple-cell case.
In *IEEE International Symposium on Information Theory*, Toronto, Canada.

Perron, E., S. Diggavi, and E. Telatar (2009, February).
On noise insertion strategies for wireless network secrecy.
In *ITA*, San Diego, CA.

Phan, C. V. and J. G. G. Kim (2007, March).
An energy efficient transmission strategy for wireless sensor networks.
In *IEEE WCNC*.

Pitic, R. and A. Capone (2008, March-April).
An opportunistic scheduling scheme with minimum data-rate guarantees for OFDMA.
In *WCNC*, Las Vegas, USA.

Rajan, D., A. Sabharwal, and B. Aazhang (2004).
Delay-bounded packet scheduling of bursty traffic over wireless channels.
IEEE Trans. Info. Theory 50, 125–144.

Rappaport, T. (2001).
Wireless Communications: Principles and Practice.
Upper Saddle River, NJ, USA: Prentice Hall PTR.

Rashid, M., M. Hossain, E. Hossain, and V. Bhargava (2009, oct.).
Opportunistic spectrum scheduling for multiuser cognitive radio: a queueing analysis.
IEEE Transactions on Wireless Communications 8(10), 5259 –5269.

Rose, S. M. (2003).
Introduction to Probability Models (8 ed.).
Academic Press.

Shakkottai, S. and A. L. Stolyar (2000).
Scheduling for multiple flows sharing a time-varying channel: The exponential rule.
American Mathematical Society Translations, Series 2, 2002.

Shakkottai, S. and A. L. Stolyar (2001, December).
Scheduling algorithms for a mixture of real-time and non-real-time data in HDR.
In *Proceedings of 17th International Teletraffic Congress (ITC-17)*, pp. 793–804.

Srivastava, V. and M. Motani (2005, December).
Cross layer design: A survey and the road ahead.
IEEE communication magazine, 112–119.

Stolyar, A. L. and K. Ramanan (2001).
Largest weighted delay first scheduling: Large deviations and optimality.
The Annals of Applied Probability 11(1), 1–48.

Striegel, A. and G. Manimaran (2002).
Packet scheduling with delay and loss differentiation.
Computer Communications 25, 21–31.

Tarello, A., J. Sun, M. Zafar, and E. Modiano (2008, October).
Minimum energy transmission scheduling subject to deadline constraints.
Wireless Networks 14(5), 633–645.

Tse, D. and S. Hanly (1998, November).
Multi-access fading channels-part I: Polymatroid structure, optimal resource allocation and throughput capacities.
IEEE Trans. Inform. Theory 44(7), 2796–2815.

Tse, D. and P. Viswanath (2005).
Fundamentals of Wireless Communications (1 ed.).
Cambridge University Press.

Urgaonkar, R. and M. J. Neely (2009, June).
Opportunistic scheduling with reliability guarantees in cognitive radio networks.
IEEE Transactions on Mobile Computing 8(6), 766–777.

Verdu, S. (2002, june).
Spectral efficiency in the wideband regime.
IEEE Transactions on Information Theory 48(6), 1319 –1343.

Viswanath, P., D. N. Tse, and V. Anantharam (2001, January).
Asymptotically optimal water-filling in vector multiple-access channels.
IEEE Trans. Inform. Theory 47(1), 241–267.

Viswanath, P., D. N. Tse, and R. Laroia (2002, June).
Opportunistic beamforming using dumb antennas.
IEEE Trans. Inform. Theory 46(6), 1277–1294.

Wu, D. and R. Negi (2005, May).
Utilizing multiuser diversity for efficient support of quality of service over a fading channel.
IEEE Tansactions on Vehicular Technology 54(3), 1198–1206.

Wyner, A. (1994, nov.).
Shannon-theoretic approach to a gaussian cellular multiple-access channel.
IEEE Transactions on Information Theory 40(6), 1713 –1727.

Xiao, H., W. K. G. Seah, A. Lo, and K. C. Chua (2000).
A flexible quality of service model for mobile ad-hoc networks.
In *IEEE VTC-Spring*, Tokyo, Japan, pp. 445–449.

Yao, Y. and G. Giannakis (2005, August).
Energy-efficient scheduling for wireless sensor networks.
IEEE Transactions on Communications 53(8), 1333 – 1342.

Zander, J. (1997, August).
Radio resource management in future wireless networks: Requirements and limitations.
IEEE Communications Magazine 35, 30–36.

Zhang, Y., H. Hu, and M. Fujise (2007).
Resource, Mobility and Security Management in Wireless Networks and Mobile Communications (1 ed.).
Auerbach Publications.

Zheng, D., W. Ge, and J. Zhang (2009, Jan.).
Distributed opportunistic scheduling for Ad-Hoc networks with random access: An optimal stopping approach.
IEEE Transactions on Information Theory 55(1), 205–222.

Zheng, D., M.-O. Pun, W. Ge, J. Zhang, and H. Poor (2008, dec.).
Distributed opportunistic scheduling for ad hoc communications with imperfect channel information.
IEEE Transactions on Wireless Communications 7(12), 5450 –5460.

Zorba, N., A. I. Pérez-Neira, A. Foglar, and C. Verikoukis (2009).
Cross layer QoS guarantees in multiuser WLAN systems.
Wirel. Pers. Commun. 51(3), 549–563.

Zorba, N. and C. Verikoukis (2010).
A QoS-Based dynamic queue length scheduling algorithm in multi-antenna heterogeneous systems.
EURASIP Journal on Wireless Communications and Networking 2010.